Designing the Microbial Research Commons: Proceedings of an International Workshop

ISBN
978-0-309-21979-2

228 pages
8 1/2 x 11
PAPERBACK (2011)

Paul F. Uhlir, Editor; National Research Council

Designing the Microbial Research Commons

Proceedings of an International Symposium

Paul F. Uhlir, Editor

Board on Research Data and Information
Policy and Global Affairs

NATIONAL RESEARCH COUNCIL
OF THE NATIONAL ACADEMIES

THE NATIONAL ACADEMIES PRESS
Washington, D.C.
www.nap.edu

THE NATIONAL ACADEMIES PRESS 500 Fifth Street, N.W. Washington, DC 20001

NOTICE: The project that is the subject of this report was approved by the Governing Board of the National Research Council, whose members are drawn from the councils of the National Academy of Sciences, the National Academy of Engineering, and the Institute of Medicine. The members of the committee responsible for the report were chosen for their special competences and with regard for appropriate balance.

This study was supported by the National Science Foundation under Grant No. OCI-082173 and by the Department of Energy Grant No. DE-SC0002579. This report was prepared as an account of work sponsored by an agency of the United States government. Neither the United States government nor any agency thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States government or any agency thereof. Any opinions, findings, conclusions, or recommendations expressed in this publication are those of the author(s) and do not necessarily reflect the views of the organizations or agencies that provided support for the project.

International Standard Book Number-13: 978-0-309-21979-2
International Standard Book Number-10: 0-309-21979-5

Additional copies of this report are available from the National Academies Press, 500 Fifth Street, N.W., Lockbox 285, Washington, DC 20055; (800) 624-6242 or (202) 334-3313 (in the Washington metropolitan area); Internet, http://www.nap.edu.

Printed in the United States of America.

THE NATIONAL ACADEMIES

Advisers to the Nation on Science, Engineering, and Medicine

The **National Academy of Sciences** is a private, nonprofit, self-perpetuating society of distinguished scholars engaged in scientific and engineering research, dedicated to the furtherance of science and technology and to their use for the general welfare. Upon the authority of the charter granted to it by the Congress in 1863, the Academy has a mandate that requires it to advise the federal government on scientific and technical matters. Dr. Ralph J. Cicerone is president of the National Academy of Sciences.

The **National Academy of Engineering** was established in 1964, under the charter of the National Academy of Sciences, as a parallel organization of outstanding engineers. It is autonomous in its administration and in the selection of its members, sharing with the National Academy of Sciences the responsibility for advising the federal government. The National Academy of Engineering also sponsors engineering programs aimed at meeting national needs, encourages education and research, and recognizes the superior achievements of engineers. Dr. Charles M. Vest is president of the National Academy of Engineering.

The **Institute of Medicine** was established in 1970 by the National Academy of Sciences to secure the services of eminent members of appropriate professions in the examination of policy matters pertaining to the health of the public. The Institute acts under the responsibility given to the National Academy of Sciences by its congressional charter to be an adviser to the federal government and, upon its own initiative, to identify issues of medical care, research, and education. Dr. Harvey V. Fineberg is president of the Institute of Medicine.

The **National Research Council** was organized by the National Academy of Sciences in 1916 to associate the broad community of science and technology with the Academy's purposes of furthering knowledge and advising the federal government. Functioning in accordance with general policies determined by the Academy, the Council has become the principal operating agency of both the National Academy of Sciences and the National Academy of Engineering in providing services to the government, the public, and the scientific and engineering communities. The Council is administered jointly by both Academies and the Institute of Medicine. Dr. Ralph J. Cicerone and Dr. Charles M. Vest are chair and vice chair, respectively, of the National Research Council.

www.national-academies.org

Steering Committee for Designing the Microbial Research Commons: An International Symposium

Dr. Cathy Wu, Chair,
Edward G. Jefferson Professor of Bioinformatics & Computational Biology
Delaware Biotechnology Institute
University of Delaware

Michael Carroll, J.D.
Professor
American University, Washington College of Law

Dr. Micah Krichevsky
President
Bionomics International
Rockville, MD

Dr. Michael Lesk (NAE)
Professor
School of Communication, Information and Library Sciences
Rutgers University

Dr. Stephen J. McCormack
Chief Executive Officer
IMI Devices
Bonn, Germany

Dr. James Staley
Health Services
University of Washington

Dr. Larry Smarr (NAE)
Director, California Institute for Telecommunications and Information Technology
Professor, University of California San Diego

National Academies Staff
Paul F. Uhlir, Director, Board on Research Data and Information

Preface and Acknowledgments

The opportunities to accelerate scientific discovery and resulting applications are made increasingly possible by technological breakthroughs and pioneering methods to process and integrate vast amounts of data, information, and raw materials. Microbial research, which is outgrowing its "small science" institutional structures, should consider building upon these opportunities in an attempt to develop a global microbial research commons to promote access to databases, literature, and materials through an open, digitally distributed network. However, the increasingly blurred line between basic and applied research confers potential economic value even upon research inputs that are far upstream. As a result, the research community is increasingly being forced to come to terms with commoditizing pressures within developed economies. These pressures restrict the conduct of public-sector research through strong intellectual property rights and related contractual restrictions on access to and use of materials, publications, and data. At the same time, restrictive policies in developing countries under the Convention on Biological Diversity complicate research uses of microbial materials held in public repositories *ex-situ*, and make it increasingly difficult to access the vast *in-situ* materials these countries control.

These trends have led to a proliferation of diverse licensing strategies and techniques, which collectively have elevated the transaction costs and other barriers for even relatively simple cooperative research projects. There is, thus, a need to focus on the obstacles to upstream, non-commercial research and the solutions to them. An early step is development of a set of design principles that address the economic, legal, and institutional dimensions of the transformation of the existing research infrastructure into what could become a globally distributed and digitally integrated research commons. The goal of this redesigned "soft" infrastructure would be to better manage publicly funded research resources, without compromising downstream commercial applications and fruitful partnerships between the public and private sectors, or between developed and developing countries.

Of course, a variety of responses is possible. Some are more conservative with respect to an understanding of the scientific "commons" as a common resource available on a non-discriminatory and non-commercial basis, whereas others may be based upon a pro-actively managed or regulated set of practices. These latter responses would compromise the conservative view in the interest of achieving greater patronage and participation of actors who have other motives and rationales for participation. A more detailed discussion of the "commons" concept is provided in the presentation by Paul David in Chapter 3 and Charlotte Hess in Chapter 25, as well as from various other perspectives of course throughout this volume.

The Board on Research Data and Information held an International Symposium on Designing the Microbial Research Commons at the National Academy of Sciences in Washington, DC on 8-9 October 2009. Organized by a separately appointed Steering Committee, this symposium expanded on prior international discussions on the same topic at a conference in June 2008 in Ghent, Belgium (see: http://www.microbialcommons.ugent.be/). The October 2009 symposium addressed topics such as models to lower the transaction costs and support access to and use of microbiological materials and digital resources from the perspective of publicly funded research, public-private interactions, and developing country concerns. The overall goal

of the symposium was to stimulate more research and implementation of improved legal and institutional models for publicly funded research in microbiology.

The International Symposium on Designing the Microbial Research Commons focused on accomplishing the following tasks:

1. Delineate the research and applications opportunities from improved integration of microbial data, information, and materials and from enhanced collaboration within the global microbial community.

2. Identify the global challenges and barriers—the scientific, technical, institutional, legal, economic, and socio-cultural—that hinder the integration of microbial resources and the collaborative practice of scientific communities in the microbial commons.

3. Characterize the alternative legal and policy approaches developed and implemented by other research communities, such as common-use licensing for scientific data and information, standard-form material transfer agreements, open access publishing, and open data networks that could be applied successfully by the microbial research community.

4. Define the contributions of new information and communication technology (ICT) tools in building federated information infrastructures, such as ontologies, data and text mining, and web 2.0.

5. Discuss and evaluate the institutional design and governance principles of data and information sharing among information infrastructures, drawing upon and analyzing successful and failed case studies in the life sciences.

6. Identify the range of policy issues that need to be addressed for maximizing open access to materials, data and literature information in an integrated microbial research commons.

The statements in this volume are those of the individual authors and do not necessarily represent the views of other workshop participants, the steering committee, or the National Academies. The symposium agenda is provided in Appendix A and the list of the meeting participants is presented in Appendix B.

On behalf of the Board, we gratefully acknowledge the support for this project of the Department of Energy under grant number DE-SC0002579, and from the National Science Foundation under grant number OCI-0821873, as well as the core support it has received from the National Institutes of Health, the Defense Technical Information Center, the Institute of Museum and Library Services, the National Institute of Standards and Technology, and the Library of Congress.

This volume has been reviewed in draft form by individuals chosen for their technical expertise, in accordance with procedures approved by the NRC's Report Review Committee. The purpose of this independent review is to provide candid and critical comments that will assist the institution in making its published report as sound as possible and to ensure that the report meets institutional standards for quality. The review comments and draft manuscript remain confidential to protect the integrity of the process.

We wish to thank the following individuals for their review of selected papers: Minna Allarakhia, University of Waterloo, Canada; Subbiah Arunachalam, Consultant; Nancy Connell, University of Medicine and Dentistry of New Jersey; Michael Carroll, American University; Melanie Dulong de Rosnay, *Communia*; Micah Krichevsky, Bionomics International; Michael Lesk, Rutgers University; Elinor Ostrom, Indiana University; James Staley, University of Washington, Seattle; and W. Edward Steinmueller, University of Sussex, UK.

Although the reviewers listed above have provided constructive comments and suggestions, they were not asked to endorse the content of the individual papers. Responsibility for the final content of the papers rests with the individual authors.

We would especially like to recognize the contributions of Daniel Cohen, on assignment to the National Academies from the U.S. Library of Congress, who assisted with the editing and the production of the manuscript. Subhash Kuvelker and Cheryl Levey of the Board staff also helped with the review process and the preparation of this volume. Finally, and not least, we would like to thank Fran Sharples, director of the Board on Life Sciences, for her assistance with the project.

Cathy H. Wu
Steering Committee Chair

Paul F. Uhlir
Project Director

Contents

1. Introduction
– Cathy Wu

University of Delaware

This symposium was organized by a committee of the Board on Research Data and Information of the National Research Council (NRC), in collaboration with the NRC's Board on Life Sciences and Board on International Scientific Organizations. The Board on Research Data and Information was established in October 2008, and it is driven by the growing impact of digital data and information in research. The board's mission is to improve the stewardship, policy, and use of digital data and information for science and the broader society.

Recent decades have witnessed an ever-increasing range and volume of digital data. All elements of the pillars of science—whether observation, experiment, or theory and modeling—are being transformed by the continuous cycle of generation, dissemination, and use of factual information. This is even more so in terms of the re-using and re-purposing of digital scientific data beyond the original intent of the data collectors, often with dramatic results.

We all know about the potential benefits and impacts of digital data, but we are also aware of the barriers, the challenges in maximizing the access and use of such data. There is thus a need to think about how a data infrastructure can enhance capabilities for finding, using, and integrating information to accelerate discovery and innovation. How can we best implement an accessible, interoperable digital environment so that the data can be repeatedly used by a wide variety of users in different settings and with different applications?

That is the objective of the symposium: to use the microbial communities and microbial data, literature, and the research materials themselves as a test case. We want to see if we can identify some guiding principles that can address various dimensions of this transformation—the economic, legal, and institutional dimensions—that could be brought about and whether the upstream microbial research inputs can ultimately be organized in a globally accessible, effective, and digitally integrated research commons.

In this symposium, we would also like to discuss models and approaches to lower the transaction cost and support access for not just digital data and literature, but also for the physical microbial materials. We intend to look at this from various perspectives—publicly funded research, public–private interactions, and the concerns of developing countries.

Thanks to the sponsorship of the Department of Energy, we will have a special thematic focus on research and applications in energy and environment. So the idea is to use the microbial research commons as a way to analyze the various issues that are affecting the potential opportunities offered by the use of the digital and material resources, and to examine how the microbial commons could be used as a model for further discussion and analysis in other research contexts.

This symposium is very timely. Just last month the National Research Council published a new report, *A New Biology for the 21st Century*,[1] in which the NRC called for a federally funded, decade-long interagency effort to harness biological technology and information. The report called for a New Biology initiative that would take an integrated, interdisciplinary approach to life science research to address some of the most pressing problems in food, environment, energy, and health. One of the major goals of the New Biology initiative would be

[1] National Research Council. A New Biology for the 21st Century. Washington, D.C.: National Academies Press, 2009.

to address the tremendous opportunities created by the massive and growing amounts of data generated from recent biological research and technological advances. Thus this symposium on the microbial research commons is timely and fitting in that microbial research affects all these other areas of interest and may contribute to this New Biology initiative.

There are several things we would like to accomplish in this symposium. First, we will look into some of the research and application opportunities that may arise from the sort of improved integration we are discussing. This is the value proposition: What do we gain by this integration?

Then we will review the scientific, technical, institutional, legal, economic, and socio-cultural barriers and challenges. We will examine some alternative legal, policy and institutional approaches, such as a compensatory liability regime for the transfer and use of microbial materials, common-use licensing of scientific data and information, open-access publishing, open data networks, and so forth. In this context, it is important to emphasize that open-access publishing is but one component of a Commons, which is a broader concept that encompasses free access with few reuse restrictions on all types of upstream knowledge resources, subject of course to legitimate countervailing policies or requirements. We also will consider information technology, including the impact of new information communication technology tools, such as social networking, data and text mining, and Web 2.0. We will look at governance and institutional designs, principles, and policies, as well.

Because the topics we will address are very broad and encompassing, we designed a program that would allow us to discuss various aspects of these issues in different sessions. In this morning session we will try to establish the context, focused primarily on microbial research, and the opportunities and the barriers. We will look into the value proposition and then also examine the industrial perspective and the concerns of developing countries. Then in the afternoon, we will first analyze access and reuse of microbial materials and then of digital knowledge resources. We will review how the commons models might work for materials and culture collections, including from the legal and economic perspectives, as well as for the digital resources, including Web applications, Web information services, and academic publications.

Tomorrow we will have the thematic focus on microbiology research and applications in energy and environment, and the opportunities they provide in terms of both materials and the digital commons. We will look into issues such as international cooperation, inter-governmental organization, and institutional design for the materials research commons, as well as management of academic journals, data standardization for facilitating interoperability, and economic and institutional issues for the digital commons.

Finally, the last session of the symposium will focus on the governance issues associated with an integrated microbial commons, and, again, there are a few existing approaches that can be reviewed.

Following the presentations in each session there will be panel discussions during which we hope to gather many different opinions and approaches. It is not the symposium's objective to come up with one consensus, but rather to provide a broad review of the issues that need to be addressed and brought to the forefront.

2. Microbiology in the 21st Century – Joan W. Bennett[2]

Rutgers University

The Commons Defined

Common:

1.
a. belonging equally to or shared equally by two or more; joint: *common interests.*
b. of or relating to the community as a whole; public: *for the common good.*
2. widespread; prevalent.
3.
a. occurring frequently or habitually; usual.
b. most widely known; ordinary: *the common housefly.*
4. having no special designation, status, or rank: *a common sailor.*
5.
a. not distinguished by superior or noteworthy characteristics; average: *the common spectator.*
b. of no special quality; standard: *common procedure.*
c. of mediocre or inferior quality; second-rate: *common cloth.*
6. unrefined or coarse in manner; vulgar: *behavior that branded him as common*
SOURCE: *http://www.thefreedictionary.com/Common*s

Commons:

The common people; commonalty.

2. commons *(used with a sing. or pl. verb)*
a. The social class composed of commoners.
b. The parliamentary representatives of this class.
3. The House of Commons. Often used in the plural.
4. A tract of land, usually in a centrally located spot, **belonging to or used by a community as a whole**: *a band concert on the village common.*
5. The legal right of a person to use the lands or waters of another, as for fishing.
6. commons *(used with a sing. verb)* A building or hall for dining, typically at a university or college.

SOURCE: *http://www.thefreedictionary.com/Commons*

A Primer in Microbiology

In microbiology we are working at a scale that is orders of magnitude smaller than what most people are used to thinking about. Many of the microbes that are studied, like bacteria, are smaller than single cells of the human body. Thousands of *Bacillus* cells will fit on the tip of a pin. Most archaea and bacteria are about the size of the nucleus of a eukaryotic cell. Viruses are smaller still, so they are difficult to visualize unless one has an electron microscope. Because microbes are so small, early microbiologists figured out ways to grow them in the laboratory so we could see populations of them growing together in colonies. Microbiologists have had to be experimental. Many of common microbial techniques were developed by 19th century bacteriologists. While the symbol of the classical microbiologist is the microscope, the symbol of the experimental microbiologist is the Petri dish. Although we often think of a microbial colonies growing

[2] Presentation slides available at *http://sites.nationalacademies.org/xpedio/idcplg?IdcService=GET_FILE&dDocName=PGA_054556&RevisionSelectionMethod=Latest.*

only on Petri dishes, that is a bit of a misconception. Nature frequently provides its own colonies. Sometimes you can see whole populations of microbes "bloom." Purple sulfur bacteria create a rather pretty colony, for example. Those of you who are skiers have probably seen "water melon snow" created by cold-tolerant algae and cyanobacteria that grow on snow and ice during alpine and polar summers. And, of course, in the world of rot and ruin, microbes are frequently visible. They make ugly colonies when they spoil fruit, vegetables, and other food stuffs. Microbes also have been known by the good things they do, such as their uses in bread making and fermentation, which go back to the early stages of civilization. Yeast is the microbe used in making bread; it provides the leavening. Similarly, yeasts are essential for the fermentation of wine and beer. People harnessed yeast metabolism for centuries without knowing that they were working with a microbe.

Microbes are best known for the diseases they cause. *Pathogen* is the general name for an organism that causes a disease, and is used to describe all microbes able to cause disease in animals and/or plants. Infectious diseases are those that spread from one person to another. An infectious disease can be more formally defined as a clinically evident disease resulting from the presence of pathogenic microbial agents, including viruses, bacteria, fungi, protozoa, multicellular parasites, and the aberrant proteins known as prions.

When we consider microbiology, we often think about the traditional scourges. One is leprosy, which has many references in the Bible and other classical literature. The causative bacterium of leprosy, also known as Hansen's disease, is *Mycobacterium leprae.* A second notorious infectious disease of antiquity is plague caused by *Yersinia pestis*. The "Black Death," as it was known, killed a third of Europe's population during the 14th century. It got this name from the black skin splotches it caused on affected people. Swollen lymph glands, or buboes, are the basis of the name "bubonic plague." The word plague, which should be used narrowly to describe just that one disease, often has come to be used as a general term for any devastating epidemic disease. Tuberculosis, for example, is sometime called the white plague.

In addition, many viral diseases have been known since ancient times. These include chicken pox, influenza, mumps, polio, rabies, and yellow fever. Most of these diseases are still very much with us. On the other hand, smallpox, in one of the great triumphs of public health and microbiology, has been eradicated. Young people are no longer vaccinated. A corollary of the eradication of small pox: If you want a good bioterrorism weapon, you do not have to do any genetic engineering—just unleash smallpox again.

Professional Societies, Journals, and Culture Collections

What about the professional societies organized by microbiologists? Microbiologists started organizing themselves into professional groups right after the research of Koch and Pasteur changed the face of medicine and public health. The American Society for Microbiology (ASM), the society that I know best, is not only the world's oldest microbiological society but is also one of the world's oldest biological societies. It was founded in 1899, well over a hundred years ago. It was initially named the Society for American Bacteriologists, and for quite a while, even though it had members who worked on viruses and fungi, its members referred to what they did as bacteriology. The Society name was changed from the American Society for

Bacteriology to the American Society for Microbiology in 1961. The ASM now has over 30,000 members representing more than two dozen disciplines of microbiological specialization plus a division for microbiology educators. It is international in scope, with about a third of the members being from outside the United States. For much of its history, the ASM has published a number of journals. The *Journal of Bacteriology* was the first one. Now several additional premier journals that specialize in different facets of microbiology are also published by ASM such as *Applied and Environmental Microbiology, Eukaryotic Cell, Journal of Clinical Microbiology*, and *Journal of Virology.* The Society also publishes a monthly magazine called *Microbe*, formerly *ASM News*, as well as an education journal.

Early in its history, ASM was deeply involved in the publication of *Bergey's Manual*, which has been the premier reference book for compiling the names of microbial strains. The first edition was initiated by action of the Society of American Bacteriologists by the appointment of an Editorial Board chaired by David H. Bergey. The first edition was published in 1923, the second edition in 1925, and a third edition came out in 1930. *Bergey's Manual* is now owned and produced by an independent trust.

Genomics and the Microbial Commons

The landmark microbiology-biotechnology patent case was *Diamond v. Chakrabarty*, decided by the U.S. Supreme Court in 1980. Diamond was the Commissioner of the U.S. Patent Office and Chakrabarty at the time was employed by an oil company. The subject of the patent case was not a genetically engineered organism, but rather was an organism that was claimed to chew up oil waste. In a close decision, Chakrabarty's side won, and the ruling established the precedent that microbial and other life forms could be patented. By the time this case came along, I was a young, recently tenured associate professor and I remember how excited I felt by the sense that we were entering a new world for biotechnology. Significant scientific and commercial advances have kept coming since that time.

The pace of DNA sequencing then got faster and faster, and the whole field of genomics was born. *Genome* is an old word used in genetics—cytogenetics actually—to describe the entire genetic content of an organism. *Genomics* was adopted in the late 1980s to describe the new sub-discipline. It is a word used many ways, but we would not have genomics without the tools developed by the study of bacterial genetics. The components underlying the ability to sequence DNA were first developed in microbiology. For example, much of the sequencing for the human genome project was done using the M-13 bacteriophage.

The Microbial Commons in the 21st century

The published scientific literature raises other questions. We have a long-entrenched tradition that many professional societies exist almost solely to publish their journals, and that these societies get much of their revenue from protecting and selling those publications. The advent of digital publishing has shifted the economics of scientific publishing. It takes a great deal of time for scientists to engage in peer review. Someone has to edit and organize the efforts into a coherent publication. If scientific societies do not do so, who will?

Thus, the tradition of peer reviewed scientific publications of research findings is not going to go away. It is scientists who keep the publication system going, with their long hours in the laboratory doing experiments followed by long hours devoted gratis to editing and reviewing. Every scientist worth his or her salt devotes a lot of time—free, unpaid time—to peer review, and not just to publications, but to grant proposals as well.

About 50 years ago, an increasing number of commercial units started publishing scientific journals. These professional journals have become increasingly expensive to obtain. I could not find a study on the price of microbiology journals, but there are several good studies on the cost of chemistry journals. According to the *Library Journal Periodicals Price Survey*, the average cost of a chemistry periodical to a library in 2009 was over $3,700. If you multiply that by the number of journals a research library is supposed to carry, the costs are crushing.

Another development comes from government. In response to congressional legislation, the National Institutes of Health now requires grantees to put their research articles online, free of charge, within 12 months if they have been supported by NIH funding. This has been a good development for more open availability of research results but the full implications for the long term sustainability of the current scientific publication system remain unclear.

In contrast to the world of the traditional scientific paper, the situation with databases is more favorable to open communication. As part of the genomics revolution, the databases are relatively new, with much of their development having happened during the past 20 years. Since its beginning, genomics has generated a number of other "omics"—proteomics, metabolomics, and so on—and there has been a convergence of the bioinformatics community with the experimental microbiology community.

Bioinformatics brings mathematical, statistical and computing methods to the analysis of the vast amount of DNA sequence information, gene expression data, and other information generated from the new biological disciplines. These are *huge* datasets. These databases facilitate studies by a new kind of biologist who does "*in silico"* research.

The pace of discovery has speeded up incredibly since the early years of recombinant DNA research and gene sequencing. In the beginning, we had to work hard to analyze a few thousand DNA base pairs. Today, however, the speed is absolutely incredible. The sequencing technologies that are used at the big sequencing centers now are generating vast quantities of DNA sequence data at a rate that is hard to imagine. Moreover, sequencing technologies evolve very rapidly. As a result, we have increasingly large amounts of data to manage, a need for increasing storage capacity, and an even more important need to develop tools for data manipulation.

What is microbiology? How do the increasingly blurred lines between basic and applied research affect the discipline? What economic, legal and institutional dimensions of the existing research infrastructure shape our ability to create a digitally integrated research commons? The issues that we are supposed to cover at this symposium are enormous.

I think I was asked to speak because I am a past president of the American Society for Microbiology (ASM). During the year I was president in 1990, almost 20 years ago, I did what was then considered a rather radical thing. I had a presidential symposium on the Human Genome Project. At the time, many microbiologists were against genomics, because they considered it "big science." How times have changed!

In establishing the context for our Conference on the Microbial Commons, I have been asked to set the stage for our discussions about the integration of digital and physical resources for clinical and environmental microbiology.

The charge was so broad and so challenging that I have decided to focus on examples from clinical microbiology to the exclusion of the environmental. Furthermore, because one cannot talk sensibly about the future without talking about the past, I am going to start with some real basics. Many of you are not microbiologists, and perhaps may need the primer.

I went to the University of Chicago, where professors always said to go back to Aristotle—start with the definitions. When you look at definitions in our post-modern world, you should remember that the words also carry their connotations, so I begin with some of the definitions of "common" below, and then I follow it with definitions of "commons."

When we speak of "commons," we generally are talking about the notion of shared public resources. If you are an educated person, it is very hard to use the term without the shadow of Garrett Hardin's famous 1968 essay on the "tragedy of the commons." This concept of the commons being associated with negative outcomes has entered the shared vocabulary of science, and I think the vocabulary of law. It occasionally even makes it into the popular press.

Now let me offer some basics for those of you who are not microbiologists. Microbiology is not small biology, but rather it is the biology of organisms that are too small to be seen by the naked eye. Microbiology encompasses the study of a whole array of life forms, especially a group that used to be called bacteria and that are now called the eubacteria and the archaea. Microbiology also deals with the study of viruses, protozoa, fungi, and algae.

The symbol of microbiologists traditionally has been that of the light microscope because of the dependence on microscopy to see individual microbial cells. The microscope often appears in the logos of professional societies. When you look under the microscope at bacteria, you do not see a great deal of morphology and this has always been one of the challenges of microbiology: Not only are the organisms small, but they do not look as different as, say, peacocks and ostriches do.

Among the biggest heroes of microbiology—and microbiology has many heroes—was Louis Pasteur, who refuted the theory of spontaneous generation. He proved that microorganisms are generated by other microorganisms. He also developed one of the first treatments for a microbial disease, rabies.

For infectious diseases, the hero was Robert Koch who first showed definitively that a particular microbe could cause a given infectious disease (anthrax). The experimental protocols he used are now called Koch's Postulates. They are a microbiologist's four-rule version of the Ten Commandments, done in a specified sequence to connect diseases with specific etiological agents.

An aside: I happen to believe that many chronic diseases such as heart disease and arthritis are going to have microbial connections, but this hypothesis is not going to be easy to prove by Koch's Postulates. That is another story, however.

Although the study of infectious disease is what brings microbiology much of its fame and scientists much their funding, it is certainly not the positive, friendly side of microbiology. Microbiology has an image problem rooted in this notion that microbes are germs that are bad and that make us sick. Even though there are many "good microbes"

that have a positive economic impact in our lives, much of the study of microbiology is associated with its medical applications to infectious diseases.

Of course, there are many other microbiological societies outside the United States. The British have the Society for General Microbiology, which also publishes a number of premier journals and a magazine. Then there are dozens and dozens of microbiological societies associated with different countries, many of which also publish distinguished journals. Some countries have multiple microbiological societies focusing on clinical, environmental, or industrial aspects of the profession.

The International Union of Microbiological Societies (IUMS) was founded in 1927. The IUMS is an umbrella organization that attempts to provide a forum for all of these international societies. If a person is an IUMS representative from a smaller country like Peru or Israel, he or she will have an equal vote in kind of a United Nations of microbiological societies, which can lead to some interesting political alliances when the society meets every three years. The IUMS is subdivided into three congresses: bacteriology, virology, and mycology. The virology congress is the most active.

The IUMS and many of the national societies are associated with culture collections of microbial strains and materials. Culture collections are physical repositories of microbes (bacteria, molds etc) and their derivatives. Culture collections contain microbial materials, deposited by scientists, which are associated with the scientists' publications and patent applications, or are used in teaching or for other purposes. The governing organization here is the World Federation of Culture Collections (WFCC). The WFCC in turn is associated with a major international biological organization, the International Union of Biological Sciences, as well as with the IUMS. The WFCC is concerned with the collection, authentication, maintenance and distribution of cultures of microorganisms. It provides liaisons, sets up information networks, organizes workshops and conferences, and publishes newsletters and other works.

The Federation watches over more than 450 individual culture collections from 62 countries, which vary enormously in size and capabilities. Here in the United States, for example, we have the American Type Culture Collection (ATCC) and the collection at the U.S. Department of Agriculture's Northern Regional Research Laboratory (NRRL), both of which we will hear about later in the symposium, and a number of other collections. The size and the range of these collections vary, and many of them have been struggling constantly, from the time they were founded, to get the funding they need to keep going.

The WFCC has an extremely ambitious agenda. According to its website, it has taken on a variety of challenges such as standardization, the financial sustainability of collections, microbiology education and outreach, and intellectual property issues. Another project is the World Data Center for Microorganisms (WDCM) developed through the activities of Professor Skerman, University of Queensland, Australia, and his colleagues during the 1960's. WCDM pioneered the development of an international database on culture resources worldwide. This data resource is now maintained at the National Institute of Genetics in Japan and its records contain information on the organization, management, services and scientific interests of the collections, as well as linked records containing lists of all species held.

Concerns such as adequate staffing and obtaining funding are a constant fact of life for culture collections. Many of the smaller culture collections run by individual scientists at universities are permanently shut down when the major investigators retire because of the expense of keeping the collections active.

As I mentioned before, I have been a cheerleader for genomics in general and, more recently, for fungal genomics in particular. My experiences in working on the steering committees of several fungal genome projects make me think that perhaps we should seek the common ground for the microbial commons in lessons learned from genome projects rather than in some of the treaties that have come out of international biodiversity concerns.

Microbiology and genomics have become very strong bedfellows. In fact, microbiology has played a central role in the development of genetics and molecular biology since the middle of the twentieth century. To offer a quick overview on the history of genetics, let me remind you that it was a microbiologist who first showed that genes were made of deoxyribonucleic acid or DNA. The so-called transforming principle was discovered in bacteria, and bacterial genetics soon superseded all the work that had been done with Morgan's Drosophila flies, Mendel's peas, and McClintock's corn plants. That microbiologist was Oswald Avery, who did this wonderful research at the Rockefeller Institute during World War II. Once it was understood that DNA was the transforming principle, the discovery of the structure of DNA became a holy grail of biology. It was elucidated by Watson and Crick, who published the now-classic double-helix model for DNA in 1953.

The breakthroughs in genetics, many of which were nurtured in a microbial womb, continued to occur, but I will skip over most of them and go right to work that happened in the 1970s—the breakthroughs in recombinant DNA research and genetic engineering. Genetic engineering very rapidly led to many forms of commercialization. I have been told that there was once a meeting at which Paul Berg presented some of the early work on gene splicing, and some young fresh-faced person in the back of the room said, "Gee, there are practical things you could do with that." In response, Berg deadpanned, "It never crossed my mind." Unsurprisingly, basic research in gene splicing led rapidly to the first recombinant product, human insulin (marketed by Eli Lilly) after which microbiology and genetics became increasingly supported by venture capital.

Another important enabling technology is the polymerase chain reaction (PCR), which some people at the Cold Spring Harbor Laboratory have described as the genetic equivalent of a printing press. The key work was done by Kary Mullis, who worked at a biotechnology company, Cetus Corporation. The company no longer exists, but PCR certainly does, and Mullis would later share a Nobel Prize for his work.

PCR was coming of age at the same time that DNA sequencing became possible. When I was a graduate student, if you wanted to figure out anything about the content of a gene, you had to do it laboriously, working back from the amino acid sequence of proteins. At the time, many people did not think it was possible to sequence something as simple as the DNA molecule, which has only four different kinds of nucleotides. However, research by Sanger, Maxam, and Gilbert made such sequencing possible, originally using laborious radioactively labeled sequencing gels to determine the different bases in a large DNA polymer.

The U.S. government generously supported the human genome project from its beginning, and the first draft of the genome was finished faster and with more detail than anyone had expected. There were lots of surprises and they continue to this day. The field has become "big science." When I was a graduate student, it was rare to see a biology paper with more than four authors, but now papers routinely appear with dozens of authors, as those of us who have been involved in genome projects well know.

Simultaneously, computers were adapted to the study of DNA. Among the high points in what is called "bioinformatics" were the founding of GenBank at the National Center for Biotechnology Information and the development of the BLAST algorithm. Several international groups have developed efficient methods for data storage and sharing that have transformed all of biology.

Now let me talk about the 21st century. Here is our charge: We are supposed to focus on improving the management of both the physical materials and the digital scientific literature and databases in microbiology. The reason for going through all this history is to show you that the field has more than a century of internationally diverse professional societies, traditions, mores, and ways of going about its work. It is not going to be possible to create a uniform new centrality out of this Tower of Babel.

Perhaps, however, we can do some civilizing around the corners. Remember that the physical materials we are talking about are largely those that are held in the culture collections. These living cultures consist of taxonomic type strains, the model strains used by geneticists, patent deposits, as well as a lot of derivative materials that come out of gene cloning and genomics research. There are also cell lines and there are phage splices. The culture collections that maintain these vast resources often have very different standards of quality control. It is also worth noting that the culture collections already are associated with digital resources, which also have varying accessibility and quality. There are some big issues concerning best practices here. Moreover, in recent years, following the anthrax scare, the U.S. federal government has been creating hurdles related to biosecurity. Finally, material transfer agreements raise many new issues within the research community. There was a time when you sent your money in to the American Type Culture Collection, and it sent you a culture. Or I could write to Cletus Kurtzman at the NRRL, and he sent me a culture at no cost. Those times are over.

A common estimate of how many microbes are known to science suggests that it is about one percent of all microbes on the planet. The existing estimate comes from various sources, but the important take-home lesson is that we are pretty sure we have not identified most microbes. As we go forward, we should be careful that we do not make it too hard to identify, study, and preserve the hitherto undiscovered microbes. Unfortunately, however, evolving regulations may be making the research on microbial systematics more difficult, not less. Some of the negative developments have arisen from various biodiversity treaties that focus on plant resources and from making sure that so-called less developed countries have access to the biological materials found in their countries. Such concerns have led to various restrictions on the collection of materials and data.

This raises an issue that it is sometimes hard to describe to people who are not scientists. What motivates scientists? Many of us work extremely hard. My husband comes from a Wall Street tradition and he does not get it. He says, "Why do you work so hard? You are not getting paid for this. You are not getting paid for that." The reward structure of the scientific culture is hard to describe to outsiders, but a microbiologist friend of mine, Simon Silver, once over simplified it this way. He said that there are three kinds of people in the world. "There are money people, there are power people, and there are fame people." Simon added that most scientists, including microbiologists, "are fame people." In general, scientists gain their acclaim through publication in the peer reviewed literature.

Here are some more statistics. Between 1986 and 2006, journal expenditures of North American research libraries increased by a staggering 321 percent as libraries

expanded access to journals by licensing bundles of journals from different publishers.[3] At the same time, the average journal cost increased by 180 percent, while the U.S. Consumer Price Index rose by 84 percent. In other words, journal costs have outstripped inflation by a factor of more than two. At Cornell University, the faculty senate passed a resolution in 2003 describing the cost of journals as "literally unbearable," "unsustainable," and "threatening to undermine core academic values." The Cornell faculty pointed to the name of one commercial publisher in particular as a driver of these huge price increases.

With the development of the software that is used to bring the cost of publishing down, publishers no longer have to set type or do much else. Basically, the author does all of the preparation work for text and figures. The reviews are done free. Manuscripts are transmitted over Internet protocols. Yet commercial publishers reap the profits and charge these high subscription rates.

Simultaneously, as the online digital revolution has progressed, publishers have tried to adapt what has been done in print to the new environment. Free, open access publications add yet another challenge to the financial stability of scientific publishing. Scientific societies are fighting back in various ways. For instance, a number of microbiology journals are available free online six or fewer months after the release of the paper. The American Society for Microbiology has been a leader in this policy. Interestingly, although review articles are not as time-sensitive, you have to wait a year to get access to them.

The psychological impact of genomics is often expressed in metaphors. One of the few positive metaphors used is "wealth" often used in conjunction with the term "data mining." More commonly, scientists describe being swept up in a "tsunami" of data. Other disaster metaphors include "explosion," "avalanche," "deluge" and "flood" of data. In summary, not only are researchers having trouble dealing with the legal, social, and economic aspects of the new biology, we also are having difficulty dealing with it intellectually. New technological systems regularly outpace our capacity to adapt.

I will conclude by mentioning a few major issues and organizations that we will be speaking about more in detail at this symposium. The Convention on Biological Diversity's Article 15 on Access to Genetic Resources gives states sovereign rights over their natural resources, and provides that "the authority to determine access to genetic resources rests with the national governments and is subject to national legislation."

Many unresolved issues arise from intellectual property treaties concluded at the World Intellectual Property Organization. Such agreements typically are developed to protect creative artists, such as the Beatles, not for microbiologists who need to access and reuse a broad range of digital and material inputs in order to conduct their research. Further, the World Trade Organization and its conventions are not developed by or for scientists. As a practicing scientist, I can attest to the fact that scientists often ignore the many newly imposed rules in exchanging microbial cultures across institutions or borders. If you can call a contact and get a necessary microbial culture for your research, you just do it informally and forget about the material transfer agreement. I have seen a statistic that some 60 percent of microbial cultures still are transferred this way.

Looking ahead at designing a microbial commons, I think our best hope lies in the fact that the genome sciences have done such a good job in developing an "-omics"

[3] See *http://library.uic.edu/home/services/publishing-and-scholarly-communication/the-cost-of-journals*

commons. GenBank has no restrictions on the use of its data. Furthermore, GenBank has cooperative agreements for the exchange of genomic data with groups in Europe and Japan, and I would not be surprised if we see open DNA databanks established in China and Korea in the near future. The European Bioinformatics Institute has great open source information. The Genomic Standards Consortium is working on doing research community outreach and developing common vocabularies. The notion of a common vocabulary is often referred to as a genomic Rosetta Stone. Such semantic interoperability will facilitate meaningful access to the information. Many geneticists are good at coming up with catchy names, and they consider the human genome as analogous to a periodic table of our genes.

Finally, to return to where I started, infectious diseases continue to emerge, and while it is bad news for humankind, these diseases make continued microbiological research essential—and fundable. Every time there is a major new human health problem it is more likely for governments to spend money on the relevant research. During my adulthood I have seen a number of diseases such as herpes and AIDS go from obscurity to prominence. Other emerging diseases such as SARS also have the potential to cause enormous harm.

Trying to end on a positive note, let me reemphasize that it is good that microbiology has entered the popular consciousness. Awareness of the need for a microbial commons will help ensure public support for our efforts. Nevertheless, we have our work cut out for us.

Reference: Hardin, Garrett. 1968, The Tragedy of the Commons, *Science*, 162:1243-1248

3. Breaking Anti-Commons Constraints on Global Scientific Research: Some New Moves in "Legal Jujitsu" – Paul A. David[4]

Stanford University & All Souls College, Oxford
& United Nations University-MERIT, Maastricht

I expect that most of those who are attending this symposium will have heard something about "the anti-commons" and that they understand that it is not a good thing. Perhaps they have also run across the article in which Michael Heller and Rebecca Eisenberg[5] argued that the monopoly rights granted to inventors under the patent system of the United States and many countries—ostensibly for the encouragement of inventive activity, or at least public disclosure of the latter's results—actually might have the perverse effect of inhibiting both invention and innovation.

Surely there will be others present who recall the message of Garrett Hardin's 1968 article "The Tragedy of the Commons,"[6] in which the it was the opposite of the "anti-commons" that figured as a decidedly bad thing. At least, the latter is the view one is left to draw from Hardin's account of historical experience with common-use arrangements, such as the common grazing rights on lands held by agrarian communes in medieval Europe that led inexorably to the destruction of valuable and exhaustible resources by "over-grazing." So, we are confronted with two disconcerting if not necessarily contradictory views about the effects of private property rights in valuable economic resources: the *absence* of the right to exclude others from trespass, as in the case of "the tragically over-grazed" village commons, leads to a bad outcome; but the same may be said about the *presence* of the patent-holder's right to exclude others from the use of the patented invention. Is either of these propositions valid as a general rule? How can both be true?

This brief presentation is premised on my conviction that is important not only for economists and lawyers, but people with the diverse range of expertise that is represented in this audience to recognize and understand the "anti-commons effect" as a general economic phenomenon, in the same way that we grasp the logic of more familiar argument that valuable resources are best held as private property, because their owners would have strong incentives not to exploit them wastefully. Although land and other tangible physical resources hardly are the same as data and information, and patent rights differ from copyrights and database rights, the economics of both the commons and its dual, the anti-commons are germane will be seen to be directly germane to subjects presently under discussion. Like the microbial commons, many of large scientific and technical databases that have been constructed either for the public domain or made available on an open access basis to qualified users, constitute "research resource commons" or "semi-commons".[7]

---

[4] Presentation slides available at *http://sites.nationalacademies.org/xpedio/idcplg?IdcService=GET_FILE&dDocName=PGA_053729&RevisionSelectionMethod=Latest.*

[5] M. A. Heller and R. S. Eisenberg, "Can patents deter innovation? The anti-commons in biomedical research," *Science*, 280 (1 May, 1998): 698-701.

[6] G. Hardin, "The tragedy of the commons, "Science 162 (1968):1243-1248.

[7] The term "semi-commons" has been employed by several previous presenters in referring to contractually constructed common-use arrangements that occupy a position intermediate between that the public domain

The Problem of the Commons in Theory and History

It will be useful to begin with "the commons," in order to put aside confusions that appear frequently in the economic and legal literature due to widely shared misconceptions about the "tragedy" of common-use exploitation of land, fisheries and other natural resources. That desertification is a tragic consequence of "over-grazing" in many parts of the world is not in doubt. Prolonged unrestricted livestock grazing in arid climates, by sheep in the Patagonian region of Argentina and by goat-herds in northern Chile has been a major contributor to contemporary desertification, just as unrestricted grazing by cattle in the rangelands of southern Texas was a factor in the region's "dust bowl" during the 1930's. Likewise, it is thought that after the 3 century C.E. the multiplying Bedouin sheep-flocks in the Negev combined with the decline of agriculture and the abandonment of the associated irrigation dams and channels to produce a reversion to the desert conditions that are found in the southern region of modern Israel.

But the exhaustion of village lands in medieval Europe due to over-grazing of their common fields, the illustration that Garrett Hardin provided as a parable supporting his argument that efficient natural resource use requires private property rights and market pricing of resource use, is just a fantasy. Its repetition, unfortunately, has served only to obscure an important lesson that the actual historical experience holds in regard to the management of common-use resources.

Europe in the Middle Ages never knew public domain "commons" of the sort Hardin imagined, and where once-cultivated lands were allowed to tumble down to grass and village settlements eventually were "lost." This was symptomatic of the retreat of agriculture from the marginal, semi-arid regions into which population had expanded in the 13th and early 14th centuries—before the mortality crisis of the Black Death. In reality, the feudalized regions of western Europe knew no territories that formally were in the public domain; "null terre sans seigneur" (no lands without a lord) was the canonical expression of the situation under which control and exploitation rights over physical property—arable, pasture, woodlands and sub-surface mineral deposits—had come to be held by one or another king and their respective vassals. Thus, the areas of agrarian settlement where "common-rights" were established were not the wilderness or "waste," but lay within the jurisdiction and governance of particular communities that regulated, inter alia, the number of animals that each of the holders of a tenure in the village could put to graze upon the common stubble-field following the harvest, or on the common meadow-lands close to the village. What is important to emphasize is that seasonally designated common-fields, and their meadow-lands were not open for all to use, not even for the villain tenants to exploit at will by bringing in additional hands from other villages, say to glean the fallen grains of wheat following the harvest and before cattle were turned into the fields to graze upon the stubble.[8]

The managed commons of the village communes in medieval (and early modern) Europe therefore exemplify the "club goods" form of resource commons—intermediate between the public domain and the regime of private property. The Commons in tangible

---

and the private domain. As will be seen in the following pages, I prefer the description of such arrangements as "club commons"

[8] The presentation slides illustrate the detailed nature of the limitations placed upon the exercise of these "common rights," in the case of Salford Manor, in Oxfordshire at the end of the 16th century.

exhaustible resources is not a defunct institution, for collective ownership of exhaustible resources did not, and does not translate automatically into a chaotic struggle for possession among neighbors, nor does it result in the egalitarian distribution of use-rights. Even in western Europe today, such arrangements based upon *de jure* common use rights (*res communas*) dating from the Middle Ages have survived in the Swiss Alps and Northern Italy—e.g., the Magnifica Comunità di Fiemme, in the valley of Aviso (Trento) —where they still govern the use of tens of thousands of hectares of alpine forests, pasture and meadow land.[9]

I have undertaken this historical digression as a means of putting aside a number of the overly simplistic and misleading preconceptions that have developed around the popular story of the tragedy of the commons.[10] Inasmuch as the "microbial commons" involve the curation and sharing of tangible research resources, the point that has been emphasized regarding the importance of user-based governance of natural resource commons is immediately germane.

But, because we are here concerned also with *digital commons* for sharing scientific information and data—some of it representing "metadata" that directly complements the organic material of the microbial commons—a different point should be emphasized: unlike land, the productive value of data is not diminished by "over-use" *per se*. Data and information are more akin to fire than to coal: one gains light from them without their being consumed in the process. This does not, however, warrant the conclusion that there is no need to restrict access to a scientific data and information commons, because it will remain "un-depleted" by repeated, intensive utilization. While the latter is true, governance arrangements cannot be discarded if the quality of data and the reliability of information is to be maintained. Data can be degraded by being mixed with other data that are inaccurate, and screening of contributed materials to minimize that form of "contamination," standard formats for data, and accompanying minimum metadata requirements need to be enforced in other to insure the widest extent of usability of the common's resources.

The imposition of management procedures with restrictions on contributed resources, rather than limitations on access to prevent "congestion" or "overuse" is the primary economic rationale for the "club goods" form in the case of scientific resource commons. For, in this case, and particularly that of digital research resources, the value of the commons actually improves with more intensive exploitation by members of an extensive community of expert users. The removal of recording and copying errors, and the annotation of data-files and reports that link the contents to the corpus of published research findings, and to related datasets with which they may be federated for further analysis, are semi-automatic consequences of the symbiotic relationship between a research community and the resource commons that it builds and exploits. Thus, we are here in a world very different from the "tragic commons" conjured up by Garrett Hardin,

[9] See also, David P. Mitigating'anticommons' harms to research in science and technology. UNU-MERIT Working Paper No. 2011-001, Analysis and Debate of Intellectual Property Issues, Forthcoming 2010

[10] These, unfortunately, continue to figure in leading economists' textbook expositions of the "common-pool problems", as, for example this one in Suzanne Scotchmer's *Innovation and Incentives*, Cambridge, MA: MIT Press, 2004 (p. 88): "The *anti-commons* is a play on words and refers to the 'tragedy of the commons' which is taught in freshman economics. In the tragedy of the commons peasants in early modern Britain overgrazed shared pastures ('the commons') because the absence of private property eliminated incentives to conserve."

and repeatedly invoked by advocates of private property rights as the necessary condition for efficient resource use.

The threat posed by the anti-commons, however, is quite another matter. In an earlier presentation to this symposium, Paul Gilna[11] noted that biological communities and microbial biology communities are entering Leroy Hood's second phase of scientific breakthroughs, where having worked on the problem of how to generate and capture new data, they now have to think of what to do with the data. What are the modes of analysis we need to handle data at this enormous scale and volume, and to do it with links connecting a distributed community? In this context, one also must take into account the fact that a part portion of the interested researchers at present, and probably for some time into the future will not have the training to write their own analytical algorithms or even to use the open-source algorithms that are already available to work with the data. There are consequently advantages of scale in use, for very large groups are more likely to draw in supplementary resources and mobilize the necessary expertise of a few members that can in the development of analysis tools.

Minna Allarakhia[12] did point out in her presentation that some open-access communities are beginning to provide complementary access not only to data and to archived publications, but also to analysis tools, search tools, and other types of research tools. The premise was that, without such analytical tools, users will not be able to benefit sufficiently from the enormous data investments that are being made.

Here is the point: If you go down the road towards user-friendly analysis tools, you will enter the part of the software world where commercial software vendors have been operating. This is off-the-shelf software. By contrast, the typical mode for large scientific work groups has been to assemble their own software—put together a lot of pieces that they already have experience with, make something that works quickly, and keep on going. Now, people in the United States and in other parts of the world where patenting of software is possible have patented many of the subroutines and algorithms, which are then embedded in black boxes—machines where you feed the data in, press a button, and something comes out.

This has been going on for a long time. It happened in physics with mass spectrographic analysis. Fast algorithms make it possible to do things in real time, but only for the people who have access to them. There have been ongoing discussions even in the journals as to whether people should be forced to publish these algorithms, but those who developed the algorithms resist. They say, "No, we are working on it. We are trying to document it. We are going to upgrade it. It could take a year. And then we are going to release it for sale."

So researchers are likely to bump up against patent thickets, where some key algorithms are not freely available. It is possible to go around that. A group of open-source people might make it a project to write an open-source version, for instance, or contact people familiar with that type of software and ask them to work on it. The result, however, will not necessarily be user-friendly. In the case of UNIX applications, this can be a useful step that will yield new and more efficient customized code. There are two problems, though. First of all, many of the users in the commons will need to have somebody repackage these custom-made programs to make them usable. That imposes an extra cost. Secondly, there will be little standardization in this approach, and

[11] See Chapter 17 within this publication.
[12] See Chapter 20 within this publication.

standardization of analysis techniques is one of the ways in which research communities increase replicability and transparency of their procedures, both of which are desirable properties of research tools.

With packaged software, people do not have to go through a long description of the published algorithms. They get a standard algorithm that is widely used—it is in the library someplace, or it is in a commercial package, and the code is "stabilized" so that attempts at replication are not frustrated by ambiguities regarding which particular "release" or customized version of the algorithm had been used in obtaining the results reported in publications. This is one of the things that eases access for the people—usually not those at the cutting edge of the new field, but those people who are following after—who are going to do the normal science in the field. The process of "black boxing" and commercialization reduces the cost of producing user-friendly techniques to the less skilled, but the impulse of the holder of a patent filed on the original prototype to share the profits from the developed commercial version(s) may not only inhibit that development, but create an obstacle to those who would simply implement the basic concept of the (patented) research tool for their own work.

The point here is that the device of a *contractually constructed commons*—the phrase introduced in the seminal 2003 publication by Reichman and Uhlir[13]—addresses the reality of the research world in which prior work has given rise to patented procedures, or other IPR and sui generis forms of restricted access (database rights, in the EU) that encumber subsequent application and extension. This source of impediments to the cumulative, incremental advancement of scientific knowledge has been a prominent concern in the present discussion, largely because our focus has been upon on "the data tsunami" —the massive wave of newly available data, much of which is not copyrightable and not patentable. But there are also new fields with new analysis tools that are emerging, and as people follow the science, eventually they will wander into some part of this terrain and they will find that others have staked out property rights there before them. During the remainder of this presentation, therefore, I will emphasize the case for the creation of digital resource commons as an *ex post* fix for those inherited problems.

There are many fields where researchers who are trying to do collaborative work are tripping over the fact that the downside of building on the shoulders of giants is that sometimes you are building on the shoulders of pygmies.[14] In this case, a researcher may find not a step created by others on which it is simple to build, but an obstacle in the path that requires paying a fee if it is to be used, or a cost in time and effort if one is trying to work around it. Leading research groups in some fields, which are to say academic researchers in many instances, have patented many results including research tools among them, and thereby have left in their wake obstacles for researchers in the same fields, and also for those seeking to transfer established techniques to new fields of investigation. Frontier researchers will prefer to proceed as far as possible by employing

[13] J. H. Reichman and P. F. Uhlir, "A contractually reconstructed research commons for scientific data in a highly protectionist intellectual property environment," *Law and Contemporary Problems* 66 (2003): 315ff. Available at *http://www.law.duke.edu/shell/cite.pl?66+Law+&+Contemp.+Probs.+315+(WinterSpring+2003)*

[14] There is a double meaning in this allusion to the now widely repeated phrase, uttered by Newton, in the course of a priority dispute with Hooke. Hooke was a very short man, and Newton, rather nastily quipped: "If I have seen farther, it is by standing on the shoulders of giants" (and not persons of little stature –both physical and intellectual).

and adapting where necessary tools that have become standard, well-known, and documented, and such tools in the future are more and more likely to come with IPR restrictions.

So what is the anti-commons problem? The anti-commons problem is like an onion—a simple onion, in which there are three discernable layers. Layer 1 is search costs, the costs of discovering whether tools described in the research literature are privately appropriated and to whom the property rights were assigned, whether as patents, copyrighted computer code, or database rights. If you have distributed inputs into a process, the inputs are not all in one place, they have been produced by different people, and it takes you a while to look at them. If this is an area covered by patents, patent searches can be very lengthy and very costly processes.

Layer 2 is transaction costs. These arise when one has identified the owner of the intellectual property (IP) and seeks a license or an agreement to transfer materials. These are different from the search costs. This is negotiation, and the key attribute is that negotiations take time. Even if your university has a lawyer whom you can use, so will the other side. There will be many lawyers, and they will have many meetings and many expenses. This is a long process. I experienced some of this when I was at Oxford University. Even when people were not holding out, there were delays. Everybody wanted to see what the contracts were going to be, so negotiations stretched out over 18 months for a demonstration project that had something like a 3-year budget. Ultimately, they went ahead without anything, but the university lawyers were complaining and saying, "No, you cannot do that." It gets even worse when more fundamental research is involved because it is impossible to know what is going to come out of it.

At Oxford University I encountered the office of Research Services, which was staffed by very competent solicitors, whose main responsibility was to do "due diligence," protecting the corporate interest of the institution from the harms to which it could be exposed by embarking on faculty initiated research projects. The lawyers took it upon themselves to worry about liability for accidents involving new and dangerously toxic materials, entanglements in the liabilities of other institutions with which the university had joined in collaborative agreements, and the possibility of suits by third parties that claimed to have been injured (commercially or otherwise) by following the advice based upon a research publication. But in addition to the hazards of untoward outcomes, there also were the risks of failing to fully exploit opportunities that might arise from successful research. Not getting the largest share possible of the income derived from commercial exploitation of research findings is no less a "risk," when viewed from the window of the Research Services offer, as failing to protect the university from a liability law suit, or a charge of patent infringement. The reality is that such failures do not simply represent the loss of a potential benefit; they carry penalties. Not obtaining strong patent rights on a discovery or invention that could turn out to be important, and moreover a major source of revenue had it been property privatized could expose the institution's leaders to the kind of response that Oxford's requests for increased overhead research funding were known to have elicited on some occasions from officials in Her Majesty's Treasury: "Had you only thought to patent penicillin, you wouldn't need to be here now, would you?"

Due diligence therefore suggested that in negotiations about collaborative research agreements it was better to seek the strongest possible IPR protections for the university's interests, or to push all the conceivable liability risks (or the costs of insuring against them) onto other parties, even if this strategy would have the result of blocking

the project in question from going forward. Considering that was no end to the number of research proposals that the faculty seems to be able to bring forward, the best (diligently cautious) stance was to be wary of those that were surrounded with greater uncertainties. The problem with this, however, is that uncertainty is in a sense the hallmark of novel, more interesting research proposals—those that typically distinguish academic science and engineering from the projects to which the bulk of corporate R&D funding is committed.

The uncertainties about the nature of the products and processes of the proposed research project, taken in conjunction with the professional incentives of those charged with performing "due diligence" and their inability to calculate the countervailing value of the losses entailed in not doing the research, tend to promote behaviors that reflect extreme risk aversion. Fears of failing to secure as large as possible gains from intellectual property rights on university conducted research appears to be a major source of protracted negotiations for collaborate agreements. This is observed not only where inter-institutional and collaborations, and university–business projects are involved, but also in cases of projects proposing grant or contract research to be conducted in different departments and schools within the same institution. In other words, the representatives of the university's corporate interests, as distinct from those of their faculty researchers, are pre-disposed to advocate and adopt a tough bargaining stance, trying to get the other collaborating party (or parties) to yield the greater part of any potential income that is envisaged to result from the research, and to bear the greater part of the potential liabilities, or the costs of insuring against them. Moreover, should that appear to be infeasible, the conscientious legal counsel will not be hesitant to recommend that the project should not be undertaken.

Understandably, that stance tends to come as an unwelcome surprise to uninitiated prospective corporate "partners" who entered the negotiations with the expectation that "the university" would be seeking a way to satisfy the interests of the faculty counterparts of their own research group, just as they themselves were under instructions from the vice president of research to find a way to "make the project happen." Disappointment of those expectations would at least account for the shocked and disparaging terms in which research directors of large, R&D-intensive U.S. companies have expressed their views about the experience of negotiating with universities over the IP rights to joint R&D ventures, such as those reported on the basis of a survey of 60 vice presidents of research that was carried out by Hertzfeld, Link and Vonortas. The consensus view was that trying to deal with universities over IP matters was much more difficult, and more frequently unsuccessful than negotiating collaborative research agreements with other business companies.[15]

[15] See the 2003 survey results reported by H. R. Hertzfeld, A. N. Link, and N. S. Vonortas, "Intellectual Property Protection Mechanisms in Research Partnerships', *Research Policy*, 35 (June-July), 2006 [Special Issue on Property and the Pursuit of Knowledge: IPR Issues Affecting Scientific Research, P. A. David and B.H. Hall, eds.]. See also P. A. David, "Innovation and Europe's Universities: Second Thoughts about Embracing the Bayh-Dole Regime," in *Perspectives on Innovation*, ed. F. Malerba and S. Brusoni, Cambridge U.P., 2007: pp. 251-278. Esp. Table 1 and accompanying text discussion.

The Core of the Anti-commons: "Multiple Marginalization"

The foregoing difficulties are further compounded by the problems encountered when one reaches the third layer, that being innermost core of the anti-commons phenomenon. It involves the condition referred to by economists as "multiple marginalization" which copyright lawyers will be familiar with as "royalty stacking." It arises when there many parties holding exclusion rights over the use of research tools, each of them asking to be paid what to them appears no more than a reasonable royalty for the license to use the patented research tool. To assemble a collection of photographs for a book, for example, may entail paying for the copyright license on every image, and when these are separately owned the copyright holders individually will ask to receive what appears to be a modestly small percentage of the revenue from the prospective sales of the book. But it mounts up: a 0.5 percent royalty charge levied for each license to reproduce 50 different color photographic plates will take 25 of the sales revenues from the print run of the book.

The analogy to the art book's photographs is the collection of different research "tools" that will be used in carrying out a proposed scientific or research project, many, if not all of them under patents or copyrights that are held by distinct parties. Even when there are no strategic holdouts in the negotiations over licenses, and even though the negotiations can be rapidly concluded, when the number of items is large what seem to be very reasonable requests for very low IP royalties to be paid on commercial sales of downstream research products, the total bill for royalties—which none of the distributed IP owners has considered—can become an obstacle to going forward with the project.

In a survey conducted among academic biomedical researchers at U.S. universities and research institutes, John Walsh, Ashish Arora, and Wesley Cohen asked whether the respondents had abandoned a research project because the costs of obtaining licenses on patented research tools was un-supportably large.[16] They found so few such instances of blocked or abandoned research projects that they report the victims of the anticommons (in the field biomedical research) to be "as rare as white tigers." While on the surface this seemed to be very good news, when considered more closely it took on a different cast. In the first place, it is unlikely that a planned project would actually be terminated once under way, whereas a preliminary investigation of the patent status of indicated tool sets that revealed a multiple marginalization problem would led to modifications in the research design, or where that was infeasible, to substituting a different project altogether—before the one initially contemplated got under way. Secondly, a follow-up survey questions asked what the research initiators might have done to avoid the impediments created by the requirement to required licensing access to numerous tools-sets and data-sets. With surprising frequency the answer was: "We just don't pay any attention to the patents." This disclosure stirred some concerns about the consequent unknown extent of exposure of these scientists' universities to patent future infringement suits.[17]

---

[16] J. P. Walsh, A. Arora and W. M. Cohen, "Research Tool Patenting and Licensing and Biomedical Innovation," *The Operation and Effects of the Patent System,* Report of the STEP Board of the National Academy of Sciences, National Research Council, Washington, D.C. National Academies Press, December 2003.

[17] Expressions of worry on that score from U.S. research university administrators increased noticeably following a 2002 ruling by the U.S. Federal Court of Appeals for the 9th Circuit in a patent infringement suit. The judgment for the plaintiff in *Madey vs. Duke University* greatly narrowed the scope of the so-

The problem with distributed claims is simply that the cost of using research tools that have been protected by IPR is that the IP owners are acting independently, rather than considering the consequences the collective impact of all of their independent actions. Quite naturally, their position is "Why should I be the one to desist or to charge nothing for the use of my patent, or database, if everybody else is going to demand royalties to grant a license? And they figure that perhaps if they hold out long enough, some of the others will reduce their charges, and they will not have to.

This has serious implications for federated databases because there are various ways that patented technologies can affect the access of materials in a database. Perhaps the data are locked up by patented encryption software, for example, or perhaps the search tools are patented. The patent holders can charge you to remove your own data from the database.

Graham Cameron at the European Bioinformatics Institute (EBI), in his contribution to 2002 EU working party report on IPR issues affecting Internet-based collaborative research, remarked that were one to try to replicate the EBI's federated databases in the then-existing environment, it would not be possible.[18] The Institute could not raise enough money to buy off the people and it would consume way too much time in the negotiations. That may be an extreme example, but Cameron was making a point. In building some research infrastructures, we are out in front of the process of those people who are privatizing parts of the public domain, but, we still have to work with the requirements of science, some parts of which are now impeded.

A little microeconomics analysis indicates that is to be expected under those conditions. When database rights are distributed among commercial owners each of whom independently set prices on the contents in order to maximize the owner's individual profits, symmetrical owners will set the same charges for access rights and the greater the number of databases that a project must consult, the higher will be the stack of access charges the project will face. Facing this elevated cost for the search activity that is an input into the planned research project, either the extent of the search will be restricted, or an alternative project that is fewer searches intensive will be substituted. You will either access less or you will substitute at the margin. You will not do extensive searches. If you need to do this kind of search, you are not going to either reverse engineer or write your own tools for ones that are very important to your work. To do that you would have to hope that you can use something that will not infringe, or perhaps you

---

called "research exemption" from patent enforcement that had been widely supposed to exist, left American research universities substantially greater risk from infringement suits that previously had been supposed. In ruling in the case of *Madey v. Duke University*, 307 F.3d 1351, 1362 (Fed. Cir. 2002), the court did not completely reject the research exemption defense, but left only a "very narrow and strictly limited experimental use defense." A patented process or device might be used without permission (license) for "amusement, to satisfy idle curiosity, or for strictly philosophical inquiry." The court also precluded the defense where, regardless of profit motive, the research was done "in furtherance of the alleged infringer's legitimate business." In the case of a research university like Duke University, the court held that the alleged use was in furtherance of its legitimate business, and thus the defense was inapplicable. The U.S. Supreme Court subsequently refused to hear Duke University's appeal, thereby allowing the Appellate Court running to stand.

[18] See *IPR Aspects of Internet Collaborations*, EC/Community Research Working Paper, EUR 19456, April 2001. Not only were most of the European genetic and proteomic and ancillary demographic databases subject to copyright and database right restrictions, or protected by clickwrap licenses granting pass-through rights, but technical compatibilities among the various digital rights management (DRM) systems that had been deployed would frustrated the "deep linking" of database contents that was required for a searchable federated database.

might just keep quiet about what you are doing. The time that you spend going around the databases, or figuring out how to build new analysis tools, or use other instruments that you build in the lab would be enormous. If you cannot perfectly substitute for the database search in the end, then the research product is going to be degraded.

In these cases, exploratory science will be most affected because you do not know how to limit the discovery space. Commercial firms are less affected because they are looking for certain targets. For example, if a pharmaceutical company is trying to produce a particular drug, it may not need to use the epidemiological data or know about protein folding. It can just key in on the molecule that it is interested in to see if it can figure out how to build the key that goes in that particular lock. For that purpose they are willing to pay the $100,000 flat access fee just to get into certain databases.

The fact that exploratory science will be most affected by this reinforces what people from the sciences have been saying—that if these federated databases cannot be put together, then a lot of the potential is not going to be fulfilled. The outcome is actually worse than if there were a monopoly of all the databases because the monopolists would be aware that if each of them set prices to maximize the revenue just from the usage fees from that database, it would reduce the number of people who will ever pay to use any given database, and unless the data in that database are critical, the total revenues would likely decline.

Responses to the Anti-Commons Problem

What is to be done? Preventing distributed IRP protections being placed on materials that would form complementary sets of research inputs is perhaps the most straight-forward line of attack on the core aspect of the anti-commons. The young field of genomic research provides an exemplar of preventive action that is feasible if people see the problem coming and can act swiftly in concert to avert its materialization: the International Haplotype Map (Hap Map) Project, which was put together by the National Human Genome Research Institute (see International Haplotype Mapping Project in 2002 (see *http://www.genome.gov/10001688*). This was the result of a coalition of publicly funded researchers and some commercial firms, all of whom wished to avoid having lots of fragmentary gene sequences protected by patents or copyrights that were held by different research institutions—because that would greatly raise the costs of working with that data. To reduce the data use costs of their research, they first reduced the number of single-nucleotide polymorphisms (SNPs) that would be needed in order to examine an entire genome for association with a phenotype. The so-called Hap Map project then followed the precedents established by the Human Genome Project, in rejecting protection of the data under copyright or database rights and establishing a policy requiring participants to release individual genotype data to all the project members as soon as it was identified.

This is a special case of *legal jujitsu*, where a "copy-left" strategy has been mutually imposed on database users by an enforceable contract in the absence of IPR ownership. In essence, "copy-left" says: I have something under copyright. I am going to give you a license which makes you not exploit this, which makes you share it on a share-and-share-alike basis. This is the sort of logic used in the contractually constructed commons.

Let us now return to the question of what happens if you have research fields where you cannot start afresh. This is the state of affairs for which the devise of a

contractually constructed research resource commons originally was intended. The core idea of the neuroscience commons project, initiated by Science Commons (*http://www.sciencecommons.org*) under the aegis of Creative Commons, was to figure out some way to enable researchers to escape from the patent thickets in which their work had become entangled. Each of the researchers held a piece of the solution, and they found they needed to work together. The negotiations that were undertaken to create a way in which they would each pay each other for the set of licenses they needed eventually led them to ask whether it would not be simpler to put their IPR into a common pool, from which the members could freely draw the items they required. [19]

Some people who are proponents of market solutions for market problems ask, "Why won't the market respond by having private intermediating organizations emerge and profit by providing a market solution for science's anti-commons problem?" This was the idea behind the Collections Society proposal. The goal was to reduce the costs of searches and transactions in the same way that other organizations have done for copyright in music and other types of content. The idea is that you make the IPR less costly and that will then encourage research production by inducing more inventions. The Collections Society would have an incentive to write contractual provisions, such as grant backs, in order to induce non-cooperating owners to share the use. This would create incentives to put content into the Collection Society.

It sounds very good when you first hear it, but there are lots of reasons to be skeptical. The main problem is that arguments by analogy in this area are really dangerous. Intellectual property is not the same thing everywhere. Authors typically want their works to be widely distributed, but inventors and researchers creating databases for their own research uses often do not seek a similar kind of wide distribution. Copyrights in songs, text, and even images are more likely to be close substitutes than is the case with patents and scientific data.

So, what is the response to this? It is this: Inside the intellectual property domain you can try to create a space that emulates the public domain by getting people to volunteer to put their patentable or otherwise protected assets into it. In exchange they can benefit by being in collaborations with other people whose patented material they want to use. There are also other sorts of incentives that may appear if this becomes regarded as a good thing. There are preemptive benefits, for example. A researcher might put something into that space at an early stage in order to have some control over how it gets used later.

There are a number of different ways in which a commons could be established. One important thing to keep in mind when you are designing something like this is that there are capabilities for abuse. The argument is that when you put a lot of resources into a club and it is not open for everybody, that can be a restriction of competition, and so competition regulators may want to look very closely at that. The defense against that is that this is an efficient patent pool, not an abusive patent pool. An efficient patent pool is one that is constructed out of elements that are complements in some desired process (here research production), because it is their complementarities that give rise to adverse

[19] Science Common's Neurocommons Project (*http://neuroscience.org),* collaboration between Science Commons and the Teranode Corporation, having created a database with open access scientific information and data – content that is digital, online, free of charge, and free of most copyright and licensing restrictions -- is using it also to build a semantic web to permit linkage of the contents and sophisticated search facilities for neuroscience research projects. (Here I should disclose an "interest," in that I have been and remain a member of the scientific board of Science Commons.)

externality effects when ownership is distributed and owners do not take account of the effects upon others of their own price-setting decisions.[20]

The "efficient" scientific resource commons therefore should not bundle together extraneous intellectual property, and the contents should instead to restricted to collections of research tools (including data and information) that will be close complements—in that they already constitute an actual patent "thicket" that could block downstream use and elaboration the research tools, or are expected to be regularly used on conjunction with one another in exploratory data search and analyses. An objective empirical procedure for establishing the likelihood that a collection of patents (or copyrights) is an obstructive "thicket" would be particularly useful in addressing this issue. It is relevant to notice the proposal and practical demonstration by Gregory Clarkson[21] of a method of using network analysis to discover patent thickets and disqualify them as ineligible for efficient pool status. Nevertheless, dual pricing policies by foundations operating research resource commons, potentially would be subject to abuse, and competition among those entities will be quite limited if they are successful in internalizing complementarities among research tools. Therefore, there seems an inescapable conclusion that there would be a need for continuing monitoring and vigorous antitrust supervision of these new institutional arrangements.

Looking Ahead

If you begin to look ahead on the path that would be opened by a coordinated program of commons formation to break the constraints imposed by extensive IPR restrictions on research tools, it appears possible that an desired outcome could be the retrieval from universities a lot of their patented material, much of which never even has a license issued on it and some of it which is used to form blocking patents. When we get further into the development of nanotechnologies, we will have entered the first major research domain where virtually all the fundamental tools will have been patented, many by universities. This will be a very different situation from that of the biotechnology revolution of the early 1970's, when Cohen-Boyer patent on restriction enzyme techniques was licensed on nonexclusive basis at very low rates—$5,000 was the flat rate for the Cohen-Boyer license. In the future, by contrast, the consequence of extensive academic and public institute patenting during the past three decades will mean that many of the necessary tools for continuing advance in the new fields of application are proprietary. Cleaning up after that parade, and thereby opening the way for future scientific advances will be important task, to which the institution of the contractually constructed research resource commons can contribute.

---

[20] The substantial literature has recently developed in economics on the topic of "efficient pools" is directly relevant in this context. See, e.g., J. Lerner and J. Tirole, "Efficient Patent Pools," *NBER Working Paper*, 2002; C. Shapiro, "Navigating the Patent Thicket: Cross Licenses, Patent Pools, and Standard Setting," *Innovation Policy and the Economy*, 1, 2000: pp. 119-150; M. A. Lemly and C. Shapiro, "Patent Hold-up and Royalty-Stacking," *Texas Law Review*, 2007. [Available at: *http://faculty.haas.berkeley.edu/shapiro/stacking*].

[21] See G. Clarkson, "Objective Identification of Patent Thickets," *Harvard Business School Working Paper*, version 3.9, 2004.

4. An Industry Perspective: Development of an MTA Harmonious with a Microbial Research Commons

– Stephen J. McCormack [22]

Exela, LLC, Las Vegas, NV

My early work was done in academia. I was involved in the sequencing and other research associated with the advent of Sanger dideoxy sequencing in the 1980s at University of Massachusetts Medical Center in the laboratory of Michael P. Czech, MD. and then in Rockefeller Institute with Robert G. Roeder, MD. My Ph.D. is in virology, and since receiving it I have intentionally moved into different fields. I spent five years here at Georgetown University, for instance. My first professional job was at the American Type Culture Collection, working for the general counsel, where I was responsible for putting together many of the early formative material transfer agreements (MTAs). The one being used now is very different than the one I worked on there. I have also been involved in the formation of many different companies, including companies focused on diabetes and cancer, and now I work for a firm that deals with medical devices.

In my talk I will not address many of the issues associated with copyright and publication because this is an area that I am not very familiar with, and there are many people here who are far more well-versed in the copyright and the publication issues.

Let us go back to 1997, when many of the MTAs were being developed for a number of collections with regard to biological repositories and DNA and other matters. It was right around the time when Human Genome Sciences (HGS) and The Institute for Genome Research (TIGR) were formed, and it was also at the time when Craig Venter had left the National Institutes of Health to start TIGR, a not-for-profit organization, that was very tightly affiliated with HGS. According to the relationship that was set up, a number of government findings were coming into each one of these programs, that were used to sequence microbes and to have a full composite of gene expression data. At the same time, human expressed sequence tags (ESTs) were being discovered, and these were then being taken to HGS, where they were being categorized and utilized according to how they were expressed and what their therapeutic potential could ultimately be.

It was within that framework and during that period that people began to recognize that there had to be some structures in place with regard to exchanges of information and biological materials between a not-for-profit or a government-funded organization, and a for-profit organization, such as HGS.

It was also right around that time that Smith Kline Beecham invested $125 million into HGS on the basis of the promise and potential of those data and the information that was generated from them. Many of the repositories, including the American Type Culture Collection (ATCC), were used as warehouses for all of the sequence data that were being collected, and, according to the *Budapest Treaty on the International Recognition of the Deposit of Microorganisms for the Purposes of Patent*

[22] Presentation slides available at: *http://sites.nationalacademies.org/xpedio/idcplg?IdcService=GET_FILE&dDocName=PGA_053664&RevisionSelectionMethod=Latest.*

Procedure. A very significant percentage of all patented microorganisms were and are currently stored by ATCC.

So, that is a snapshot of where we were about 12 years ago. Let me switch gears and talk about the commercialization of microbial resources. Microbes have been of tremendous societal and commercial value for millennia. Brewer's yeast is a good example, as it has been used for baking and brewing for at least 6,000 years. Since the initiation of modern biotechnology, which takes us back to 1980 or so, these microbes have formed the basic underpinning of both generalized non-applied research and applied research. We have just scratched the surface of the commercial potential of these microbes because relatively few have been identified. In recent years, the genomes of several of these microbes have been sequenced, and we have just begun to harness the value of that new information.

Today, some global standards and principles exist, but are being applied unevenly to the characterization, access, and licensing of these microbes and collections. So, just as the world culture collections all have minimum scientific and technical standards, a microbial commons approach—whether it is based on a compensatory liability rule, or on other approaches—is also a potential means of global legal standardization, which is essential for the propagation and development of this potential.

It has been suggested that diverse licensing strategies and techniques have elevated transaction costs and other barriers for relatively simple, collaborative research projects. I think that we can all agree that there are higher transaction costs and that there is also an investment of time associated with the discussions and negotiations that take place. Furthermore, according to the liability rule principle, providing access to all microbial resources and collections will eliminate any of the competitive advantage that arises from keeping these materials and data from other organizations.

However, I have run several companies, and I know my duty is to represent the interests of the shareholder. If I am running a microbial company, the essence of my duty is to extract as much value as possible out of what I am working on. If that means withholding that information from the general public or identifying a collection and utilizing that collection for the best purpose that I can to maximize my competitive advantage, it would be my fiduciary responsibility to do so.

So, there is a potential caveat: A substantial amount of the data in the collections are in private organizations and in companies that are very well funded, such as pharmaceutical companies that are using the data for their own discovery work or for other purposes. It is clear that the limitation of access to these microorganisms does not permit a level playing field on the fundamental research but the actual sequence data is probably of greater value. The primary point to a research commons is that the data and the collections will not be kept locked away inside private organizations, but that the majority will be out in the public domain.

When you look at what is being done with the Brain Research Project, two things are clear: first, that the applied research that will come out of it has not fully manifested itself as it has in microbiology and, second, that the genesis and the progression of the research in brain sciences has been within what is fundamentally a very collaborative government-sponsored research program. While the brain itself is extraordinarily complex, the policy issues surrounding the data are less so, and the neuroscience community is taking a far more linear and straightforward approach than is required in creating a microbial commons.

One other thing keep in mind with regard to the postulate is that it is a seller's market for those who have control of these microbes. As a buyer, if I am negotiating to acquire a microbial collection for commercial purposes, whether it is a world culture collection or a collection from another company, I am going to negotiate on the basis of what I think the fair market value is at that particular time.

As a case study, let us examine ATCC. It is an independent, private, non-for-profit 501(c)(3) biological resource center, and it is part of the World Federation for Culture Collections (WFCC). Its primary goal as a biological resource center and as a research organization is to provide reliable, qualified, and low-priced biological materials for the advancement of basic research. That was the fundamental founding principle when it was established in 1925, when a committee of scientists recognized the need to have a common repository in the United States and a means to exchange microbes among the scientific community for scientific research. The purposes included the continuation of the advancement of that research, the validation of some basic research funding decisions, or other related goals.

However, a lot has changed since 1925. In 1949, the first patent culture deposit took place at ATCC. In 1981, it began to accept patent materials from any country that had signed the Budapest Treaty. In 1997, ATCC initiated the first of its special collections and moved to Manassas, Virginia.

As a case study, ATCC is reflective of many of the other culture collections throughout the world, so I did a little research to see what the mission statement is because such statements are indicative of the priorities within an organization. The mission statement on its Web site states that it is a global, non-for-profit, bioresource center and research organization that provides products, technical services, and educational programs to private industry, government, and academic organizations. Its goal is "to acquire, authenticate, preserve, develop, and distribute biological materials, information, technology, intellectual property, and standards advancement and application of sciences."

Its brochure offers a slightly different mission statement: After saying that its goal is "to acquire, authenticate, preserve, develop, and distribute biological resources and knowledge to scientific researchers," which is very similar to the Web-based statement, it adds, "We strive to be the preferred provider of high-quality biological reference standards which, along with products and services developed in-house, enable science to touch people's lives." This is very different. It is saying that ATCC is a product-based organization.

By contrast, the microbial commons will work in a context that is not product-based. It will not be a situation where people are sharing their products for open dissemination so that a competitor or a researcher in Ghent or a company in Tokyo can use them to create their own products. This is something that needs to be considered as we move forward.

Another key point is that ATCC's holdings, which are almost identical to the ones throughout the world, are growing exponentially. However, while the number of holdings is growing very rapidly, it is on the order of tens of thousands of samples, perhaps a couple of hundred thousand at the maximum. When you look at the cloned genes collection, however, it contains 8 million cloned genes, and the number is continuing to grow at a rapid rate. In setting rules of commons, whether we consider it a microbial commons or a biological resource commons, the same principles apply. In creating

research commons we have to be sure to allow for value creation to be recognized and patent protection to remain in place.

ATCC, for example, has a number of special collections, including MR4 for malaria, the mantle cell lymphoma cell bank, and a yeast genetic research collection. These special collections are microbial semi-commons that hold special value to the researcher in the field of malaria, lymphomas or yeast research. ATCC, by characterizing and combining these groups of cells or microbes, has created value that may warrant some royalty payment out of the discoveries from these special collections.

There is a disincentive to having a research commons. From the industrial perspective, the sequestering of biological resources and data allows for a perceived monopoly on the downstream application of research discoveries. If I am running an organization that is seeking to obtain funding, whether it is a large laboratory where I am looking for research funding or a large, private organization where I am seeking investor support, then I definitely want to emphasize the value of this perceived monopoly. I am trying to sell as a value proposition that I have this great collection I am working with.

It is only when sharing and broad distribution will provide a greater possible upside to the owner that this material will be disseminated, whether that material is held within a private corporation or within a culture collection itself. For example, I may want to get my results out into the public domain as quickly as possible because I am going to then get the best publications, the research grants, and the talks at the major conferences that take place.

Consider a commercial licensing program for available microbes. If I want to create an environmental company and license microbes for, let us say, wastewater treatment or some environmental purposes, would I sign onto a microbial commons? Not unless some type of protection would sanction the research investment to commercialize the discoveries out of this environmental companies R&D laboratory. In almost every instance, I could not justify an investment into research program that would be equally accessed and "owned" by my competitors.

The final two issues are what "commercial use" is and how a MTA would be applied to it. Here is the definition of commercial use in the standard ATCC material transfer agreement:

> "Commercial Use" means the sale, license, lease, export, transfer or other distribution of the Biological Materials to a third party for financial gain or other commercial purposes and/or the use of the Biological Material:
> (a) to provide a service to a third party for financial gain;
> (b) to produce or manufacture products for general sale or products for use in the manufacture of products ultimately intended for general sale,
> (c) in connection with ADME testing;
> (d) in connection with drug potency or toxicity testing which does not include either screening multiple cell lines for potential inclusion in a screening assay system or screening multiple compounds in a system for internal research purposes only;
> (e) in connection with proficiency testing service(s), including but not limited to, providing the service of determining laboratory performance by means of comparing and evaluating calibrations or tests on the same or similar items or materials in accordance with predetermined conditions; or
> (f) for research conducted under an agreement wherein a for-profit entity receives a right whether actual or contingent to the results of the research.

We can conclude that the definition of commercial use is not very simple in application or determination. In essence, a commercial use refers to the sale, license, lease, export, transfer, or other distribution of biological materials to a third party for financial gain or other commercial purposes and the use of the biological material. The first two uses described in the definition, (a) and (b), are very clear because they point to financial gain, either from services provided to a third party or products manufactured for general sale.

The ADME testing in (c) refers to absorption, distribution, metabolism, and excretion, which are the essential elements one examines in the early stages of product development in the pharmaceutical industry. When someone ingests a pill or a compound or a chemical, you want to know what happens—how it is absorbed, how it is distributed, whether it is altered within the body, and then, ultimately, how it is excreted. There are some contract groups that charge quite a bit of money to do this in early testing, so perhaps some businesses have been formed out of it, but is it a major commercial use? I do not think so. I would argue that this scenario probably goes right into the category of "may not be too valuable." So we could remove those microbes from the MTA and place them into that microbial commons.

Concerning (d), drug potency or drug toxicity testing, there are some very basic and general research components and activities that are done in animal studies long before any product comes into existence. If I am running a company and I have to pay to use a microbe for which there is a probability of only 1 in 10,000 that it may ultimately result in a product, I would not use that, and I would not be inclined to buy into a program under those conditions. It is too early within the process.

The final one, (f), refers to research conducted under an agreement where a for-profit entity receives a right, whether actual or contingent on the results of the research. This example is a little bit more open-ended. It supposes that someone is doing sponsored research: I, Stephen McCormack, sponsor this research, and a laboratory wants to obtain all the rights and data that come out of that and then eventually may want to build something out of it. The conclusion that we can draw from this is that the definition of commercial use is not very simple, either in its application or determination, and it is also a moving target.

There are various technologies available for licensing from ATCC, the world's largest repository of biological materials. These include, for example, materials in ATCC special collections. If you take a series of microbes or a series of cell lines and bundle those, can you gain a proprietary value or a perceived proprietary value out of that? The ATCC also offers pre-1980 cell lines that are not subject to the terms and conditions of the Bayh-Dole Act.

Another issue to keep in mind concerning liability rules is what happens if the WFCC organizations pursue value-added research and bundling strategies. In that case, the collections will move up the value chain, approaching the level of commercial products, and, indeed, they could be sold directly on the market for general application. This may invalidate a liability rule approach.

Although the proper balance needs to be found, I do not think these culture collections should deviate from the basic purpose of preserving, maintaining, and distributing biological materials. That is why I started out talking about what a mission statement is and what a founding principle is—it is in these areas identified by the

mission statement that these culture collections and these biological and microbial commons will grow.

To wrap up, here are some things to consider as we develop a microbial research commons. First, in certain cases, the mere characterization of a microbe can create immediate commercial potential for the products. Consider H1N1. Sequencing and identifying these epitopes and picking them out is part of their fundamental characterization because you do not know what you have unless that is done. However, once one identifies the changes that have taken place, the information is immediately of value..

Second, microbes and microbial connections should meet certain non-commercial qualifications for the entry into this microbial commons. The timing is important, however. Discoveries and advances in scientific research will regularly move the line on what is eligible for these liability rules because as you learn more about what is commercially important and what is not important, the point where that liability rule should be applied will change.

In conclusion, the commercial use of microbial cultures is very difficult to define because the value changes over time and is subjective to begin with at the time of their appraisal. Different people may look at the same thing in a different way and value it differently. So a multifaceted system may be required to form a microbial commons that will enable broad and effective access to data and to biological materials. The culture collections will have to continue to lead with MTA agreements that will work as desired in the core of these microbial commons.

5. Developing Country Perspective: Microbial Research Commons Including Viruses – Ashok Kolaskar[23]

University of Pune, India

In recent years the BRIC countries (Brazil, Russia, India, and China) have been experiencing rapid industrial growth, and innovations have been occurring at a much higher rate than in many other countries. This trend is not limited to any one specific area but encompasses much wider spectrum of economic sectors. I am referring not just to research and development per se, but also to innovations at every level—in particular, the processes. This rate of innovation has ramifications for all aspects of science and society.

When we look at specifically India, we see companies like Biocon which are trying to develop more biologicals and getting into the largest applications as well as developing products that are important to improving the general health. A second example is the vaccines developed by the Serum Research Institute, which are administered to every third child in the world. There are companies involved in energy and, particularly, biorefining. These companies are learning to use not only the edible crops but also the non-edible crops such as castor bean. Appropriate microorganisms are critical for the process of biorefining whether it involves edible or nonedible biomaterials. A third area includes companies and industries that are looking at producing new types of biomaterials using microbes. So, the challenges are not very different from those in the United States or European countries.

There are also real differences, however. What we are seeing in these developing countries is a sudden increase in basic research as well as applied research, which is happening mainly with the support of government agencies. The private sector is not providing financial support to the same extent. For this reason, there is a need for microbial and genetic resources. Where are these resources going to come from? Right now, most of the resources are coming from the BRCs, including the American Type Culture Collection (ATCC).

India too has started in the recent past, creating similar resources. Microbial and other biological culture collections have been established in several universities and research institutions. As in the United States, we have lost some of these collections. Some of them were transferred as special collections. Ways to recirculate and reuse them have been established, but in a country like India these systems are still in the process of development. Lack of well established systems such as these is one of the main differences.

There are several culture collections in India. The Microbial Type Culture Collection (MTCC) at Chandigarh is the one that is recognized as part of the World Intellectual Property Organization, and it gets the deposits of cultures for patenting purposes. We now have laws similar to those in the United States as far as the patenting is concerned. The MTCC performs a role in India much like ATCC does in the United States.

[23] Presentation slides available at: *http://sites.nationalacademies.org/xpedio/idcplg?IdcService=GET_FILE&dDocName=PGA_053663&RevisionSelectionMethod=Latest.*

There is also the National Collection of Industrial Microorganisms at Pune, and this is completely different from that in the United States. That is, there is a separate culture collection for those microbes that are industrially important and that can be used more for development or in a commercial fashion.

We also have a separate culture collection for viruses, and we have a national facility for animal and tissue cell culture along the lines of the ATCC, although it has not grown to the same extent.

In addition to these main collections at the national level, we have several specialized collections, and they do not get very much support from the national agencies. These are being done more at an institutional level or as specific projects. As long as those projects are ongoing, these culture collections will grow. For all of these culture collections, most of the information is available only in printed formats. Very few of them have their own websites where the full information is available. That makes getting information from these localized culture collections very difficult.

These culture collections have also developed material transfer agreements (MTAs) —not quite to the level that we see at ATCC and similar collections, but they are following similar lines to those established by the ATCC. The result is a similar set of restrictions on distributing the cultures, with advantages and disadvantages that are similar to those with the ATCC. As seen in the United States and elsewhere, the majority of scientists generally try to get materials from these culture collections on an informal basis. As a consequence, it is difficult to be certain about the quality of the cultures used in performing scientific research. This lack of quality assurance makes it difficult to verify research results.

Thailand has a similar network of culture collections. One part of the network, the Biotech Culture Collection, has nearly 3,430 cultures. The Department of Medical Sciences has its own separate collection, as does the Department of Agriculture and the Thailand Institute of Scientific and Technological Research.

These distributed culture collections generally do not communicate with each other. This is the case not just in India, but in many countries. How good is that interaction in the United States, for instance? Human nature is the same everywhere, and scientists want to hold on to what they have, even though we keep saying that scientists like to publish. Yes, they do like to publish, but only the final results. Once scientists have developed those collections and materials, they want to hold on to them.

There are a variety of issues facing culture collections in developing countries. In most of these culture collections the characterization is very minimal. Few culture collections are characterized at the DNA-fingerprinting level, and we do not know whether a given culture in the collection is the same as or different from some other culture. The data are not fully computerized, and information about the cultures is not easily available, as I have already mentioned. Moreover, in most culture collections there is duplication: The same cultures are available in different base collections, and.this leads to higher maintenance costs. The material transfer agreements are similar to those used by ATCC and by most repositories.

There is no system in place to detect or prevent misuse of an MTA. An MTA may be signed, but we do not yet have a system to detect whether the MTA has been used or misused. Nor are there any good answers to the question of what needs to be done if the the MTA is misused. That aspect of the system is still very weak.

Another major difference between the developed and developing countries is that, at least in India, very few scientists are conversant with taxonomic classifications. Even

at the national culture collections, there is a paucity of skilled human resources. We keep boasting in India that we have a demographic advantage, and we do have a young population below the age of 25 years—nearly 550 million, out of 1.2 billion total population. If we train them properly, the members of this population segment can do wonders. However, there are several obstacles that come in the way of exploiting this population advantage; current state of training is not of acceptable level; remuneration is not attractive to bright youngsters; and finally the repetitive nature of the work does not offer intellectual stimulation. Inability to recruit new generation of taxonomists to replace the older retiring taxonomists is going to be a huge issue. Finally, concerns related to biosafety and national security are still not given due emphasis in developing nations compared to the developed countries.

There are issues that go beyond the microbial commons. At the current time the desire to understand the functioning of an organism extends beyond the gene level. There is great interest in looking at metabolic pathways and figuring out how to describe them - what language should be used to describe the pathways - and how to describe the results obtained from analyzing the metabolic pathway information. It is important to deliberate about what sort of commons we will acquire for this purpose. We need to double what we have been doing last few years, try to collect the metabolic pathway information on all those microbes whose genomes are fully sequenced as well as study the data that are available in the public domain, and then curate them.

Most of the metabolic pathway information that is available is not in the public domain and is not necessarily fully curated. Since labor costs in India are lower than in developed countries, our group at Pune University made use of that asset and tried to curate this information and build a metabolic pathway database. In a fashion similar to what the previous speaker described regarding brain research, we are trying to integrate all the relevant types of metabolic information into this database. This includes information on genes, enzymes, and various types of chemical and biochemical reactions in addition to information on the organisms themselves, and that is where we will need a proper commons.

In addition, we have added the data acquisition and integration tools to see that all the data are entered in a systematic fashion. You can imagine the knowledge that can flow from such a resource—and this includes everything from visualization tools and structure predictions to the simulations of various structures and of the organism itself. We have indeed developed a visualization mechanism within the system for small molecules, the metabolites. It is also possible to develop models to look at large complexes. We would like to extend the analysis of the metabolic pathways to the level of patterns. However, the number of identified pathways is much smaller compared to the total number of pathways present. It is still a major challenge to identify all the pathways.

Examining metabolic pathways is one of the major means to understand how the microbes are related to each other, rather than simply looking at their genomes and calculating the phylogenetic distances from each other based on the genome information. Indeed, the two approaches lead to two completely different trees representing the relationships among the microbes. This has a potential to change our whole understanding of the science. The study of metabolic pathway similarities amongst different organisms is also essential to engineer new pathways or new microbes with specific properties.

Some pathways are absent in one particular bacteria versus other, so we can start looking at whether this is what differentiates one species from the other species. Most of

the times we look at what the species have in common and not how they differ. If we use this molecular-level information and look at the differences that will probably tell us better how to even name them properly. We have given them names according to their macro properties, but we have not really taken micro-level details into account. Now that we are able to see the micro aspects, I think we should re-look at all of these things and decide how we should distribute, divide, and further classify them.

One of the major issues in developing such a database is taxonomic classification, which has been done according to whatever is available from the National Center for Biotechnology Information. If this information has any errors, these errors will get translated into most of the secondary databases. In addition, there is no standard system to represent the metabolic pathways themselves, and we will have to develop that as well because, as we start looking at bringing in and integrating the information, we cannot go further unless we have certain standards. We create these standards mainly to coordinate activities and exchange information amongst different groups. Those who are specialists or who are working on a special project probably do not require these standards because they understand exactly what is being done. But when people come from different backgrounds and want to understand what is being said without any ambiguity, these standards become necessary.

We must ask, however, how far to go with standardization. If we standardize too much, we may lose necessary flexibility. We have to bring in the systemization only to that level that is necessary for exchanging views, information, and in such a way that everyone can understand what is being done without any ambiguity.

In many cases, the chemists have tried to put together some type of commons. We will have to see how well these can be integrated with the metabolic pathways database in development. The other issues that need to be considered for integrating the information generated by the chemists with metabolic pathways database are the accuracy of the data and also the accuracy of data entry. This is particularly so in the case of different isolates and different strains because there can be variations in chemical structures of metabolites and other pathway components even amongst the different isolates and of course different strains.

Depending upon which commons we use, which tools we use, and the way we look at the microbes, we are going to describe the system in very different ways. We may require different words. We will, therefore, have to start asking the questions: What is the best way, how do we say it, and how can all these things be integrated into the system as we develop commons and semi-commons?

Viruses are one set of organisms that will need to be examined because viruses are part of the microbial world. We tend to concentrate mostly on bacterial systems and do not include viruses at that level. However, since we have the full genome sequences of many viruses and their isolates, we can look at and identify them right at the protein level. We can look at which proteins are specific to a particular species and use that information as an identifier to classify and identify viruses.

Our group at Pune University has developed a database of animal viruses. The database is entitled VirGen. In the process of developing this database, we found out that it is much better to identify the viruses first using the macro features and then confirm by using molecular properties. In order to do this, however, we need to develop many more appropriate tools and that is where the field of genomics and data mining of the viral genomes becomes very important.

The genomic data must be structured in such a fashion that it is straightforward to get from family to the isolate level. With the full genomes of viruses being available, it is possible to examine and compare them. Fortunately, the International Committee on Taxonomy of Viruses (ICTV) has developed a universal taxonomic scheme for viruses, and one can make use of that. However, there is a long interval between the publication by a scientist and the assignment of standard nomenclature. Correlating the work published before and after the assignment of standard nomenclature could be an issue that needs to be addressed in developing the microbial commons.

We included the whole genome phylogeny and the method to predict B-cell epitopes, based on that information in VirGen database. This sort of additional value-added information offers an example of how we can start integrating and creating information for the commons in ways that can be then used in other area of the microbial commons itself.

As I mentioned, the ICTV classification information available in the published literature does not always match what is being published by the researcher. There is no standard method to describe viral isolates or the various strains. Electron micrographs and other image data are not readily available, which makes identification difficult and inaccurate. Only now are data in the form of images and, particularly, high-resolution graphic images being slowly added to the public-domain databases.

There are other databases that have been created by various experts, but they are not publicly available. We need to include information from those databases as well, and, once that is done, there will be the new challenges of how to describe these databases in a consistent fashion and how to use them for extracting knowledge.

Recombination occurs much faster in viruses than in bacteria and other microbes. Therefore, that will also need to be taken into account. Host and vector information will need to be described in standard language. At a minimum, immunological properties and therapeutic options will also need to be added to the databases.

Having raised these various issues, it is fitting that I should also suggest some solutions. First, we will have to devise a means to build trust among developing nations that the developed nations are not looking to exploit their resources. Today, there is a huge gap in understanding and awareness between those in the developing nations and those in the developed world.

Most of the scientists and culture collection people in developing countries feel that they wil be losing huge resources, financial and otherwise, if they make their collections available, even if their country's administrators feel it should be done. This is where education is required. We need to show them that by sharing their materials and data, it will lead to a situation in which all will receive the benefits of that sharing and that it will be a win-win situation for both the developing nations and the developed nations.

This is a critical issue that we will have to address in order to create the microbial commons and make it workable. Otherwise, the commons will not work, and only this small group will be talking about it. To really make it workable, we will have to communicate the value of cooperation to those in the developing nations. Networking and consortia among scientists can play a role, and the curators of culture collections as well as policy-makers from developed and developing countries need to understand the importance and get involved. Unfortunately, most of the time everyone is talking at cross purposes. This meeting is one of the few efforts to bring together all of the people who

are interested in creating a microbial commons, and the policy makers are probably less well represented among the attendees than we might wish.

Material transfer agreements should be standardized, and they must take into consideration national security and the biosafety issues. If we do not take these issues into account, those MTAs will not be accepted by most countries, including the United States.

We will have to create awareness about open access and open educational resources. This is something that has been discussed in some detail in India. The National Knowledge Commission, which was established about three years ago by the current prime minister to look at what is necessary to transform India into a knowledge society, recommended that India move towards more open education resources and open access.

The commission had seven members, and two advisors. Our role was to suggest methods to improve the total knowledge infrastructure, including access to and dissemination of knowledge. Instead of worrying about the philosophical definition of knowledge, we decided to look at the practical issues and try to work on those: How can we improve the creation of knowledge? How we can improve access to knowledge? How we can build the human resources to conceptualize the knowledge and knowledge network as well as the applications and services?

We divided into groups to focus on and identify the areas related to each of these issues. We identified and consulted with various experts both formally and informally, and then we created background research and analysis groups, or working groups. After they deliberated, these working groups offered their findings and recommendations in a report, which was deliberated upon by the National Knowledge Commission.

The final recommendations covered a variety of areas, including higher education research, how to improve the quality of Ph.D.'s, how to improve primary education, and how to improve the national knowledge network. The recommendations were given to the prime minister in the form of summaries, each about two to two-and-a-half pages long and containing about 10 to 12 specific recommendations. We kept them brief because we believed that the prime minister does not have more than 4 to 5 minutes to read this type of information.

Many of the recommendations given to the prime minister have been now accepted by the government, and their implementation is in progress. One such action was the establishment of the National Science and Social Sciences Foundation, because we found that the scientists and social scientists in India do not work together. They have been kept in two different rooms, so to speak, and the wall has been very thick. So we decided to bring them together.

The U.S. National Science Foundation has at least a small group working on the social sciences side, and it provides a certain amount of funds to the social sciences. In India, it does not work in the same fashion. The Department of Science and Technology works only on the scientific aspects. So we decided to create a new foundation to facilitate cooperation, and it is now being established.

Similarly, we recommended the creation of a National Knowledge Network because improving access to all of the work we do is important. If the work is not accessible, then it is not going to serve anyone. So we are creating a gigabit network that will connect a large number of institutions within the country and will also connect to institutions in the United States. The funds to establish this network have been allotted both on the institutional side and on the Indian government side, and the connectivity is getting established.

Five years ago we did an experiment to connect India and the United States at 660 megabits per second. Now, we are increasing the speed to one gigabit per second, and, probably, it will very soon be increased to 10 gigabits. The main point here is that the National Knowledge Commission suggested to the government that it should create open educational resources and open access systems, and now, as a consequence, all this work is getting funded. The government in India will provide significant funding for open access, just as it happened in the U.S.

We have also started to provide open education resources, with courses developed with the various Indian Institutes of Technology and the Indian Institute of Science, the topmost institutes in the country. These courses have been made a public resource.

In the context of the research commons, each country will have to establish policies to be followed concerning research papers published in the microbiology area. In India, we are just getting ready to do that, and this sort of symposium should help to push that effort further. Policy makers are being lobbied to make the outcomes of the government-funded research publicly available. A bill on open access has just been drafted and will be probably passed in the next parliament.

Various steps are being taken to encourage scientists to publish in open access journals. The institutions are being given funding for that purpose in recognition of the fact that whenever open access is instituted, somebody has to pay for it. Individual scientists will not be able to pay. The Indian government is providing funds so that whenever a paper is accepted and published in an open access journal, the appropriate institution will get funding to be applied towards the costs of publishing; certain additional money will also be provided to these institutions. As a consequence, the institutions and the scientists will both have an incentive to publish in the open access journals.

We also need to organize training programs by international experts to improve the quality of culture collections and the databases. In this way, I feel that we could really have much better interactions.

Finally, we need to improve access to the specialized culture collections. These collections exist, but we need to improve access.

Panel Discussion

PARTICIPANT: I wanted to ask Stephen McCormack this question. You put up an interesting slide talking about perceived disincentives to sharing. I wanted to know if you were aware of any actual data on the degree to which there is any reality to that or how much of it is truly just a fear or an exaggerated perception as opposed to something that is been ground-truthed with real numbers.

DR. McCORMACK: That is such a difficult number to come to up with and to be able to identify. I have given a lot of time and thought behind that to try and identify how a privately held collection or a privately held body of researcher data would be identified. I only know, from various experiences and in looking at things where the legality would come into play, that if it resulted in a product or it is used for research purposes and the proper license was not taken out, it could be a problem. This is really difficult information to get,so I cannot answer that definitively.

PROF. WU: Relative to that, there is a very interesting opinion piece recently published in *Nature Reviews Drug Discovery* where three pharmaceutical companies—I think it was GSK, Merck, and Pfizer—together wrote about tearing down the firewall of data protection. They want to change the model from the proprietary nature of the data to the proprietary understanding of the data. They think that for all this potential data integration there is a need for them to do discovery with all the data. Instead of trying to generate all the data and keep that within the firewall, they prefer to have this public-private partnership to take the drug discovery to the next step. It is a very interesting opinion piece if you have not read it.

PARTICIPANT: This question is also for Stephen McCormack. You mentioned that materials in the ATCC Special Collection that were deposited prior to 1980 were not subject to the Bayh-Dole provisions, and I was wondering if materials that were deposited prior to the Convention on Biological Diversity were also exempt from those requirements.

DR. McCORMACK: I think you would have to ask the ATCC. I cannot make statements or representations for that organization.

DR. SIMIONE: Since I work at the ATCC I will answer the question, but I am a little puzzled as to the pre-1980 distinction. The pre-1980 material is useful because it does not have bovine spongiform encephalopathy (BSE) implications, the prion "mad cow disease." So it becomes useful for that, but from the Bayh-Dole standpoint—that one I cannot answer. I will be happy to try to find out. We have experts at ATCC who would be able to answer those questions.

PARTICIPANT: I have an overview question for both Joan and Mark. I thought your presentations were really fabulous and interesting. I used to be a scientist, although not a biologist. My question to both of you, speaking with my law professor hat on, is: Where are the problems? What are the issues? It seems like you are working it out. What are the issues that somebody like me, speaking as a law professor, might want to be able to think about?

Also, I had a comment on the point that Stephen McCormack made about how to define commercial use and when it is commercial use. I certainly would acknowledge that it is never going to be easy to define commercial use, but at least in my view, it is preferable generally to push these definitions downstream, and that is one of the nice things about a take-and-pay regime as opposed to the type of regime that would require you to determine whether the user is a commercial entity.

When you can talk about activities at the point of use, it is advantageous from a legal point of view because there are always going to be fuzzy boundaries, but if you are at a point where the commercial potential is more concrete, there is more incentive for people to negotiate, to come to some sort of agreement and avoid a law suit. I wanted to know if you had a reaction to that.

DR. McCORMACK: Yes, I think that that obviously is going to be the right approach. It is just in the creation of these proposed commons, right? If the understanding were that the take-and-pay regime would be the dominant one, and the repositories of these

materials themselves were all held dominantly within such space, but with some partitioned in a semi-commons or in a completely privatized area, then, yes, I think that that would obviously be the right approach.

PROF. REICHMAN: I would like to comment on the general discussion. Not being in your community officially, I have to say that I would rather see this discussion start at a much higher level. What I mean by that is, we are at a very interesting time in history when we have the capability to mash together huge amounts of information for the benefit of mankind. If we spend our time focusing first on how someone makes money, how you nibble around the edges of this commons, then we are not going to accelerate human understanding, we are not going to achieve the rate of progress that we have the potential to do. We will get bogged down in incrementalism.

I think that the starting point for a reengagement and how one develops any kind of a science commons is to ask what the changes in thinking are in either the litigious community or in the sociology of science. We need to try to get people to share at the moment they know something new, and we need to invent within that framework a way that protects the potential opportunity to benefit in a fair way for the individuals who contribute, not knowing the value of what they are contributing at the time they contribute. So I would like to say we are doing this wrong, trying to build on what has been going on in the past, so let us start thinking about what are the real goals here, not making money.

PROF. BENNETT: I have felt I am living in a parallel universe in that most of us scientists are more into the curiosity of the research than some of the other goals. That is why I told that little story about the money people, the power people, and the fame people.

I also try to make the point that the genomics community and the other "omics" communities have had the chance to more or less start over. New kinds of information exchange have been developed that are so much better than some of the approaches that was developed early in the 20th century. We should be using the new capabilities as a model. I am more or less echoing what Mark has just said, but in a slightly different way.

PARTICIPANT: Can I add a comment to that? First, I agree, but I would make one other supporting point, which is that in genomics in particular there has been an explosion of technological capability. This has made genome sequencing into a category of data that is now so easy to generate and is generated in such massive quantities that even if you tried to protect it or keep it behind a firewall, it is literally no longer practical to do that. Somebody with a new 454 or Illumina machine could simply reproduce everything you have done within half an hour tomorrow morning. So, getting to Mark's point, I ask: Might you be able to separate data according to their nature, as in data that are easy to produce in such massive quantities versus those that must be processed in some more direct, intellectual sense to add value?

PARTICIPANT: This project has such a broad charge. Why do we not add this as well? Ideally you would want to get rid of all transaction costs, but there are reasons that some stipulations should be put in place if everybody is going to be able to contribute and play, and there are a lot of reasons you want everyone to do that. And, yet, from what you said, and from what was quoted in the paper, if there is a barrier, the scientists are going to go

around it and they will do as they like within their informal communities instead. So, are you thinking that the only situation that will really be acceptable is having no transaction costs, that there will not be any records where you have to refer back to the original source, or else people simply will not participate? How is that going to work out?

PROF. BENNETT: You probably need a social scientist to answer a question like that, if you want any kind of answer at all, but I think there is a critical point where regulations become so burdensome and so expensive in time and energy that people start circumventing them. When I look at some of the documents that come out of certain international activities, they fail in that way. They really do, and at some level I think they are unnecessary.

One of the most famous discoveries in the history of microbiology was penicillin, which came from a microbe. There was a time before penicillin was manufactured as a medicine, when people in Peoria were collecting penicillium strains, the fungus, from all over the world. What is wonderful about this story is that there were no barriers: The strain that finally was selected was from a cantaloupe in Peoria. It was not from some exotic place, and, yet, that process of discovery would not have gone on as easily as it did if there had been the kind of barriers there are now for accessing and using organisms. It would have been impossible.

PARTICIPANT: I also wanted to pick up on what you were just saying about the way in which scientists are motivated by fame. Within the patent laws, there is what is referred to as the research exception. Maybe people do not always rely on it, but it is available as a way to allow the freedom to operate and do research. For data and materials, however, I do not know that we have something comparable, especially for those early stages of research when a private company might presume that everything is valuable or commercial, and it really does not know where the research is going to go.

It just seems that in designing the commons, if there are some incentives and ways to get that early push of creativity going, then we know what the real values are likely to be.

PARTICIPANT: I would like to ask the panel to address the education and training dimension of the commons. Maybe Joan Bennett and Ashok Kolaskar can address that. The question is: If you create a commons like this, there is an additional value in both directions for engaging the commons with the education and training process. For one thing, if you have a diverse community engaged in using the commons, and they have a diversity of expertise, then there is a tendency for the most sophisticated and well-funded members of the community to take greatest advantage of that and for the less well-known members of the community to be left behind and lose their funding.

So, there is a major advantage in training all members of the community to take maximum advantage of the microbial commons, and, by doing that, you lower the fear barrier for those disadvantaged members of the community to contribute to the commons. This can be the case both within a particular country and also internationally, so there is a great value in making sure that all members of the community are well trained in using the commons. There is also a major benefit to the quality of the data and the resources in the commons to having a much larger number of trained eyeballs examining what is in there. It may be scientists from diverse countries around the world, or it may be

undergraduates at diverse educational institutions throughout the United States or other countries. I was wondering if any of you would like to comment on that.

PROF. BENNETT: Great idea.

PARTICIPANT: First, I guess I am a little confused about your emphasis on commons. The commons is a vehicle to achieve something significant in the community. So training is something that needs to be done in order to create within the entirety of the communities that participate in microbial biology more very well-educated individuals who can work across disciplines, who understand enough about one subdomain of microbial biology, let us say, and enough about one or two other domains to be conversant and to advocate for the future of the programs that you hope to foster. The result would be the creation of a more evident, interactive environment.

So, on the training side I would say the focus should be on finding a way to bring the brightest people who are attracted to aspects of your field to projects around this kind of an international effort. They need to be given enough resources for a long enough period that they can be trained and be creative and have an impact. We should make sure that they are trained in more than one discipline.

PROF. WU: Yes, there has been a lot of discussion regarding the under-appreciation of data scientists, those people who are involved in maintaining the quality of the data and now managing the data. So we need to encourage both training and this kind of cultural change in terms of appreciating the value of people who are doing this kind of work. That would be a separate issue that needs to be addressed as well.

PROF. KOLASKAR: I think there is an absolute need to provide training, especially with reference to the developing countries. The scientists who are doing high-quality research work will not necessarily participate in these types of activities. The gains are very few. Those who publish good, high-quality research papers get much better recognition than those who are working in the commons infrastructure activities. That is why training becomes very important—and not just for those who are involved in culture collections or in database creation. Those are definitely essential. But even for these scientists, the training is important, and that is why I feel that we will have to really give a very high priority to do this, and only then will we be able to develop better commons.

PARTICIPANT: Let me make a concrete suggestion since the word "concrete" has been used. In the six or so years of trying to put together this international neuroinformatics facility, training has been a topic, and the international participants involved in trying to form something that would be practical there have wandered all over the map in terms of things that one could do. A blue ribbon review panel for OECD looked at this last May, and the conclusion was that one should do something very much like a Cold Spring Harbor course. The idea is that a limited amount of money engaged in something practical like that would result in the biggest impact. Perhaps one could create the equivalent in Europe of a Cold Spring Harbor course for two weeks and encourage the member nations to have a competition for their best potential trainee. The trainees would need to be multi-disciplinarily prepared persons, and the curriculum of the course should cover a broad enough spectrum so that all of the participants from the member nations went away feeling a broad sense of camaraderie. Perhaps within about a year or two

years, a syllabus could be produced that could be used to train people in other countries through the standard curriculum and departmental type functions.

That was a formula that we came up with. Then the idea was to go after EU funding and funding from various other sources to support that course. It was not easy to get everybody to agree to that sort of thing and I am not sure it will actually happen, but that is what we are trying to do. A similar sort of approach might be useful to you here.

6. A Compensatory Liability Regime to Promote the Exchange of Microbial Genetic Resources for Research and Benefit Sharing – Jerome H. Reichman

Bunyan S. Womble Professor of Law
Duke University School of Law

Scientists know only about 1 percent of the world's microbial resources. A good selection of known and scientifically validated microbial resources are held in hundreds of public culture collections around the world, which have accumulated these precious resources over a long period of time. Many other semi-public collections are held by government departments, especially in the U.S., and by universities around the world, who assemble materials for specialized research purposes. About 600 of the public collections are loosely organized under the World Federation of Culture Collections (WFCC) which operate under agreed quality and security standards. The original principle underlying the establishment of the WFCC was that their cultures belonged to the common heritage of mankind, in the same way that plant genetic resources were initially treated by the United Nations Food and Agriculture Organization (FAO). The members of the WFCC make these resources publicly available for research purposes, with the holdings listed in open catalogs and, of particular importance, the member collections are obliged to authenticate all their resources and to track all uses.

By the late 20th century, there was a tug of war to propertize these resources, some of which had acquired commercial (and strategic) value. One response was that some public culture collections began to devise material transfer agreements (MTAs) that progressively restricted access to, use of, and redistribution of their microbial materials for research purposes. A leader in this trend has been the American Type Culture Collection (ATCC), but many others within the WFCC have followed its example by imposing restrictions on access and use that one would expect to find in the private sector. For example, the ATCC's standard MTA for materials available to non-profit affiliates contains the following prescriptions, according to the ATCC model contract:

1) use in a single laboratory only;
2) no redistribution without permission;
3) limitations on derivatives;
4) a built-in reach-through claim on any and all derivatives (which replicates the copyright law's derivative work model); and
5) a need to negotiate and obtain permission for each pending transaction in connection with any given material.

Moreover, this highly proprietary model, which applies to not-for-profit and for-profit researchers alike, has been imitated by the other collections to varying degrees, with a view to addressing their potential relations with commercial clients. While some culture collections have tried to preserve space for public research, even here the tensions between making microbial resources available and limiting both access and use remain visible. One result has been the progressive privatization of upstream resources that are needed for research.

A second response was that, after 1992, the Convention on Biological Diversity (CBD) established the principle that genetic resources of every kind found within the territorial boundaries of nation-states are subject to the sovereign rights and control of those states. Under Article 15, access must be obtained on mutually agreed terms, subject to prior informed consent and to further obligations concerning benefit sharing and the transfer of end-use technology. Over time, the trend has been for developing countries to restrict access to their microbial resources for virtually all uses, including public research uses, and to assert ownership claims to resources held in the public collections of developed countries that were not obtained in conformity with the principles set out in the CBD.

Let us now ask what is fundamentally wrong with this picture? The growing trend on both sides of the development divide is to view each microbial resource as if it were potentially as valuable as gold. Sometimes, of course, when the proprietor has a known or likely commercial application for a specific microbe, strong restrictions on access, use, and reuse make sense. Some of the special collections deposited at ATCC provide examples of such a situation.

Generally speaking, however, the bulk of all microbial materials residing in public culture collections all over the world have no known or likely commercial applications whatsoever. In reality, the only value that the bulk of these materials possess is to serve as research materials, as inputs for basic scientific research. The hoarding and proprietary tendencies that increasingly predominate thus undermine and risk defeating the research potential of university research scientists everywhere. Academics depend on their ability to screen large collections of raw materials against leads developed in their laboratories either by phenotypical observations or by genetic analysis, or by some combination of the two. Needless to say, as we observed in an article regarding the hoarding of small molecule libraries (Rai et al. 2008), narrowed access to these upstream research resources actually leads to fewer commercial payoffs in the end, a situation in which everybody loses.

Not surprisingly, serious researchers have reportedly devised informal means of their own to ignore or avoid these restrictive practices. Such measures are especially prevalent at universities, which may hold a large amount of microbial resources at varying states of validation. To this end, single laboratories or research units informally exchange biological resources among themselves for public research purposes on the basis of mutual trust and reciprocally recognized quality controls, without entering into any formal legal undertakings. In effect, this informal network, which antedates the use of MTAs and reportedly still accounts for approximately 60 percent of all microbial materials exchanged, converts the private goods of the single participants into a type of "club goods" available to trusted members. This informal system thus tries to maintain the original sharing norms of the WFCC within a carefully selected group of likeminded researchers.

The result is an informal, closed semicommons based on “group loyalty” and reciprocity gains that drive a large but shrinking part of the research domain. Some characteristics of these informal exchanges are summarized as follows:

- Ad hoc verbal agreements are usually preferred to MTAs; and the standard norm is “use in lab only." This practice likely violates the rules of university technology transfer offices and transnational regulations. There is little or no tracking or independent authentication (with concomitant risks).

- These cooperative networks have generated some push-back against formal restrictions on research as well as some interest in more science-friendly MTAs, such as the one on which the Science Commons has been working.
- On the whole, without any unified MTA, these informal exchanges have diminished over time as more laboratories tend to restrict at least so-called commercial uses for fear of lost opportunities, especially a lost blockbuster.
- The end result is unacceptably high transaction costs and growing restrictions on freedom of research with the concomitant risk of fewer commercial payoffs for both researchers and the public at large.

To break out of these two existing options—overly restrictive formal legal MTAs or loosely organized informal exchanges based on mutual trust—we offer a third option. Adopting a formula Paul Uhlir and I first put forward in relation to data exchanges in 2003, we propose to "formalize the informal sector" on a more research-friendly basis, by adopting standard contractual templates, i.e., to devise a contractually constructed research semicommons for publicly held microbial resources all over the world. (Cf. Reichman & Uhlir, 2003).

Designing a Third Option

In order to build a third option on a solid legal foundation, there must be a standardized material transfer agreement that contractually regulates the relations between all the participating microbial research communities and their members. Enforcement of such a standard-form agreement would be the province of a governing body or trusted intermediary that is generally responsible for oversight and management of the projected microbial research infrastructure.

A key premise underlying this initiative is that any deposit in the proposed materials semicommons does not forfeit all rights to benefit from downstream commercial applications, if they should emerge later on. This premise is necessary to heighten the potential reciprocity gains from participating in the research pool by directly addressing fears of losing commercial opportunities later on.

Participants will not normally contribute materials that have known or likely commercial potential. Why not? Because if a participant's microbial material already has a known or likely commercial application, he or she stands to gain more from holding out than from the basic research opportunities flowing from participation in the semicommons. (See Minna Allarkhia's presentation summary in this volume at Chapter 20.) Those high-value materials will logically flow to ATCC and other entities equipped for this purpose.

Hence, the third option that we envision would receive only deposits of microbial materials that lack any known or likely commercial uses at the time of deposit. This premise would capture the bulk of all materials in the public culture collections, not to mention the unvalidated resources outside those collections in universities, government agencies, or *in situ*.

There would be various other criteria for materials to be included in the pooled collection we envision. For instance, high-quality standards should be set and maintained for admission, including measures for authentication, validation, and tracking. To qualify,

participants would have to meet these standards (through which the semicommons continues to expand). However, many university laboratories would not qualify without upgrades.

Within the semicommons, there would be no restrictions on upstream public research functions with respect to deposited materials except that specimens could only be redistributed to member collections meeting the agreed quality standards. That is, we start with a built-in, absolute research exemption to achieve the broadest possible upstream research domain with no unworkable distinctions between commercial and noncommercial research. Instead, we would address future commercial outputs directly, as will be seen.

There would be strict and careful tracking of all uses of materials from the pooled collection in order to maximize the scientific verifiability of results, as already occurs in some form at most WFCC public culture collections. The attribution and reputational benefits of all depositors would be preserved to the fullest extent possible, and, of course, biosafety and security regulations must be fully observed.

The economic logic underlying this model is that the providers of microbial materials would presumably obtain more potential reciprocity benefits from the vast upstream research opportunities generated by the semicommons than would accrue from operating in isolation. But fears of losing unknown future commercial opportunities could undermine the prospect of these potential research gains, so we address this concern directly with a built-in provision for benefit sharing from unknown future downstream commercial applications. That is, we would build a so-called “compensatory liability rule” for downstream commercial applications into the system, yielding equitable compensation for the providers or their designate representatives under international law, while fulfilling international obligations under the CBD. This component thus gives providers a means of securing equitable compensation from future commercial applications, unknown or unlikely at the time of deposit, that ultimately resulted from research uses of the deposited materials. A liability rule means that one may freely take the materials for any research purpose, without need of any permission to use, on condition that a duty to pay equitable compensation arises if and when the application itself accrues commercial gains. It is neither an “absolute permission” rule (i.e., comparable to an exclusive property right) nor a pay-per-use rule; it is a “take and pay” rule, as I will explain in a moment.

Recall that only about one percent of the world’s existing microbial population has actually been identified. Most of those identified microbes possess largely unknown properties and characteristics of no commercial interest. Some of these are held by private industry in collections whose contents have not been publicly certified. One may accordingly describe the bulk of the holdings in existing public microbial collections as "pre-competitive" in the sense that they have elicited little active scientific interest at the present time and—perhaps for that very reason—have no known or likely high payoff commercial applications.

In other words, the bulk of the microbial materials currently and prospectively to be held in public collections may properly be characterized as building blocks of future knowledge. Yet, the existing MTAs applicable to these materials tend to impose restrictive conditions on use and reuse that make scientific research costly and difficult to conduct. Such constraints particularly impede collaborative research involving large and diverse microbial populations that may be subject to high-throughput screening or other

advanced research methods (especially when computational biology and the use of automated knowledge discovery tools is envisioned).

The payoff from our proposal is that any scientist authorized by dint of his or her connection to a participating institution could roam and explore a vast expanse of known microbial research space, with a view to maximizing future abilities to add to, identify, and develop value-adding contributions about specific contents. This would include the linking of federated, distributed collections in a virtual semicommons by digital means.

Nevertheless, it remains possible that access to the pooled resources, which are made available to scientists under a regime of minimum research restrictions, could lead to the discovery of a later commercial application of a given material that was neither likely nor foreseeable in advance. Such discoveries are welcome contributions to social welfare. They are a product of skilled efforts and the investment of time and labor, which should not be confused with parasitic or free-riding uses that undermine incentives to innovate.

To internalize and capitalize on these gains, we seek not only to develop a broad research commons, but also to encourage investment in downstream commercial applications by careful use of *ex ante* liability rules that do not impede downstream patents on end products. Liability rules are "take and pay" rules (like the non-exclusive licenses covering the Cohen–Boyer patents), not absolute permission rules, like exclusive rights. The latter only work well when the values of potential uses are known relatively well *ex ante*. They do not work when each party over-values his or her property because nobody knows its true worth (which is what we have here). (See T. Lewis and J. H. Reichman, 2005).

With a liability rule, the message is not, "You cannot use my microbial materials for commercial purposes." It is instead the opposite: "Please find commercial uses for my research materials, and, when you patent the end results, please pay me a reasonable royalty from your gross sales." Notice that this is not a compulsory license *ex post*. It is a built-in automatic *ex ante* license to use and pay—a pre-existing obligation to share a small percentage of any eventual economic returns with depositors and with the culture collections that maintain and regulate them, both of which contributed to the downstream commercial payoff. (Cf. J. H. Reichman, 2001). Notice, too, that there is a built in possibility of lottery effects if many downstream commercial applications spin off from any given microbial resource as actually occurred with the Cohen –Boyer patents.

The end result we propose is thus a kind of built-in public-private partnership. The culture collections from which the microbes were taken would manage any resulting income streams from downstream applications. We envision relatively low royalty rates—2 to 4 percent (comparable to those used in Canada when medicines were subject to a license of right). A part of this revenue stream should go to the collection to help defray its costs. The rest would go to the upstream scientists and laboratories that provided the materials (thus enabling downstream commercial applications) or to designated authorities in developing-countries for materials deposited under the CBD. No depositor would be able to write or limit research restrictions, however. Full use of deposits for all scientific research is the *sine qua non* of participation and cannot be bargained away. Such an approach would thus improve on the Food and Agriculture Organization's International Treaty on Plant Genetic Resources for Food and Agriculture (2001), which was the first international convention to codify a "compensatory liability regime" along the lines I first advocated in 2001. Finally, governance rules, including mediation and dispute resolution, would be needed.

To recapitulate our basic thesis, start with the notion that liability rules—"take and pay" rules—go beyond current applications of club economics to the formation of knowledge semicommons by avoiding the limitations of exclusivity. At the same time, such rules still provide enough economic incentives to contribute to the pooled resources from which scientific researchers would derive more benefits than they could obtain working alone. This approach provides an intermediate zone, where Creative Commons licenses are insufficient but exclusive rights and concomitant restrictions on research would impose unnecessary overkill in relation to the still uncertain value of the upstream inputs. Nevertheless, the liability rule is triggered via standard-form licenses that keep transaction costs low, in the manner of a Creative Commons license. And the use of liability rules fully preserves and promotes the benefits of the collaborative research model. In fact, it enables that research model by alleviating fears of lost downstream gains.

The approach we propose must further be reconciled with three other unresolved problems:

1. What Fiona Murray calls the "Big Refrigerator Problem." How much capacity can the public culture collection muster, and how does one reach more of the unvalidated materials held by academics outside of the public culture collections? We envision a federated network of distributed collections building on the World Federation of Culture Collections prototype that would be managed within a single, overarching governance scheme.

2. What will be the future impacts of genetic research techniques on the collections and on the scheme as a whole? We think it likely that upstream genetic resources can be pooled and managed under a similar framework. In any event, the evidence suggests that genetic research results must be squared with living and evolving cultures over time.

3. Will we need an international treaty, like that of the International Treaty on Plant Genetic Resources for Food and Agriculture, which is governed by an international organization under the auspices of the Food and Agriculture Organization? We hope it would not become necessary to go outside scientific circles, but some, such as Michael Halewood at this symposium, think an intergovernmental organization will be necessary.

In the future, the long-term goal should be that culture collections become a means of combining materials, plus relevant literature and data, into a digitally integrated whole. This would be the true long-term payoff from creating the materials semicommons, and it is discussed in the later chapters of our forthcoming book.

REFERENCES

Helfer, Laurence R. 2005. Using intellectual property rights to preserve the global genetic commons: The International Treaty on Plant Genetic Resources for Food and Agriculture, in *International Public Goods and Transfer of Technology Under a Globalized Intellectual Property Regime* (K.E. Maskus & J. H. Reichman eds., Cambridge U. Press, 2005).

Lewis T. and J. H. Reichman 2005. Using Liability Rules to Stimulate Local Innovation in Developing Countries: Application to Traditional Knowledge, *in International Public Goods and Transfer of Technology Under a Globalized Intellectual Property Regime*, Ch. 13 (K.E. Maskus & J. H. Reichman eds., Cambridge U. Press, 2005).

Rai, A.K., J. H. Reichman, P. F. Uhlir, and C. Crossman. 2008. Pathways across the valley of death: Novel intellectual property strategies for accelerated drug discover. *Yale Journal of Health Policy, Law & Ethics* 8:53–89.

Reichman, J. H. 2001. Of green tulips and legal kudzu: Repackaging rights in subpatentable innovation. *Vanderbilt Law Review* 53(6):1743–1798.

Reichman, J. H., and P. F. Uhlir. 2003. A contractually reconstructed research commons for scientific data in a highly protectionist intellectual property environment. *Law and Contemporary Problems* 66:315–462.

Reichman, J. H., T. Dedeurwaerdere, and P. F. Uhlir. Designing the microbial research commons: Global intellectual property strategies for accessing and using essential public knowledge assets (Cambridge Univ. Press, forthcoming 2013).

Question and Answer Session

PARTICIPANT: With regard to the royalty amounts and the decision between having something on the order of 2 to 4 percent, how is that decided, and then what happens when other technologies intervene or other licenses come into the finalized product? How do you accommodate the royalty stacking provision?

PROF. REICHMAN: Well, there would have to be a royalty stacking provision. There is something at the end of our draft chapter that talks about the model, but we have not yet actually worked through the model to do it. You need a royalties stacking provision, and you need a mediation and dispute resolution arrangement. It may well be that different communities would have different royalty values.

The International Treaty for Plant and Genetic Resources put a very low royalty on their pooled genetic resources, but they broke new ground. This was the first time it had ever been done at the international level, and I think they decided on a 0.5 percent royalty. I think that was a mistake, but if you are serious that the primary object is

research, not gain, the point is to not be left out of the game so that you can support the research. You do not want to encumber these downstream licenses with a heavy reach-through amount, otherwise you get the pharmaceutical sector on your back and you may have lots of problems.

I published an article in the Vanderbilt Law Review in 2001 called "Of Green Tulips and Legal Kudzu" (Reichman, 2001). It is about repackaging rights in sub-patentable innovation generally, and I introduced the "compensatory liability regime" there. So, in different contexts you might have different percentages of royalties, but I do not think they should ever be very large because, by definition, the materials in question are so far upstream that you have not added enough value at this point to justify a commercial payoff. If you are fairly certain that there will be a commercial application, then you will need a tailor-made license, and the microbe will not enter the research pool. You will know what it is worth. You will have enabled it later on. So, we think 3 to 5 percent is not too much, but not too little either. Increments above a baseline royalty of 3 percent could depend upon the amount and quality of valuable information, if any, that is disclosed with the microbe. You will notice that in government use licenses of patented inventions, the U.S. government usually offers 4 percent as a starter. It generally never goes above 6 percent, although one or two licenses went to 10 percent. I am an expert on government use licenses. So, in that range, we think ours is a pretty fair estimate *ex ante*.

MR. HALEWOOD: I am with Bioversity International, which is one of the international agricultural research centers supported by the Consultative Group on International Agricultural Research (CGIAR). I have two questions, or one point and a question. One has to do with, under the treaty, the swapping of the right to keep things open for research downstream and the requirement to pay. In the plant-breeding community there have traditionally been two forms of intellectual property rights. One is plant variety protection rights, which allow downstream research, and the other is patent rights, which do not. People who were very much involved in the negotiations of the treaty wanted to, in a sense, punish the patenters and give support to those who were still resisting that. So that is why they wanted to have a mandatory benefit-sharing clause that complemented the pre-existing division, if you will, on one side.

The other thing I wanted to ask was something you wrote about in your paper and hinted at in your presentation. You talked about membership or conditions of entry into the pool, and you linked that to quality management criteria and said that that would limit, initially at least, the number of organizations that could be members. I wonder if you could not split that in two, looking at this from a development perspective. You could have a limited pool of entrants as suppliers, given their need to respect and meet high standards, but recipients could be global and anybody. I do not know why you could not make it that way.

PROF. REICHMAN: That is a very interesting proposal. I am not against it, especially if you were thinking of developing countries. I can see some problems about free riding though, and you have to make sure that the people who go in think they are getting enough out of it. There is some recent research that came up at the COMMUNIA workshop in June of this year. There were some economists from Germany who had done very good empirical research verifying the real importance of this reciprocity hypothesis, and I am a little nervous about undermining the reciprocity gains expected from membership in the pool at the beginning, until it is established. Once it is established,

then everybody can see the payoffs, as is the case with so many of these other genomic commons, and then you can relax the admission standards. I certainly think your goal is desirable, and I would endorse it. I just think it has to be handled carefully because, at the moment, you are trying to get the people who are already in it to stay in it.

Another limiting factor is that the pooled microbes can be widely distributed for in-lab research, but they cannot be widely redistributed without quality and security controls, including authentication, validation, and the tracking of all uses. This limitation is built into the present-day microbial research model, to preserve the purity of research results.

PARTICIPANT: Could you quickly clarify how you would handle third-party transfers? Are they allowed?

PROF. REICHMAN: Third party transfers would be handled entirely by the downstream people in the normal way that it is being done already. If you are in the stream and you come up with something that now is known or likely to be profitable, then you are going to negotiate that deal with third parties out there. Of course, the tracking of the microbes in question must be used. But when you are transposing into the downstream world, you will still owe the commons under a "reach through" liability rule. You do not owe them anything else, however, and you are free to do what you want with your work.
Then what do you do? You go to the pharmaceutical company, you go to the fertilizer company, or you go to the beer company, and you negotiate your own deal, and they are going to go through the patent process, the clinical trial process, the whole nine yards. Everybody is aware, however, that there is a built-in reach-through agreement that must be respected in the end. That agreement does not say that you have to negotiate with me. It says you have to pay me a reasonable royalty from your ultimate gross returns, and that passes on all the way down the line.

PARTICIPANT: I come back again to the question about royalties. It is a royalty-stacking question. There are two kinds of royalty-stacking issues. One is that there will be other technologies that do not involve the use of microbial material, which will be under patent or copyright, and there will be charges. But let us put that aside. It is a question of the science: In how many cases is the model of innovation that you go from a single microbe to a commercial use? That is sort of the model in conventional chemistry.

PROF. DAVID: If you need to use an ecology of microbes, then the question of royalty-stacking arises.

PROF. REICHMAN: Yes, that is right. You have to address it.

PARTICIPANT: Who would have jurisdiction? When I first read about this provision that you put in, I said, No, this is not an *ex-ante* liability rule. I am not a lawyer, but from what I have learned from working with lawyers, the liability rules work on the basis that you show essentially that you have been injured by the use.

PROF. REICHMAN: Well, that is the tort law origin of them, but we are adapting it to intellectual property. It becomes an *ex ante* entitlement, i.e., a built in, automatic license rather like a non-exclusive license, and it has to do with notions of equity, not injury.

PARTICIPANT: Well, no, it is an adaptive view, but what it is, is a pass-through license without claiming intellectual property.

PROF. REICHMAN: Yes and no. It is a form of intellectual property rooted in liability rules rather than exclusive rights. You underestimate the fact that it remains an *ex ante* entitlement, like all other intellectual property rights, some of which are non-exclusive from the get-go (e.g., trade secret protection).

PARTICIPANT: So, you have a pass-through license. Now, the question is: Since you want to encourage downstream use, why are you not using a flat fee? Because, first of all, incentives like flat fees are less distorting. It is a lump sum payment. It is predictable. It gives greater incentive, on the one hand, to people to go for a use which will be very commercially attractive because it is a fixed fee. The fixed fee can be justified on the basis of the costs of meeting the quality standards, of devoting time to the curation and to the running costs of collections. It also has another positive feature that it is administratively easier. You do not have to have these *ex-post* negotiations to deal with stacking problems because you are not seeking a percentage of the revenues. And that is economically more efficient because it reduces the pricing of the end products, and therefore tends to increase the ultimate social benefit because you do not have the dead weight.

PROF. REICHMAN: Yes, it has some advantages. These are friendly comments.

PARTICIPANT: Yes, and we want to make it better.

PROF. REICHMAN: These questions are coming from one of the world's leading economists in this area, so I am very grateful for them. On the second question, the problem is that you do not know enough to assess a flat fee. It will likely be either too little or too much, like literary authors who should not accept a flat fee for their books. You really do not know in advance, and you have this "lost blockbuster" complex. The flat fee might work in the example that you gave this morning about the specialized collections. Flat fees sometimes work when you are just making a research tool available, like the Cohen-Boyer patents. You would probably get a lot more action and a lot more science done with them. This is practical, however, because you know the value of the research tool in advance.

But, here, how do you know at the beginning what value to put on microbial resources? It is like the .5 percent in the treaty on plant genetic resources. Well, 0.5 percent in certain cases could be all right, but in most other cases it looks too low. So, you are worried about the blockbuster. Well, if your microbe turns out to be part of the cure for cancer, then you will want a piece of that action.

PARTICIPANT: Well, that is a different case because you are trying to get a piece of the action and not promote downstream use.

PROF. REICHMAN: You are trying to promote it, but I do not think you are hindering it. You do have a possible royalty stacking problem, however. We found that in our "Pathways Across the Valley of Death" article, as well, and I think we have some

innovative solutions. In that context, you could have a group of small molecules of which only one is going to be the winner, but all four were necessary to arrive at the winning result. We go out of our way to state in that article that all four of those molecules participate in the liability rule, but you cannot stack the royalties, while the winner gets a tailor-made license with the patentee.

There is not a separate valuation for each relevant microbe. They divide the proceeds from the liability rule, but the winner gets a negotiated contract with the patentee, and the losers get a share of the return based on the *ex ante* liability rule. In our "Pathways" article, we suggest that that is a workable way to defer a lot of the risks, to spread the costs of the risk premium that people are getting because if you had one of the relevant small molecules but not the big one, at least you would get a small piece of the action, and that helps you cover the losses when your clinical trials go bad.

In some of these and other cases, we will need the mediation and dispute resolution mechanisms discussed in our book. I do agree with you that the stacking problem will require an express provision. So, I thank you very much for that and thank you for your questions.

7. The Agricultural Research Service Culture Collection: Germplasm Accessions and Research Programs

– Cletus P. Kurtzman[24]

National Center for Agricultural Utilization Research, USDA/ARS

In this talk I will describe our collection and some of our operations in order to give you a perspective on what we are doing, and it also may address some of the issues raised by this symposium. The Agricultural Research Service (ARS) Culture Collection, as it is formally called, which many of you know as the NRRL (Northern Regional Research Laboratory) collection, was established by the U.S. Department of Agriculture (USDA) in 1940, when our laboratory opened. Its mission is, basically, to collect, maintain, and utilize microbial germplasm for agricultural and agro-industrial uses. Let me emphasize the "utilize" part of that because our laboratory was set up to utilize agricultural products, and the driving force both for the Center and for the collection has been to utilize microbes to convert agricultural commodities into higher value products.

The collection started out quite small, in 1940. Most of the collection was brought to us by the original group of curators. It started with just a few thousand strains, but over time it has grown quite a lot, and we have about 9,000 actinobacteria, about 10,000 of the "standard bacteria," about 53,000 filamentous fungi, and about 15,000 yeasts. We also have a patent collection of about 6,000 strains. The U.S. Patent Office asked us to accept patent cultures in 1949, which we did, and it was at about that same time that the American Type Culture Collection did as well.

Within the U.S., these are the two official patent culture depositary authorities under the Budapest Treaty, NRRL and ATCC. We at NRRL distribute about 4,000 strains annually. Our web site went up several years ago, and when it was put online we expected that it would receive a great deal of attention and we would be overwhelmed with culture requests. In fact, however, it did not much increase the number of requests; it simply clarified the requests.

Over time, we have accumulated cultures from a variety of sources, some representing abandoned collections. One source was the Charles Thom Collection, for instance, and if you are a mycologist, you will immediately recognize that collection as a source of *Aspergillus* and *Penicillium*. That collection came to us in Peoria, in part, and to ATCC, in part.

Harvard University contributed a very nice collection of mucorales. The N.R. Smith Collection of bacilli went to both of our collections, I believe. The A. J. Mix Collection of *Taphrina* species, a group of plant pathogens, was received and is often sent to plant pathologists. We also received a collection, which Howard Dulmadge assembled over his entire career, which contains about 2,000 bacilli, including *Bacillus thuringiensis* and *Bacillus sphaericus* strains, which are the microorganisms used for biocontrol of insect pests on crops.

---

[24] Presentation slides available at: *http://sites.nationalacademies.org/xpedio/idcplg?IdcService=GET_FILE&dDocName=PGA_053611&RevisionSelectionMethod=Latest.*

Jack Fell from the University of Miami recently retired and sent his collection of about 4,000 or 5,000 marine yeasts to us. In the early 1970's, we received about 6,000 biodegradation fungi that the U.S. Army Quartermaster had collected. Those were collected during World War II, mainly in the South Pacific, where tents and clothes and other fabrics were falling apart faster than they could be manufactured. For a while there was little interest in these fungi, but recently, with the interest in biomass conversion, these strains are attracting attention as a group of organisms that could be really useful for breaking down cellulose and other fibers.

Over the years we have also had a number of research programs that netted us literally thousands of cultures related to food safety, microbiology of cereals, and so forth. And, finally, a number of our cultures have been contributed by scientists who asked that we maintain them because they are part of their publication process.

BOX 7-1

Issues for Germplasm at Risk

Abandoned Collections
- Who will decide their value?
- Who will take them?

Research Materials – Deposit of strains
- Key strains should be deposited in culture collections and distributed without restrictions because these cultures represent part of the materials and methods of the published research and are therefore essential for verification and extension of the findings.
- Will journals enforce this concept by requiring that subject cultures be deposited in a recognized culture collection and free of restrictions on distribution?

Research Materials – Undeposited strains
- How does one ensure that cultures cited in a publication will be available to other investigators when the culture is available only from the investigator who published the paper? What if the culture is lost by the investigator? What if the investigator will not share the culture after publication?

A couple things are clear about abandoned collections (Box 7–1). One is that a huge amount of money was spent gathering these collections, and each may have taken someone's whole career to assemble, often using quite a lot of support from the National Institutes of Health or other agency. But once the scientists retire, their collections are candidates for the trash heap. This is incredible waste, but this seems to be a common problem nationally and internationally. So, where it was possible, the ARS Culture Collection has taken some of the more prominent abandoned collections.

The other observation that can be made about abandoned collections is that with their varied history—their varied investigators, substrates, and contributors—it is not clear who owns these cultures. The U.S. government certainly does not own them. We at ARS maintain them, but could the heirs of Charles Thom claim them, for instance?

So, there is an interesting dilemma if one of these abandoned cultures becomes important biotechnologically: Who should get the payoff? How do you deal with this?

Our philosophy is that we are here to maintain these cultures the best that we can and to distribute them to requestors, no strings attached in terms of any biotechnological application. If the requestor makes a brilliant discovery that brings, a large financial return, we wish them well. Obviously, that is a different philosophy than we see in certain other collections.

From our general collection, we distribute strains per request, but not more than 24 strains per year per person. We ensure that the Animal and Plant Health Inspection Service (APHIS) permits and other necessary permits are provided. The reason for the restriction on the number of cultures distributed is that, according to ARS-USDA policy, we cannot charge for this service, so our resources are limited, and this restriction allows us to live within our budget. The patent collection is a different matter. Since it is covered under the Budapest Treaty, requests for cultures from the patent collection are governed by the rules of the Treaty.

Our cultures are preserved primarily by lyophilization, a simple freeze-drying process. If the cultures do not survive well under lyophilization, they are preserved with liquid nitrogen. For patent cultures, we use both preservation methods.

Among our staff, curatorial duties take up about 10 to 20 percent of each scientist's time with the remainder of the time devoted to research. This model, which has been in place for quite some time, allows us to have professional microbiologists providing oversight to the particular collections instead of having less-trained people looking them over.

One conflict within our agency—and, I suppose, everywhere else—is whether we are spending too much time on the collection. Why are we not spending 100 percent of our time on research, given that we are a research agency? The argument for this arrangement is that it allows us to provide professional microbiological oversight to the collection. It is generally very difficult to find good people to look after a culture collection, especially since the number of taxonomists has been declining, and they were probably the primary source of curators in the past. So when we make a new hire, we explain that staff members can have 80 to 90 percent of their time to work on research but they must spend 10 to 20 percent of their time maintaining the scientific aspects of the collection.

One result of this arrangement is that, over the years the collection has been linked to developmental research programs and has been responsible for a number of discoveries, such as finding the production strains for penicillin on a cantaloupe, as was mentioned in an earlier presentation. Other important finds included large-scale production of xanthan gum using *Xanthomonas campestris*, and use of *Leuconostoc mesenteroides* for production of dextran gum, which is used in emergency rooms for quick fluid buildup in accident victims. Riboflavin production from *Eremothecium* came out of our cooperative interactions with NCAUR chemists, as did production of beta-carotene from other fungi. The first yeast known to ferment pentoses (D-xylose) was discovered at NCAUR, and that was important in conversion of biomass to fuel alcohol. Finally, much of our recent work has been on diagnostic gene sequences, which I will discuss shortly.

Figure 7–1 shows a typical storage of microbial cultures and a refrigerator where the cultures are stored at about 4 degrees Centigrade. Each of the little boxes has about 10 to 12 lyophil tubes in it.

FIGURE 7–1 Typical storage arrangement at the ARS Culture Collection.
SOURCE: Cletus Kurtzman

Note, these examples are from research conducted by the scientific staff of the ARS Culture Collection, but that point seemed to have been lost, hence the significant rewrite of what follows. You may recall that I showed photos of staff members during these particular examples.

As an example of how the collection and the ongoing research may be used, about three years ago there was a recall issued for a contact lens cleaner produced by a prominent pharmaceutical company because users were getting corneal infections caused by fungi. It turned out that the company had reformulated its cleaner, and, unfortunately, the new formulation was a good growth medium for *Fusarium*. The question was, which *Fusarium*? Kerry O'Donnell , of our group, and David Geiser at Penn State, who both work with *Fusarium*, developed an enormous database of gene sequences to study plant pathogens, and using this database, they were able to quickly identify which *Fusarium*

species were causing the problem. That allowed development of a treatment and a solution to the problem.

Todd Ward, also a member of our group, developed an extensive multigene database for *Listeria monocytogenes* and combined this with Luminex Technology for rapid diagnostics. The interest in this rapid diagnostic technology has come not only from the food safety group within ARS but also from the food safety group at the Food Safety and Inspection Service and from the CDC because of the variety of gene sequences used.

Alex Rooney, in our group has worked with *Clostridium*, *Bacillus*, and *Salmonella*, and has played a role in trying to characterize the source of the 2001 anthrax attacks known as Amerithrax. I do not know how widely it is known, but there was a *Bacillus* sp. contaminate in the *Bacillus anthracis* that was released. We have an enormous *Bacillus* collection, and the Federal Bureau of Investigation contacted us and asked us to provide multi-gene sequences for all of our *Bacillus* strains in order to determine if the contaminant was something unique. As it turned out, the strain that contaminated the anthrax preparation was not unique, so, we could not track the contaminant based on population genetics, but it could have proved quite valuable.

In short, the culture collection holdings and its interactive research have a lot of possible uses. My work is primarily on food safety in the context of food contaminant organisms, but I am also involved with biocontrol organisms. We have also developed a barcoding system for yeasts that seems to have triggered greater interest for its use in clinical diagnostics than in agriculture.

The recurring theme here is that most of the organisms we work with are dual purpose. Many of them are important in agriculture and biotechnology, but many within the group are also human and animal pathogens, so our work on them draws quite a bit of interest from the medical community as well.

Many of our challenges and concerns are similar to those facing a microbial research commons. One challenge is cost recovery for strain distributions. Of course, we are hampered more than others because we cannot charge, but even for those who do have a fee, the question remains of whether the charges can be set high enough to recover the costs or whether some type of supplementation will be required.

Costs for long-term maintenance—refrigeration, liquid nitrogen, and so on—are not cheap. Getting sufficient funding for qualified staff is another issue. It is hard to get people who are well-trained and who are willing to work in culture collections. Funding to characterize the germplasm can be a problem, as well. Of the approximately 90,000 strains that we have, about 11,000 can be put out on our website because they are either type strains or because we characterized them from at least one gene sequence, so we feel we know what they are. The remainder has not been genetically characterized, so there is no simply no point in putting them out and misleading people. The solution to this would be for us to get additional funding to identify all of those strains properly.

There are also the costs for backup sites for collections and strain data. Such backups are essential, but this space is a challenge for everybody.

Abandoned collections raise still another set of issues. Who is going to decide their value, and who is going to take them? What collection has the capability to do that?

Finally, research materials and published strains pose a different sort of challenge. Most scientific journals ask that the authors provide upon request the germplasm that is the subject of their paper. The germplasm is actually part of the materials and methods. Furthermore, in 1949 the U.S. Patent Office decided that microbial patent applications needed to be accompanied by deposited germplasm in order to substantiate their claims.

For chemical processes, we can develop some sort of a formula that conveys how that product is made, or for mechanical patents we can have a drawing of a little machine, and that serves as full disclosure. But for microbials it is not workable to describe on paper everything that is known about the organism and expect others to be able to reproduce it. So a culture must be deposited as a part of microbiological process patent.

Indeed, the key strains should be deposited for every publication. If that does not happen, the strains may be lost by the investigator, or the investigator may decide not to share the strains after all, and then there will be no validity to that research because it will not be possible to reproduce.

I do not know if this is something that the discussions within this group can help with, but I would like to think so. Will the journals force this concept? It is hard to say. I was at a meeting in January organized by the American Phytopathological Society, and the people in that organization are very concerned about where to put their germplasm. They are also concerned about whether they can get a variety of plant pathogens for their own research. So I raised this issue with them. In principle they liked the idea of depositing all the strains reported in their publications because they would like the cultures to be available, but they were worried that if they make their cultures generally available someone will "steal" their research.

In short, this is a universal problem and concern: Depositing strains in a public collection will make it possible for others to profit unfairly. This really is unfortunate because that is counter to the idea of publishing to begin with. Clearly, though, this is something that we need to deal with.

Question and Answer Session

DR. RAINEY: Fred Rainey from Louisiana State. NRRL is my favorite collection because who is going to complain about a collection where there is no paperwork and no request for payment? But I have a question for you. Why does NRRL not have an MTA, while all of these other collections do? Is it something to do with it being a government agency? How did that decision come about?

DR. KURTZMAN: Yes, that is a good point. We actually do now have a simple MTA. It was developed about a month ago, and it came up because USDA was concerned about the safety aspects of culture distribution. For the few BSL-2 organisms that we maintain, such as *Listeria*, we had already asked requestors for certification that their lab was equipped to handle these cultures. The thrust of the MTA is the requirement that the recipient of the culture is a competent microbiologist who would handle the culture safely and not to pass the culture along. Passing it along does not bother us, but one reason for not passing cultures from any collection is that the person you get it from may not have faithfully transferred it. But, other than that, it is not a problem. We do not ask recipients to sign an MTA, but we simply say that by sending the culture with this short MTA, the recipient accepts the conditions by opening the package. The MTA puts no strings on any technology that might arise from using those cultures.

Not being a lawyer, I have no idea whether this simple MTA would stand up in court. I suspect it is a little dicey, but it is simple and very transparent. I did not write the MTA and it may be subject to future revision.

PARTICIPANT: I am from the Fungal Genetics Stock Center. We have talked about this a little bit before. You mentioned that the U.S. government does not claim ownership in these materials, and you asked if the heirs to the people who collected them would own them, but, presumably, they were also originally scientists paid by a public entity.

We have the same issue with our collection. We do not really claim that we own these materials, but we are responsible for them. I was wondering if you had any further thoughts on that.

DR. KURTZMAN: No. I suppose anybody could challenge anything in court, can they not? And Charles Thom's heirs may come along and say that because somebody in Peoria was trained by Charles Thom, they went out and recognized the right culture from a molded cantaloupe and saved the world, so they should somehow get a payback. I know that sounds extreme, but there are many possibilities, and I do not want to go into all of them. From my own perspective, I think it is kind of silly because, in most cases, the advances and discoveries that are made from these cultures are ones we could not predict. Now, if you and I are contacted by somebody not for the cultures *per se*, but for suggestions on the research, I would say we might be co-investigators on a project, but not because we simply supplied the germplasm.

Post symposium note: In November 2010, ARS decided for budgetary reasons that technical operation of the ARS Culture Collection was to be only by scientific support staff.

8. American Type Culture Collection: A Model for Biological Materials Resource Management – Frank Simione[25]

American Type Culture Collection

I am not going to talk about what we do at the American Type Culture Collection (ATCC), but rather about the question of why we do it. I have been with the organization for a long time, and I have watched the evolution of what we do. In my opinion it is a successful model because we have been able to survive and thrive as an organization by using it, although some will contend it is not an ideal model

I will begin with a brief overview of the ATCC to show you how we evolved to this point, and then I will describe our operational focus, which offers some insight as to why we operate the way we do. The ATCC is a non-profit 501(c)(3) organization under the U.S. Code, and we are not part of the government, although some people who do not know us well think that we are. ATCC was founded in 1925. However the core collection really originated with the Winslow Collection at the Museum of Natural History in New York prior to 1925, so we can claim nearly 100 years of experience and expertise.

As mentioned in an earlier talk, in 1899, at the first meeting of the Society of American Bacteriologists a group of bacteriologists met and had a discussion about creating a collection of biological materials that the microbiologists could contribute to and draw upon. C.E.A Winslow started that collection at the Museum of Natural History, which eventually became the ATCC. It is now a diverse collection in life sciences that does not consist only of microbes, but also includes cell lines and other biological materials, because we are continually trying to keep up with what is going on in science.

We are, along with the Agricultural Research Services Collection (formerly NRRL), one of the first International Depository Authorities (IDAs) officially recognized in 1981 as a depository for materials used in support of life science patent applications. We are also a global distributor with the capability to move materials all over the world. For example, we have a contract with the Centers for Disease Control and Prevention (CDC) to develop an influenza reagent resource and manage it, and part of our service included providing support for the response to the H1N1 virus outbreak when it hit in May of this year. In response we assisted the CDC and the World Health Organization (WHO) in delivering diagnostic kits to laboratories in 133 countries in a very short time period. The WHO noted that this was the fastest response to a global disease outbreak it had ever experienced. ATCC did not develop the kits, however we assisted in their manufacture and assembly, and we distributed them.

ATCC is ISO 9001 certified, and ISO Guide 34 and ISO 17025 accredited. ISO Guide 34 establishes the requirements for reference material standards manufacture, and ISO 17025 establishes the requirements for laboratory testing. These accreditations support the activities of ATCC as a material standards provider. We are also accredited by the American National Standards Institute (ANSI) as a Standards Development Organization (SDO). ANSI accredits organizations that develop written consensus

[25] Presentation slides available at *http://sites.nationalacademies.org/xpedio/idcplg?IdcService=GET_FILE&dDocName=PGA_053677&RevisionSelectionMethod=Latest*.

standards, and we believe that we are the first organization accredited to develop written consensus standards for biological materials. Although there are other organizations engaged in managing biological materials, few of them are developing written standards.

We receive no government subsidy for our collections, and all of our financial resources are generated either through the distribution of the cultures and related products, or from other activities in biological materials management. For a time in our history we did receive some subsidies; however ATCC no longer receives any support from the government for any of its collection activities. ATCC started in Chicago, however it has moved to several other locations in its history including three locations in Washington, DC, until a more permanent home was established in Rockville, Maryland in 1964.

In 1973 ATCC nearly went out of business, and since I joined in 1974 I was not there at the time, but I have heard many of the details. Financially ATCC was in trouble, the director was fired, and the government stepped in and bailed us out, however noting that the support was temporary and that the organization needed to get its act together. Despite this caution we did continue to get assistance for some of the later programs we developed.

With help from a government grant we were able to expand our facilities in 1976, and in the 1980s we began to wrestle with the issue of protecting intellectual property and equitably transferring the biological materials. In 1982 we received some additional government funding in the form of a grant to assist with getting some of our scientific capabilities updated. Data management, particularly with regard to the handling of molecular biology materials, became a critical need at this time. By 1993 our facilities were in bad shape, and we did not have the financial resources to upgrade or relocate. This is typical in a non-profit model where income gets used as it is received; therefore you put your hand out and ask somebody for money.

That is when Dr. Cypess arrived at ATCC, and began turning the company around by capitalizing it through better fiscal and operational management, and we eventually ended up with new facilities in Manassas, Virginia. We are now financially very strong and our products and service offerings include not only the traditional cultures, but also derived materials like DNA as well as other related products. We also have a major division that manages government and commercial contracts, generating additional revenue for support of the collections by leveraging our core capabilities in biological materials management. These activities provide much needed infrastructure support, allowing us to pay our bills and enabling the construction of new facilities.

There are three areas of operational focus for the ATCC that are of particular interest here. First, we seek to protect our biological assets to ensure their continued availability in support of the advancement of life sciences. As some earlier speakers mentioned, a number of biological collections have disappeared over the years because they lost funding, and I too could offer a number of other examples of where that has occurred. Our goal is to not allow that to happen to the biological materials we have been entrusted with; some of them are over 100 years old.

We have an uncompromising commitment to quality, standards, biosafety, biosecurity, and regulatory compliance. We want to make sure things are handled safely, and we want to make sure that we offer the best-quality materials. We also want to ensure standardization for global research activities so that researchers in one location are working with the same material as someone else in another. If the materials they use are different, then their research is not comparable. And, as I noted earlier, we have strains

that have been around since 1925, and because some of these were materials that came to us from the Winslow collection they are actually older than that.

We have to ensure that our resources meet changing scientific needs. In particular that means that while we continually add to our collections, we do not however take just anything because doing that would strain us financially. Instead we have to make sure we are adding things that are going to continue to be used as the needs of science change.

ATCC has expanded its unique position as a knowledge and technology transfer broker between research and commercial entities. This is not my area of expertise However we have staff at ATCC who deal with this every day, and whose job is to manage the licensing opportunities, the material transfer agreements, and other activities that ensure equitable sharing and use. Our business is managed today in a way that ensures the ATCC enterprise continues in operation and that the biological materials under our care are not lost. Everything we do and every dollar we earn gets plowed back into the organization to ensure that ATCC will be around 100 years from now. In my opinion there are not many government-funded collections that anyone can say will be around 100 years from now, as funding priorities change. Similarly if a private collection is getting funding from an external source, it may not be around in 100 years either for the same reason. Our model for operating the ATCC is focused on ensuring its continued existence and support for scientific advancement.

Quality is also important at ATCC and we maintain our cultures as close as possible to the original lines, so when you get an ATCC culture you know it is pretty close to what was put there. A seed stock system was developed at ATCC to ensure that we have low-passage material as close to the original as possible. Since living cells and organisms are perpetuated by replication in culture, genetic changes can occur and it is important to minimize the number of times a culture is sub-cultured. Our brand ensures high-quality standards with low passage numbers, and that is why we do not want others putting the ATCC label on cultures that we have no control over. We did experience problems with that at one time, but we have been able to address that problem and are now in a position to deal with the violators. As some of the microbiologists in the audience are undoubtedly aware, there were companies such as the DIFCO that sold cultures obtained from ATCC that were then used for nefarious purposes, and we would later hear from people that an ATCC culture was used in these instances. While those cultures were not obtained directly from us, the use of our name and trademark implied that we were the supplier.

We constantly apply new technologies to our AAPPDD activities, that is, to acquisition, authentication, preservation, production, development, and distribution—all the activities involved in maintaining these materials. We are also continuously updating our operating procedures, and we have a robust quality systems program that manages and monitors our standard operating procedures. Planning for renovation and new facilities is an ongoing process because it is essential as we continue to grow to remain viable.

ATCC's primary focus is standardization, and our goal is to provide standard reference materials to enable scientific continuity and advancement. We are now working with official standards (certified reference materials or CRMs) under our ISO Guide 34 accreditation however we have always focused on providing standards in all of our offerings. If you obtain biological materials from the ATCC today and the same materials again 20 years from now, you can be assured they will be the same.

Recently we developed a proficiency testing standards program which originated out of a need for proficiency testing standards following the H2N2 incident, in which kits were sent out for proficiency testing containing the wrong virus. This was not discovered until an astute diagnostic lab in Canada discovered the mistake, and the CDC alerted us to their concerns and we responded by putting this program in place. It is not a mandatory program and it is purely voluntary; nevertheless a lot of the proficiency testing manufacturers are now coming to us and obtaining reference standards that can be traced back to ATCC to ensure that there are no similar problems in the future. The suppliers that do not participate in this program accept the risk that they may experience a problem in the future.

ATCC provides quality control reference standards, some of which are mandated or recommended both in FDA regulations and in the standards put out by such organizations as the Clinical and Laboratory Standards Institute (CLSI). Because we are now an accredited standards development organization, we also develop written consensus standards in support of the material standards we provide.

One of the key concepts underlying what we do is termed "added value." Suppose an organization has some materials that have no value to them, perhaps, for example, a pharmaceutical company has done a study that generated a set of clinical specimens, and the company has no reason to believe the specimens have any further commercial value. While they should consider sharing them with others, their concern is how they can receive some return on their investment in the materials. On mechanism is to offer them through a broker like ATCC that can find new uses for the materials.

One thing that should be kept in mind concerning the microbial research commons is that if you do not have a way to support it, it is not going to last. The ATCC is a real example of a national resource that almost disappeared; however, we are not going to let that happen again.

Question and Answer Session

DR. KURTZMAN: Can you share with us what percent of your income may be derived from the sale of cultures versus other services?

DR. SIMIONE: I can give you a high-level answer. The products and services generate about 50 percent of our revenue, and the services area generates the other 50 percent. But within the products and services, I do not know the details of the income breakdown. The cell cultures are probably our biggest seller, and probably about 80 percent of the revenue is derived from the distribution of the cells and microbes. The other 20 percent comes from the sale of culture media, reagents and other items.

DR. KURTZMAN: It does. I think it also emphasizes once again that culture collections, except under your situation, are clearly not self-sustaining, and if we expect to have these cultures last for decades in the public domain, something really does need to be done beyond what we are doing now.

PARTICIPANT: Have you given thought to disaster backup? I ask as someone who lived in New Orleans for many years and kept my backup strains at the Department of Agriculture in New Orleans. There was no power in either place after the Katrina hurricane disaster, and we lost them all.

DR. SIMIONE: I meant to mention that. We have backed-up our entire collection since 1979, but we had it in various places starting with the government. For a while it was at Fort Dietrick, and then we moved it to a private company. Even those locations were not safe because the government said eventually, "We need the space, take your stuff out." The private company, while it is still in business, had financial problems and presented a potential risk.

We now have our own facility 60 miles west of the ATCC that we maintain with all the bells and whistles—backup generator, everything. We store our collections there in the event we have a physical disaster at the main facility, and we are working on a business continuity disaster plan right now that goes even further to ensure that if we do have a disaster at the main facility, we can continue at least to supply some of the top products .

PARTICIPANT: Two questions. One is: At what rate are your collections growing?

DR. SIMIONE: Good question. Steve mentioned that they are growing exponentially. Actually, they are not. The areas that the government contracts support— like the MR4, which is the malaria collection—are designed to build collections for specific uses. Those are growing rapidly, however the government is paying for them. The ATCC collections are not growing at the same rate.

The main reason is not that we do not have the resources to take on additional materials. We choose very carefully because we do not want to just take a lot of new materials that are not going to be useful, and then just sit in a freezer. It was about 20 years ago we started to hear, "I am not going to give you the material anymore because it might have commercial value. And if I give it to you, you are going to make it available to everybody." That is when we had to start changing how we operate, and while I do not know how fast they are growing, I do know that they are growing. The cell collection probably now contains about 4,000 lines.

PARTICIPANT: The point of the question is not just about the ATCC, but for any collection. There are obviously financial implications as a collection grows that are key to the rate at which it grows. My second question is, are there particular components of the collection that are of greater community interest in terms of accessing them than others?

DR. SIMIONE: That is an excellent question, but I cannot get into the specifics because I am not the expert. The problem with providing an answer is that it is a moving target. That is the problem with a public resource like this and I want to emphasize it is public. If we go out of business, these belong to the public; nobody owns them. But, yes, we do pay attention to what is needed over time. We have a marketing group, and while I do not like to call it that; that is what it is, people who look at this issue and ask what is needed by our customers.

So, we are adding materials, but I do not know what the hot stuff is. I know that the cell lines are in great demand, although I do not know which ones. The molecular materials, such as the clones, were hot for a while, and then slowed down.

DR. KYRPIDES: A very quick question. It is about the bacteria culture collections from which, if I am not mistaken, number about 15,000 or 20,000.

DR. SIMIONE: Yes, I think that number is close. It is 18,000 on the slides.

DR. KYRPIDES: For how many cultures do you really know what it is, in the sense that you have genomic sequencing?

DR. SIMIONE: I can tell you we do that, but not for the whole 18,000. As with my answer to the previous question, we are going to focus on those that are needed by the scientific community. Now, most of them had been characterized in some way when they came in, but that was prior to the genomic sequencing. We have that capability now, but there is no way we can go back. The issue is the cost; that is it would not be cost effective.

DR. KYRPIDES: Half or more than half?

DR. SIMIONE: I do not know the answer. I could find out for you though, because somebody at the ATCC knows the answer.

9. Contracting to Preserve Open Science: Lessons for a Microbial Research Commons – Peter Lee[26]

University of California, Davis, School of Law

In this talk, I would like to address three topics. First, I would like to present some ongoing research on the use of "private ordering" mechanisms to broaden access to the fruits of publicly-sponsored science. I would then like to apply some lessons from this practice to the challenge of designing a microbial research commons. Finally, I would like to explore some additional principles and considerations for constructing such a commons.

Let me begin by providing some context. Much of my recent research focuses on the role of "public institutions" in creating a noncommercial research commons in biomedicine. In particular, it focuses on efforts by government, academic, and nonprofit entities to enhance access to patented biomedical research tools. Obviously, enhancing access to intellectual property is very different from enhancing access to physical resources, such as microbes, that are ordinarily subject to material transfer agreements (MTAs). However, there are some striking similarities between these two scenarios and some lessons to be learned. With that in mind, allow me to turn in greater detail to my first topic, the use of public norms and private ordering to create a biomedical research commons.

The problem that these efforts address is familiar to many, and it has to do with proprietary claims on scientific inputs operating to inhibit valuable research. I focus on patented biomedical research tools, which can encompass anything from extracted and purified human embryonic stem cells to genetically modified organisms, including genetically modified microbes. The relevant theory here—which, of course, must be empirically verified—is that patents on research tools constrain access to these resources, thus inhibiting basic scientific research and the development of valuable technologies and industrial applications.

The present tendency of parties to patent research tools gives rise to the challenge of how to provide appropriate access to such resources. In exploring this challenge, let us start with some first principles. Wide access to research tools enables more parties to conduct scientific research, thus generating significant positive externalities. In general, society is better off when scientists have ready access to resources like human embryonic stem cells and microbes that are critical to conducting basic research.

This state of affairs suggests a particular set of policy responses. Given the desirability of maintaining wide access to research tools, perhaps we should simply commit them to the public domain by prohibiting parties from patenting them. This, however, would be problematic for several reasons. First, as commentators have noted, many research tools are actually dual-status inventions: research tools, such as extracted and purified human embryonic stem cells, are often precursors to commercial products in addition to being valuable enablers of basic research. Policy considerations thus weigh in

[26] Presentation slides available at *http://sites.nationalacademies.org/xpedio/idcplg?IdcService=GET_FILE&dDocName=PGA_053668&RevisionSelectionMethod=Latest.*

favor of granting exclusive rights on these research tools, even publicly-funded research tools, to encourage their further development and commercialization. Indeed, this was in large part the intuition underlying the enactment of the Bayh-Dole Act, which allows federal grant recipients to take title to patents arising from taxpayer-funded research.

Consequently, we do not want to categorically commit research tools to the public domain. Rather, it would be preferable to create a mediated semicommons where such resources are subject to context-specific access and exclusivity. Unfortunately, traditional policy levers such as legislation, judge-made common law, and regulation have proven inadequate in this regard. In large part, these instruments are too blunt and cumbersome to facilitate mediated access to research resources on a case-by-case basis.

However, where the law fails to provide for optimal resource management, interested parties often resort to private ordering. We see this in the real property context where communal management of environmental resources, such as fisheries, has proven highly efficient. We also see this in the intellectual property context. In the copyright sphere, for example, collective-rights organizations such as ASCAP and BMI have created pools of licenses that allow interested parties to access performance rights for proprietary musical works.

In the technology sector, private ordering has taken the form of massive cross-licensing and patent pools in patent-intensive industries, such as software and telecommunications. In addition, we see similar attempts to use private law mechanisms to enhance access to protected content in the emergence of open-source licensing, such as the General Public License for software and Creative Commons licenses for a variety of copyrighted works. In all of these contexts, norms and contracts are driving enhanced access to proprietary material.

Furthermore, "private ordering" is not simply the domain of private institutions; public institutions are also fruitfully engaged in private regulation. This is particularly apparent in the life sciences sector, a field in which public institutions enjoy enormous leverage.

Take, for example, the National Institutes of Health (NIH), which provides about $30 billion per year in funds for biomedical research. As noted, under the Bayh-Dole Act, federal grantees can take title to patents arising from taxpayer-financed research. Some have criticized this law as providing a "double subsidy" to grantees, who receive both taxpayer funds as well as patent rights. That being said, these public funds come with certain strings and expectations attached.

Along these lines, in 1999, NIH issued principles and guidelines for obtaining and disseminating NIH-funded biological resources. There are two principles in particular that I think are very relevant. First, the NIH guidelines advocated the wide availability of NIH-funded, grantee-patented research tools for noncommercial uses. At the same time, however, these guidelines allowed for targeted exclusivity of such resources for commercial development. While the Bayh-Dole Act complicates and arguably prohibits direct enforcement of these guidelines, NIH considers compliance with its principles and guidelines in reviewing grant proposals and awarding research funds.

The model that arises is one where NIH provides some sort of consideration, in this case money, to a downstream resource developer, and in return that downstream developer is expected to provide qualified access to proprietary resources for research purposes. That access applies not only to NIH scientists, but extends on behalf of NIH to the wider research community as well.

Of course, NIH is not the only game in town. I teach in California, and there the California Institute for Regenerative Medicine (CIRM) will provide about $3 billion over

ten years for human embryonic stem cell research. And here again, public funds are embedded in a quid pro quo. As with the federal Bayh-Dole Act, recipients of CIRM funds can patent the results of publicly-financed research. However, CIRM imposes certain access requirements for state-funded research resources. Grant recipients must agree to make their patented, CIRM-funded inventions readily accessible on reasonable terms to other California organizations for use in noncommercial research. In addition, grantees must make CIRM-funded materials described in scientific publications widely available for research purposes. These access requirements are embedded in the terms of funding agreements and are legally enforceable.

CIRM places strings on public money in another way as well. Unlike NIH, CIRM collects royalties on inventions arising from public funding. These royalties are deposited into the state's general fund and can be used for a wide variety of public expenditures. If multiple funding sources contributed to an invention, the state is entitled to a share that is "proportionate to the support provided by CIRM for the discovery of the invention."[27]

In addition to the federal government and state governments, universities are also major contributors to biomedical research. Increasingly, universities are using contracts—specifically, technology transfer licenses—to enhance access to patented research tools. For example, here is some boilerplate language from an exclusive license at Harvard University: "Harvard will retain the right, for itself and other not-for-profit research organizations, to practice the subject matter of the patent rights for internal research, teaching and other educational purposes."[28] So, when Harvard exclusively licenses out some patented invention, it retains the right to use that invention for research purposes and to allow other nonprofit institutions to engage in similar activities.

In addition to demonstrating how licenses can enhance access to proprietary resources, university practice can inform the design of a microbial commons in other ways as well. As a general matter, resource owners tend to systematically overvalue their assets. We see this in the physical property realm, and we also see such overvaluation in the intellectual property context. Based in part on the expectation of generating significant revenues, universities have dramatically increased their patenting activities over the past several decades. However, university intellectual property actually generates relatively little income. Income from licensing is largely hit or miss, and it is overwhelmingly miss. In fact, a large proportion of university technology transfer offices lose money. It appears that universities systematically overvalue their patents, a phenomenon that is likely to apply to other owners of research resources as well.

From the perspective of designing a microbial commons, universities are also informative in that they demonstrate a high degree of normative plurality. Norms can diverge widely among seemingly similar institutions, as illustrated in various universities' approaches to technology transfer. For example, Columbia University and the University of California have been very aggressive in patenting and licensing, while other leading research universities, such as Johns Hopkins, have been much less active in this realm. In addition to differences *between* institutions, we also see important normative plurality *within* institutions. In general, research faculties and individual scientists tend to be quite committed to the norms of open science. However, this

[27] CAL. CODE REGS. tit. 17, § 100308(c).
[28] Harvard University Office of Technology Development, Licensing Harvard Patent Rights: A Guideline to the Essentials of Harvard's License Agreements, available at *http://www.techtransfer.harvard.edu/resources/guidelines/license.*

normative orientation may diverge sharply from that of senior university executives and technology transfer offices, which may place greater emphasis on asserting patent rights and maximizing licensing income.

Returning to our survey of "public institutions," nonprofit funding organizations are also significant sources of scientific venture capital. One very important player is the Howard Hughes Medical Institute (HHMI), which provides approximately $600 million per year in research funding. Consistent with the theme of placing strings on research funding, HHMI "expects all HHMI research tools to be made available to the scientific research community on reasonable terms and in a manner that enhances their widespread availability."[29] This policy is consistent with NIH's principles and guidelines on sharing biomedical research resources.

In these examples from government, academia, and the nonprofit sector, we see attempts to formalize the informal norm of open science. Public institutions are using quid pro quos and contracts to leverage their provision of "upstream" scientific capital to ensure access to "downstream" proprietary research assets. In this manner, public institutions are utilizing private ordering to create a biomedical research commons. Drawing from these case studies, there are several lessons that can be applied to designing a commons for microbial resources. First, it may be useful to consider an integrated approach that provides access to both intellectual and physical property. The first generation of the microbial research commons will likely focus on enhancing access to physical resources, namely microbes themselves. However, in future iterations, it may be useful to include patents related to these resources within the commons. After all, scientists often have to clear patent rights in addition to obtaining physical materials in order to conduct valuable research.

The next lesson for designing a microbial commons relates to how parties go about formalizing informal norms. Here, issues of authority and process are critical. For instance, who speaks for the scientific community, and how does a community arrive at normative consensus? Here, the scientific community can learn valuable lessons from NIH's development of its principles and guidelines. These principles and guidelines arose through a consultative process that included formal notice and comment, which may serve as a helpful model for designing policies governing a microbial research commons.

Within this endeavor, institutional buy-in will obviously be important. In achieving such buy-in, it is crucial to consider normative plurality, that is, the fact that different institutions, such as various culture collections, may have different norms. And there might be normative plurality even within particular institutions, as we see in various universities.

So, how can we encourage participation in the microbial research commons? As noted earlier, the experience of universities suggests that parties systematically overvalue their assets, which may discourage them from contributing their resources to a commons. One way to discipline this tendency is to leverage upstream, scientific capital to mandate or encourage compliance with access objectives. Many public institutions enjoy significant leverage in this field, such as government agencies, universities, private foundations, and even scientific journals. One promising model would involve public institutions providing some sort of material support (or, in the case of journals, publishing

[29] Howard Hughes Medical Institute, Research Policies: Research Tools (SC-310) 1 (2007), *available at http://www.hhmi.org/about/research/sc_310.pdf*.

opportunities) to downstream parties, who must in return provide access to proprietary resources to the research community

Instructive in this regard, there is an existing track history of calibrating property regimes to accommodate dual purpose assets—assets that are both directly used in noncommercial research and have high potential for commercial application. Past experience shows that a single framework can accommodate these different uses. The potential for commercial exploitation does, however, give rise to difficulties regarding apportioning upstream versus downstream contributions in calculating appropriate compensation. Consider the following hypothetical: party A transfers some material to party B, who finds a way to commercialize it. What is the obligation of B to compensate A, and how is this complicated when parties C and D also contributed to commercializing this product? As we have seen, CIRM has already addressed this problem when determining royalties owed to the state of California where CIRM is only one of several funding sources for some commercialized product. I suggest that the proportionality principle used by CIRM can be a useful guide to allocating compensation when one party commercializes assets contributed by another.

In designing a microbial commons, institutional considerations are of course paramount. In any type of initiative encompassing a broad array of institutions and practices, centralized coordination and standardization can play a valuable role in lowering information and transaction costs. In the patenting and licensing sphere, NIH has played a central role in standardizing the ways in which various entities promote access to patented biomedical research tools. Is there an analogous body in the microbial research commons area, and to what extent are these challenges exacerbated because of the global dimensions of this initiative?

Finally, I want to explore some additional design principles and considerations for constructing a microbial research commons. In doing so, I have loosely modeled my observations on the objectives outlined in the draft monograph by Reichman, Dedeurwaerdere, and Uhlir. First, how can we shore up incentives to share proprietary resources? In terms of benefits, it is important to emphasize communal norms as well as the language of self-interest. Generally speaking, nobody likes giving up his property for free. In some cases, however, such behavior can actually advance one's self-interest. In this respect, patent pools in the software industry represent an influential model, as firms are voluntarily donating their patents to a common pool because it enhances innovation for everyone.

Second, it is useful for parties to assess realistically the value of their microbial assets. As we have seen, parties systematically overvalue their assets. However, rational expectations concerning the profitability of these resources may encourage greater participation in the sharing system.

Within any attempt to build a microbial commons, distributing duplicates and derivative resources will be very important. In many ways, this need to broaden access to "downstream" iterations of microbes distinguishes this initiative from traditional MTA-mediated exchanges. Here, open source and viral licenses that ensure access to derivative works can provide a useful model.

Maintaining high-quality standards for microbial resources will require dealing with a variety of governance issues, which I understand will be the subject of a later panel. At this time, however, let me just say that there is a strong need for institutional leadership and community consensus to define standards, assign monitoring

responsibilities, and delineate sanctions. There may be organizations that are already well-suited to provide these functions.

Reputational benefits are critical to the normative economy and, in some sense, provide a solution to the incentives question. Why should parties voluntarily contribute proprietary assets to a commons? One answer is that participation makes it possible to establish priority, increase citations, and obtain communal recognition.

Along these lines, it is interesting to consider reputation as both a carrot and a stick. In many ways, the effort to build a microbial research commons reflects an attempt to create a community of strangers. Creating such a community gives rise to the challenge of cultivating trust and identifying bad actors. Here, the literature on peer production can offer a useful guide. User-generated ratings and "distributed accreditation" offer a new twist on peer review. Consider, for example, eBay, another community of strangers. How do I know that I can trust another party on eBay? The answer is that everybody on eBay is rated by everybody else.

Finally, I would like to address the challenge of securing equitable compensation in a commons. If I contribute some asset to a commons that another party eventually commercializes, how much compensation, if any, should I receive? I am a big advocate of liability rules, and I have argued in favor of them in other contexts. Such regimes, however, give rise to a number of challenges related to valuation and apportionment. Who determines equitable compensation, and how exactly is a reasonable royalty calculated? While considering these questions, it is also important to consider eliminating or mitigating the threat of exit. Is it going to be possible for an entity to simply opt out of the system once a product becomes highly profitable? Perhaps the party will think it can do better than equitable compensation and choose to leave rather than commit that resource to a commons.

Along these lines, it is useful to look forward, beyond merely sharing physical materials, to consider commercialization and patenting. In this regard, it may be helpful to have structured rules whereby parties who obtain resources from the commons must inform those who provided them of plans to patent any related inventions. Among other functions, such notice might initiate discussions regarding co-inventorship and co-assignment of inventions.

To summarize, I believe the microbial research commons is eminently feasible, and I have suggested that it can benefit from related experiences to create a biomedical research commons in the intellectual property context. We already have relevant experience dealing with multiple-purpose assets that are useful in both basic research and commercial applications. In designing the microbial commons, it would be helpful to integrate access to all relevant intellectual, physical, and informational resources. Within this project, we can leverage public norms and private ordering to formalize an informal ethos of scientific sharing. Finally, there are, of course, a host of institutional challenges regarding standardization, coordination, and determination of equitable compensation. If these institutional hurdles can be overcome, the creation of a microbial commons promises to greatly streamline and enhance microbial research.

Question and Answer Session

PARTICIPANT: I would like a better understanding of your metaphor of the new public license. We are all trying to deal with the difficulty of intellectual property.

PROF. LEE: The basic idea is a contractual understanding that "derivative works" based on some asset will also be disseminated in an open-source manner. The analogy to software is not perfect, but essentially the concept is to have an embedded obligation that all future iterations of some asset obtained in the commons will also be widely available to others, particularly for noncommercial research use. To the extent that we want to encourage openness, I think that we can use viral mechanisms to facilitate that outcome.

PARTICIPANT: I also am curious about what CIRM's experience has been with this obligation to negotiate reasonable terms *ex post* for access to materials. I am wondering with CIRM (a) if there is much history of how successful they have been with this process, and (b) whether they can be held confidential by the party.

PROF. LEE: There are actually two components to CIRM's royalty hierarchy. One involves *ex ante* definitions. Here, there is an established regulatory framework for determining what the royalties will be, and they define certain blockbuster categories. For example, if a grant recipient has $500,000 or more in sales, that triggers a certain percentage of royalties, and then there are graduated categories going up from there. So, those royalties are based on an established, *ex ante* schedule. The difficult part is apportioning royalties when a CIRM grant recipient combines state money with funding from other sources. CIRM recognizes that there are often multiple inputs to a commercial product, and CIRM's claim relative to other upstream funders tends to be determined on an *ex post* fashion based on proportionality analysis.

Regarding the other component of your question, there is not much track history here in apportioning royalties. CIRM was established relatively recently. It is doing a lot of grant work, but not a lot of commercial products have been produced yet to provide test cases for these rules.

10. Designing the Digital Commons in Microbiology—Moving from Restrictive Dissemination of Publicly Funded Knowledge to Open Knowledge Environments: A Case Study in Microbiology

–Paul Uhlir[30]
National Research Council

I think that everyone would agree that the rate of change in new technological systems often outpaces human capacity to adapt to the technological advances—and, even more so, the ability to exploit those advances for maximum social and economic benefits. This is particularly true for transformational technologies that displace their antecedent ones and their associated organizational paradigms.

In such cases, not only is it necessary to adopt new management approaches in response to technological progress, but it is necessary to overcome the substantial resistance to change by entrenched interests whose business model is based on the superseded technology. Such a transformation has been taking place over the past couple of decades as a result of the technological revolution brought about by the combination of digital information technologies and global communication networks.

Table 10–1 presents a comparison of the characteristics of publishing under the print paradigm with those of disseminating information via global digital networks.

Comparison of some key characteristics of the print dissemination and digitally networked paradigms:	
PRINT	**GLOBAL DIGITAL NETWORKS**
(pre) Industrial Age	post-industrial Information Age
fixed, static	transformative, interactive
rigid	flexible, extensible
physical	virtual
local	global
linear	non-linear, asynchronous
limited content and types	unlimited contents and multimedia
distribution difficult, slow	easy and immediate dissemination
copying cumbersome, not perfect	copying simple and identical
significant marginal distribution cost	zero marginal distribution cost
single user (or small group)	multiple, concurrent users/producers
centralized production	distributed and integrated production
slow knowledge diffusion	accelerated knowledge diffusion

TABLE 10–1 Print versus digital network paradigms.[31]

Although these comparisons may be familiar, it bears emphasizing that the magnitude of the changes made possible by the shift from print to digital technologies and networks cannot be overstated, either quantitatively or qualitatively. The explosion in the production of digital bits is now well known as a function of Moore's law. Digital networks also have well-known quantitative advantages over the previous print paradigm in time, geographical extent, and cost; that is, digital networks can provide instantaneous,

[30] Presentation slides available at: *http://sites.nationalacademies.org/xpedio/idcplg?IdcService=GET_FILE&dDocName=PGA_053717&RevisionSelectionMethod=Latest*.
This presentation is based in large part on the draft monograph, Reichman, J.H., T. Dedeurwaerdere, and P. F. Uhlir. Designing the microbial research commons: Global intellectual property strategies for accessing and using essential public knowledge assets (Cambridge Univ. Press, forthcoming 2013).

[31] Uhlir, Paul F. (2006) The emerging role of open repositories for scientific literature as a fundamental component of the public research infrastructure. In: Open Access: Open Problems. Polimetrica Publisher, pp. 59-103.

concurrent, and global availability at near-zero marginal cost of access by each additional user. These quantitative improvements make possible, even if it has not yet been realized, the universal availability of information.[32]

The qualitative advantages of digital technologies and networks in accelerating the dissemination of information and the diffusion of knowledge are just as important as the quantitative ones. Because networks provide the opportunity for non-linear, interactive, and asynchronous communication with multimedia capabilities, the potential to improve the dissemination and diffusion processes has been greatly magnified. The digital nature of the information imbues it with flexible transformative properties, making it subject to easy manipulation and straightforward integration with other types of information, which in turns allows the creation of new knowledge that was either not possible or much more difficult in the print context.

Moreover, the network makes possible entirely new forms of collaborative knowledge production on a broadly distributed and interactive basis, transforming or dismantling the hierarchical and centralized organizational models through which information was produced and knowledge diffused in previous eras. Perhaps most important, digital networks make possible entirely automated approaches to the extraction, processing, integration, and organization of vast amounts of information, which can in turn be transformed into unlimited new discoveries and products, eclipsing the capabilities of purely human information production, dissemination, and use.[33] As both the principal inventors and pervasive users of the Internet, scientists have a great deal at stake in fully exploiting the potential of this new medium for accelerating scientific progress and its benefits to society.

Table 10–2 offers a summary of some of the advantages to science of open access to—and unrestricted reuse of—publicly generated or funded data and information on digital networks.

Advantages to science of open access to and unrestricted reuse of publicly generated or funded data and information on digital networks:

- Promotes interdisciplinary, inter-institutional, and international research
- Enables automated knowledge discovery
- Avoids inefficiencies, including duplication of research
- Promotes new research and new types of research
- Reinforces open scientific inquiry and encourages diversity of analysis and opinion
- Allows for the verification of previous results
- Makes possible the testing of new or alternative hypotheses and methods of analysis
- Supports studies on data collection methods and measurement
- Facilitates the education of new researchers
- Promotes citizen scientists and serendipitous results, enabling the exploration of topics not envisioned by the initial investigators and the primary research community
- Permits the creation of new datasets when data from multiple sources are combined
- Promotes capacity building in developing countries and global research
- Supports economic growth and social welfare
- Generally provides greater returns from public investments in research

TABLE 10–2 Advantages of open access to an unrestricted use of digital information.

[32] Ibid.
[33] Ibid.

If one were to start over and construct a new institutional regime for scholarly communication on digital networks, what should the guiding principles be? I would suggest the following:

1. Maximize public-good aspects of publicly funded research data and info;
2. Avoid monopolies and artificial markets (service, not captured product);
3. Take advantage of zero marginal cost for global dissemination;
4. Support freedom of inquiry and collaborative research;
5. Optimize content for automated knowledge discovery tools; and
6. Maintain the traditional characteristics that are essential to the research community and the progress of science (quality control, reputational benefits, research impact, speed of publication, ease of access, and long-term preservation and sustainability).

The bottom line is that open access online and the unrestricted reuse of research data and information produced from public funding is, in most cases, far superior to proprietary and restricted dissemination, as it maximizes value for the content producers and the user community rather than for the intermediaries who perform the dissemination services. The question is: How to get there?

As part of our study, we analyzed the access and reuse policies and licenses of both the microbial journal literature and of some databases used in microbial research. The traditional practice for researchers publishing scientific articles is for the authors to assign their copyrights to the publishers, who are either commercial entities or learned societies and other not-for-profit scientific organizations. As a result, it is the publishers rather than the authors who initially determine the conditions for access to these articles and for reuse of the information and data they contain.

Today, access to the contents of microbial journals is usually regulated by two sets of contracts. First, the publisher's contract with the author will determine what the publisher owns and—to some extent—what it can do with the material. In the pre-digital age, this contract was usually the only one at issue, because readers' and users' rights were determined by statutory intellectual property laws (i.e., copyright laws) and, since 1996 in the European Union, by database protection laws.

In our empirical research of the journal literature, we assessed the copyright and access policies of publishers responsible for journals containing primary research articles and reviews in the field of microbiology. We also selected science journals from other areas, such as immunology, that regularly publish articles in the field of microbiology.

Most of the open access journals were obtained from the Directory of Open Access Journals (DOAJ) and from individual publisher websites, such as that of Horizon Press. The hybrid and subscription journals were selected primarily from the publisher websites and a few other Web resources. Sixty-four percent of the selected journals include articles about microbiology only, while the remaining journals publish articles from other areas as well. We analyzed a total of 303 journals dedicated in whole or in part to microbial research results. Some of the highlights of our findings include:

- About 30 percent were full open access (OA), including hybrid (both purchased immediate OA and subscription); 20 percent were openly available but read-only; and 50 percent were subscription based.

- 80 percent of subscription journals allow author self-archiving on personal websites, but almost 90 percent do not allow archiving on the author's institutional websites and most are silent on external repository deposits (e.g., on PubMedCentral).
- 98 percent of subscription journals require transfer of copyright, although we do not know the number that would approve an author's request to retain copyright and grant only a nonexclusive license to publish.
- About 75 percent of all journals surveyed are published by for-profit publishers.
- 96 percent of subscription journals give no direct discount to developing country subscribers (but some may participate in group discounts to libraries through the INASP or HINARI programs).

We also briefly analyzed the scientific databases used in microbial research. This survey was less comprehensive or rigorous than the one we did for the journal literature, in part because the information about these databases is less standardized and more diffuse. We found that:

- Many molecular biology databases (genomic, proteomic) and taxonomic databases are openly available and free to use.
- Molecular biology data in a lot of specialized research (e.g., energy and environment) are not deposited and not available.
- There are many legal, policy, economic, and cultural pressures for the researcher to keep data secret, either because of the data's commercial potential or strategic advantage or because of the burden of making the data useful to others.

The intention of latter-day intellectual property laws is to secure rents from specified end uses of relevant knowledge goods, such as music, films, and software. The beneficiary industries do not contemplate uses, reuses, or redistribution of their products beyond those income-producing activities regulated by these laws, although the state may require them to tolerate some uncompensated uses in the larger public interest. Courts have traditionally narrowly interpreted the limitations and exceptions in favor of strengthening the right holders' exclusive rights and the incentive effects they are supposed to provide.

This approach conflicts directly with the needs of science, however. This is particularly true for public science in the digital domain, whose norms favor maximum use, reuse, and redistribution by third parties of the knowledge that publicly funded researchers generate. In the pre-digital epoch, legislation—and copyright legislation in particular—did contain some measures that attenuated this conflict in the interest of science, but the digital revolution that has created such promising opportunities for scientific research has also generated intense fears that publishers of literary and artistic works generally would become vulnerable to massive infringements online and to other threats of market failure. In response, publishers have pushed legislatures to recast and restructure copyright law in the online environment so as to preserve business models built around the print media.

Thus copyright laws in Organisation for Economic Co-operation and Development (OECD) countries and database protection laws in the European Union are on a collision course with some of the most promising scientific movements in history. These impediments to the global exchange of basic scientific information are then magnified by the ability of intellectual property rights holders to override relevant exceptions and limitations by a combination of technological protection measures and even more restrictive contractual conditions.

In this legal environment, the continued ability of scientists to access, use, and reuse essential upstream knowledge assets depends increasingly on their willingness to disregard—consciously or unconsciously—the legal and contractual constraints on their everyday research. However, the implicit assumption that proprietary intermediaries will not detect violations of statutory or contractual restrictions on their continued treatment of these assets as public goods or, if they detect those violations, will not enforce their rights is neither tenable nor desirable. Sooner or later, there could be a clamorous case involving academics after which risk-adverse universities and university technology transfer offices would shut down the secret or arguably unintentional infringing activities now going on at many universities and scientific laboratories.

The existing system thus offers only three unsatisfactory pathways for making available the basic building blocks of digitally integrated microbial research. The first is to continue to muddle through by ignoring a hostile legal environment, with all the attendant risks of civil disobedience generally. The second is to embrace the tendencies to privatize public goods by adopting the commercial and restrictive practices that are thought necessary to generate both research funds and revenues from downstream commercial applications. This commercializing trend will increase the costs of publicly funded research, which depends on access to general purpose research tools. It will also severely restrict, if not make entirely impossible, the exploitation of automated knowledge-generating opportunities through a proliferation of legally contrived thickets of rights and restrictive licensing conditions.

The third pathway attempts to build an alternative open-access infrastructure, which could generate important payoffs in terms of enabling cumulative public research. However, a lack of coordination with respect to intellectual property provisions intended to maximize these different expected payoffs hampers the further development of any such alternative infrastructure. For this reason we examined a number of top-down and bottom-up responsive measures for making the current legal environment more science friendly.

It is worth considering what sorts of legislative changes at both the domestic and international levels would be needed to improve the prospects for digitally integrated research. We have suggested several legislative changes to help balance the intellectual property (IP) regime between rights holders and public-interest, publicly-funded research and education users. Legislatures could provide more robust limitations and exceptions to traditional copyright law for not-for-profit, publicly funded research, for example. Laws could allow for greater access to and use of public research in digital copyright. And research funders could (with enabling legislation) mandate such things as author deposits and copyright retention with the authors.

In our opinion, however, most of the needed legislative reforms have little or no chance of being enacted under the existing political–economic situation, or until new forces emerge, perhaps in the developing world, to rebalance the system. The balance of

such political forces remains decidedly contrary to such efforts, and the drift of bilateral relations, at least, is towards even higher levels of protection.

Looking beyond these unlikely legislative solutions, there are numerous encouraging bottom-up initiatives that are already underway, where some progress has been made in achieving higher levels of access to relevant scientific literature and data. The open-source software movement is one example; the establishment of open repositories for publications in a specific area is another.

The challenges in deriving maximum scientific value from still under-exploited technological opportunities lie largely in changing the social systems—the institutional, legal, economic, and sociological aspects—rather than in the technological advances, which will continue even without advances in the social systems. To make progress on these human behavioral aspects, all of the stakeholders involved worldwide in public research and in the process of communicating research results should take part in the unfolding debate, at some level, because they have a vested interest in its outcome.

Up to this point, most of the advances that have been made toward opening up the information created by publicly funded research have come from the bottom up, from the work of many dedicated and visionary individuals and institutions. These actors have been the pathmakers in developing a broad range of initially disparate, but related institutional and policy initiatives in diverse information types, disciplines, and countries. As these projects proliferate and become better established, they are coming together to form a nascent, interoperable global information commons for public science.

Those who fund and regulate public science from the top down are beginning to take notice. They are starting to build upon the tactical successes of the pathmakers and integrating them into broader national and international strategies for the investment and management of public science. A gradual restructuring of the scientific information sector and of the processes of scientific communication is thus now well underway, with the aim of taking more complete advantage of the transformational capabilities of digitally networked technologies.

In light of the clear benefits to the research enterprise and to society from the open availability of publicly funded scientific information in the digitally networked environment, it is not surprising that a variety of new models have already been developed within the research community. As I noted in *Past, Present and Future of Research in the Information Society,*[34] the common element of all these different types of initiatives is that the information is made openly and freely available digitally and online. In many cases, the material is made available under suitably reduced proprietary terms and conditions through permissive licenses (e.g., the GNU license for open source software, or Creative Commons licenses for open access journals or for some works in open repositories), or else the material is put into the public domain. In other cases, such as the delayed open availability that some publishers use for their journal articles, the works remain protected under full copyright, but eventually they become freely and openly accessible on a read-only basis.

Just as the desirability of providing open availability to publicly funded scientific information online was substantiated in our survey of the microbiology literature and

[34] Olson GM, David PA, Eksteen J, Sonnenwald DH, Uhlir PF, Tseng S-F, Huang H-I. International Collaborations Through the Internet. In: Shrum W., Benson, K.R., Bijker W. E., Brunnstein. K., editors. Past, Present and Future of Research in the Information Society. Boston, MA: Springer US; 2007 p. 97-114.[cited 2011 Aug 15] Available from: *http://www.springerlink.com/index/10.1007/978-0-387-47650-6_7.*

databases, the many different models that have already been established attest to the feasibility of doing so. These various examples now provide valid proofs of concept for all information types, for most disciplines—including microbiology, which has been done in many countries—and for all types of institutions, including government agencies, universities, not-for-profit organizations, and even for-profit firms.

Taken together, these activities can be seen as part of an emerging broader movement in support of both formal and informal peer production and dissemination of publicly funded scientific (and other) information in a globally distributed, volunteer, and open-networked environment. These activities are based on principles that reflect the cooperative ethos that traditionally has imbued much of academic and government research agencies. Their norms and governance mechanisms may be characterized as those of the "public scientific information commons" rather than of a market system based upon proprietary data and information. The activities of such information commons activities respond—either explicitly or tacitly—to the needs of science and scientists.

Although much industrial microbiology is conducted in private laboratories, the bulk of research in this area takes place at universities. This research has become increasingly computational and data driven. Universities already host many culture collections, and they also hold a vast amount of microbial materials in research collections outside the formally constituted culture collections. University research on all these materials has increasingly become a networked digital process linking distributed thematic communities.

As the digital component increases in importance, the research becomes more interdisciplinary and dependent on inputs from bioinformatics, computational science, genomics and proteomics, environmental science, agriculture, and health. These interdisciplinary activities, although emanating from a core thematic group based at one or more university centers, operate across university boundaries—and even national boundaries— in order to pursue the thematic interest on an increasingly global basis. In successful cases, the research outputs of these knowledge hubs are usually the fruits of resources that the networked participants have voluntarily pooled from the outset. These outputs are made available for use and reuse to an ever-expanding open community of interested scientists on terms determined by the thematic community. The productivity of these thematic communities is then further enhanced by a growing array of digital and computational tools and techniques, which are put to a common purpose.

When these joint research activities reach the point of yielding published research results, however, they are typically outsourced to a professional society or a publisher, and this step then normally triggers all the legal constraints and restrictions we have described. This customary institutional arrangement in turn limits access to and use of the knowledge assets that the digitally sophisticated scientific community has at its disposal, even when it is the source of those very same assets.

The logical response is to cut the Gordian knot by retaining ownership and control of all knowledge assets produced by the relevant research community with public funding within the public science framework itself, rather than assigning them to external publishing intermediaries. Although this was customary in the past, when the print medium dictated high front-end costs, it is not necessary in a digital world. Once possessed of ownership and control, the scientists and their universities will be in a position to do two things: (1) to avoid all the technical and legal restrictions described above, and (2) to organize the use and reuse of these knowledge assets by means of new

institutional frameworks that are specifically designed to promote collaborative research within fully integrated digital networks.

Such an institutional framework would, for example, give universities the power to determine the conditions under which research results were disseminated and reused, in a manner consistent with the needs of microbial research and education. In this approach, if external intermediaries were used, these intermediaries would operate as service providers on science-friendly terms and with open access prerequisites, as prescribed by the universities. The quid quo pro would be the provision of efficient services that the universities, for various reasons, did not wish to undertake.

Another option would be for the university to integrate the publishing function into the work of the emerging knowledge hubs themselves. In such a case, the funder's support would enable interdisciplinary collaboration in the production and rapid dissemination of research results that were themselves publicly funded, thereby magnifying the social benefits of the public investment. At the same time, the knowledge hubs could evolve into a more solid institutionalized platform, with a view to integrating and systematizing all the knowledge resources needed by the community and all the digital services that made access to use and reuse of these resources as easy and efficient as possible, while also stimulating related educational activities and downstream commercial applications. In this scenario, public funds would remain within the circle of knowledge creators and would nourish all the relevant services, with very low transaction costs and without dissipation to unnecessary external information brokers.

Furthermore, taking microbial journals back to universities and certain other public research institutes would also make it possible to exploit the interdisciplinary resources and inputs of different departments, including, for example, computer science and engineering departments, medical schools, public policy institutes, environmental institutes, and library information services and resources. Moreover, these advantages might be compounded if a consortium of universities pooled their resources to manage and produce a given journal or a set of journals organized on thematic lines. Scientific control over contents through the universities should ensure that high-quality standards were maintained and that the journals would be open access from the start and optimized for network exploitation.

Indeed, once the opportunities of digital networks are taken into account, placing microbial journals in the universities would appear to offer many more advantages than keeping them at commercial publishers or even at professional societies. For example, the societies cannot provide the educational and research opportunities that already exist at the universities, and so they would remain essentially extrinsic, semiautonomous bodies that depend on services provided by individual scientists. Nor can the professional societies make available the kind of interdisciplinary resources available at the universities without transforming themselves into quasi-universities themselves, which, even if otherwise feasible, would be a wasteful and duplicative use of the relevant funds.

An even more powerful argument for preferring the universities to either professional societies or commercial publishers is that microbial science journals should no longer be seen as ends in themselves. Rather, by repositioning them within the universities, the journals could become cogs in—and stepping stones to—the realization of digital knowledge hubs in which journals are but one component.

From this perspective, all the microbial journals thus repositioned should become open access by definition, and all their contents should become available for harvesting by others, for thematic re-integration in other collections, and for various forms of digital

manipulation. More broadly, the publishing function that supports the journals would logically be expanded to support specialized knowledge environments built around the relevant user and research communities and themes. By thus deconstructing the print publishing model and moving the journals or the articles in them into an academic environment, one begins to reconstruct a digitally networked scientific communications model, in which the content providers are the communicators, the intermediaries, the users, and the governors of a dynamically constituted knowledge environment.

We call this digitally networked scientific communications model an "open knowledge environment" (OKE). Over time, these knowledge environments, although hosted by different universities, could be linked together in an integrated knowledge ecology that would enhance the reputational benefits of the participating universities and yield scientific payoffs greater than any single source could produce.

Integrating openly available scientific information resources with open-source collaborative tools online would enable the formation of OKEs for the creation of new knowledge, the enhancement of educational opportunities, and the stimulation of downstream applications. Such an approach would harness the social and technical power of the network which, if properly managed, could greatly increase the value of the knowledge in ways not currently possible with the traditional information production and dissemination processes, and it generally could do so at a much lower cost than the traditional approach.

At the core of an OKE are interactive portals focused on knowledge production and on collaborative research and educational opportunities in specific thematic areas. Ideally, OKEs would be developed around one or more thematically linked, open-access journals and would be augmented by openly available reports, grey literature, and data. Various interactive functions (wikis, discussion forums, blogs, post-publication reviews, and perhaps distributed grid computing) would be added to stimulate discussion and contributions related to specific issues.

The OKEs we envision could readily be hosted at single universities, or their components could be distributed among a consortium of universities having a strong interest in the relevant subject matter. They could also be based at other not-for-profit research centers or at government agencies, although this would compromise the educational function that we also seek to promote. In every case they would be multidisciplinary in character, not only bringing in experts with the appropriate subject-matter expertise, but also involving computer engineers, information scientists, librarians, and other potential contributors to help establish and manage the OKEs and to learn from operating them. Such a knowledge-production project not only would involve senior faculty and experts in its development and application, but also would serve as a mechanism for teaching students in the related departments at the university and as a vehicle for involving the students in the management of the OKE itself.

At the same time, the thematic OKEs could integrate information beyond the conventional disciplinary boundaries, making them tools that are especially well suited to interdisciplinary environments. The OKE concept proposed here would thus build upon a number of recent, but already tested, advances in the online peer production of knowledge and participative Web 2.0 techniques.

Such capabilities are virtually impossible under the proprietary journal model. Indeed, within our proposed open knowledge environments, the narrowly stove-piped, print-paradigm journal model would be transformed into a truly interactive networked initiative. Nonetheless, we stress that these OKEs should maintain the highest-quality

standards of scholarly endeavor, and they should strive to promote the reputational benefits of the participants and of their universities.

Most of these thematic knowledge hubs would also provide essential digital infrastructure functions in support of the microbial research community. Such service functions could include high performance search engines that would enhance the possibilities for finding relevant information in publications and would allow for cross-linking and text mining based on standardized metadata.

While these collaborative functions of the OKE may seem futuristic, they are already being implemented in some microbial science communities as well as in other disciplines. What makes the concept seem futuristic is the existing condition of publishing. The legal terms and conditions in many of the publishers' contracts, buttressed by the larger statutory environment, aim tacitly to protect the print model against the challenges—perceived as risks rather than opportunities—of the digital networked environment. It is this limited vision and obstructive legal culture, in addition to certain other challenging problems, such as obtaining sustainable funding, that makes it difficult to broadly realize OKEs. Nevertheless, there are some examples of the OKE concept already operating.

The move towards an integrated microbial research commons requires linking the materials, digital data, literature and other information resources available from a globally distributed open-access infrastructure and providing interactive platforms for scientists to build on those resources and contribute to them. Effective links between the different open-access components of the material and digital commons are needed to improve the efficacy of cumulative research and to increase the speed of the entire research cycle. Moreover, in specific cases, the combined use of *in vitro* and *in silico* biology offers new opportunities for research, as we noted above. For instance, the task of searching for sequence similarities between the results of high-throughput screening and similar sequences with known properties available from public databases has become a key tool of metagenomics research. Without the aid of computers, the full genome sequences, which are sometimes several hundred pages in length when printed, are not interpretable. Hence, in genomics, advances in computing and in molecular analysis go hand in hand.

Under the larger framework we envision—with a federated network of interactive portals to all the materials, databases, and literature made openly available—it would become possible to establish a registration system administered by a governing body or a trusted intermediary (or an international database collaboration agreement). The World Federation of Culture Collections (WFCC) already hosts different open-access components of the research infrastructure, such as the World Data Center for Microorganisms and the StrainInfo.net bioportal for data and access to the materials held in the culture collections. Moreover, many individual scientists who are active within the WFCC also play key roles in sister organizations, such as the International Union for Microbial Sciences, that also promote open access, especially for research results in the scientific literature. Hence, the WFCC could play a key role in catalyzing the establishment of a governing body for the fully integrated system, which could grow out of the StrainInfo experiment and be established under its own umbrella or within a new organizational and collaborative framework.

In addition to its publishing aspects, this restructuring should considerably augment the scientific payoffs by accelerating the diffusion and reuse of research results, by integrating disparate knowledge components into a dynamically evolving whole, by

facilitating automated knowledge discovery, and by making published research results openly available to nontraditional users or reusers in other disciplines and in developing countries.

This restructuring would prove particularly beneficial for microbial science as a whole, which seems poised to enter a "big science" framework but remains hindered by a disaggregated "small science" heritage and corresponding mentality. By embracing the open knowledge environment vision, microbial science could break out of the organizational limitations inherited from the past and move to the forefront of life science research. The likely result would be a more powerful collaborative approach that would expand the existing knowledge base while fostering greater technical and intellectual capacity to exploit.

Moreover, this restructuring could produce the critical mass needed to self-organize in a way that limits the undue influence of commoditizing pressures on public and upstream research, while creating mechanisms for greater cooperation in pre-competitive and noncommercial research activities; such cooperation has to date been lagging in microbial science. We have in mind the example of molecular biology in the late 1980s, which self-organized and developed a big science infrastructure and became a leader in the life sciences open access movement.

More broadly, the OKE model could have far-reaching implications for the work of universities and research policy institutions, both for targeted problem solving and for the dissemination and impact of high-level reports and research results. This approach could eventually become an integral part of many research plans and budgets. In addition, it is easy to envision many other organizations, at both the national and international levels, applying such methods to developing their knowledge inputs and outputs.

Finally, these insights also suggest why open knowledge environments provide a promising solution to the hard problem of hoarded data. Viewed in isolation, a data pool is only as good as its single components. But an OKE puts all the strength of the microbial research community behind the pool, in the sense that the data pool is itself just one component of a larger whole that combines the data with the literature, materials, and technical services in one community-managed resource. In the context of OKEs, the exchange process is established on a solid and reliable foundation, one that makes full use of automated knowledge tools that are geared to community-determined goals. While these goals evolve and shift over time, in keeping with the relevant sub-communities' own research needs, an ever-expanding infrastructure supports and magnifies all of the reciprocity gains from "formalizing the informal process" of the exchange of data, information, and materials.

Question and Answer Session

PARTICIPANT: Having myself, on occasion, offered wonderful visions of the future, I applaud another great vision. As a longtime university professor, however, I say to myself that this is one more social problem, one more societal need that is being put in a truck and driven over to the nearest university with the instructions, "You solve it."

Now, the simple question I ask is: Universities have been already encouraged to spin off the results of research projects into commercial ventures because that was regarded as a social good. Why should they not also spin out initiatives that come out of the research communities, such as StrainInfo, into a not-for-profit corporate organization, in which university professors could be allowed to participate as they do when they are working across the street in their commercial lab.

Why cannot universities and foundations raise funds for this? Because this is one more task that deflects from others, unless the reason includes a way to generate more funds for university. If we could have more funds to support these activities within the framework of research, we would have many more documented usable and early-released databases that we now have, and part of the problem is the funding agencies do not want open-ended commitments to support the infrastructure.

So, the question is: It is a great idea, it can work on a small scale, but you are using the marginal resources of the university to do something that really should be done properly and recognized as an important infrastructure. You are talking about changing the model of publication, and that should be done on an experimental basis to see if it works with foundation funding.

MR. UHLIR: I agree with all these comments. I left out quite a bit that we have in our draft monograph that addresses some of these issues. First, there is a model already existing in universities—the law reviews, which are run by students and which are effectively open access and published at very low cost within the university, generally without any extra funds.

So, there is a proof of concept already in a different context. Now, we recognize also that there are imperfections of analogy there, so the model that we have been developing for science is somewhat different than for the law journals, but it is related. In particular, there are three examples that I did not have time to get into, but which will be discussed by different people. Peter Dawyndt will be talking about StrainInfo, and the CAMERA Project has already been noted by Mark Ellisman and will be discussed tomorrow as well by Paul Gilna. The Genome Standards Consortium (GSC) will also be discussed to some extent, I believe.

So, there is some experimentation going on, and it is coming from the bottom up. The GSC has an open-access journal that it just launched as part of what I would call its open knowledge environment, or open interactive portal.

There are some proofs of concept in the science field, as well. The infrastructure aspect is really fairly low cost. Our model depends on a lot of existing expertise and labor within the universities—within, say, the libraries, the computer departments, and the information schools—which would all be brought into creating such environments. The students would be involved in the management. It would be part teaching tool, part knowledge production and dissemination. It would also generate interest by funders to provide grants and attract collaborations because it would be a new kind of thematic hub relating to a certain area of research. And, so, it would become, I think, a much more

vigorous and attractive knowledge production and educational tool with fairly low costs for implementation.

But it remains to be tested, and I agree that it needs pilot projects that would be funded by let us say NSF or foundations. Certainly we do not expect all the journals to be superseded by this kind of process and it would all be done in an incremental way. It would be a way to get away from the stovepipe print-paradigm journal system, with journals that have a bunch of unrelated articles in each issue that are not optimized for automated knowledge discovery.

11. The Web-Enabled Research Commons: Applications, Goals, and Trends – Thinh Nguyen[35]

Creative Commons and Science Commons

Today we have heard about some of the problems that material transfer agreements (MTAs) have posed for the sharing of materials for research. Almost 10 years after the NIH, under Harold Varmus, issued a challenge to universities funded by the NIH to simplify the way that we share materials, this problem has not gone away, and may would argue has become worse.

Norms that led to these problems took a long time to develop. Once entrenched, they are difficult to reverse. Because these are legal rather than professional rules they are not informed by the needs of bench science, as they must be.

There is a risk that we will replicate that problem for how scientists share data. To avoid doing so, we need to create a robust, scientific consensus that serves the needs of science and of the public. The legal modes of sharing that we choose should be driven by that consensus.

In the first part of my talk, I will take a specific example—a bioinformatics project undertaken by Science Commons—to provide some motivation and context for a set of goals that I want to propose for data sharing. Second, I will map those goals against three, broad categories of legal regimes that typically we see in data sharing. In the final part, I will discuss some possible convergences: a possible emerging consensus among scientists about how to share data, particularly in the genomics area, and how to build a system for data sharing that promotes, rather than inhibits, scientific discovery.

The example that I will use is the Neurocommons Project from Science Commons. We heard this morning from about using ontologies and the semantic Web to link different kinds of data and to use computational bioinformatics to make systematic discoveries based on the corpus of knowledge we already have. Consider just a few genes that are involved in Huntington's disease, or that are thought to be involved in the development of the disease. For each of these genes, there are in some cases of up to 40,000 or even 100,000 papers attached to each particular gene or protein involved in this network.

If you are a scientist trying to understand how these interactions work and you are faced with hundreds of thousands of papers, you must narrow your search, for you cannot possibly read all those papers. The process by which you narrow your search is driven by heuristics that have worked well for you, but that involve discarding the vast field of data that is otherwise available. So, the bioinformatics challenge is how to tackle using computational approaches the information processing problem that would overwhelm human beings trying to make sense of a sea of data. Fortunately, computers are very good at this kind of task. All they need is access to content. But that is where social and legal constraints, rather than technical limitations, come into play.

In the Neurocommons Project, we draw out all these different types of sources. These are different databases or sources for information about genes that might affect neurological diseases, and the challenge is that all of them are in different formats, they

[35] Presentation slides available at: *http://sites.nationalacademies.org/xpedio/idcplg?IdcService=GET_FILE&dDocName=PGA_053707&RevisionSelectionMethod=Latest*.

use different terminology, and they are not really built to be compatible. So, how do you put them all together and start to do federated searches and queries?

The approach that we have chosen is to use available ontologies and, specifically, we use a technological tool called RDF, Resource Description Framework, to traverse these data sources on the Web. To provide an analogy, on the Web a URL is basically just a link between two Web pages. However, you can also conceive of URLs as definitions of things, and then the links could have meaning. Rather than just saying this page is linked to that page, we can say this receptor is located in that cell membrane and use these terms to connect different data sources. That is to say, that the Web can be used not only to link pages, but also concepts, and by doing so, to merge definition and knowledge.

So, for example, each of these concepts could exist as a separate link or a separate resource on the Internet. Then, when you want to make that connection, you just link other things to these networks of definitions. When you study the genes related to Alzheimer's, the networks of biological pathways are extremely complicated and the only way that you can begin to elucidate them using computational approaches is to build the skeleton of meaning on which the flesh of knowledge can be attached.

The challenge is that, right now, we are limited to using sources that are open access, that are public domain, but there are a lot of journals with primary sources that are not available for text mining., There are consequently lots of holes in our ability to do this kind of research because of the closed status of some journals. Many databases, including government databases, are built upon restrictive licensing models that make data integration impractical or impossible. Thus, the challenge is how do we reformat what we already have stored in databases and journals and other sources of knowledge into a digitally networked commons that we can connect together, and then how do we also get the materials that are related to these digital objects into the emerging research Web so they could be accessible to those who need them most?

To explore the question of what kind of data-sharing protocol enables the kind of research we have been discussing here, I want to first describe three "licensing" regimes. The first broad category of data is those that are in the public domain. They have no restrictions on their use, no restrictions in distribution, and if there is any copyright, it is waived. The last is sometimes called "functional public domain"—because can be treated as public domain even if the legal status is different.

The good news is that there are fields of research where the functional public domain is the norm. The human genome research community is one example, which evolved from the very deliberate consensus formed by the Bermuda Principles.

The second kind of regime is of community licenses, such as open source licenses, like GPL, and the Creative Commons licenses. What they have in common is that they are standard licenses that everyone within the relevant community uses. They sometimes offer a range of different rights, with some rights reserved. So, this is not the public domain because there are some restrictions, but, generally, the information or the resources are available to everybody under the same terms.

The third regime is private licenses, and, by that, I mean custom agreements that are specific to particular institutions or providers. These of course are the norm for commercially available sources of data, and they range wide broadly in terms of the rights provided to the user. However, in general, they are fairly restrictive in terms of redistribution or sharing of data, because of the need to protect a revenue stream.

Based on the models that I have been talking about and the need for bioinformatics, I want to propose a set of goals against which we measure these legal regimes.

Goals
Interoperable: data from many sources can be combined without restriction
Reusable: data can be repurposed into new and interesting contexts
Administrative Burden: low transaction costs and administrative costs over time
Legal Certainty: users can rely on legal usability of the data
Community Norms: consistent with community expectations and usages

The first goal is interoperability. The question is: Can data from different sources be combined? We have seen that the ability to combine the data is really very important for bioinformatics. You cannot link together knowledge that leads to new discoveries if are not aware that such knowledge exists. While the growing costs of scientific periodicals have been widely discussed, the most important issue to scientists is not only cost, but accessibility and searching. In other words, the problem is interoperability of knowledge.

Interoperability
Public Domain ****
Can be combined with other data sources with ease
Community Licenses *** / **
Depends on type of license: share-alike or copyleft are unsuitable, but attribution-only licenses are less problematic
Private Licenses * / **
Depends on restrictions, but not scalable; permutations too large

Transaction costs and the administrative burden are significant barriers to data integration. What are the costs not only for any specific transaction, but over time? Even something as simple as an attribution requirement, when you are required to give citation, can become a huge burden if you are looking at thousands of different data sources or millions of data elements.

Administrative Burden
Public Domain ****
No paperwork or legal review needed
Community License ***
Little paperwork, but some legal review needed (attribution stacking issues)
Private Licenses *
Large amounts of paperwork, frequent legal review needed

We saw that in particular with a recent addition of Wikipedia from Germany that was accompanied by 20 pages of attribution in tiny print that nobody could read it, but it is legally required to be there. There were just these useless additional pages that served no purpose other than to comply with a legal requirement. So, that is not a problem that we want to saddle scientific projects with over time.

The final goal is: How does the legal status map against community norms? What are scientists actually doing in the research lab? Is the activity consistent with what scientists are doing or does it require them to change their behavior in some way? This is critical because we want scientific norms to drive legal rules, not the other way around. In addition, there has to be legal certainty. If you are going to put the materials or the data into the public domain, they have to stay there. If people cannot rely on a stable set of rights over time, then projects which build upon other projects become untenable.

Legal Certainty
Public Domain **** / ***
Clear rights; generally irrevocable; (copyright should be addressed)
Community Licenses ***
Generally credible, good track record with open access and open source licenses
Private Licenses **
Must be considered individually; few private licenses tested by time

How do the different licensing regimes compare in terms of these goals? I would argue that public domain, at least for scientific work, is clearly the best fit. That should not be surprising, because that has been the prevailing norm among scientists since the first scientific journals were published. That is not to say that scientists are not competitive or that they do not hoard pre-publication data. That goes with the territory, which involves intense pressure to publish and fierce competition for tenure. But at the end of the day, when results have been published and discoveries have been claimed, that information should be freely available to all, and not least because of the need to verify the validity of claims made. That need for proof in science is a crucial scientific norm, wonderfully summarized by W. Edwards Deming in the phrase, “In God we trust, all others must bring data.” The availability of data in the public domain is crucial for its operation of this norm.

Public licenses, like open source licenses and Creative Commons licenses, come in a distant second. While these licenses have served many useful purposes in the field of

computer programming and the sharing of artistic content or Web content, they are relatively new legal inventions that are foreign to many scientists. The lack of understanding of how these licenses work, and their legal jargon, may deter widespread adoption. In addition, even those of us who design these licenses do not yet understand how to adapt these types of licenses to the scientific enterprise, and so they can present hidden dangers. For example, the embedded attribution requirements discussed above, which is a feature of all open source licenses and Creative Commons licenses, may seem perfectly reasonable to a computer programmer or artist who only cares about a single work, but for a scientist who must integrate data across many sources, such legal rules quickly become burdensome, if not impossible to follow.

In addition, almost all of these licenses change over time. Common open source licenses like GPL, BSD, and Mozilla have gone through multiple revisions, as have the Creative Commons licenses. Such revisions incorporate best practices and changing community norms, and for cultural sharing, they are perfectly workable. But to build a scientific infrastructure that changes, if at all, in time spans measured by decades and not years, on such licenses would be like playing a ball game whose rules are revised every inning.

Finally, commercial or proprietary licenses—whether expressed in click-through agreements or Web site terms and conditions of use—are proliferating widely throughout the Web. Even some government Web sites share data using customized data licenses that are restrictive and burdensome. Of course we cannot avoid such licenses entirely, particularly for commercial sources of data, where they are a necessity. But there should be no reason why universities, government agencies, or other public institutions that are charged with the dissemination of data for the public good should embrace such onerous mechanisms. At least the argument needs to be made that in these contexts, "over-lawyering" is unnecessary and harmful.

Because of all these reasons, we have started to see some convergence recently in the data community. One that I have been involved with is CC0. It is a result of a three-year policy discussion within Creative Commons and with our community. Technically, CC0 is not a license, but a waiver of copyright and certain related rights, including database rights that exist in Europe and other jurisdictions. In essence, CC0 allows a data owner to guarantee to the public forever the right to use the data in the functional public domain. But if science has been operating for centuries in the functional public domain, why is such a tool even needed? The reasons have to do with the recent (by historical standards) expansion—by courts and legislatures—of the boundaries of copyright to encompass more and more of what has been traditionally considered unprotected by copyright. This puts many categories of data in a minimal state of "borderline copyright." That is, you may have a collection of data that is mostly factual data, so it is questionable whether or not there is enough creativity to qualify for copyright protection, but nevertheless there is sufficient residual doubt that you cannot entirely rely on such a conclusion. This is where CC0 is useful to remove that last residuum of doubt.

Another reason is that in Europe, there is a *sui generis* database directive that gives database owners additional rights in addition to traditional copyright in databases. CC0 also can be used in Europe to waive those database protection rights.

Even within the copyright system, varying countries have different standards for what qualifies for copyright protection. Australia has a different standard from Canada, which has a different standard from the United States. It is very hard, consequently, for international collaborations to figure out who has what rights and when. That is why CC0

is another very useful tool to use in that context to restore data to the functional public domain.

Figure 11-1 shows the deed for CC0. This is what the summary of it looks like.

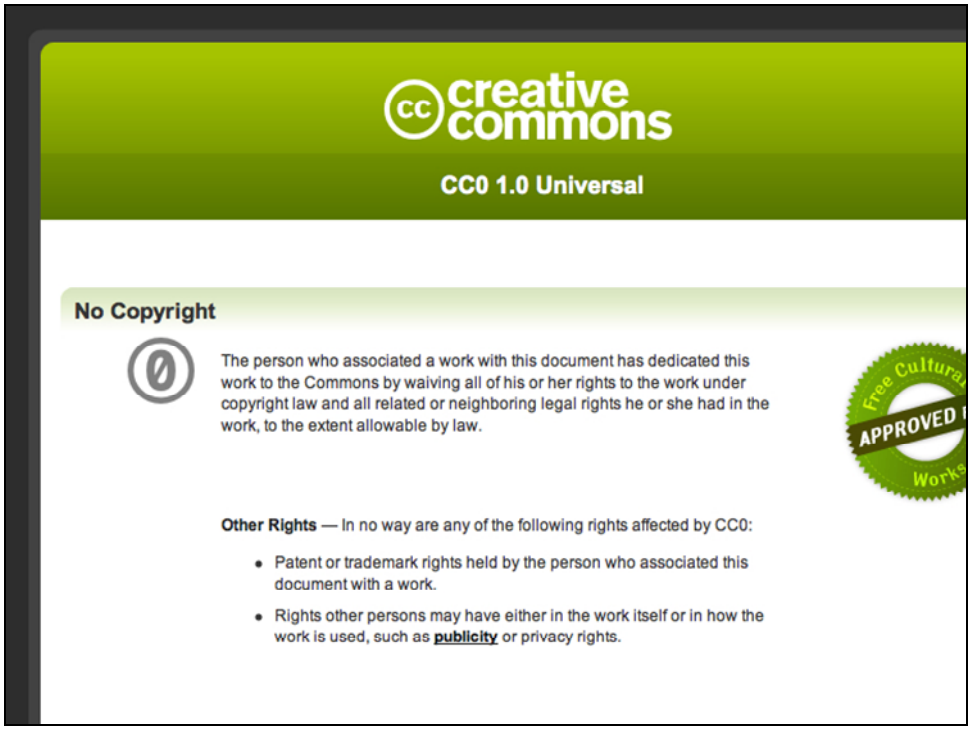

FIGURE 11-1 Summary of deed for CC0.
SOURCE: Creative Commons, *http://creativecommons.org/publicdomain/zero/1.0/*

Beneath it is actually a legal document that spells out the legal effects in more detail. As you see, the goal is that whoever associates this with their work is releasing it into the public domain. They retain patent and trademark rights and publicity rights, which cannot be waived, but they waive their copyright and related rights.

The goals of scientific research, now more crucially than ever before, demand a degree of transparency and openness that are undermined by restrictive legal rules. We need a consensus among the stakeholders of public science regarding common goals and infrastructure, so that the right rules may be chosen. Without such a consensus, a fragmented landscape of norms and legal rules will make data integration and sharing difficult or impossible. Of all the possible rules, the public domain remains the best choice for most public sources of data.

12. Comments on Designing the Microbial Research Commons: Digital Knowledge Resources

–Katherine Strandburg[36]

New York University Law School

In this presentation I will offer a few comments on some of the proposals in the monograph. At the outset, though, I will say a little about my background.

My interest in this topic stems from two things. First, I have done a fair amount of work on patents and research tools and on how the social norms of scientists might promote the sharing of research tools, how that is affected by patenting, and so on. The second thing is that before I went to law school and became a lawyer, I was a physicist at Argonne National Lab. So I used to be a scientist.

My comments today are not so much from the legal perspective. Instead, I want to emphasize the importance of social norms and of what in economics would be called "preferences," or what scientists really want to do. In other words, I will be looking at the proposals less from the institutional perspective and more from the scientist's point of view. In doing that, I use a concept that I refer to as *Homo scientificus*, playing off the economics concept of *Homo economicus*, or the rational, self-interested actor.

One can develop a typical preference profile for scientists based on various empirical studies, which in essence attempt to answer the question, "What do scientists care about?" The contours of such a profile would not be surprising to anyone in this room.

Scientists really care about doing science. They care about performing their research. They care about being able to do the research they want to do, and they want to know what other scientists are doing. In order to be able to do all this, to satisfy those preferences, they need access to various scarce resources. The scarce resources that they need most are, first, funding and, second, some sort of attention from other scientists so that they have the ability to go to conferences, talk to people, find out what is going on, and be in on the action. And although it is something that everybody knows, it is still important to emphasize that access to these resources is very strongly mediated by publication. If open access is to succeed, it will have to align with these preferences and, in particular, with the importance of publication in science.

In discussing open access, I am going to talk first about the issue as it relates to journals and then about it as it relates to data, because I think that these two issues are somewhat different when looked at from the researcher's perspective.

For journals there are at least three different ways that one could achieve the goal of complete open access. One way would be to try to create new open-access journals, perhaps having them based at universities, as Paul Uhlir discussed. A second path would be to prevail somehow upon existing journals to adopt open access. A third path that might lead toward open access—and which is the one that I want to emphasize—is a parallel path in which these proprietary journals continue to exist but, at the same time, one promotes open-access manuscript repositories.

---

[36] Presentation slides available at: *http://sites.nationalacademies.org/xpedio/idcplg?IdcService=GET_FILE&dDocName=PGA_053669&RevisionSelectionMethod=Latest.*

Let me explain why I think we ought to focus on that approach. My argument relies on the importance of something called "impact factor." Impact factor is a measure of the impact or importance of a particular journal, and the existence of such a measure exacerbates the emphasis on high-status publications. This emphasis has always been around, of course, but it has been magnified by the recent trend to try to quantify publication records using impact factor. The importance of a journal has become something we can measure objectively, and the things we can measure matter.

This emphasis on measuring journals' impact points to what may be a major problem for open access. The impact factor for current open-access journals is not bad, on average. The average impact factor is 4, and open-access journal impact factors range up to 9. If you compare this with the restricted journals purely on the basis of the average, it does not seem like much of a difference, as their average is 5.77. But what is more important is that the impact factors of the restricted journals range up to 50. In essence, all of the journals with the largest impact factor are restricted. I would submit that an impact factor of 50 trumps almost every other consideration for almost any scientist who is deciding what to do with a paper. A long-term altruistic belief in open access is just not going to win.

Thus efforts to encourage open-access models cannot depend on somehow getting scientists to forego publication in high-impact journals. We need a long- term strategy for either establishing open-access journals while researchers can still publish in the existing high-impact journals or else somehow increasing the impact of open-access journals. But this second approach in particular will not be easy. You can explain it in various ways—path dependence, or network effects, or preferential attachment—but once certain journals have been established as having high impact factors, it is very hard to dislodge them from that position or, conversely, to build up the impact factor of new journal. Scientists are unlikely to vote with their feet for the open access model as long as we have this issue with impact factors.

So, let us consider the first option of achieving open access: starting up new journals, perhaps based at universities. I have to be a bit of a wet blanket on the idea of using law reviews as a model. From personal experience of having published in both law reviews and scientific journals, I do not think the laws reviews offer a good model, for a simple reason. With law reviews, there was a tremendous proliferation of journals. Because every journal was associated with a university, it came to be that every university had to have a journal. After all, Harvard has one, and Yale has one, so we ought to have one, too. The result was an overly fine-tuned ranking of the journals, so that it became very important for an author to figure out exactly in which journal his or her article is going to get placed. The result is an overemphasis on placement.

Furthermore, graduate students are not law students. Law students are not doing research, and they need a publication venue. Graduate students, on the other hand, do not need a publication venue—they publish in the regular journals—and they generally do not have time for the journal editing functions which are provided by law students for law reviews.

I also doubt whether law review publication is really faster. I know from my former experience and also anecdotally from people I have talked to that in physics you can get something published in about six months. I do not think law reviews do much better than that. Finally, I am also not convinced that there will be synergies between the university's educational mission and the publishing of journals in science the way there

may be in law. While universities might have a big role to play in open-access publishing, I do not think that law reviews offer the right analogy.

The second option would be to get existing journals to adopt open-access policies. I believe we can go some way down that route, and I think we have seen that already with Springer Open Choice and other programs. But overall I think this option is unlikely to succeed because the journals, particularly the high-impact-factor journals, have tremendous bargaining power. Intellectual property laws protect these journals' proprietary approaches, and it will be difficult to put direct pressure on them. Even the open-access tiers, such as the Springer Open Choice method, are problematic because of the issue of where the money is coming from. Suppose I want to publish my article in an open-access journal. It will cost me, and the money will have to come from somewhere. If universities were to get directed funds to do that, initially that sounds like a great idea, but where is that money coming from? Is it going to come out of people's research budgets? Few scientists are likely to make that choice, given how tight research budgets are today.

The third option—and what I think is the more promising possibility—is author self-archiving in digital repositories. To establish manuscript repositories, you need only get the journals to acquiesce, so they will not sue you if you put your manuscript in the repository. They do not have to do anything different or change their mode of operation, which is a big advantage to this approach. The repository approach also lets the universities do those things that they can do easily and well, such as putting out manuscripts and getting the computer scientists and other researchers together to figure out how to mine the data in the manuscripts, while not asking the universities to take on those things that are either more difficult for universities to do, such as hard copy printing, or that are hard to dislodge from existing journals, such as the credentialing function. Finally, it would be possible to get funding agencies to mandate that their grant recipients deposit their manuscripts in these repositories, which would solve the collective action problem and align the incentives. Experience with the National Institutes of Health indicates that journals do not prohibit depositing in these repositories.

Last year a bill was introduced in the Senate, the Federal Research Public Access Act, which would mandate that all agencies ensure open access deposit for most federally funded research. It did not go anywhere the last time it was introduced, but I think there is some hope for something like this now because the Obama administration is making a big open government push, and this is totally consistent with that. The manuscript repository concept also tends to mitigate concerns with database protection statutes in Europe because even if the journal maintains its own database, it is not sole source anymore.

Of course, one would hope that this could be integrated with material in data repositories. You could even require, for example, that any researcher who used data from the repository would have to deposit any papers that resulted from that data. You could even go further and require researchers who used data from the repository not only to deposit any papers that came out of use of the data, but also to deposit the data and the materials associated with the papers. This is somewhat analogous to an open source software General Public License (GPL).

If such open-access manuscript repositories were successfully established, there would be at least two possible fates for the proprietary journals. One possibility is that the journals would adapt and move into a service provider role. They might have to finance this with page charges, but plenty of journals are financed with page charges now

anyway. They might make hard copies or archival versions of the journals, or perhaps they might create better or "premium" database services that competed with the open-access repository.

Or perhaps in the long term it might turn out that the proprietary journals are not commercially viable. If so, then the scientific societies or the universities or the knowledge hubs could essentially replace them or take them over, inheriting their impact factors. Or perhaps these institutions might simply partner with the proprietary journals. There are various possibilities for the future of these journals. Of the three possibilities, I think that manuscript repositories are probably the most practical path to an open-access world.

Switching gears, let us consider the data depositories. In many ways the issues regarding these data depositories are similar to the issues relating to material and research tool sharing. In particular, the major potential problem with data depositories is the collective action problem. This refers to the temptation to withdraw data or materials from the pool without contributing to the common pool.

The situation is similar to the classic prisoner's dilemma. Suppose there is a group of scientists each deciding independently whether to share or not share their data. Let us focus on one particular scientist, Scientist A, who is trying to decide whether to share or not share. If everyone shares, then everyone gets whatever the value of the database is once all the data is in it, which, of course, depends on how many scientists contribute. Scientist A will get a bit of first-mover advantage regarding his or her own data, perhaps from knowing the data better than others or perhaps because the repository gives Scientist A six months to use it exclusively. Scientist A also gets some reputational value from contributing, which might come in the form of attribution when someone reuses the data. Finally, Scientist A must take into account the fact that there is a certain cost to contributing—not just actual costs, but also opportunity costs.

On the other hand, if Scientist A does not share but everybody else does share, Scientist A still gets essentially the same benefits—access to everyone else's data plus exclusive use of his or her data. There might also be some cost to not sharing—a penalty imposed by the granting agency, for example, or reputational cost.

The bottom line in the economics approach to understanding the situation, which assumes Scientist A is a rational actor, is that Scientist A will do whatever offers the greatest return—share or not share, depending on a rational calculation of the benefits and the costs of each alternative.

In this overly simplistic, rational choice model the value of the database does not play a role, because Scientist A does not think that his or her choice is going to affect what everybody else is going to do. Either everybody else will not share, in which case there is no database, or else everybody else will share, and Scientist A can have a free ride—get the benefits of the database without the costs of sharing.

This is the typical free rider problem, but modeling it explicitly with an equation involving the benefits and costs to Scientist A, emphasizes that the success of a depository depends on increasing the benefits and reducing the costs of sharing. Even if we believe that people will naturally want to do the right thing, that they are going to feel guilty if they take data and do not share their own, it is still important to make the economic case as attractive as possible. One of the best ways to do this is to reduce the cost of contributing.

An interesting article in *Nature* from a couple of weeks ago called "Empty Archives" described exactly this phenomenon. Everybody said it would be great to have

this archive, but when it was set up at a university, at a cost of perhaps $200,000, nobody contributed. Why did they not contribute? Probably they just did not have time. They were too busy. It was too costly to them to contribute.

It is also important to provide rewards for contributing. This is why I believe the attribution aspect is important. We ought to think carefully about how to structure this incentive, because the best approach might not be the same mechanism as the usual citation mechanism or the usual collaboration mechanism. Furthermore, when we are considering what the best reward mechanism might be, we should keep in mind that rewards for contributing to the database are competing with the rewards for just sharing informally with collaborators. On the one hand, if I put my data in the repository, I may get rewarded by getting citations from everybody who uses it. On the other hand, if I keep my data for myself, it might help me get collaborations with other people. That could be quite valuable. So it is important to think carefully about how to do the rewards.

One other factor that must be taken into consideration is the value a researcher gets from depositing data versus not depositing them. If the data are very interdependent—that is, if a scientist's set of data is not worth very much by itself—then the scientist is much more likely to contribute the data than if it is possible, for example, to write 10 papers based on those data alone. Clearly, depositories are likely to work better for interdependent data, so it would be a good idea to look for opportunities to use them for interdependent data.

Finally, what about people who are not academics, such as scientists who are in industry? We should think carefully about what to do about industry scientists because of the free rider problem of people withdrawing data without contributing. Within the academic community, that may be difficult to do. Once you publish, people know that you have accessed the data, so it is hard to hide what you have done. If industry scientists have open access to the data, however, they are probably much more likely to be free riders because they may not be publishing everything. They also may not care as much if people are talking about them behind their back, and they may not rely on funding from the same funders. Furthermore, it is just much easier for them to keep what they are doing secret.

This is a problem. Should we put some fences around data repositories to keep industry scientists out? I do not know. Maybe, maybe not. Perhaps we will decide that because public money has gone into making these data, we should encourage private actors to do whatever they want with them. If you put a fence around a resource, however, you have it available to trade with people on the outside. You can get them to pay for the data or perhaps trade their own data for them. It is something we should think about.

13. Toward a Biomedical Research Commons: A View from the National Library of Medicine at the National Institutes of Health – Jerry Sheehan[37]

National Library of Medicine

I was asked to represent the perspective of the federal information policy community. There are numerous agencies across the federal government, each with its own practices and policies, so I am pleased to see that one of tomorrow's presentations will provide a broad cross-agency view. I am going to focus my remarks on my own small part of the world and give you a view of some of these issues from the perspective of the National Library of Medicine (NLM) of the National Institutes of Health (NIH). It will be a combined NLM-NIH perspective because the NLM is often the organization that sets up the repositories that respond to NIH policies.

We at the NLM have a mission to collect, organize, make available, and disseminate biomedical knowledge in order to improve health, medicine, and well-being. As such, the NLM is a variety of things to a variety of people. We are a library, with more than 8 million artifacts of different types. We are also a research and development organization, with intramural research labs that do work on data mining, data search, retrieval, presentation, image archiving, and so on. We are home to the National Center for Biotechnology Information, which not only provides data and information services, but also conducts a great deal of research on bioinformatics, improving the ways that we link, find, and do research with biomedical information. Our Specialized Information Services provide information resources related to environmental health, toxicology, and disaster information management. NLM also funds extramural research and training in biomedical informatics.

The NLM has a number of different kinds of databases, data sources, and information sources that it makes available to the community as a whole. They are, for the most part, publicly available databases, and they encompass a broad range of types of information and data sources. MEDLINE and PubMed Central, for example, are literature databases that provide access to journal citations and to full-text journal articles, respectively. MedlinePlus offers consumer-oriented health information. Two other NLM databases are GenBank, which is a relatively well known archive of discovered human genes, and dbGap, which is the Database of Genotypes and Phenotypes. It serves as a repository for data produced by NIH-funded genome-wide association studies, which link genotypic data to phenotypic data. It aids in answering such question as to what extent variations in genes are associated with variations in the expression of a particular disease or a condition, such as diabetes or obesity. We also have a small molecules database (PubChem), a hazardous substances database, and ClinicalTrials.gov, which is a registry for ongoing clinical trials and, as of about a year ago, became a repository for summary results of some of those clinical trials.

These databases are not static, but rather continue to grow. As of October 2009, MEDLINE had 16 million citations from more than 5,000 different biomedical journals, and we add about 700,000 new citations a year, representing new peer-reviewed literature

[37] Presentation slides available at: *http://sites.nationalacademies.org/xpedio/idcplg?IdcService=GET_FILE&dDocName=PGA_053665&RevisionSelectionMethod=Latest.*

from those journals. PubMed Central, which is a bit younger than MEDLINE, had about 1.8 million full-text peer-reviewed journal articles, and it gets about 300,000 users each day who are either accessing or downloading copies of those articles. There has been phenomenal growth in GenBank, which had on the order of 100 billion base pairs and about 100 million full sequences. Its rapid expansion reflects the deluge of information that must be captured, collected, curated, and maintained over time. As of October 2009, the clinical trials database had descriptive information on about 80,000 registered trials with information on 340 trials being added each week. We now have details on the results of these trials coming in at the rate of about 200 results records a month, so over time this will grow to be a fairly substantial resource for different kinds of comparative effectiveness research and for other kinds of evidence-based medicine research. With all of these databases, we notice that as we add content, the amount of use goes up.

Most of the databases that I have mentioned so far contain information that spans the spectrum of biomedical research and is accessed by a broad range of users—researchers, care providers, and the general public. We also have databases with information that is tailored for particular types of research and/or specific audiences. For example, our Influenza Virus Resource database (Figure 13–1) pulls literature from PubMed and PubMed Central as well as a variety of genome sets, some of which are generated by researchers associated with the National Institute for Allergies and Infectious Diseases. Thanks to their influenza genome sequencing project, we have now about 90,000 influenza genes and 2,000 full influenza sequences in the database.

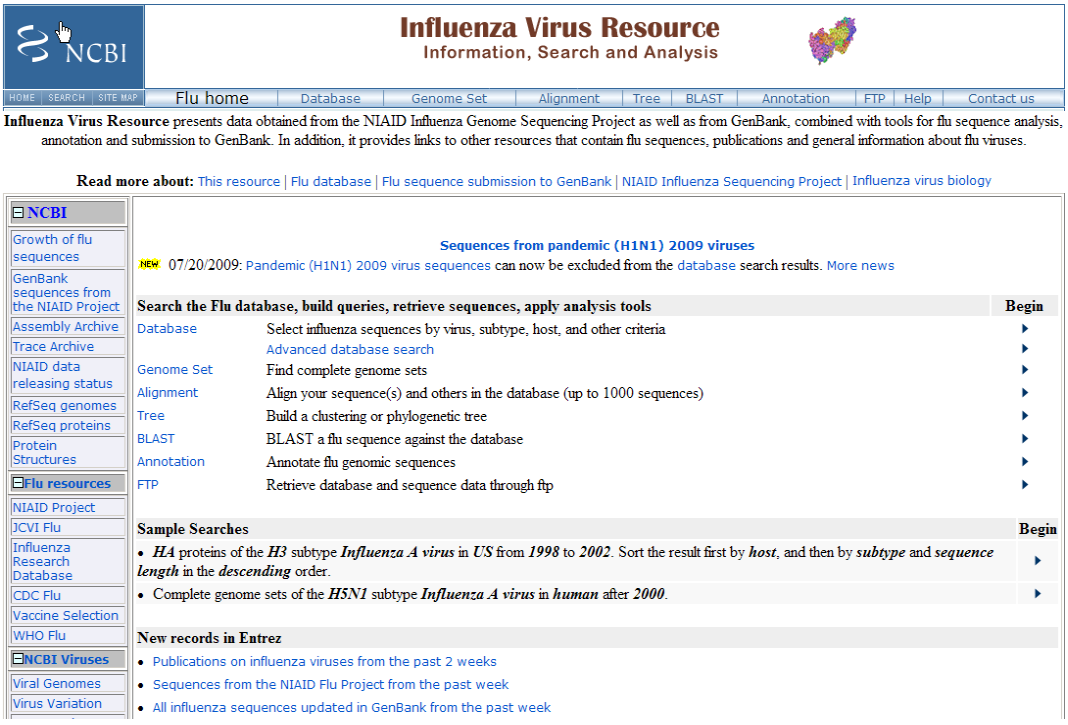

FIGURE 13–1 Screen shot of the Influenza Virus Resource database.
SOURCE: National Center for Biotechnology Information, National Institutes of Health

NLM has also been working to develop channels for getting out information about H1N1 influenza faster than typically occurs through traditional publication channels.

NLM worked with the Public Library of Science (PLoS), which developed a new type of publication, called *PLoS Currents*, to speed scientific communication. The first phase of the program focuses on influenza. The information in *PLoS Currents: Influenza* differs from that traditional journals in that it is not fully peer-reviewed; instead, a governing board comprised of experts in various aspects of influenza examines incoming contributions to make sure they are relevant and based on sound analysis. Articles are posted in a matter of weeks, rather than months or years, with the expectation that the reported research may eventually be published as a standard, peer reviewed publications

NLM initially developed a new service Rapid Research Notes to serve as an archive for *PLoS Currents: Influenza* and other fast turn-around research communication mechanisms that may be developed. Over time, it was recognized that much of the content of *Currents* took the form of short journal-like articles that could be archived in PubMed Central and benefit from the enhanced search capabilities build into that platform and the integration of PubMed Central with other NLM resources. Hence, *PLOS Currents* is now a full contributor to PubMed Central, depositing its full content into the archive, where it is assigned a unique identifier and can be easily accessed by researchers, clinicians, and the public.

All of the services I have described are essentially databases that collect, organize, and make accessible particular types of information, often for a particular community of users. While they have considerable value as stand-alone resources, their real value—to NLM and the user community as a whole—comes from linking them together into what could be considered an integrated, online biomedical knowledge resource.

To illustrate what we have in mind, imagine doing a MEDLINE search for cancer treatments. You find the abstract of an article that looks valuable. By analyzing the text of the abstract you find valuable and your original search string, we can generate a list of related articles that you might also find to be relevant. If any of them are available as full-text articles in PubMed Central, you can click on the link and retrieve it. If the retrieved article discusses a drug being studied in a clinical trial, you can scroll down to the bottom of the abstract and find an identifier called an NCT number. The NCT number is a unique clinical trial identifier that the NLM assigns to trials registered at ClinicalTrials.gov. By clicking on the NCT number, you are brought directly to the clinical trial registration record in ClinicalTrials.gov, which may also contain summary results information from the trial, including adverse events. If you look at the bottom of that ClinicalTrials.gov record—because we have standard formats and a process for putting these identifiers on citations and journal articles—you can link back to the original citation, which would take you back to that first article you found.

Where this gets more interesting is where this sort of linking can work across all NLM resources. Imagine that after searching PubMed for articles on treatments for influenza, you found an article in PubMedCentral that discusses the potential role of different drugs in treating the disease, e.g., oseltamivir and zanamivir. You could then follow a link to the PubChem database of small molecules to see the structures of these drugs and find out what is known about their chemical properties, be presented with a list of PubMed links to other articles with more information about the role of those chemicals in blocking the production of certain proteins, then link to three-dimensional views of the protein structures that show how the chemicals bind to them and even manipulate the images in various ways, and so on. This is the vision for the infrastructure we would like to create by integrating and linking among the multiple databases and information resources we have at NLM.

Bringing that vision to reality requires advances on multiple fronts. It requires the creation of unique identifiers for all of the elements involved and widespread use of those identifiers across the relevant communities, including among publishers. It also requires good vocabularies and terminologies to enable intelligent linking of related materials from across databases. At its simplest, such vocabularies can ensure that when a user performs a search on a key word, the system will not only know its various synonyms but will also know of various relationships involving that word, such as the relationship between a disease and agents used to treat it. These capabilities are among those in which NLM has strengths.

Data and information sharing remain a priority for NIH. Our efforts to promote data access and linking were boosted by the recent appointment of Dr. Francis Collins as the new NIH director. When Dr. Collins assembled the NIH staff on his first day on the job, he listed a set of areas where he thought there were significant opportunities for NIH. He identified an opportunity in applying high-throughput technologies to help enhance understand fundamental biology and uncover the causes of specific disease states. He sees such technologies as offered a way to ask questions that, as he put it, have the word "all" in them: What are all the transcripts in a cell? What are all the protein interactions? We should do it all, he said, because we have the ability to do that.[38]

Those of us who work on the data access were quite happy to hear how Dr. Collins followed up that opportunity with this quote. "Those kinds of questions are now approachable, especially if we do the right job of making really powerful databases publicly accessible to all those who need them and empower investigators in small labs as well as big labs to plunge into that kind of mindset." In short, I think you can expect to see a lot more development of these kinds of resources from NIH and development of a lot more of the data that will populate these kinds of databases.

NIH already has in place a number of agency-wide policies to promote data and information sharing. These include the NIH Data Sharing Policy, the NIH Public Access Policy, the NIH Genome-Wide Association Study Policy, and emerging policies (and regulations) governing clinical trials registration and results submissions. According to surveys, researchers support the idea of sharing data with others in the research community. In practice, we find that supporting data sharing does not always translate into active data sharing. We can build databases to house the data, but it is not enough to simply encourage voluntary contributions of data, for many of the reasons that have been discussed today. Thus, in a number of cases, the NIH has stepped in and put in place policies that either require the submission of information and data or else come as close as we can to requiring that without actually using that word. All of this is done with a great deal of consultation, public notices, and public comment in order to come up with a consensus or, at least, well-informed policy options.

Two policies in particular are standard for NIH-funded research. First, there is the NIH Data Sharing Policy, which imposes a requirement that any researcher who receives more than $500,000 in one year must provide with the grant application a plan for data sharing. We expect that the data will be made available in a timely manner, and the guidelines indicate this should happen no later than when the manuscript is accepted for

[38] *http://www.usmedicine.com/articles/new-director-at-national-institutes-of-health-outlines-goals-for-fy2011-funding-.html.*

publication. There are certain exceptions to this requirement, such as if the data can be identified as coming from particular individuals or if there are national security concerns.

Another requirement, expressed in the NIH Public Access Policy, is for NIH grantees to submit to PubMed Central any peer reviewed publications resulting from NIH-funded research. The publications must be submitted upon their acceptance by a scientific journal, but public release can be embargoed for up to 12 months. This embargo period addresses concerns that making the publications publicly available might affect the subscription-based publication models of a number of the journals used by NIH researchers. We have no evidence to-date to indicate that availability of articles in PubMed Central up to 12 months after their publication date has resulted in cancelled subscriptions to journals.

The NIH Public Access Policy applies to about 80,000 to 85,000 papers a year, but that is only a fraction of the papers that are deposited into PubMed Central every year. We work very closely with a number of publishers to collect other published papers, beyond those funded by NIH. We have developed mechanisms whereby several hundred journals provide us with their full journal content, sometimes with an embargo period, but often without. In other cases a journal may submit the final printed version of only those articles that were funded by the NIH, again with up to a 12-month delay in the release.

Certain types of studies have their own data sharing requirements. The NIH genome wide association study (GWAS) policy, for example, requires that researchers funded by the NIH for a GWAS must put the resulting data in a publicly accessible database where it is available to other researchers for subsequent years. We have built a database into which they can provide that information, dbGaP. As with depositing articles into PubMed Central, there is a delay period: A researcher can have 12 months of exclusivity to generate the first publication based on that data, even if other researchers are granted access to the data before that embargo period has expired.

The GWAS research generates both genotype and phenotype data, and the existence of the genotype data in particular leads to concerns about the subjects in the studies being identified. Thus we have a process to minimize the chances of the subjects being identified. The data are not publicly available, other than some metadata that cannot be used to identify individuals. There is, however, a procedure by which a researcher can request access to these datasets for secondary research use.

The clinical trial datasets have their own requirements for contributions. Results information must be submitted for certain phase 2 through phase 4 trials of FDA-regulated drugs and biologics and for non-feasibility studies of FDA-regulated devices. Results are required to be submitted within 12 months of the completion of the study if the drug, biological product, or device has been approved, cleared, or licensed for use. There are penalties for noncompliance with these requirements that are specified in the law. Congress also instructed the NIH to consider whether to require the submission of data for trials of unapproved products and the timeline for submitting such data, if required. As part of our efforts to determine whether to propose such a requirement we recently held a public meeting to solicit input on that topic, and others.

These policies demonstrate that there are a number of issues to consider about how to populate a commons or a publicly available database or an information-sharing repository. The first issue is how to get people to participate and submit data.

One way to do it would be to create an expectation within the scientific community that such data are shared as a normal part of the scientific enterprise. In the

biomedical sphere, the publishers have sometimes been helpful in creating such an expectation. For example, publishers will generally ask for a GenBank accession number when manuscripts are submitted that deal with genomic information. Something similar is true for articles reporting clinical trials: The International Committee of Medical Journal Editors announced that articles submitted for publication should have the data registered at inception in a publicly accessible database. Our database was the only one that met their criteria at the time, and publishers look for our NCT number in submitted articles as verification that the trial has been registered. The lesson is that there are groups other than funding agencies that can put pressure on the community to submit data.

Another issue to consider is how to monitor compliance. How do you make sure that people fulfill their requirements? When the NIH Public Access Policy was voluntary, compliance rates were quite low, less than 5 percent by one measure. When the policy became mandatory, there was a large increase in the number of manuscripts that were submitted to the database each week and in the compliance rate. Then, the first time that progress reports were due to the NIH for the projects subject to the policy, our project officers had a chance to look through the lists of referenced publications and ask for the PubMed Central ID numbers for those subject to the policy. More manuscripts were deposited into PubMed Central and the compliance rate jumped again. The lesson is that closing the loop on compliance—by identifying lack of compliance and informing those responsible for submission—is important if you want to ensure equitable submission of information and data into these repositories.

Simplifying the process is another way to encourage—or not discourage—submissions. We have done a great deal of work to try to simplify our systems for depositing, both for PubMed manuscripts and for other data.

We also have thought about ways to develop incentives to reward and recognize those who contribute their data and their publications. We do not have the answer there, but one approach would be to develop better way of tracking citations or other types of metrics so as to be better able to give people credit for what they have done. As noted, we assign identifiers to publications or data sets submitted to NLM. What is needed are standard practices for citing data sets and for recognizing the collection and sharing of data sets as a valuable scientific activity that is rewarded by the community and taken into account in hiring and promotion decisions.

There is a lot to think about in the design of policies governing these databases. Different kinds of data might warrant different kinds of approaches, even if the objective in the end is to get as much data as possible into a repository as quickly as possible. It is important to take into account the concerns in the research community about wanting to hold onto data, at least until a first publication. For certain types of data, such as clinical trial data, there may be concerns about releasing the data before a product or device is approved.

The lesson may be that policies need to be flexible. It is not necessarily the case that "one size fits all" when you are talking about different kinds of data. I am not familiar enough with the microbial datasets to understand the different ways that you might need to treat them, but Paul Uhlir talked about how different thematic communities might develop somewhat different rules for data submission.

Finally, it is important to facilitate interoperability. Putting data into a repository or archive is only the first step. The second step is making the data useful, which means making it possible for users to find what they are looking for, to understand what it is (i.e., appropriate use of metadata), and, where possible, to be able to find other data that

will add value to that original dataset. To that end, the NLM does a lot of work with a larger community of people on terminologies and vocabularies. There is an international group meeting in Bethesda today that is working on vocabularies for clinical medicine. Persistent digital identifiers can play a major role simply by helping to connect various information and records. The NLM has also worked on the metadata standards and data descriptions that are going to be used. We have been trying to provide ways to help people understand which kinds of standards exist for describing data and which formats data should be provided so others can easily make use of them.

We would also like to facilitate having data in a good form, archivable, and well described. One approach would be to use data scientists to prepare the data, but there may also be ways to embed good data sharing and data curation practices into research training or education processes so that people know how to prepare data well and can do it more quickly and more efficiently.

Ending on a positive note, I do think we are making progress in improving data and information sharing in the biomedical community. There are a number of successful efforts, some of them represented in the room here today. The number of conferences and meetings and activities indicates that there is a growing interest in making information more easily available within the biomedical research community in order to advance the science and make better use of the research dollars that are provided by the NIH and other funding organizations. I also believe there is an increasing recognition of the need for various types of infrastructure and resources.

How do we actually make this happen? We at the NIH build or fund the development of many places to store data. As I mentioned, we put a great deal of effort into standards and reference vocabularies to make the data more easily shareable. It might take awhile to realize the vision that is being articulated at today's symposium, but we are taking some good steps.

14. Academic Publications
–Fred A. Rainey[39]

Louisiana State University

As a bacterial taxonomist, today's discussions have been somewhat surprising for me. I was under the impression that someone who publishes a paper about, say, a mouse with a new trait has to make the mouse accessible to everyone who might want to work on it, but this seems not to be the case. We bacterial taxonomists are a bit more civilized in that if we describe a bacterial species, we have to deposit it in two culture collections in two different countries in the world. This is mandated by the Bacteriological Code, which is overseen by the International Committee of Systematics of the Prokaryotes. When you submit a paper for publication describing a species, your paper cannot be published and your species name cannot be validly published unless you provide a certificate of deposit from the culture collections. Perhaps this is something that could be applied in other areas of biological science.

Today, however, I am going to describe an academic publication, *Bergey's Manual of Systematic Bacteriology*, which is the work of many bacterial taxonomists and which is published by Bergey's Manual Trust. The trust is a nonprofit private organization whose role is to produce updated classification and descriptive information about the species of bacteria and archaea. All of this work is done by volunteers. There are no paid members. There are trustees, associates, and all of the authors who contribute the information.

Our editorial office is currently at the Department of Microbiology at the University of Georgia. In addition to our main goal of providing up-to-date descriptive information on bacteria and archaea, we also provide an unofficial classification of the bacteria and the archaea using a phylogeny based on the 16S ribosomal RNA gene, which is the gene that is accepted as the hierarchical phylogeny in the prokaryotes. One of our aims has been to provide the scientific community with an inexpensive resource on bacterial taxonomy—books that are not as expensive as most academic publications, which would be accessible even to graduate students and maybe even, in the case of one of our publications, to undergraduates. We also promote bacterial and archaeal taxonomy through publications and scientific meetings.

The trust was formed in 1936 as an outgrowth of the Society for Bacteriology, which is now the American Society for Microbiology. The founding trustees were David Bergey, Robert Breed, and Everitt Murray. I am not quite sure how the trust came to be named after Bergey. The trust signed a contract with the publisher Williams & Wilkins to publish *Bergey's Manual of Determinative Bacteriology*, of which nine editions have been published so far.

The activities of the trust are supported totally by royalties from these publications. The current trustees come not only from the United States but also from the United Kingdom, Germany, Belgium, and Korea, so it is truly an international organization. We have two current publications. One is the ninth edition of the *Bergey's Manual of Determinative Bacteriology*, which was last published in 1994, so it is

[39] Presentation slides available at: *http://sites.nationalacademies.org/xpedio/idcplg?IdcService=GET_FILE&dDocName=PGA_053721&RevisionSelectionMethod=Latest*.

somewhat out of date. Interestingly enough, it is still in print and sells very well. It is only out of date in that it does not include information on new organisms that have been described since then. However, it has a lot of valuable information on the older organisms, including many of the organisms that people deal with in clinical situations, in industry, and especially in university teaching laboratories.

In the last eight or nine years, we have been involved in producing a second edition of *Bergey's Manual of Systematic Bacteriology*. This is a much more substantial publication, and this second edition will comprise five volumes. Volume 1 was published in 2001, Volume 2 was published in 2005, and Volume 3 was published last week. Volume 4 is in press and should be also published by early 2010. We hope to have Volume 5 completed in 2010 as well.

This second edition is being published by Springer. It is an interesting publication deal because we do everything up to the point of the typesetting and then deal directly with the commercial typesetter, which prints it, and Springer distributes it. In short, it is not the typical academic book situation where the editor collects the manuscripts and sends them to the publisher, which takes it from there.

This five-volume work has approximately 600 individual authors involved. The authors, who are from many countries around the world, are each experts in a particular taxonomic group. They assemble the information on each organism described in the literature and then write about it for the publication. This approach guarantees that the information comes directly from the experts, the people who have done the most work with the particular organisms and may be presumed know the most about it. It is quite an achievement to get all these people together to write for the publication. The book is aimed at a global audience of microbiologists and other professionals who work in such areas as the biodiversity of microorganisms and the animal and human health community, as well as at undergraduate and graduate students.

The trust has a variety of other activities as well. We publish a taxonomic outline, as I mentioned before. It is available on our website (www.bergeys.org) and shows a total hierarchical structure from the high-level taxa down to the genus level for all of the bacteria and the archaea. We also give out a number of awards. The Bergey's Award is given each year to a young to middle-aged scientist who has made a significant contribution to bacterial taxonomy. The Bergey's Medal recognizes senior scientists who have had a lifetime commitment to the field of systematic bacteriology. And we promote the field of bacterial taxonomy by sponsoring sessions at meetings and having experts from various places participate.

To give you an idea of what the five-volume *Bergey's Manual of Systematic Bacteriology* has to offer, Volume 3 describes 240 genera and 1,346 species belonging to the phylum Firmicutes, which are also known as the low G+C Gram- positive bacteria. This includes a number of well known bacteria, such as *Bacillus* and *Clostridium* and *Streptococcus* species. More generally, the book describes many medically and industrially important organisms. I know that many people have been waiting for this book for quite some time.

Each volume is organized with a taxonomic outline at the beginning, followed by descriptions of all taxa that fall within the part of the phylogenetic tree covered by that particular volume. We describe the upper-level taxa—the phylum, the classes, the orders, and the families—and then the lower-level taxa, the genera, species, and subspecies. In some cases there are serovars and pathovars described as well.

Volume 3 has 1,450 pages and is available on the Springer Web site for $249, which is a bargain for what you get in terms of the number of pages and the amount of information. It is a high-quality, hard copy book with glossy paper and many pictures, photomicrographs, and diagrams.

Each of the chapters describes all the information available on a particular genus and on all of the species of that genus, and that information includes all of the phenotypic data as well as some genotypic data. The chapter will also describe how to differentiate each particular species from the other species of the genus as well as how to differentiate particular genera from related genera.

As useful as all this is, however, we are still in the situation of publishing a book—a paper thing that is basically like a doorstop and is not something you can easily carry around with you. Nor is it easy to access if you are away from your desk or your bookshelves. So at some point we will need to move from paper to a digital format, but doing so will require us to deal with a variety of issues.

First is the amount of material and information that we have. Each of our species descriptions has probably between 150 and 200 characters. We have 8,000 described prokaryote species, so that is a lot of characters. I know that this may not sound like a lot compared with the amount of genome data that is being accumulated, but in bacterial taxonomy terms it is a lot of characters.

The second major problem is that we are dealing with a moving target. Every day there are papers being submitted describing new species. Each month when the *International Journal of Systematic and Evolutionary Microbiology* is published there are perhaps 10 additional genera and 20, 30, or 40 additional species in each new issue. Keeping this body of information updated is a major task. An author who writes about a particular genus may not be enthusiastic about updating the genus chapter just because one additional species has been added. However this updating is important in the context of comparative taxonomy. We need to come up with some workable way of updating the publication on a regular basis. Volume 1 was published in 2001, and there are now probably about 580 genera that were discovered too late to be included in the relevant volume.

There are various options for dealing with this issue that we could go ahead with right now. We could, for example, have Web-based access to the information, either in the form of PDFs of the individual chapters or as Web pages in which everything could be searchable and fully linked. We have also considered having some sort of Wiki-type format in the future. It would not be as totally open that anyone could write about anything and change anything they wanted to, but rather the authors of those chapters could continually update their chapters, so that it would be a living document.

We are also considering the question of how to get all this information about the characteristics of the organisms, which exists as chapters on the genera and the species, into a database format. This is something we probably have to move to in the future.

Then there are also some issues associated who actually owns the data. Even though the volumes are published by Springer, Bergey's Manual Trust owns the copyright on the printed books, but all of the factual data come initially from the primary literature. The original papers are all referenced in the volumes, but the information is being lifted from the primary literature by the authors of the chapters.

Finally, we must determine how we will fund the updating and curating of all of these data in the future and how often we should be updating the data. We are quite

happy to have all of this be open access, but we will have to have some sort of funding to keep the activity moving along.

15. StrainInfo: Reducing Microbial Data Entropy – Peter Dawyndt[40]

Ghent University, Belgium

In a keynote speech [5] given at the 2002 O'Reilly Open Bioinformatics Conference, Lincoln Stein of Cold Spring Harbor Laboratory compared today's bioinformatics landscape with the old Italian city-state model:

> "During the Middle Ages and the early Renaissance, Italy was fragmented into dozens of rival city-states that were formed by legendary families, such as the Estes, Viscontis and Medicis. Although this era had some positive aspects, the political fragmentation was ultimately damaging to science and commerce because of the lack of standardization in everything from weights and measures to the tax code, the currency, and even the dialects the people spoke. And because a fragmented and technologically weak society was vulnerable to conquest, Italy was dominated by invading powers from the 17th to the 19th centuries.
>
> The old city-states of Italy are an apt metaphor for bioinformatics today, as this field is dominated by rival groups that each promote their own Web sites and Web services and data formats. And while this environment has led to some creative chaos that has greatly enriched the field, it has also created a significant hindrance to researchers who wish to fully explore the wealth of genome data.
>
> Eventually, the nation of Italy was forged through a combination of violent and diplomatic efforts, so that it is now, despite its shaky beginnings, a strong and stable country. It is also a component of a larger economic entity, the European Union, whose countries share a common currency, a common set of weights and measures, and a common set of rules for national and international commerce. The hope is that one day bioinformatics will achieve the same degree of strength and stability by adopting the same universal code of conduct."

If you look at the consortium of culture collections that are members of the World Federation of Culture Collections from a bioinformatics point of view, one striking observation is that fully integrated information about microorganisms is not immediately available. In contrast to the data organization mechanisms put in place by the National Library of Medicine (NLM), that allow to easily follow the links between genes and the diseases they cause and various publications on these genes and diseases, you cannot easily do the same sort of research on the strains in the culture collections. Suppose, for example, you would like to find all 16S rRNA genes of *Pseudomonas* strains that have been isolated from soil and get information on their taxonomy, ecology, genomes, and so on. There is no easy way to collect this information without time-consuming searches.

At Ghent University, we decided to perform an experiment to see how far we could get with the information as it is made available from the culture collections today, in order to try to build something like the infrastructure constructed at the NLM. The idea was to build a software platform that accesses information from the culture collections, so that somebody could direct questions to a single point of access instead of having to visit

[40] Presentation slides available at: *http://sites.nationalacademies.org/xpedio/idcplg?IdcService=GET_FILE&dDocName=PGA_053718&RevisionSelectionMethod=Latest.*

the catalogues of the culture collections one by one to get an answer. Collecting online information coming from autonomous and heterogeneous data providers is the sort of job a *web spider* does, so we decided to look into building this kind of infrastructure. We also decided not to build the software platform as a monolithic structure, but make it flexible in the sense that it could take into account regional projects that had already established portals for a number of culture collections, in various countries or in regions like Asia.

The idea is that if a researcher has a question about a microorganism, instead of having to go to the online catalogues of the individual culture collections, the system would do it for the researcher. The Internet is conceived as a collection of data that is linked together by hyperlinks. These hyperlinks indicate connections points within and between various datasets. But the Internet does not lend itself very well to discovering new links and new ways of finding compatibilities between different datasets.

The approach we took to link microorganisms with all their downstream information is inspired by the "knuckles-and-nodes" model described by Lincoln Stein [6]. The idea is to organize nodes of information into a number of thematic networks, each with its own hub, or knuckle, that interconnects with all the other networks through the knuckles. Some of these knuckles have already been established, so we simply needed to integrate them. The Bergey's Manual, as the previous speaker described, could serve as a taxonomy knuckle, and a variety of bioinformatics knuckles are also available (e.g., public sequence databases bundled into the International Nucleotide Sequence Database Collaboration; INSDC). But what was missing was the organism knuckle, which would provide access to all the bacterial, archaeal and fungal resources that are in the culture collections and, by extension, in all public and private research collections.

So here's another look at Lincoln Stein's idea: a number of people put their data online in databases or simply as text documents. To bundle all this information together in what he calls knuckles requires the construction of some sort of integration network that helps discovery across these disparate data sources. One possible approach to accomplish this could be to build an infrastructure on top of the disparate data sources where globally unique identifiers are assigned in an ongoing discovery process of pointers between autonomous and heterogeneous data sources.

By following this approach, you can test various hypotheses and answer different questions about the data. One particular question we focused on was to estimate how many organisms for which the complete genome sequence is available from public databases are also available from public culture collections. To get an answer on this question, we took the integrated information from the culture collections and simply linked it with the Genomes OnLine Database (GOLD; *www.genomesonline.org*) [2]. What we found was a tremendous gap between the availability of genomic information and the availability of the sequenced organisms in public culture collections.

In bacterial taxonomy there is a rule that states that if you want to describe a novel species, you have to deposit its type strain in at least two culture collections in two different countries. This to safeguard that the species remain available for further research. A similar rule is not required for when depositing and publishing the complete genome sequence of an organism. It seems natural that researchers would make the biological material available in order to add value to their publication of a whole-genome sequence. However, the results of our investigation show that more than 50 percent of the complete genome sequences that have been deposited in the public sequence databases do not have a publicly accessible organism.

Let me now give a more detailed description of the StrainInfo bioportal (www.straininfo.net) that we developed [1]. One major purpose of the bioportal is to increase discoverability of the biological material preserved within a global network of culture collections. One possible way to access the bioportal is to enter the scientific name of an organism. You can also enter an accession number assigned to a sequence record by the INSDC or a strain number of organism assigned by whatever culture collection. The smart search feature of the system will figure out what your search terms mean, so there is no need to specify what type of identifier was used. For example, if you enter a strain number like LMG 6923, the system will display all available information about that particular microorganism by collecting strain information from all culture collections. The system collects all this information on the fly, and tries to associate it with related information from taxonomic databases, sequence databases, publication repositories and so forth. From the resulting information one can easily find the different collections that have a copy of this organism, simply by looking at their geographic distribution on a world map. In addition, for example, the resulting information also includes all genome sequences from this organism and pointers to the published literature that made us of this organism.

If you want to drill down to more detailed information, you can, for example, visit the individual online catalogues of the culture collections. Deep links to related information in these online catalogues are provided, obviating the users' need to know all the strain numbers that have been assigned to the same organism by the different collections. Say, that you know the strain number of an organism assigned by the American Type Cultures Collection (ATCC). In the public sequence databases you may find sequences that are linked to that particular ATCC number. What we add are links to all sequence, regardless of the strain number being used when the sequence was deposited.

The StrainInfo bioportal offers a way to discover information by fetching all related data and trying to make sense of it using different integration strategies. For example, the bioportal provides the entire genealogy of a strain, from the initial isolate down to its distribution from one culture collection to another or from one researcher to another. Integrating this genealogical information is extremely difficult. First of all because the way it is encoded in the catalogues of the culture collections is completely unstandardized. Secondly, quite a lot of implicit distribution information is missing from the catalogues of the culture collections. For example, collections change names as they move from one funding agency to another, as two or more collections merge, or for other reasons. Whereas people that are intimately familiar with the past history of culture collections might known that Collection A has changed its name to Collection B and then to Collection C, the broader community is not aware of such reorganizations. This might introduce uncertainties in the distribution history of reference strains.

As a countermeasure, we developed an algorithm that can automatically reconstruct the strain distribution history from unstandardized and incomplete textual descriptions [8]. My original idea was to put a Ph.D. student to work on building an editor so that end users could reconstruct strain distribution histories by manually gluing bits and pieces of information together. But during the development process the student came to me and said, "I think I found a way to build these histories automatically". I did not believe him at first, and made him convince me that the automatic predictions actually were correct.

To prove the success rate of his automatic reconstruction algorithm, the student first undertook a fully manual curation experiment. He took the collection of all 8,000 bacterial type strains as a working data set for which the strain distribution history needed to be reconstructed, and went to the Laboratory of Microbiology at Ghent University, trying to convince the local microbiologists to reconstruct all 8,000 strain histories by hand using the information that was made available from the StrainInfo bioportal. At the same time, he processed the data with his reconstruction algorithm to construct the strain distribution histories in a completely automatically fashion. After three of four weeks, enough histories had been manually reconstructed (about 60 percent) in order to evaluate the success rate of the automated predictions.

A comparison of the manual and automatic history reconstructions showed that 98 percent of the histories that were manually created could be rebuilt automatically, with only a minor number of inconsistencies. After inspecting those inconsistencies, it often turned out that the manual curation was wrong. The automated reconstruction algorithm uses all available strain distribution information at the same time, which overall makes it more robust than the manual reconstruction. Only related to the lack of completeness of the reconstructed strain histories, the automated reconstruction algorithm was outperformed by manual reconstruction. The reason for this observation is that some of the strain distribution information is not explicitly available from the online catalogues of the culture collections. Manual curators can compensate this lack of implicit data using their background knowledge about the problem domain. For example, they may know some of the relationships between the culture collections or the people working in those culture collections.

The bottom line is that we were able to motivate some experts to make a manually created data set and use it as a benchmark to prove that we could automate the whole process. As such, we could automatically reconstruct the strain distribution history for more than 700,000 strains of microorganism that are available from a global network of culture collections. In order to counterbalance errors made in predicting the strain distribution history, the StrainInfo bioportal allows its end users to make corrections if they find mistakes and to make updates whenever they have additional information.

This is one example that shows how we were able to use a semi-automatic approach to conquer a problem that seemed impossible to automate at first sight. We found that we could approximate human curation with automatic prediction while allowing end users to make annotations and corrections to further enhance the quality of the information.

As a second example, I will demonstrate the ontogrator experiment (tools.envotestsite.org/ontogrator) that makes use of information extracted from the StrainInfo bioportal. Usually, ontologies are used when autonomous and heterogeneous data sets need to be integrated into a single portal. This is the general approach taken by the ontogrator, that automates the integration pipeline for a given set of data sources and a given set of ontologies. As an experiment, the researchers that developed and implemented the ontogrator used the following data sources: CAMERA [4], PubMed, GOLD [2], SILVA [3], and StrainInfo [1]. In addition, they used a series of controlled vocabularies, or ontologies, related to ecology, geographical locations, habitats, and so forth. Next, he integrated the different data sources based on the fact that they can be linked through a common vocabulary. The integrated interface resulting from the ontogrator approach for example allows one to search for all entities that relate to dairy products. The underlying knowledge base knows that yogurt, cheese and ice cream are all

dairy products, so the user does not need specify this. And yes, it is also possible to specify exactly what kind of dairy product you are looking. Searching StrainInfo this way produces hits on organisms in the public culture collections that were isolated from dairy products, along with their descriptions. From those organisms it is possible to jump directly to related information in CAMERA, the published literature or to get their complete genome sequences. Simply by using a shared vocabulary, we can allow users to use faceted browsing as a way to relate pieces of information that were not explicitly related to one another.

As a final comment, when we were building our initial prototype of StrainInfo, we deliberately decided not to put any extra burden on culture collections and their staff. Knowing that culture collections overall have limited information technology resources, we took the challenge to work with the information as it was available and see how far we could go in our integration experiment. We simply screen scraped the data from the online catalogues of the culture collections and indexed it in somewhat the same way Google is doing, as common data exchange formats have simply not been adopted to in the field of culture collections. This approach worked initially, until we came to the point that we were indexing more than 60 culture collections. By that time, however, it had become clear to the culture collections that we had built a software platform that gave them more visibility. This increased their willingness to make some additional effort in helping us to scale up the integration process. Instead of simply screen scraping the HTML-formatted data from the online catalogues, we now offer the culture collections the export of their data in a standardized exchange format called the Microbiologial Common Language (MCL) [7]. This allows us to index the culture collections more frequently, extract and integrate more detailed information, and scale up the number of culture collections being indexed.

We were aware of the fact that it would be extremely difficult to convince culture collections to export their data in a standardized XML format, knowing that this might seem quite straightforward for a computer expert. But because the culture collections could directly see the added value from the initial prototype, more and more they started to provide us their data in the MCL format, and more culture collections wanted to become members of StrainInfo as soon as possible. Gradually introducing a data exchange standard thus will allow us to scale up the integration experiment behind the StrainInfo bioportal from the five culture collections we initially had in mind to more than 500 culture collections that are member of the World Federation of Culture Collections.

REFERENCES

[1] Dawyndt, P., Vancanneyt, M., De Meyer, H. , Swings, J. (2005). Knowledge Accumulation and Resolution of Data Inconsistencies during the Integration of Microbial Information Sources. IEEE Transactions on Knowledge and Data Engineering 17(8):1111-1126.

[2] Kyrpides, N. (1999). Genomes OnLine Database (GOLD): a monitor of complete and ongoing genome projects worldwide. Bioinformatics 15:773-774.

[3] Pruesse, E., Quast, C., Knittel, K., Fuchs, B. M., Ludwig, W., Peplies, J., Glöckner, F.O. (2007). SILVA: a comprehensive online resource for quality checked and aligned

ribosomal RNA sequence data compatible with ARB. Nucleic Acids Research 35(21):7188-7196.

[4] Seshadri, R., Kravitz, S. A., Smarr, L., Gilna, P., Frazier, M. (2007). CAMERA: A Community Resource for Metagenomics. PLoS Biol 5(3), e75.

[5] Stein, L. (2002). Creating a bioinformatics nation. Nature 417:119-120.

[6] Stein, L. (2003). Integrating biological databases. Nature Reviews Genetics 4: 337-345.

[7] Verslyppe, B., Kottmann, R., De Smet, W., De Baets, B., De Vos, P., Dawyndt, P. (2010). Microbiological Common Language (MCL): a standard for electronic information exchange in the Microbial Commons. Res Microbiol 161(6):439-445.

[8] Verslyppe, B., De Smet, W., De Baets, B., De Vos, P., Dawyndt, P. (2011). Make Histri: Reconstructing the exchange history of bacterial and archaeal type strains. Systematic and Applied Microbiology.

16. Research and Applications in Energy and Environment – Daniel Drell[41]

Department of Energy

In this presentation I will describe some of the Department of Energy's programs, particularly those related to sequencing. The Department of Energy (DOE) is generating more and more data in ever larger amounts. Our missions include developing biofuels, understanding the potential effects of greenhouse gas emissions, predicting the fate and transport of contaminants, and developing tools to explore the interface of the physical and the biological sciences.

These first three missions are not new. For years we have had high-throughput computing used to simulate climate processes. We are also the inheritors of the Atomic Energy Commission and its legacy of the nuclear weapons programs. Many of the nasty contaminants developed and used in those programs got dumped in the ground and ignored for many years. Now we have to deal with them.

The Biological Systems Science Division, where I work, has a genome sciences program. We also have three large bio-energy research centers, some imaging and radiobiology research programs, and a very small program on ethical, legal, and social issues. And then we have one user facility in our division called the Joint Genome Institute.

The parallel division, the Climate and Environmental Sciences Division, has programs appropriate for that division, looking at modeling climate processes and characterizing subsurface biogeochemical processes. I am currently the chair of an interagency group with a diverse collection of member agencies all with an interest in microbial research. That has led to a charter, which is to maximize opportunities offered by this science, as well as one primary direction to fulfill that charter: to generate large amounts of data and to get the most out of these data.

The DOE's Joint Genome Institute was started in 1997. The Facility was built to carry out the DOE's obligations to the Human Genome Project. We assembled the sequencing and processing facilities in one place in order to take advantage of economies of scale and do the job faster, better, cheaper. A major aspect of the Joint Genome Institute is the community sequencing program, which is an outreach program to the wider community to provide a high-throughput, highly capable sequencing facility. Its goal is to provide sequencing and analyses services to anyone who has some tie to one of the DOE missions in bioenergy, biogeochemistry, or carbon cycling and who passes its peer review process.

The four areas of genome science within the community sequencing program are plants, fungi, prokaryotic isolates, and metagenomes. The outputs of the sequencing runs performed at the JGI are put into the Integrated Microbial Genome (IMG) system. The throughput from these machines has absolutely revolutionized biological science in a very short period of time.

This is one of the reasons that this meeting is critical—because the front end of data production is quite literally the tsunami that several people referred to yesterday. My presentation is already out of date, since it is four days old, but as of four days ago there

[41] Presentation slides available at: *http://sites.nationalacademies.org/xpedio/idcplg?IdcService=GET_FILE&dDocName=PGA_053727&RevisionSelectionMethod=Latest.*

were 1,110 published complete genomes in the public literature. There are also 111 archaeal complete genomes, 3,342 ongoing bacterial projects, 1,165 ongoing eukaryotic genomes; and 200 metagenomes, for a total of nearly 6,000 sequencing projects of biological organisms that are in various stages of completion. It will be a big challenge to deal effectively with all this.[42]

In the future, single-cell projects will provide another major source of data. It is extraordinarily exciting to be able to sequence the genome of a single cell without growing it. It will also be another source of microbial data however, with which a commons is going to have to deal.

The data flood is not stopping. It is not leveling off. It is increasing. Potential future projects that the Joint Genome Institute is talking about are in the terabase range—trillions of base-pairs. The institute is also engaged in some international projects.

All of this information is deposited in the Integrated Microbial Genomes (IMG) system. The IMG is a data management and analysis platform designed to get value from the sequence data produced by the Joint Genome Institute and other places.

Another facility that we support is the Environmental Molecular Sciences Laboratory (EMSL), which has high-throughput capabilities in nuclear magnetic resonance, mass spectrometry, reaction chemistries, molecular sciences computing, and so forth. We are aggressively exploring ways of putting these two facilities together.

In the future, we hope to issue a call for projects that entail both Joint Genome Institute sequencing and EMSL proteomic analyses—the kinds of projects that neither of those two facilities could do by itself but which, if they work together, can be tremendously valuable and provide yet another kind of data that a commons would want to include.

Our data sharing policies state that any publishable information resulting from research that we have funded "must conform to community recognized standard formats when they exist, be clearly attributable, and be deposited within a community recognized public database(s) appropriate for the research conducted." There is no time element here, and it is left up to the community to determine what the standards should be. In sequencing, we have moved to the immediate release of raw reads, and reserved analyses of more than 6 months are discouraged. Twelve months is the absolute maximum we will hold onto data without releasing it. A reserved analysis is anything that would compete with the stated scientific aims of the submitter of the project. We are also launching a knowledge base initiative to accelerate research and integration and cross-referencing of data.

To sum up, there is just so much data being produced so rapidly that you feel that the rest of biology is not keeping up. I think this effort by the National Research Council is critically important.

---

[42] As of the end of February, 2011, there were 1,627 published complete genomes in the public literature. There are also 211 archaeal complete genomes, 5,790 ongoing bacterial projects, 2,002 ongoing eukaryotic genomes; and 308 metagenomes, for a total of nearly 10,000 sequencing projects of biological organisms that are in various stages of completion. Source: Genomes On Line database, *http://www.genomesonline.org/cgi-bin/GOLD/bin/gold.cgi*. This only underscores the challenges that collectively we (and a microbial commons effort) face.

17. Large Scale Microbial Ecology Cyberinfrastructure – Paul Gilna[43]

University of California, San Diego

A few years back, I spent some time with the GenBank Project when it was at Los Alamos and before it moved to the National Library of Medicine. It was during this time that the project initiated the concept of direct data submission. Prior to that, all the data that entered the database were essentially lifted from the printed page and manually entered by a curatorial staff based at Los Alamos. There was broad recognition that this kind of approach would not scale up, particularly because of the growth in genome data that everyone knew was coming.

The idea, then, was to convince the members of the scientific community that they should submit their own data, preferably in advance of publication and preferably in electronic form. I remember getting a call from an author after we had asked if he would mind submitting his data in electronic form, and he said, "But I faxed it to you." So it was a hard-fought battle to install that paradigm, and we were helped out by the scientific journals, which were the primary architects of what is now a relatively standard policy of requiring that authors submit their data to the databases and present evidence of that submission as part of the publication process.

For a while the community was quite resistant to the notion of submitting and releasing data, with the standard arguments against release being that researchers who had spent a lot of time generating data needed time to exploit the data themselves before releasing them to others. Keep in mind that at this point we were talking not about whole genomes, but about single genes. It was natural that a researcher who spent a considerable amount of time isolating the necessary materials and performing the sequencing would want time to do the science on the gene, so folks would hold back on releasing or submitting their data.

The submission process did eventually catch on, at least partly because of the policies instituted by the journals. There came a turning point, however, where suddenly it seemed to be in a researcher's interest to submit data and have those data released in the researcher's name in GenBank, rather than have them held in confidence because there had been many instances where a scientist was essentially outpacing his or her competitors by having released the data. Researchers therefore came to see that protecting their data was, in fact, against their own interests because competitors could use something like GenBank to not only deposit but release—and, in a sense, show prior evidence of publication—of their data.

Today, we have reached a point where it is relatively easy to sequence not only a gene but an entire genome. I believe we are rapidly approaching that same point where it is in the interests of everybody to have their data available and in their own names—and citable in their names in the public collection.

Today I work on the Community Cyberinfrastructure for Advanced Microbial Ecology Research and Analysis project (CAMERA), which was created to serve and perhaps promote the creation of a community around the general discipline of microbial

[43] Presentation slides available at: *http://sites.nationalacademies.org/xpedio/idcplg?IdcService=GET_FILE&dDocName=PGA_053723&RevisionSelectionMethod=Latest*.

ecology. It is a global project, with approximately 3,100 researchers from more than 70 countries who are registered, daily users of the CAMERA project.

We are now moving towards CAMERA 2.0, the goal of which is to provide a metadata-rich family of scalable databases and to make those available to the community. This represents a major change in how we perceive genomic data. In the past, for the large part, we paid scant attention to information about the environment from which a particular genome was isolated. Today, of course, we spend a considerable amount of time, particularly with metagenomics projects, sampling environments. As a result, data about the environments from which those genomes come take on a significant scientific importance.

So, in part, the purpose of CAMERA is to collect and reference the increasing volume of metadata on environmental genome datasets and to provide the ability to query based on the metadata. The underlying assumption is that the metadata are just as important to the scientific process as the data themselves.

No one system is going to be able to generate and create the necessary armament of tools needed by the scientific community to analyze the coming tsunami of data. CAMERA is a platform that the community can use to integrate such tools into a system. One of the key features of the project is a semantically aware database that is designed for storing and making available the environmental parameters, with the goal of facilitating the observation and management of sequence data.

For any set of data in a database, there are often relevant data that exist in other databases or other repositories, and it is important to be able to connect seamlessly to those relevant data as well as to connect to and utilize ontologies that are available. It is also important to be able to query these data. New query methods include graphical and geographical methodologies. One of the things CAMERA has been working on is providing an easy way to query geographically. CAMERA also has data submission capabilities, and the community is encouraged to submit metagenomic datasets. There are thousands of metagenomic data collections waiting to be submitted and made available in some form or another.

To reiterate, CAMERA has been designed to collect the various metadata associated with a given sample and to make those types of metadata conditional upon the environment. Even though we have various standards for the metadata, the system also permits the user to add new metadata that might not be considered by the standard system. Indeed, the whole submissions paradigm has changed and evolved. Metadata are now collected before the sequence data. We are capturing important data even before the core or anchor data have been generated, after which there is a series of steps along the way.

Over time, as genomics and sequence generation has evolved, the appearance of data in the online electronic databases has become somewhat decoupled from the traditional journal publication process. Many data are appearing in the scientific databases with no reference whatsoever to a publication or the scientific literature. In many cases, that publication may arrive after the fact. We strive to conform to data standards where they exist. Where they do not, we take part in consortia that are designed to generate those standards. Although our initial focus was on marine microbial science, it was always understood that it makes no sense from a scientific perspective to limit the project so narrowly, so CAMERA contains data from soil, from hosts, and from air sampling.

An important part of the project is to generate a user-friendly computing environment. Thus a great deal of effort is put into making the system and the interface easily usable for the community of researchers. This involves consideration of workflows and workflow architecture. Various parts of industry and academia have been using workflows for a while now, but it has taken time for them to come into widespread use in genomics and bioinformatics. An example of this approach is a simple annotation pipeline that is available to users, who can customize components or actors in the workflow and so tailor the annotation process to their needs. In the past, the whole process of annotation was a black-box effort that was done almost offline by systems. Now we are giving the user control over what and how data are annotated.

Another issue that we have been working on related to workflow is the concept of provenance. That is the ability to provide the information needed to be able to replicate or repeat an experiment.

The basic reason for CAMERA's existence is that we believe we can make a major difference. This is one of the most exciting times for genome biology and genomics. We have reached a stage where the ability to peer into a genome is no longer rate-limited by the ability to generate the data for that genome. We have worked out how to generate the data—and now the community needs help to work out what to do with it. Moreover, we are not just dealing with growth in the amount of data; we must also deal with growth in the number of investigators who are generating the data. The ability to sequence large amounts of information is now available to a vastly broader segment of the scientific community than has been the case to date.

Up until now, the ability to generate large amounts of data was largely the purview of a small, elite set of groups in the United States and Europe: the Joint Genome Institute, the sequencing centers at the National Institutes of Health, the Sanger Center, and a variety of other centers in Europe. Now, a machine capable of replicating the outputs of these facilities can be bought for around $500,000. The number of scientists who need access to these data, who generate the data, or who need systems and tools to be able to analyze those data is growing as fast as the amount of data itself.

Thus we no longer have the luxury of time to learn how to be a bioinformaticist in science. That places a great responsibility on everyone here to make sure that the data we generate—and the tools we create to analyze those data—are far more usable and easy to understand than has been the case in the past. This is a significant community responsibility. For a long time we have been in the business of generating software for people who know how to use it, who understand the basics of what is going on, and who can tolerate the "UNIX-speak" of most of our software tools. But now that is changing. Our systems and our approaches need to address a broader audience.

Question and Answer Session

PARTICIPANT: CAMERA is a fantastic project. What I would like to know more about is some of the organizational aspects. You mentioned a foundation. How does it make decisions? How do you put this all together? And what relations do you have with the university? I imagine you have external funding of some importance, but does the university support you? Are you an integrated part of it? Do you get an advantage from that? And how do you make decisions and govern the project?

DR. GILNA: The project is funded by a grant from the Gordon and Betty Moore Foundation to the University of California, San Diego. So it is staffed at UCSD, at the California Institute for Telecommunications and Information Technology, and at the Center for Research and Biological Systems. It is essentially academic staff that operates the CAMERA project. Decisions about how to assign resources and how to set priorities are made by the staff with the aid of either external advisory bodies or foundation-commissioned advisory bodies. We have a science advisory body. We also spend a lot of time in the community gathering input. So a lot of our decisions are based on our sense of the voice of the community, and by listening through various systems and sessions we hear what the greatest needs are from this community, for example, for the next analysis tool we should be delivering.

PARTICIPANT: Does your staff teach as well?

DR. GILNA: Some of them do; some of them are professional. A lot of the staff are professional programmers hired through the staff-level system at the university and dedicated to the project; some are folks in the more traditional academic side of the University. However, I would say largely the project is populated by professional staff dedicated to the project itself.

To run a project of this size requires more than the funding from the Moore Foundation, which itself is not a small amount of funding. So the project lives on the back of several large projects and institutions that also are involved in biology and computer science. There are other projects funded within the group by the National Institutes of Health, the National Science Foundation, and the Department of Energy. We draw from a pool of quasi-stabilized professionals and academics within the San Diego Supercomputer Center, the California Institute for Telecommunications Information Technology, and the Center for Research and Biological Systems. We use whoever is needed for the project goals to achieve what are determined to be the milestones to reach our outcomes on a quarterly or annual basis.

It is a very tightly managed project. Twice a year we have scientific advisory board meetings and we go over very carefully where we are with our deliverables. We have quarterly management meetings where representatives from our main funder, the Moore Foundation, sit with us for a day and go over those details. We are not talking about the budget at those meetings; we are talking about scientific details. Every year we generate a very carefully prepared strategic plan and tactical plan.

PARTICIPANT: Do you have plans to track forward uses of data that are released prior to a publication? That is, is it possible to look at subsequent editions to the research literature that cite deposits, and, if that is not being done within this project, do you think

you could plug into some other activity that would be doing that? I think the reasons you might want to do that are fairly obvious: to enhance the value of the data for subsequent researchers, as well as to give feedback credit and allow people to be able to annotate the deposited data.

DR. GILNA: There are two answers to that. In practice we do not do that, but that is not to say that we could not or that the capabilities do not exist for that. There is a tradition, if not an ethical expectation within the scientific community, that if you are going to use data, whether or not they have been published, you will cite them. So any dataset, whether it is in CAMERA or the National Center for Biotechnology Information, travels with a unique identifier—an accession number, for example, or something else. The number is expected to be used as a citable entity in the work that is being reported. It should be searchable and indexable and therefore would allow us, or anyone for that matter, to track the general trends and usage of the data.

18. Proposal for a Microbial Semi-Commons: Perspectives from the International Cooperative Biodiversity Groups – Flora Katz[44]

Fogarty International Center, National Institutes of Health

The Fogarty International Center is the only component of the National Institutes of Health (NIH) that is specifically mandated by Congress to work internationally. We are now in our 40th year. We work in about 100 countries, and in the past 15 years we have mainly focused on low- and middle-income economies, including the so-called developing countries. In my presentation I will discuss the issue of increasing access to microbial samples from these countries from the perspective of the International Cooperative Biodiversity Groups, a program we have been running for the past 16 years.

First, let me address the question: Is access to microbes from developing countries important? We actually do not know. The presumption is yes, but more than 99 percent of global microbiodiversity is still unknown. We do know that microorganisms are not uniformly distributed on the earth—one of our groups did a study comparing New Jersey and Kyrgyzstan, and there was little overlap—and we also know that developing countries are the most biodiverse on a macroscopic level. Furthermore, many of the microorganisms track with those macroscopic organisms, like endophytic fungi. So it is likely to be the case that developing countries will be an important source of microorganisms. We have already discovered a number of important new chemicals from microorganisms in these countries, and we have discovered new species in the genera of actinomycetes, which has been sort of the mother lode for drug discovery. We also know that biodiversity is threatened and that culture collections in these countries are not secure, which is another reason to pay attention to the contributions from these countries.

Many developing countries are working on bioenergy projects that rely on microorganisms. Examples include a new species of alga in Thailand, a Patagonian tree fungus that seems to expel hydrogen gas, a biomass-degrading fungus from the Solomon Islands that was noticed because it ate through the canvas and other materials used by the U.S. Army when it was stationed there. That fungus was recently sequenced by the Department of Energy and found to have some very interesting genes.

These countries desperately need new sources of fuel, so they are investing in biofuels. Unfortunately, they are getting biofuels by cutting down their rainforests, destroying their wetlands, and planting, for example, the oil palms that are now seen all over Indonesia and Borneo. This is not a sustainable practice. The countries thus are very interested in finding alternative ways of producing biofuels, and micro-organisms may be an important resource in this regard.

With respect to the semi-commons idea that we are discussing at this meeting, there are two relevant models at NIH that you might want to consider. One of them is from the National Cancer Institute (NCI). For 20 years NCI has supported contracts to collect natural materials (plants and marine organisms) from countries all over the world. It uses a simple letter of collection. The collections are targeted specifically for discovering agents active against cancer and HIV, so they are very focused. The materials are completely managed by the U.S. government. NCI uses a standard memorandum of

[44] Presentation slides available at: *http://sites.nationalacademies.org/xpedio/idcplg?IdcService=GET_FILE&dDocName=PGA_053724&RevisionSelectionMethod=Latest*.

understanding and a material transfer agreement (MTA) with a commercialization trigger—that is, the benefits must be renegotiated if something becomes a lead compound and moves on to commercialization. Because of its focused nature and the fact that the materials are completely controlled by the government, I do not think this is the most adaptable model.

The second model is represented by the International Cooperative Biodiversity Groups (ICBGs). These are investigator-initiated grants for biodiversity collections and biodiscovery research. The materials are managed by the grantees. However, each Group has very high transaction costs in the form of unique memoranda of understanding and material transfer agreements.

The ICBG program, which began in 1993, has a philosophy that is very similar to the Convention on Biological Diversity (CBD), although those two efforts were parallel and independent of each other. We started with three interdependent observations. First, we knew at the beginning of the ICBG program that nature is a rich source for new drugs. For example, about half of the FDA-approved drugs currently on the market are based directly or indirectly on natural products. Second, discovering natural products requires accessing biodiversity, but biodiversity is threatened globally. Finally, we felt that countries should own their own biodiversity and that they should receive some benefit from its use, which could in turn serve as incentives for the further preservation of biodiversity. Biodiversity might then become a sustainable source for future products from biodiscovery.

The novelty of this program as it was originally conceived was that we would ask groups doing research in this field to address all of the goals in one integrated program; that is, biodiscovery, biodiversity conservation, and the development of models to provide appropriate benefits for access and use of biodiversity. It was thus a highly ambitious project.

NIH's charge was simply not broad enough to do this. We had a congressional mandate to do health discovery, but not to do biodiversity conservation or economic development. So we formed a funding consortium with three partners. Drug discovery was represented by NIH and included approximately nine Institutes and Centers at NIH with an interest in a broad array of therapeutic areas.

The U.S. Agency for International Development (USAID), which had experience with development conservation projects, represented expertise and authority for providing economic benefits. Finally, the National Science Foundation had a mandate for biodiversity conservation and bioinventory.

Over time, this funding consortium expanded further. There is a very high cost associated with sending a team to a remote rainforest. Furthermore, all natural materials potentially have multiple applications. When we take a leaf off a tree, we can test it for drug discovery, for use in bioenergy solutions, for agrochemical and animal health technologies, and for a variety of other purposes. So, the program has evolved over time to become even more complicated for the sponsors.

USAID eventually dropped out because its funding schedule did not match everybody else's. Since then, however, we have brought in the U.S. Department of Agriculture to support discovery of agrochemicals. The Department of Energy joined us last year. We work with Dan Drell, who is at this meeting, on bioenergy solutions. Our most recent partner is the National Oceanographic and Atmospheric Administration, which is interested in products from the seas and healthy oceans. The rationale for bringing the different partners together is severalfold. Not only is there some economy of

scale, so it is more cost effective, but it also mitigates risk both for the sponsors and for the grantees because, if you are looking in multiple areas, it increases the probability you will find something. We have also found that it greatly increases the impact of the programs. It protects the resources for everybody in all of these areas.

To make things even more complicated, however, in order to address these very diverse goals, the teams are very multidisciplinary and multisectoral. There is always a partnership between academic institutions in the United States and academic or research institutions in the developing country. Most ICBGs are public–private partnerships, so there is usually a pharmaceutical, agrochemical, or biotechnological company involved as well. There are usually government entities from the foreign country involved, and there are often nongovernmental organizations and local communities too. If you wish to access something from a coral reef, for example, and a village owns that coral reef, you are going to have to negotiate with that village. Consequently, there are very diverse sets of stakeholders with different cultures and goals for the project all working together in one consortium.

We have now worked in 18 countries, which are listed in Table 18–1.

Plant-based Collections:	**Microbial Collections:**
Chile	Costa Rica*
Argentina	Panama
Mexico	Fiji*
Vietnam	Uzbekistan
Laos	Tajikistan
Nigeria	Kyrgyzstan
Cameroon	Madagascar
Peru	Philippines*
	Indonesia*
	Papua New Guinea
* Bioenergy collections	

TABLE 18–1 Countries Involved in International Cooperative Biodiversity Groups

When we started in 1993, the initial projects focused on plants, particularly tropical plants. In the third round of five-year awards, which was begun six years ago, we encouraged participants to obtain microbial and marine collections, and now all of our Groups are doing that.

As you can see from the list in Table 18–1, the countries involved vary widely in terms of the size of their economies and their scientific sophistication. Because we just began bioenergy collections in 2008, we have less data on these projects, but they look very promising.

There is a variety of transaction costs involved in getting this program to work and in successfully using and benefiting from these microbial collections. We provide

guidance, but we do not provide any standard format for access and benefit sharing agreements or for obtaining prior informed consent. One of the most important things to make clear in negotiating agreements or in the process of obtaining prior informed consent is that royalties are not going to be the major benefit of programs of this sort. This was a mistake made in the early days of the CBD. The countries thought they were going to make billions of dollars from new blockbuster drugs, so they decided to protect their biodiversity based on that expected return, and when that did not materialize there was a huge backlash and disillusionment. So now, as part of the prior informed consent process, we encourage grantees to talk about not only the intended uses for the materials but also the probability of various outcomes. That transparency has made this whole thing work.

In addition to access and benefit-sharing agreements, there are permits, MTAs, and other government documents, depending on the particular country we are working in. The laws vary from country to country, and sometimes it is very difficult to discover what those laws are. In some cases, a country may have no laws at all governing the use of these bioresources. Each country should have a person who is the main CBD contact, but often that person is not very informed about what the country's laws are, so it may be necessary to do a great deal of investigation to discover how to be compliant with the laws.

Shifting regulatory landscapes add to the complexity of the situation. In the middle of a project—as happened to us recently—a president can suddenly decide to change the existing law or add new laws, and our project then grinds to a halt because we no longer can export organisms and we may have to rework everything to comply with the new law. The scientists we work with serve as advisors to these governments, however, and when they say we cannot go forward with a particular project under a new law, there is some push and pull in that. In this case, the ICBG project itself was the reason that the new law was eventually reversed.

The political landscape may also shift. For example, there have been two separate coups in Madagascar during our project there. In Fiji, the president was about to issue a major policy decree based on the work of the ICBG that would have been a major victory for biodiversity conservation when the government was overthrown. You have to deal with these sorts of events.

The easiest way to negotiate these agreements is to do them as academic research agreements and not worry initially about the possibility of commercial products being generated from them, because that is a low probability. The agreement should, however, include a commercialization trigger—a statement that if some discovery does move towards a commercialization pathway, that will trigger additional good faith discussions on how to pay for the use of that discovery.

On average, it has taken the ICBGs one to two years to negotiate these agreements, which must be finalized before any collections can leave the country. Generally, some work can be done during that period, but that is a long time to wait before being able to take the materials to outside labs. We did speed up the process this past round by announcing that anyone who did not have an agreement within a year would lose the funding. Everybody got their agreements within a year.

The ICBGs have set precedents for access and benefit-sharing, and they have been used as case studies for the CBD. Perhaps the most important thing is that they have allowed on-the-ground experiments. It is easy to sit in a room with a number of lawyers and try to discuss what should be a benefit. What seems to be the most effective

approach, however, is for people to go out and negotiate these agreements, institute them, and then see the reaction of the stakeholders and get that feedback. It serves as an experiment in benefit-sharing.

We have also contributed to national policy in a number of countries—indeed, most of the countries in which we have worked—either directly or indirectly. Often the scientists in these countries who are working with us are the ones writing the laws.

There are two basic models for how access to microbial resources is provided. The first is that the microbes cannot leave the country at all. This is the case in Madagascar, for example. By law, anything that is self-replicating cannot leave the country, so it is necessary to import the technology into the country, which forces us to engage in some sort of technology transfer, a significant benefit for the country. In some cases, the samples are allowed to leave the provider country if they are accompanied by a scientist from the country of origin. Indonesia, for example, has come up with an agreement recently that is intended to ensure technology transfer, but there is some wiggle room allowing samples to be removed in cases when it is not feasible to do the analysis in the country—if they are accompanied by an Indonesian scientist.

The second model allows isolated and identified microbial cultures to leave the country only under the terms negotiated, for the purposes described, and to the parties designated. There is no third party access or release of information without prior agreement. The chain of custody must be documented and usually there is a time limit after which the samples have to be destroyed or returned unless the terms are renegotiated.

All of the access and benefit-sharing agreements negotiated by the ICBGs have two types of benefits. The more important of the two are the low-risk, near-term benefits. There is a low probability that a blockbuster drug will come out of the work done under any given agreement, so there have to be some immediate and concrete benefits for the countries. The most important benefit is the building of research capacity in the countries. This is what the countries want and what they need in order to exploit their own resources. We provide that in the form of training, technology transfer, and some infrastructure, particularly equipment.

The countries also have the benefit of participating in a research collaboration, which seems to attract other research collaborations, so there is a leveraging effect. Furthermore, when we work with local communities, there are local economic benefits. These take a variety of forms, from creating jobs to helping the development of micro-enterprises.

The high-risk, long-term benefits relate to the commercialization of a product developed from something discovered under one of these agreements. If that happens, the benefits may include such things as milestone payments or royalties. Many of these funds are contributed to a trust fund dedicated to conserving biodiversity for the country overall. Often the pharmaceutical and biotech companies participating in the project contribute to that trust fund, so the country is getting something in the bank. In addition to financial payments and the protection of biodiversity, the other major long-term benefit that arises from these agreements is the development of products that increase the public health or provide benefits to society.

With regard to the participation of these countries in microbial semi-commons, several issues should be considered. The first question that arises is: Whose benefit? It is often presented as a global good for everyone to have access to microbial cultures—and it is—but it is a very unequal playing field. Most of these countries do not have the capacity

to make use of their own resources. What they really need is the scientific capacity to explore those biological resources so that they are better able to take advantage of them.

It is a problem if these countries see themselves mainly as providers and not as users of the commons. If that is the case, then the open-access premise of mutual benefit does not apply—these countries cannot benefit as much from their own resources or the resources of the commons as other countries. In this case, perhaps the act of provision should be balanced by a nonequivalent benefit.

I would suggest that these countries should receive a combination of both near-term and long-term benefits. The long-term benefits would be commercialization-triggered benefits; if something goes on to commercialization that is derived from their collections, there will be further discussion about appropriate compensation. The immediate benefit could involve helping these countries build their scientific capabilities. For example, there is a demonstration project underway, called the Global Biological Research Center Network, which involves 15 countries, a number of them developing countries.

The countries with established biological resource centers (BRCs) are paired with developing countries that would like to develop BRCs to help them build capacity for their culture collections. This is something that I think the countries would really welcome. I have looked at their culture collections. It gives me nightmares. We spend a great deal of money helping them collect the samples, and they are put in refrigerators without backup generators, without any backups anywhere. The cultures are not being maintained adequately, but these countries would like to maintain the cultures and bring them up to recognized international standards.

Indeed, if they are not brought up to a minimum standard, the samples are not going to be of much use anyway, since they will not be at the appropriate quality level. If you want these countries to participate, therefore, it will be necessary to help them build the capacity so that they are able to participate in a meaningful way and they would see that as a very significant benefit.

A second set of issues is: Who owns biodiversity? Who has the authority to provide cultures? It is not going to be the individual researcher in Indonesia. Permission will have to go through the government, because the government owns those resources and that is how they view it.

Similarly, who is going to receive the benefits? One could argue that the benefits should flow to everyone from the villager who let you work in his coral reef all the way up to the people who collected the organisms, isolated them, did the chemistry, and so on. In the near term, monetary benefits from use of microbial cultures could go into building a national resource center, as in the ICBG Trust Fund model. For the long term, it will be more problematic to identify the appropriate stakeholders and each country will have to answer these questions for themselves.

Intellectual property rights also raise a number of questions. For example, in the ICBGs, we have very rich associated datasets for the microbial collections. For each micro-organism, we can tell you the associated ecology, what else was collected with it, its partial or complete DNA sequence, and so on. It has gone through a lot of bioassays. We know the chemical composition. The countries may be able to offer a very comprehensive dataset for some of their collection. How does that translate into valuing the materials? Is one standard benefit enough?

Finally, to use some benefit system as an incentive to develop and as a means to support a BRC, it is necessary to track and acknowledge the origin of materials. I believe

that documented chains of custody will be critical. For these countries to take part, they will need to feel secure. A shared microbial commons also represents a community of trust. However, this community of trust is extremely tenuous. It has been violated many times. There have been some major bio-piracy cases, and these transgressions are happening all the time on a smaller scale, so it will be necessary to build this community of trust slowly and incrementally. Before these countries can participate, they are going to have to develop the quality controls and quality of materials that the community expects, and in the process of doing that, they will begin to become engaged. That, at least, is what we have seen so far. Finally, it is necessary for the global community to weigh these transaction costs against the need to protect the resource because these transaction costs may be the incentive for conserving the resource.

I will close with a success story from our Panama ICBG, which has been ongoing for over 10 years. The collaborators in Panama started with very little scientific infrastructure for natural products discovery, but over the past decade we have supported the development of a first-class parasite drug discovery lab and a first-class chemistry facility there, and we have provided funding for the first nuclear magnetic resonance machine in the country. They have sent about 30 students on to get higher degrees as a result of this program.

The Panama ICBG research consortium had to pick a conservation project and some kind of benefits scheme because we require those things as part of the award. So during the last round, which started five years ago, they decided to focus on a fairly large island called Coiba Island, which is along the country's Pacific coast. Because Coiba had housed a penal facility for decades, it had never been developed, so it had beautiful primary forest and was surrounded by almost untouched coral reefs and mangrove swamps. It is a gorgeous area. As soon as the penal colony closed, the bidding started. A number of developers wanted to move in immediately and start building condos and vacation homes.

The ICBG group decided that its conservation project would be to protect the island and get it UNESCO World Heritage status so that it could not be violated. This was really ambitious. It was just a small group of researchers.. Nonetheless, this is what they set out to do. They did all of their collections on Coiba Island and the surrounding reefs. They trained their students there. When they did an inventory of the entire island, they found many endemic species that were new to science. They also discovered many very promising molecules.

They assembled all their data and took it to the Panamanian Parliament. They said, "You cannot develop this island. This is an incredible resource. Look at all of these new species we found. They are nowhere else in the world. Look at all these incredible potential drugs we found. This is going to be a gold mine for this country because once you get World Heritage status, the tourists will start coming. You can start building a real industry. It will be very prestigious and will boost the economy of the entire country."

And the Parliament agreed. First they set up a national park there with all the appropriate controls and legal enforcement. Then they went to UNESCO to ask that it be made a World Heritage site. UNESCO told them it was not big enough, even though it was already huge. So they went back to Parliament, and Parliament doubled its size to set aside a very large marine area surrounding Coiba Island. They went back to UNESCO, and two years ago UNESCO designated it as a World Heritage site. The outcome of this innovative, integrated approach was something that was much larger than the sum of its parts.

19. The International Treaty on Plant Genetic Resources – Shakeel Bhatti[45]

Food and Agriculture Organization

I will be offering the perspective of intergovernmental organizations and processes, with a focus on one intergovernmental process in particular, the International Treaty on Plant Genetic Resources for Food and Agriculture (ITPGRFA). After a short introduction about the international legal architecture that governs genetic resources and, in particular, plant genetic resources, I will describe what has already been achieved in the context of the international treaty, which is the creation of a multilaterally governed gene pool of more than 1 million accessions of plant genetic material and which constitutes a commons for biological materials. Next, I will discuss the steps that the governing body of the treaty decided on in June of this year, which is developing policy for the intangible elements of this global gene pool. Finally, I will end with a few reflections on operationalizing such open knowledge environments as they are foreseen in this treaty.

In the ITPGRFA you already find something resembling an integrated commons linking both material and digital data and databases within a distributed and open access framework. Furthermore, it has a larger framework in its federated network of gene banks and of information portals that is now being established, which is being administered by a governing body made up of the 121 contracting parties of the treaty.

With that in mind, let me begin by locating the treaty within the relevant international legal instruments and institutional frameworks. First, is the Convention on Biological Diversity (CBD). This international treaty creates a multilateral system of access and benefit-sharing and implements those access and benefit-sharing principles on a multilateral basis. As we heard, it takes about one or two years to negotiate a well-done bilateral access agreement, and they must be negotiated on a case-by-case basis.

By contrast, under the ITPGRFA we have more than 600 transfers every day of plant genetic material related to agriculture. It would be impossible to negotiate access and benefit-sharing on a bilateral case-by-case basis for all of these transfers. Instead, the multilateral system laid out by the international treaty provides a low-transaction-cost, pooled commons of genetic material. Other components of the international governance architecture include the 1961 International Union for the Protection of New Varieties of Plants (UPOV); the 1994 Agreement on Trade-Related Aspects of Intellectual Property Rights (TRIPS) and the various intellectual property agreements that are incorporated by reference under the TRIPS standards, such as the Paris Convention; and the 1985 Budapest Treaty on Deposit of Microorganisms for Purposes of Patent Procedure. The Consultative Group on International Agricultural Research (CGIAR) is not a legal instrument, but rather the largest network of agricultural research centers and gene banks of plant genetic resources in agriculture, and it plays an important role in the treaty.

The timeline leading up to the ITPGRFA can be traced back to the CBD, which was adopted in 1992. In 1994, a request was transmitted to the Food and Agricultural Organization (FAO) of the United Nations to revise a preexisting soft law instrument into

[45] Presentation slides available at *http://sites.nationalacademies.org/xpedio/idcplg?IdcService=GET_FILE&dDocName=PGA_053725&RevisionSelectionMethod=Latest*.

a binding framework that would be in harmony with the access and benefit-sharing principles of the CBD. The negotiations lasted for seven and a half years, and finally, in 2001, the treaty was adopted by the FAO Conference, coming into force in 2004. The first governing body session under the treaty was held two years later, and within three years following that session, 40 instruments of ratification were deposited.

Then, in 2006, two years after the ITPGFRA came into effect, the first session of the governing body adopted the Standard Material Transfer Agreement (SMTA). It was the fastest rate of ratification of an FAO-administered treaty in the history of FAO. The SMTA is the standard contract for transferring genetic material within the multilateral system. The third session of the governing body was held just a few months ago (in June 2009) in Tunisia, and the next session will be held in March 2011 in Bali.

Figure 19–1 offers a very simplified diagram illustrating how the multilateral system of the treaty works.

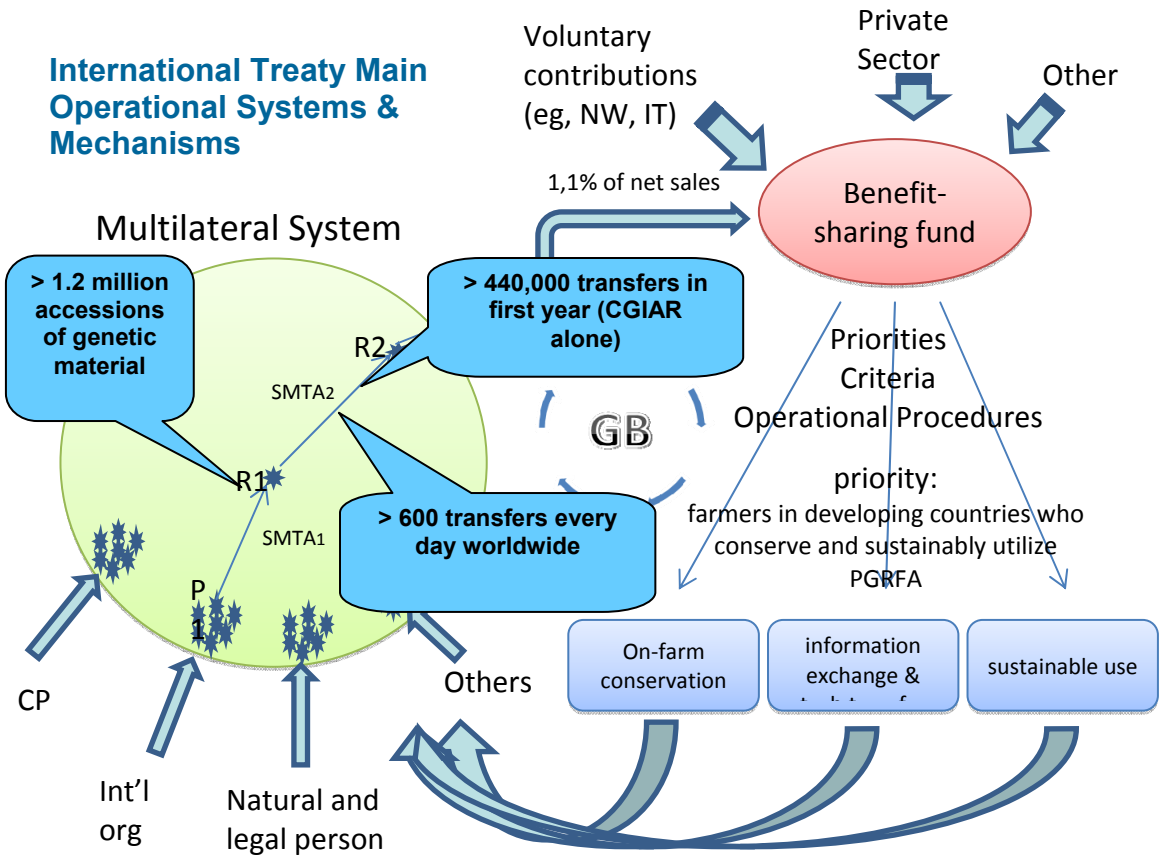

FIGURE 19–1 The main operational systems and mechanisms of the International Treaty on Plant Genetic Resources for Food and Agriculture.
SOURCE: United Nations Food and Agriculture Organization, *http://www.planttreaty.org/*

In essence, the treaty creates a multilaterally governed gene pool. Genetic material is put into that gene pool by various actors. These include the nations that have ratified the treaty and also international organizations, such as the International Atomic

Energy Agency, which has included its mutant germplasm repository in this gene pool and is now also being governed by the terms of the treaty. Individuals and organizations also contribute genetic material, including private sector entities and indigenous and local communities, such as the Quechua communities from Peru, who deposited their own germplasm into this gene pool. There is a broadening range of stakeholders who are including material. A major contributor is the group of international agricultural research centers of the CGIAR, which have so far deposited the bulk of material that we know of and about which we have solid documentation.

Once material is included in the gene pool, its use is governed by a chain of SMTAs. It begins with the provider of the material transferring it under a standard material transfer agreement, call it SMTA1, to the recipient. This recipient can now become a provider and transfer the material under a second agreement, SMTA2, to a second recipient. By receiving the material under the SMTA, the recipient takes on an obligation to transfer this material to other parties only under the same terms and conditions specified in the original SMTA. This builds a contractual chain which eventually leads—after a long series of transfers that, in plant breeding, often takes 5 to 10 years—to the development of a commercial product.

If that product is not available without restrictions for further research, training, and breeding—for example, if the product is under a patent claim—then the SMTA specifies that the recipient shall pay 1.1 percent minus 30 percent of net sales of that product to the beneficiary fund of the ITPGRFA. This is a multilaterally created, governed, and administered trust fund that receives proceeds from products that incorporate material received from the gene pool. This beneficiary fund also receives funding from a series of other channels, such as voluntary contributions from contracting parties. A number of governments, including Norway, Spain, Italy, and Switzerland, have given voluntary contributions. At the opening of the Svalbard Global Seed Vault in Norway, Prime Minister Stoltenberg along the Minister of Agriculture at that time announced that Norway will each year contribute 0.1 percent of all national seed sales to the beneficiary fund of the treaty. Contributions also come from the private sector, including philanthropic grant-making institutions, foundations, and individual philanthropists.

The accumulated funds are dispersed according to multilaterally agreed-upon funding priorities, selection criteria, and operational procedures. For example, the treaty specifies a priority for funding to farmers in developing countries who conserve and sustainably use plant genetic diversity. At the second session of the governing body, three funding priorities were set by the governing body: on-farm conservation of plant genetic diversity, information exchange and technology transfer, and sustainable use of plant genetic resources, including through characterization, research, and participatory plant breeding. These funding priorities are intended to further the conservation and maintenance of genetic diversity, which in turn feeds the global gene pool that is established by the treaty. In short, the treaty is intended to create a virtuous circle, overseen by the governing body of the treaty. That governing body now includes 121 governments, with additional ratifications being underway at present.

The system started up in 2007, and since that time there have been more than 1.2 million accessions of plant genetic material that we know of. That last qualifier is important because in reality there have certainly been far more accessions, but we do not know of all the material that is included in the system. Getting complete and reliable datasets specifying which material is included in the system is itself a major undertaking

that is now under way. At present, we have data—again, incomplete data—that indicate that in the last 12 months, there were about 440,000 transfers by the CGIAR alone. This does not include transfers of material from regional and national gene banks. There is a total of more than 600 transfers each day worldwide.

On the benefit-sharing side, in June of this year the benefit-sharing fund disbursed half a million dollars in grants for the conservation of crop genetic diversity and the sustainable use of genetic resources. The governing body has adopted an objective of raising $160 million over the next 5 years for the benefit-sharing fund, with a planning target of $50 million. It also adopted a strategic plan, which lays out the mobilization of these resources over the next five years.

Operationally, we have almost completed the first phase of treaty implementation, which consists of putting the multilateral system into operation, developing the Standard Material Transfer Agreement, and starting up the commercial benefit-sharing activity through the trust fund.

To provide a sense of how the commercial benefit-sharing is going, let me offer a few details. The first call for proposals for benefit-sharing projects under the trust fund of the treaty was solicited for one month starting in early December 2008. Within that period, more than 450 applications were received. Keep in mind that this included the Christmas–New Year period. The submissions were from countries around the world.

The treaty specifies three noncommercial benefit-sharing mechanisms: exchange of information, technology transfer, and capacity building. Of these, the first two are particularly relevant.

Concerning the exchange of information, the treaty states that access to plant genetic resources that are protected by intellectual and other property rights shall be consistent with the relevant international agreements and minimum standards for the availability, exercise, and enforcement of intellectual property rights, and shall also be consistent with relevant national laws. The treaty also states, however, that recipients shall not claim any intellectual property rights or other rights that limit free access to plant genetic resources for food and agriculture or to their genetic parts or components in the form received from the multilateral system.

Thus, the treaty is consistent with international intellectual property (IP) standards, but it sets out a particular model for acquisition of IP. Those terms are also reflected, of course, in the Standard Material Transfer Agreement and are passed on contractually to each recipient of genetic material from the gene pool. Then, if a recipient commercializes a product that is itself a plant genetic resource and that incorporates material from the multilateral system, and where such product is not available without restriction to others for further research, training, and breeding, the treaty specifies that the recipient shall pay 1.1 percent minus 30 percent of the sales of the commercialized product into the trust fund established by the governing body for the purpose of benefit-sharing.

There is, consequently, a patent-based, benefit-sharing trigger here. The treaty also calls for the governing body to review the operation of this entire system and the terms, for both access and benefit-sharing, five years after coming into force. It further states that the governing body may review the levels of payment with a view to achieving fair and equitable sharing of benefits and may also assess whether the mandatory payment requirement in the material transfer agreement shall apply also in cases where such commercialized products are available without restriction to others for further

research and breeding. This exercise was originally scheduled to be done this year, but it was postponed to the next session of the governing body.

Research results are also governed by the SMTA, which requires that the recipient of material shall make available to the multilateral system—through a global information system that will be created under the treaty—all non-confidential information that results from research and development carried out on the material. The recipient is also encouraged to share, through the multilateral system, any non-monetary benefits that result from research and development. Finally, the SMTA provides that after the expiration or abandonment of intellectual property rights, the product that incorporates the material should be placed back into one of the collections that are part of the multilateral system.

The next steps in the implementation of the treaty will include the implementation of Article 17, which concerns a global information system for plant genetic resources. This Article foresees that the contracting parties will develop a global information system to facilitate the exchange of information on scientific, technical, and environmental matters related to plant genetic resources, making information on plant genetic resources for food and agriculture available to all contracting parties. Thus, in addition to the material, which is pooled and governed by the terms of the multilateral system, the system will also include an information component to facilitate the exchange of all the data and information that results from the use and the exchange of material within the gene pool.

In June of this year, the governing body requested the secretariat to develop a vision paper that would take stock of existing information systems and outline a process for developing this global information system. So the next phase is moving from the tangible part of the system—the material—to the intangible, information-oriented side. A number of provisions in the treaty will then be operationalized through this process, including the two noncommercial benefit-sharing mechanisms, exchange of information and technology transfer.

Concerning exchange of information, the treaty states that parties agree to make available information that encompasses catalogues and inventories, as well as information on technologies and all results of technical, scientific, and socioeconomic research, including the characterization, evaluation, and utilization of the plant genetic resources within the multilateral system. All this information will be included in this information system.

With regard to technology transfer, the treaty specifies that parties facilitate access to technologies for the use of plant genetic resources, recognizing that some technologies can be transferred only in the form of genetic material itself. Again, the applicable intellectual property standards must be respected.

The treaty then sets out a set of very specific measures for tech transfer implementation, such as the establishment of crop-based thematic groups; the use of plant genetic resources for food and agriculture; and various types of partnerships in research, development, and commercial joint ventures. The transfer of technology must be consistent with IP standards.

Figure 19–2 shows the ratifications and other measures of progress on the international treaty over the past decade. The green line shows that the number of SMTAs has really skyrocketed. Requests for capacity building have also increased sharply. Given that we are working with 121 jurisdictions, which have different legal frameworks, different languages, different institutional settings, different ministries

responsible, different gene bank exchange practices, and different institutional actors, the implementation of a global multilateral system that functions coherently is quite a formidable task.

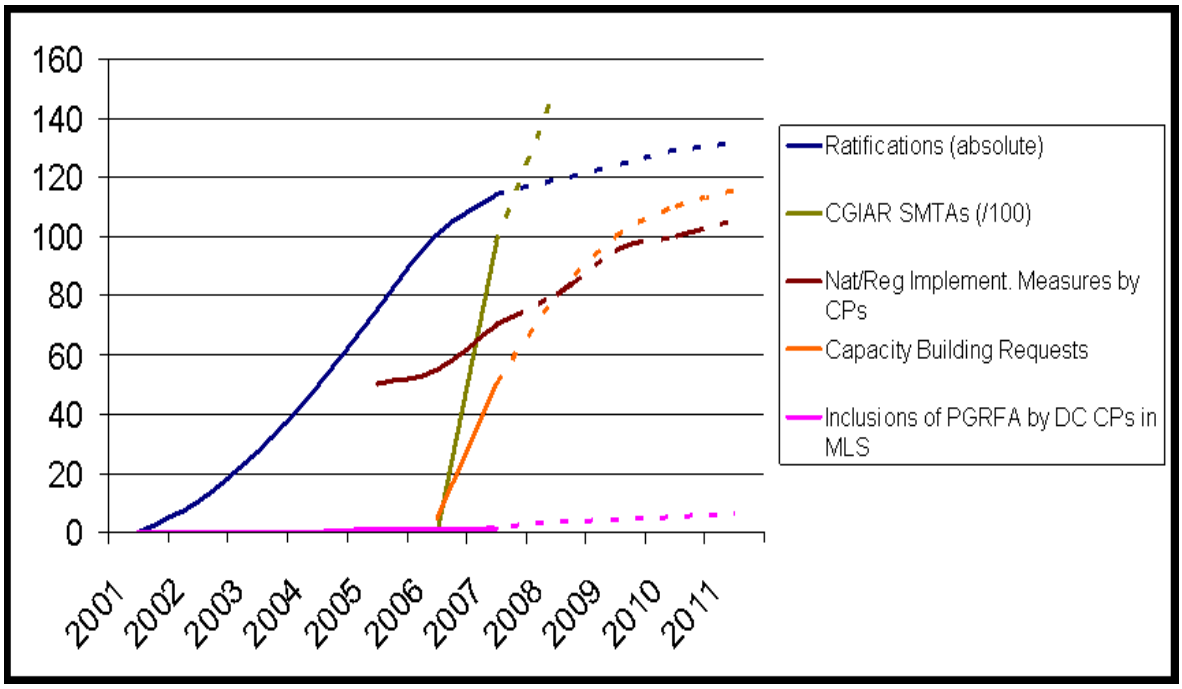

FIGURE 19–2 Progress on the international treaty.
SOURCE: United Nations Food and Agriculture Organization, *http://www.planttreaty.org/*

Seen from the point of view of open knowledge environments, the treaty has essentially established a global materials commons that is multilaterally managed. This differentiates it from the bilateral access and benefit sharing regime under consideration in the context of the CBD. The materials commons is now fully operational through the globally applied Standard Material Transfer Agreement, and we have recently adopted a dispute resolution procedure for any disagreements that might arise from these SMTAs. There is a four-stage escalating alternative dispute resolution procedure that begins with amicable negotiations and progresses through mediation, arbitration, and, finally, binding arbitration under the International Chamber of Commerce.

By beginning of next year we will also have a global information infrastructure to support the operation of this system consisting of a data store and a PID server, which serves unique permanent identifiers (PIDs) for users of the multilateral system. The treaty then calls for a global information system, a step on the path to an information commons based on noncommercial benefit sharing in which the exchange of information itself is considered to be a noncommercial benefit of the gene pool.

In moving towards that information commons, the treaty now has sustained operational funding by governments plus an existing secretariat based in Rome and housed in the FAO. The treaty's governing body has become one of the main convening forums of the plant genetic resource community. Furthermore, the treaty is increasingly

being taken as a model by such institutions as the World Health Organization. We are working closely with them to develop virus-sharing and benefit-sharing MTAs that will be based on those provided for in the treaty.

The treaty still faces various challenges. Further operationalizing and stabilizing the multilateral system is one. While the system has been successfully started, it is not yet at the stage where it should be. Significant challenges and problems remain to be addressed. For example, we need to focus on the project cycle for the benefit-sharing fund and orient this towards facilitating scientific research on the plant genetic material in the gene pool.

The treaty has a twofold nature, being both an intergovernmental political process and an operational system. Maintaining policy and operational coherence within a daily operation across so many countries and jurisdictions is a major task and requires a very significant investment of resources.

We are running a capacity building program with Bioversity International, and we are providing a great deal of assistance to countries working on their national implementing legislation. It is also important to facilitate interactions between the contracting parties who govern the system and the users of the system, that is, the gene banks, the researchers, universities, companies, and so forth.

Finally, the treaty addresses access and benefit sharing for genetic resources specifically in the context of agriculture. It is important to have the specificity of agricultural genetic resources recognized in other policy forums, not only the CBD, but also the United Nations Framework Convention on Climate Change and the various food security efforts that are going on in the field.

20. Microbial Commons: Governing Complex Knowledge Assets – Minna Allarakhia[46]

University of Waterloo, Canada

In talking about the governance of the microbial commons, I will apply a strategic level of analysis and a knowledge perspective. I would like to leave you with three messages:

First, biological knowledge structures are evolving, not only in terms of complexity, but also in terms of their value for future discovery and commercialization. We need to understand what belongs in the commons from a dynamic perspective. What might not have belonged in the commons yesterday may belong in the commons today.

Second, we need to understand clearly the motivation of participants in any commons. It is not just a matter of public sector participants. When my colleagues and I from the University of Waterloo and Wilfrid Laurier University studied 39 open-source initiatives developed after the completion of the Human Genome Project, we found that in many cases private-sector participants were involved, and a few actually catalyzed the formation of those initiatives. We need to understand both the positive and negative consequences of this participation.

Third, we should document these lessons from the commons so that they can be transferred across disciplines and even across markets, as we see researchers from different areas seeking to enter this domain, looking to participate in the commons that are being proposed as well as their own internal commons.

Commencing with the knowledge perspective, the Human Genome Project advanced the view that biological information operates on multiple hierarchical levels and that information is processed in complex networks. It is no longer sufficient to look at just the genomic and proteomic levels. We need to understand the interactions among genes and proteins—how to modulate systems, minimize malfunctions, and optimize for positive functions. From a knowledge perspective, then, biological information has become complementary. Downstream product development relies strongly on upstream research inputs. Furthermore, there is high applicability across biological systems and for our purposes microbial systems.[47]

Complicating the matter from an intellectual property perspective is that patents can exist at any level of the hierarchy of the research process and, depending on the breadth of those patents they can greatly influence the incentive for follow-on researchers to examine or use elements of such systems. Too broad a patent can inhibit the incentive of users to look at the underlying knowledge assets. So we need to find solutions to manage those incentives, both for first innovators and for follow-on incremental innovators. The proposal of the commons and the liability regime from this symposium is one possible solution.

---

[46] Presentation slides available at *http://sites.nationalacademies.org/xpedio/idcplg?IdcService=GET_FILE&dDocName=PGA_053726&RevisionSelectionMethod=Latest.*

[47] See Allarakhia. M., Wensley, A. Systems biology: melting the boundaries in drug discovery research [Internet]. In: A Unifying Discipline for Melting the Boundaries Technology Management: Portland, OR, USA: [date unknown] p. 262-274.[cited 2011 Aug 15] Available from: *http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=1509700.*

Turning to motivation, we do need to understand the incentives for participating in the commons. As noted above, our research on 39 initiatives found strong involvement from private-sector participants. There are six classes of incentives:

1. The development of a collegial reputation as a reward for working in open science. This is no longer restricted to public-sector participants; private-sector participants want to signal their quality as allies, particularly for downstream product development. Beyond catalyzation of several initiatives in our study by private sector participants, the recent open donation of compounds and continued creation of open source discovery initiatives by several multi-national pharmaceutical organizations such as GSK, Pfizer, Eli Lilly, and Merck, provide evidence of this need to signal quality and openness to further collaboration for downstream development.
2. To generate general reciprocity obligations. I mentioned the complexities and the complementarity between knowledge assets. Both public and private sector participants may want to create reciprocal obligations signaling that they are willing to contribute to the commons so that in the future they can access other external knowledge assets. The creation of open patent pools with multiple contributors can signal this reciprocal obligation assuming equitable contributions and fair access terms.
3. To influence adoption of a technology or a technology standard through increased diffusion of knowledge. We saw this in our research with the microarray providers participating in the biotech commons in order to influence adoption of their technology as a standard. However, we must note that there may be positive or negative consequences when you influence the adoption of a premature or insufficient technological standard.
4. To improve the aggregate performance of an industry in order to increase safety or regulation associated with that industry as we discussed yesterday with reference to microbiological materials and outputs.
5. To preempt rivals. We clearly saw this after the mapping of the human genome, when 10 of the world's largest pharmaceutical companies came together and formed the Single Nucleotide Polymorphism (SNP) Consortium to ensure that rivals would not encroach on this territory and build patent fences around critical areas necessary for future product development.
6. To share the risk associated with knowledge production. It is important to examine not only the issue of shared implementation from open-source software development and fair access to technology or biotechnology development, but also the way in which a commons serves to enable collaborative knowledge production. For example, in the pharmaceutical industry, the complexities associated with drug discovery are very intense. The risks are frequently too high and many pipelines for new products are

currently empty. The commons can provide the incentive to share the risk for collaborative knowledge production where there are complexities associated with new knowledge and its association to products.

Furthermore, it is important to understand the incentives to participate as well as to be able to predict when participants may exit from any commons. Here the concept of transition point is of value. The transition point is defined as the point in discovery research when researchers come to believe that unilateral gains from private management of knowledge including appropriation activities are greater than shared gains from open or shared knowledge with the subsequent outcome exit from the commons. Therefore, from a strategic perspective, we need to look at when appropriation will take place—when materials will be removed and no longer deposited.

Because the value of knowledge is changing, we are uncertain at any point what value today's knowledge will have in terms of discovery and product development. A commons can reduce that uncertainty and make it less likely that premature mistakes will be made— specifically through the enclosure of knowledge, such as occurred with the gene races.

In our research, we sought to analyze 39 open-source initiatives developed after the mapping of the Human Genome Project. We looked at the structure and characteristics of the knowledge at stake. What was being produced? How could you classify or characterize that knowledge? What rules were established to govern both data and materials and for downstream appropriation strategies?

Overall, we have learned the following:

- Participation rules existed in the majority of cases. Consortia had established entry rules, with screening by executive or steering committees. Often commitment policies, membership fees, or large upfront research payments were established or required to enable both cooperation and research.
- Knowledge access policies exist. In the majority of cases, information is released not only to members, but also to the public at large. There were a few initiatives that were somewhat more closed, which shared information only with members that had made upfront monetary commitments.
- There were rules to manage both data and materials. The decision whether to deposit data into an open or a closed repository depended on the knowledge access policy. Where knowledge dissemination was open, peer-reviewed publications and deposit into databases permitted not only the release of data but encouraged the validation of the data or deposits. In the majority of cases, consortia advocated the use of nonexclusive royalty-free licenses for noncommercial use of materials and discovery tools.
- The consortia generally avoided issues related to commercial use and product development. It was assumed that this would be handled at the individual transaction level.
- It was absolutely important to create transparency with regard to motivation. In this case, upfront commitments could provide that clarity i.e. monetary support, human capital support, or open donations of knowledge-based assets.

> Motivation for participation given the objectives established by the initiative clearly had an impact upon knowledge dissemination and access as we discovered when comparing the open initiatives to the more closed initiatives in our sample.

How do we apply these observations to the microbial commons? We have limitless capabilities for applying microbial knowledge to the energy and environmental sectors, in the development of alternative biofuels, the generation of biodiesel, and bioremediation. We need to approach this knowledge from a whole systems perspective, however. Hence, we should look at communities of interacting microbes.

Research and applications require integration and analysis of data to discern patterns of communities of microbes. The continued sharing of microbial information will be critical, as will linking literature, databases, and user communities. This is important not only for collaborative discovery, but also for the validation of the data and the results. In addition, given the sophisticated nature of the visualization data that is emerging today, we need to enable the joint representation and standardization of the data. Some of the governance mechanisms we might need to consider are the timing of data deposits, access and use, exemption clauses for noncommercial use, transfer of management from depositors to collective organizations, and commercial use clauses.

Several open access journals, databases, and supporting tools were discussed in this symposium yesterday. I want to add that it is not just a matter of the data, but we also need to have a supporting tool infrastructure in order to create queries to gain benefits from that data and pursue further discovery. For example, the Global Biodiversity Information Facility (GBIF) is an information-based infrastructure for connecting users to a globally distributed network of databases. Here, we use the notion of linking knowledge assets from an information technology infrastructure perspective.

The incentive to share microbial data is also manifested in the private sector. The Helicos BioSciences Corporation has opened up its microbial datasets, as well as a query tool. What is their motivation? We discussed disincentives yesterday, but what would be the incentive in this case? Most likely it is to showcase their genetic analysis system—an attempt to encourage the adoption of their system by displaying the value and the sophisticated nature of their data.

In yesterday's talks, we discussed the linking of both materials and information in biological resource centers. The goal is to promote common access to biological resources and information services—we see that StrainInfo is providing electronic access to the information about biological materials in repositories.

BioBricks is quite interesting from a materials management perspective. It was developed as a nonprofit foundation by the Massachusetts Institute of Technology, Harvard, and the University of California, San Francisco. With BioBricks we are moving into what I consider the convergence paradigms, as now it is necessary to manage the complexities associated with synthetic biology. BioBricks makes DNA parts available to the public free of charge via MIT's registry of standard biological parts. This is a collection of approximately 3,200 genetic parts that can be mixed and matched to build synthetic biology devices and systems. The commercial or other uses of these parts are unencumbered—without the assertion of any property rights held by the contributor over users of the contributed materials. However, novel materials and applications produced using BioBricks contributed parts may be considered for protection via conventional property rights.

Beyond the issues of data and materials management, we must also deal with the changing value of knowledge, and we find the commons model being used to manage downstream assets. One example of the latter is the Eco-Patent Commons. This is a project by the World Business Council for Sustainable Development whose mission is to manage a collection of patents pledged for unencumbered uses—even for proprietary purposes in products, processes, and composition of matter—that are directed towards environmentally friendly applications. As of 2008, 100 eco-friendly patents had been pledged by private-sector participants, ranging from manufacturing processes to compounds useful in waste management. What is the motivation there? Perhaps the participants recognize that, in dealing with premature technologies, they need to assure that there will be interoperability between technologies that develop downstream.
The AlgOS Initiative is very premature and not particularly coherent yet, but I thought it was worthwhile to mention. It is an open-source initiative seeking ways to produce biodiesel from algae. The group is attempting to aggregate research inputs from a variety of experts in order to arrive at a full-cycle design for biodiesel production from algae, allowing for modification based on the open source software GNU General Public License approach, in which modifications are permitted as long as one complies with the requirements to pass on the source code to any recipients of one's modifications, to provide them the same freedom to modify it, and to provide notice of those terms of the license.

Finally, in parallel with these other efforts, stakeholders are discussing so-called "green" licensing. Such licensing is directed towards developing countries that are looking to develop green technologies. An international licensing mechanism is under discussion that would have developing countries pay a fee in order to access this technology, while at the same time protecting the innovating firms. It will be interesting to see what form they choose for their fee mechanism. The underlying goal is to allow countries at different stages of development to be able to access the same technology.

Clearly, as knowledge characteristics change, the governance structures may need to change with them. Early on, for example, some experts suggested that data would not be deposited in the commons. Today, there is a question whether tools—pharmaceutical tools or biological tools—should be placed into the commons. Table 20–1 summarizes various examples from the microbial commons, looking at their characteristics and at the various governance strategies that are currently being employed.

Managing the Microbial Commons	Data	Materials Management	Downstream Assets
Example:	MannDB; GBIF; Helicos Microbial Data	WFCC; BioBricks	EcoPatent Commons; AlgOS
Knowledge Characteristics:	High complementary, Non-substitutable, High applicability	High complementary, Non-substitutable or Substitutable, High applicability	High complementary, Non-substitutable or Substitutable, High applicability
Knowledge Governance Strategy:	Open Access; Use of Supporting Open Access Tools	Open Access; MTAs; License Agreements	Non-Assertion Clauses; Green Licensing; GNU-General Public Licenses

TABLE 20–1 Governing the Microbial Commons

To summarize the pragmatic outcomes, managing knowledge assets has become critical, not only in the systems paradigm, but in the convergence paradigm. Here we see biological sciences, chemical sciences, physical sciences and information sciences increasingly coming together to address the complexities of health product technology, nanotechnology, green technology and energy based technology development. We need to determine what really belongs in the commons and what governance strategies are most appropriate so that researchers in multiple markets can pursue product development opportunities. These goals imply certain policies. The need for greater transparency of motives during knowledge production suggests that there should be an establishment of and commitment to rules regarding knowledge production. Different conventions regarding knowledge dissemination and appropriation imply the need to establish early in a research project what should be disseminated and in what format. Here, our research provides some indication of the commonality of rules for both knowledge production and dissemination. Follow-up research also looks closely at the transition point and its application to varied knowledge assets.

The National Research Council report *A New Biology for the 21st Century* (2009) advocates large teams converging and varied disciplines working together, whether that is promoted through federal policy or other means. We need to keep in mind that as scientists, information technology professionals, and other experts work together, they each have differing conventions regarding knowledge dissemination and appropriation. Some of them may find value in pure disembodied knowledge. Others may find value and appropriate embodied knowledge with the final goal marketable product development. We need to bring together these disciplines under a common framework, which is what a structured commons can offer to them. Extending our analysis and understanding of knowledge based activities to the convergence paradigm should lend insight into how best to structure a commons with varied disciplinary participants. Consequently, it will be important to analyze new case studies involving open-source innovation that targets the energy and environment sectors in order to look at evolving models of innovation. What types of participants are in those sectors pursuing those models? How do they handle the increased complexities as knowledge assets converge and are increasingly linked together and as new rules and perceptions of value emerge?

Finally, it would be valuable to create a repository of governance strategies for knowledge assets as is currently being undertaken by the BioEndeavor Initiative (www.bioendeavor.net), including any licensing templates or tools, so that we can apply those across commons, across disciplines, and across markets. As new markets choose to develop their own commons, they can have the benefits of the lessons we have learned about managing knowledge based assets and the development of open access journals, open data networks, and the supporting IT infrastructure.

21. Digital Research: Microbial Genomics – Nikos Kyrpides[48]

Lawrence Berkeley National Laboratory

I will be talking about the future in microbial genomics: where I would like the future to be, where we are trying to go, and where I think we are going. As of September 2009, there were 4,319 microbial—in particular, archaeal and bacterial—genome projects under way around the world. As shown in Figure 21–1, five big sequencing centers are responsible for more than 50 percent of the world's production. This is very different from the situation just one year ago, when only two sequencing centers, the Joint Genome Institute (JGI) and the J. Craig Venter Institute (JCVI), were performing more than half of the world's sequencing. Just in the past year, those two have dropped to well under 50 percent of the world's production, and the total production has increased by quite a lot. It is possible that this is an indication that we are already seeing the so-called democratization of genome sequencing, as more and more sequencing is taking place in smaller and smaller places, and a number of universities are now buying sequencing machines and starting to produce data.

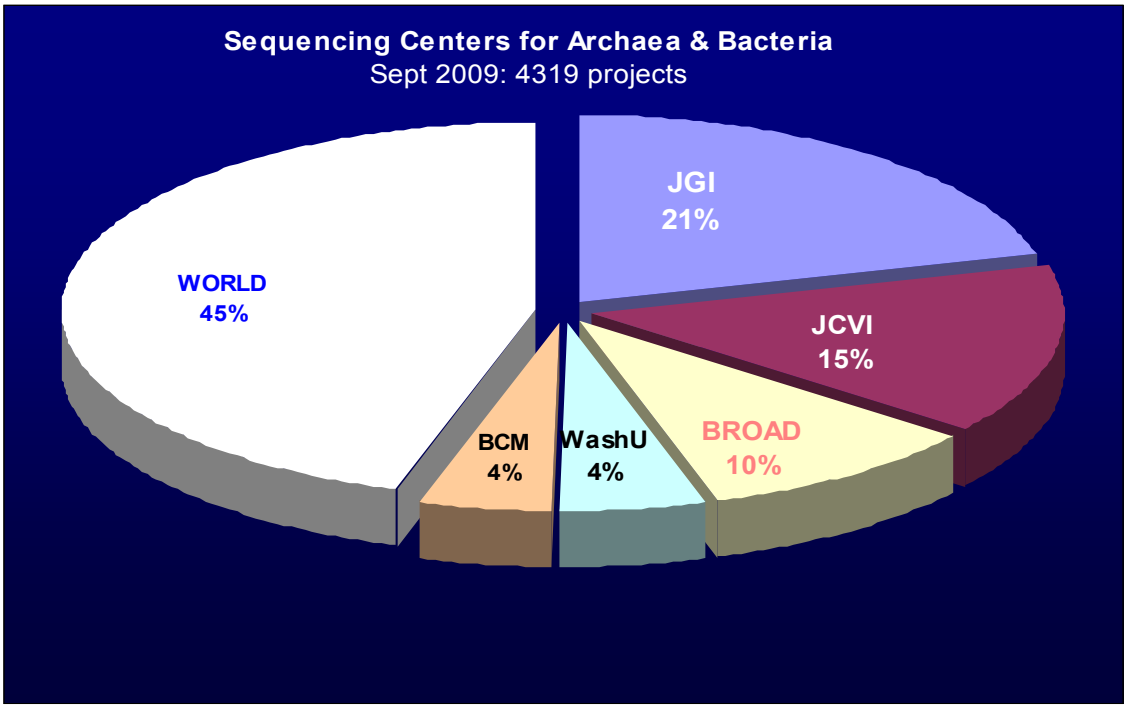

FIGURE 21–1 Sequencing centers for archaea and bacteria.
SOURCE: *http://www.genomesonline.org/*

Everybody is talking about the data deluge (Figure 21–2). However, the big question is whether we need more sequencing? My non-microbiologist friends tell me

[48] Presentation slides available at: *http://sites.nationalacademies.org/xpedio/idcplg?IdcService=GET_FILE&dDocName=PGA_053955&RevisionSelectionMethod=Latest.*

that there are now over 4,000 microbial genomes mapped. Is that enough? When are we going to be happy?

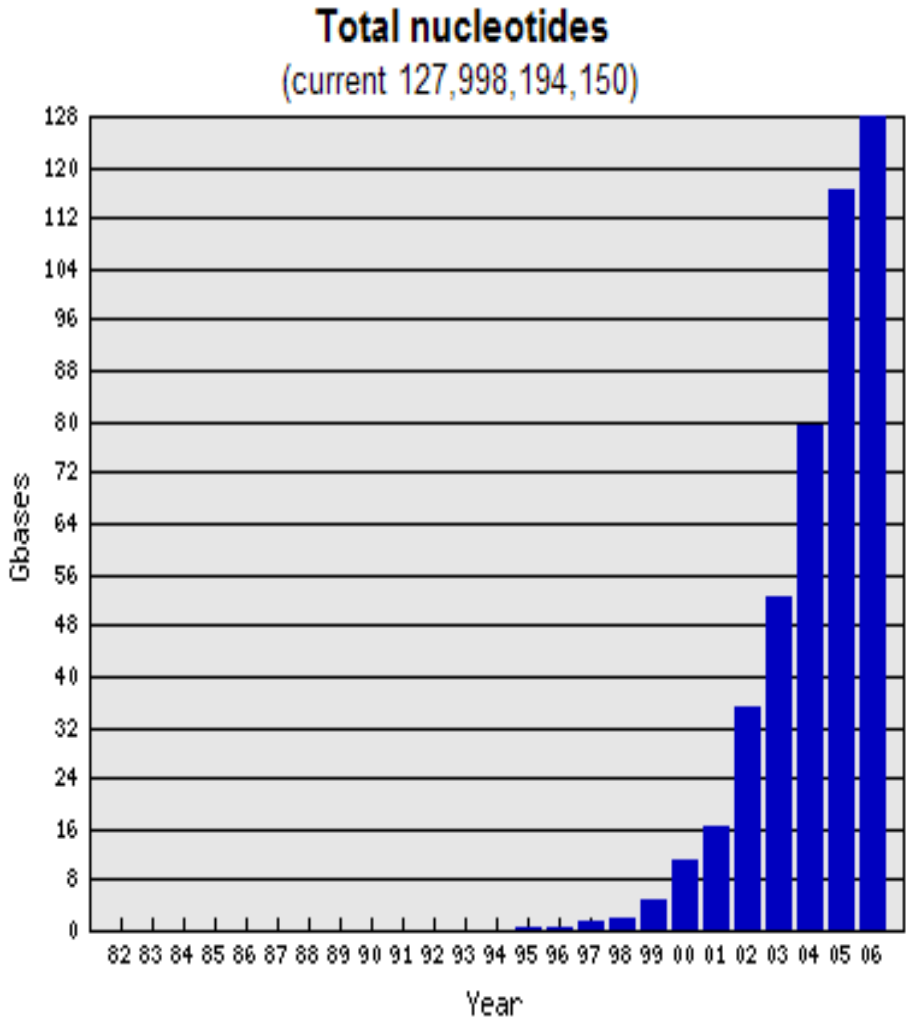

FIGURE 21–2 Total sequencing by year (in billions of base-pairs).
SOURCE: *http://www.genomesonline.org/*

A number of people believe that the more information we get, the less understanding we get. For bioinformatics, of course, this is not true. There the limiting factor is always the quantity of data. We want more data, and we keep on saying that there is no such thing as enough data—although in the last few months, or perhaps the last year, we have started to question whether we really want to keep saying that.

Of course, the issue is not just a quantitative one; it is also qualitative. We are generating more data, but it is important to look also at the types of data. In 2000, the three phylogenetic groups—Actinobacteria, Firmicutes, and Proteobacteria—accounted for 75 percent of all sequencing projects. Eight years later it was actually worse, with the whole "other phyla" portion covering much less—18 percent. We keep on sequencing more or less the same things. We know already that this is not what we see in nature. There is much greater diversity, and we know that from the rates of occurrence observed

directly in the environment and the known phyla that have no culture representatives at this point.

What do we do about that? And why do we keep on sequencing the same things? By analogy with what we heard yesterday about the brain, up to this point we have been seeing mainly genomics driven by medical applications or by energy and environmental applications, not by to the need to understand the whole diversity of nature. To offer an analogy with the brain, it is as if we do not really want to understand the brain, we just want to cure Alzheimer's and diseases of the brain. Understanding the brain, which involves doing the fundamental studies, is totally different from learning how to cure its diseases.

As we will see, however, this picture is changing. The way that we address this question is changing from both the academic side and the industry side. Together, industry and scientists have come to realize that we need to go after the uncovered areas even though there may not be any obvious direct applications there.

What should be done about the uncultured majority—the 99 percent of microorganisms that are not able to be cultured with present methods? I am saying "uncultured" instead of "unculturable" because unculturable means we can never culture them, and that may not really be the case. We do not know if we can culture these organisms or not; there have been no systematic efforts to go after the uncultured majority.

One approach would be to go from genomes to metagenomes. Rather than isolating an organism and sequencing it, we will go directly to the community. We will take a pool of the community, sequence the whole pool, and try to understand what is there. This move from genomes to metagenomes will be one of the major transitions in the second decade of genomics.

The field of metagenomics was spearheaded just five years ago by two studies. One was by JGI and the other by JCVI, which collected samples from the Sargasso Sea. Figure 21–3 summarizes what we know about the complexity that exists out there. We can organize different types of metagenomes based on the species complexity.

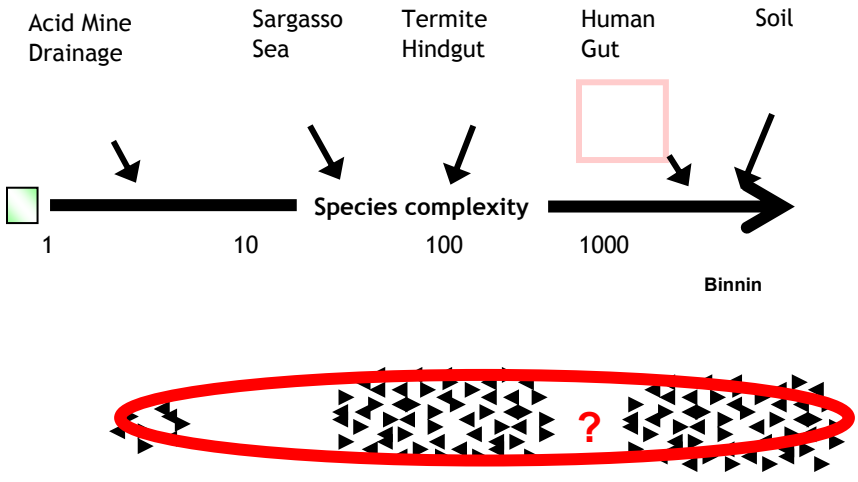

FIGURE 21–3 Species complexity.
SOURCE: *http://www.genomesonline.org/*

Very simple environments have just a few organisms; very complex environments, such as the human gut and soil, have several hundred or several thousand different microbes. Some 200 metagenomic projects are now ongoing, and it is likely there are several hundred or a few thousand other metagenomic projects that we do not know about because they are private or because the data and information have not yet been released.

In a simple environment, if we have enough sequencing we can actually cover and assemble the entire genome. As the complexity increases, however, we have more and more of what we call "unassembled reads," and we do not really know what to do with them. The big challenge here is, How can we study all of these microorganisms?

One way is the ancient Roman approach of "divide and conquer." Using a technique called binning, researchers attempt to understand the function of a microbial community by determining the individual organism or group of organism it consist of. This is the major tool we have to understand these environments. It does however require having enough isolated genomes—enough reference organisms—to provide a reference for the binning.

This leads us to the second major transition, which is moving from individual genome projects—what we have seen over the past two or three years—to large-scale projects. For its first decade, genomics projects were initiated by a single principal investigator and focused on a single genome. This is changing. A project funded by the Moore Foundation looking at 180 microbes from marine environments was, as far as I know, the first example of one project massively sequencing a large number of microbes. A year later, the Human Gut Microbiome Initiative started. And just a year after that, the National Institute of Health launched the Human Microbiome Project with a goal of sequencing 1,000 microbes isolated from different parts of the human body. This major effort is not targeting specific applications, in the sense of looking for organisms involved in a particular disease or condition. Instead, its goal is to achieve as much coverage on the phytogenetic tree for organisms that we know are living on the human body.

With funding from the Department of Energy, JGI has recently initiated a project called GEBA, the Genome Encyclopedia of Bacteria and Archaea. The goal of this project is to systematically fill in the gaps in sequencing along the bacteria and archaea branch of the tree of life.

Although there have been some projects funded by the National Science Foundation to sequence 10 or 20 different microbes, isolated and identified in diverse parts of the tree, GEBA is the first systematic effort to go after large number of genomes and fill in the gaps in the tree. We started with 150 organisms; now we have a cumulative total of 255. About 60 of them are finished, and another 60 are in the draft stage. The paper describing them will be published soon.

The project is being co-led by Hans-Peter Klenk of DSMZ, the German Resource Center for Biological Materials, which is collaborating with JGI. This project would not have been possible without the contribution of the Culture Collection Center at DSMZ. The main limiting factor in such a project is not sequencing or annotation or downstream analysis, but getting the DNA. So the big breakthrough was partnering with the Culture Collection Center, which has been growing all of the strains and providing them free to JGI.

The project has already produced a number of very important discoveries that will lead to applications, including a number of cellulolytic enzymes. Nonetheless, the main

purpose of the project has always been basic research. We constantly say that we do not even know what is out there, and that is the major driver. We must go into the environment, sequence more organisms, and try to identify novel chemical activities, normal enzymatic activities that exist in nature.

The third major transition that is occurring is moving from populations to single cells. Large-scale genome projects are very good at covering the phylogenetic tree, but their limit is that the organism must be cultured, which represent about 1 percent of all the microorganisms. What can we do for the rest of the microorganisms that cannot be cultured? New technology that allows us to sequence single cells, and therefore bypassing the need for culturing, has been initiated just a few years ago. The important thing is that there are a variety of methodologies of ending up with a single cell, and once you have that, a technology exists to amplify its genome and perform the sequencing.

This is one of the most promising technologies for the near future, but it is not yet where we want it to be. We would like to get a complete genome or an almost complete genome of a single organism, but at this point we get something like 50-70 percent coverage of a genome and often not just a single organism. There is a significant presence of other organisms as well. Still, all of the problems are considered solvable, and we expect that with the next two years, we will have the single cell technology that can give us the entire genome of a single organism.

Once that is done, however, we will have to solve some additional problems. For one thing, this is a single event. You can isolate a cell, you can amplify the DNA, and you can sequence it, but you cannot store it or replicate it, and this leads to a number of issues concerning how to save the information. How can we store the genetic material? In this case we cannot. We isolate the cell, we sequence it, and that is it. It does give us a little information about uncultured organisms, but it is something we cannot repeat.

As I mentioned earlier, we are facing an avalanche of data. By the end of this year we will have more than 1,000 finished genomes. We already have almost 1,000 in draft form, with a total of about 8 million genes. Based on that, we can make some projections about what to expect over the next 15 years.

In five years we will have at least three times as many finished genomes and at least 10 times as many draft genomes, with a total of about 52 million genes. This is a conservative estimate. Assuming a linear increase at the rate presented in a paper by Patrick Chain that just appeared in *Science*, we will reach these numbers by 2012.

How can we start comparing a million genes against 52 million genes? This is just from the isolate genomes. If we consider the environmental studies, including only what is publicly available; we have approximately 30 million genes. If we add the environmental studies that we know are ongoing and the Human Microbiome Project studies, we expect that we will have at least 300 million genes over the next 3 years. These are really staggering numbers. And even being in bioinformatics, where people generally hope to have too much data, I am starting to worry about the magnitude of this.

So where do we go from here? We will need new technology, since we cannot handle all those data with today's technology infrastructure. However, we also need conceptual breakthroughs in data processing, data comparison, and data browsing. We will need better ways to present multiple organisms; store and present data, and compute the similarities.

The fourth transition will be going from thousands of genomes to pan-genomes. The term pan-genome was coined a few years ago by Claire Fraser. Here is the issue: In the IMG comparative analysis system we often have a large number of different strains

from the same organism. We have 10 strains of *Prochlorococcus marinus*, 17 strains of *Listeria monocytogenes*, and 15 strains of *Staphylococcus aureus*. Each of these species has a lot of diversity, but most of the genes are the same from strain to strain in a single species. Do we really need to keep all of the instances of the identical genes? We are therefore collapsing all of those different strains of a single species to a single organism, which we call Staphylococcus aureus pan-genome.

Originally, the idea was to do this in order to save both in disc space and computation. However, by doing that throughout the whole phylogenetic genetic tree, we will start getting a totally new picture of microbiology.

A few years ago we thought that all we needed to do in order to understand a microbe was to just sequence a single strain. We now know that is wrong. We need to sequence several different strains. For example, by sequencing different strains of *E. coli* we find there is significant diversity among the strains. Or consider *Staphylococcus aureus*. Its genome is 2.7 megabytes, 2,700 genes. If we take all the different strains, but collapse them all to a non-redundant dataset—which means that every unique gene would be counted only once—we have a total of 14,000 genes. This is five times more than the average genome. In the case of *Pseudomonas aeruginosa* it is much less, a factor of just 1.8.

There are microbes that they are constantly acquiring new genes thus resulting in a much larger pangenome. We call that an open pangenome compared to other organisms that are not so eager to grow their genetic content, which have a close pangenome. This is what we expect to understand if we sequence all of those organisms. For the first time we will have a true understanding of microbial diversity.

The need for a definition of the microbe—the definition of a single microbial cell or single microbial organism—has resulted in a series of paradigm shifts. During the first decades of sequencing, from 1960 to 1990, we were using 16S RNA genes to construct the tree of life, and, accordingly, it looked quite simple (Figure 21–4).

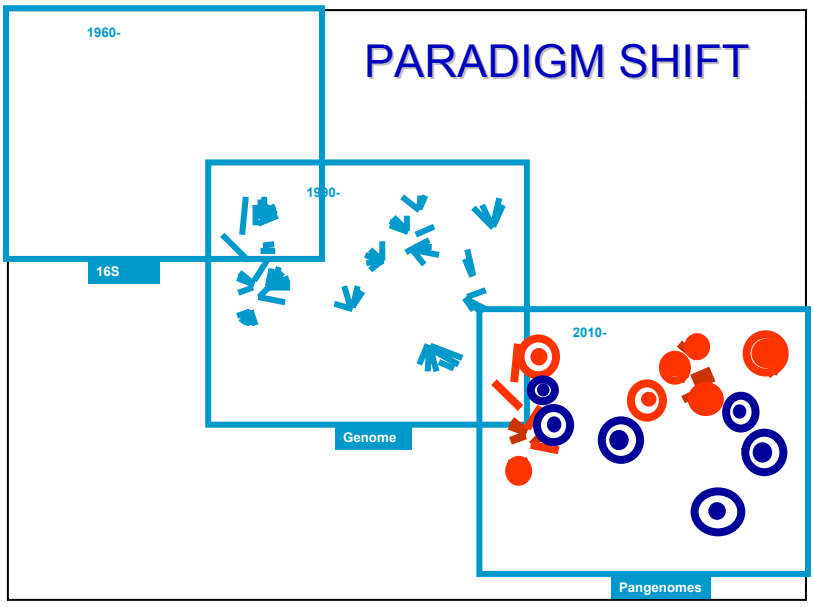

FIGURE 21–4 Changing paradigms for the tree of life.
SOURCE: *http://www.genomesonline.org/*

For the next 20 years, from 1990 to 2010, we were growing the tree. This is what is happening through the individual genome projects—adding details and branches. We expect that the next 20 years will be the era of pan-genomes, during which we develop a totally different understanding of the definition of the microbes and also of the relationship between the pan-genomes.

At this point I am going to change gears and talk about standards and how everything we have discussed until now is related to genomic standards. The Genomic Standards Consortium (GSC) was initiated by Dawn Field a few years ago with the goal of creating standards for metadata: standards for habitats, for DNA sources, for isolation, and for phenotypes. Its mission is to implement new genomic standards, methods of capturing and exchanging metadata, standardization of metadata collection, and analysis efforts across the wider genomics community.

The consortium's first major publication, which appeared last year in *Nature Biotechnology*, was a demonstration by the working group of achieving standards and representing metadata using what is called the Minimum Information about a Genomic Sequence (MIGS) specification. On the GSC website you can find the list of MIGS fields used to specify a sequence, and one of the fields is the source material identifier. This gives the information necessary to find the sequenced organism or the biological material in the culture collection where it has been deposited.

Unfortunately, too few of the sequences from the genome project have been deposited. Of the complete microbial genomes we have so far, 53 percent have been deposited and 47 percent have not. For incomplete genomes it is much worse: Only 35 percent have been deposited. The hope is that since this is ongoing work, the researchers have not yet deposited it, and by the time they release the sequence, they will also deposit it. However, the GSC has emphasized that it should not be left to the discretion of the researchers. The funding agencies and the publications should mandate the deposit of the biological materials, at least at the time of the publication.

It is a different story concerning the release of sequencing data. Sometime around 2001 or 2002, there was a change in the release policy, and the funding agencies—with the Department of Energy (DOE) taking the lead—have been forcing researchers to release the data as soon as possible. From the moment that the scientists get the data, they have three months to release it. This has resulted in an increasing number of genome public releases.

The scientific community does want to have the data as soon as possible, even without an associated publication. The complete genome analysis is not trivial; it requires a tremendous amount of effort and quite often it takes more than a year or two, or even three. We thus expect to see the number of complete genomes put into public databases without a corresponding publication to go up.

It is necessary as well to have at least a minimum amount of information about the organisms that have been sequenced, including the metadata associated to the organisms. For that purpose, the journal *Standards in Genomic Sciences* was launched in July 2009, with George Garrity as editor-in-chief. The journal, which was funded by Michigan State University, is an open-access, standards-focused publication that seeks to rapidly disseminate concise genome and metagenome reports in compliance with MIGS standards.

The journal provides an easy method for scientists to report their results. They can just download one publication, change the introduction or the information about the

organism, and submit the information about their organism. This allows the community to have access to the additional metadata in addition to the complete genome.

Metadata is not the only interest of the GSC. It also focuses on issues related to data processing, such as sequencing standards, finishing standards, assemblies, and gene predictions, as well as on annotation issues. It is likely that there will be a workshop by DOE early next year that focuses on standards and annotation. A paper just appeared in *Science* that provides the first standards for genome finishing quality. This resulted from an effort led by Patrick Chain from JGI. And there are similar upcoming efforts that will deal with both gene findings and function predictions.

To sum up: Microbial diversity remains largely uncovered. The vast majority of current genome projects do not cover novel ground. To understand an organism, we need to sequence a reasonable number of closely related strains and incorporate standards.

The question is where do we go from here? How does one use all this information to spearhead or initiate a new project that will address those problems and comply with all the standards on which we have been focusing?

Over the last few months members of the GSC have begun discussing the possibility of launching a global genome sequencing effort for microbes. The idea is to imitate, but on a much larger scale, the Human Microbiome Project effort. By funding different sequencing centers in a single project, NIH has achieved something that seemed almost impossible a few years ago. It has gotten competing sequencing centers to work with each other and share not only metadata, but also pipelines and other resources.

This is one of the things we will try to do by expanding this idea into an international consortium. If you look at a map of who is doing sequencing, based on the number of genome projects per country the United States has complete dominance, although there are many sequencing projects around the world. There are about 20 countries that have significant sequencing efforts. Instead of having a single sequencing center, such as JGI, or even four sequencing centers, as is the case with the Human Microbiome Project, we want to organize a bigger international effort and ask the different countries of the world to contribute to an international genome project.

What will that project be? We want to sequence at least one representative of every cultured microorganism at the Genus level. At this point we have only about 30 percent coverage. This means that we do not have a sequence representative for about 70 percent of all genera—or about 1,500 different genera for which there is a type species. At the species level, the coverage is only about 10 percent. The remaining 90 percent corresponds to about 10,000 species for which we need a sequence representative.

We therefore have an immediate broad goal that cannot be possibly achieved by a single researcher or supported by a single funding agency. This is of global importance, which is why we provide the list of genome targets—the information for the reference points we need in order to support the metagenomic analysis that everybody is doing. For the first time, we will be able to have a reference point for every branch in the tree of life.

Among the Archaea, we have genomes for 86 out of 108 genera and for 98 out of 513 species. So, the remaining genera and species provide us with the immediate targets for the project, but this is really the low-hanging fruit. We are only talking about sequencing already characterized organisms.

It will not be enough, however, simply to sequence type species. Instead, we need to sequence enough strains for each species to fully characterize them and generate a species pan-genome. This is the absolute minimum we need to do in order to have a clear understanding of microbial diversity and what is out there.

Furthermore, there have been only minimal efforts to understand the effects of geographic distribution on species dynamics. We have sequenced at most 30 different strains of the same species from different geographic locations. We need to do this to a much larger extent. This will be another goal of the project.

The key partners in the project will be the GSC, which is definitely the major partner; culture collection centers, which will provide all the biological material for the project; representatives from Grand Challenge projects, including the Genomic Encyclopedia of Bacteria and Archaea, Terragenome, and the Human Microbiome Project; and other participants from large sequencing centers and country members. The response so far has been enthusiastic. A few months ago, there was a meeting of the European Culture Collections Organization, and they all said that they want to contribute. Country members like China, Korea, and Japan have already been invited, and they are very strongly supportive of such a project.

Progress in the future will depend on collaborations across national centers rather than simply between individual researchers. Fortunately, the funding agencies have finally said we will not support you unless you will start working together. This is a critical step, particularly for an area like bioinformatics, where traditionally researchers have thought they could do everything by themselves. Different groups that were competing until now have actually started working together. For the first time we are talking about sharing pipelines, sharing computations, and sharing methods of analysis.

22. Accessing Microbiological Data: A User's Perspective – Mark Segal[49]

Environmental Protection Agency

My purpose today is to demonstrate that there is a potential user community for the microbial research commons that goes beyond researchers—that there is a cohort of us who are primarily users, not suppliers of data. At the same time, however, some of our needs may be the same as, or similar to, the needs of researchers.

I will begin by giving you some examples of people, like myself, who are included in that cohort. I am a scientist doing science, but within government. I work within a regulatory organization, and I am part of a scientific support group for the people who actually do the regulation writing. There are other scientists who are not also primarily researchers yet are potential users of the kinds of data and information that the commons can make more accessible. Some governments or parts of governments hire scientists to provide analyses, rather than employ them directly. These scientists may provide similar functions to mine while under contract. Besides analysts who support governmental actions, there are scientists responsible for funding research who could benefit from improved data access. Outside government, there are a number of other analysts, including those at commercial think tanks and non-governmental organization staff or NGOs who may benefit from use of consolidated microbiological information. Finally, those employed by various media to report on science issues may find it necessary to get deep into the details of given projects in order to present the results in an accurate manner to the public.

Box 22–1 illustrates the range of data types that make microbiologists remarkable in the diversity of information they must utilize.

Box 22-1

Range of Data and Information Types Routinely Used by Microbiologists

- Text
- Numerical
- Binary
- Graphical
  - Images
    - Macroscopic (e.g., colony morphology)
    - Microscopic (e.g., cell structure)
  - Charts and graphs
  - Diagrams and Cartoons
  - Molecular structures
- Sequence

At some point in our careers we use just about everything that is on this list, so the commons will have to deal with as wide a range of data as is ever encountered in science.

[49] Presentation slides available at: *http://sites.nationalacademies.org/xpedio/idcplg?IdcService=GET_FILE&dDocName=PGA_053678&RevisionSelectionMethod=Latest.*

I am going to use myself as an example to illustrate where the microbial commons can be useful. Box 22–2 lists some areas of microbiology in which people in the categories previously described could be interested.

BOX 22–2
Areas of Interest in Microbiology

- Public health and pandemics
  - Analysis of outbreaks
  - Evaluation of drugs and vaccines
- Food security
  - Evaluation of products of food biotechnology
  - Diagnostics
  - Antiterrorism
- Bioremediation
  - Evaluation of microorganisms used for cleanup
- Biofuel and bioproducts
  - Evaluation of microorganisms used to make biocatalysts, enzymes
  - Evaluation of microorganisms used to make fuels
  - Evaluation of microorganisms used to make chemical substances

Specifically, bioremediation and biofuels or bioproducts are products and processes in which I am closely involved. In particular, the items in these categories are examples of products or services provided by microorganisms that are subject to oversight by my organization. You can see that there is a wide range of potential commercial uses for which microbiological data made accessible through a commons could be used. I want to discuss the kinds of data and information that we have to deal with on a routine basis that could be made more accessible to us if the commons did exist and was in operation.

One of the things that we constantly have to deal with is knowing exactly which organism is being worked with when a submitter provides us with information on an organism. Has the submitter obtained an accurate species identification using the tools available to him? More often than not, commercial organisms belong to that collection of open-genome organisms in which there is a broad range of entities falling within a genus or within a species, with lots of apparent gene exchange and a consequently diverse gene pool. These taxa would appear to have tiny core genomes compared to many genomes in genera that are less diverse. They often have lots of mobile genetic elements. Because of this diversity and especially if determinants used for identification reside on these elements, trying to identify the species of such an organism is a challenge. But, since much of the pan-genome gene pool is sharable, this can at least tell us the range of potential functions that may be expressed, regardless of the species name applied to the strain. Knowledge of the content of this gene pool is something we can work from. We understand about the utility of metadata—how it enables us to know where an organism came from, trace it back to its origins, and figure out what it did, or at least what its precursor did, in the natural environment. Because we deal with health and safety, environmental effects, and those kinds of things, there are different types of information that are useful to us: Where is the organism from? Was it part of an outbreak? Is it is known to be relatively safe when it or its precursors are used commercially? What else could the organism be used for besides what we are being told it might be used for?

We get our data from a variety of sources: the open literature, grey literature, company files, public data banks, and other Web resources. We are interested in various

issues concerning the sources of microbiological data. The participation of private-sector parties in a consortium raises issues, such as having data held confidentially. Classified data also would not be included within the commons. Nevertheless, we need to be able to integrate those data with what we can get from public sources.

Concerning the open literature, subscription costs may limit the number of subscriptions to journals and other sources that potential users can readily obtain. As journal costs to libraries increase, this circumstance may become critical for many parties. Language can be an issue. Some of the older articles are in languages that we may not be able to translate. Recently, I had to deal with an article in Portuguese. Fortunately, I had enough French to enable me to understand the key issues I was looking for. If the article were in some language that is outside the set of language skills possessed by our group of scientists, we would have to send it off for translation. That takes time we often cannot afford.

Grey literature poses problems, particularly in finding it. It often is not catalogued. Yet it may contain valuable information. When present on the Web, it often resides on obscure sites. The fact is, we ourselves generate grey literature. The assessments of our group become grey literature. Some portions are made public, but much of it is confidential because it may contain proprietary data and information. Only a few are permitted to see it. It is not easy for others to find our reports. So, anything that makes it easier for us to make our work available and to find work that is similar to ours elsewhere is going to help us.

We need to use databanks, but we know they may not be complete. We also know there may be accuracy issues. Many databanks need to be better curated than they are. Also, who is doing the annotations? Who is printing the information? How old is it? Sometimes we have the skill to recognize the errors, and sometimes not. In some cases, we heard the data were stove-piped, which can be a problem since the data are not connected with potentially related data, leading to a limited perspective. My group integrates a lot of different types of information, and so we tend to go across disciplines a lot. Getting past those barriers is critical. There were earlier examples in this symposium in which people are trying to break down these barriers. We encourage that, but wish there was more of it.

How can we, the data users, benefit from the commons? Overall, having access to researchers and other data users is certainly going to help us. If we were able to have one-stop shopping—having portals that allow us to move back and forth among the range of data sources that we routinely use—that would be great.

We use many different digital information resources. Many of them are linked or are becoming linked, but sometimes the linkages are very awkward. It would help us tremendously if there were a way to navigate through the maze of data sources that are now out there, so that we could deal with them more easily than is possible now.

In what ways can we exert some influence on improving the situation? Can we do a better job, for instance, of getting our grey literature posted and accessible on the Web so that you can locate it? Can we find a way to limit the amount of data that is treated as confidential? We are trying to facilitate information sharing, as appropriate, so that our analyses can be made more transparent and so that the way in which we do our work can be better understood by others. Some of this is changing, but hopefully we can do more.

23. The Microbial Commons: Journals and Professional Societies – Samuel Kaplan[50]

University of Texas at Houston Medical School

I was chair of the Publications Board of the American Society for Microbiology (ASM) for nine years, so I am very familiar with ASM journals. The nine peer-review journals and two review journals were publishing annually approximately 55,000 pages as of several years ago. This is a lot of information. Digitally, the materials are posted on HighWire Press and in PubMed Central, as well as through Google, so this content is very well available. It is also available through PubMed International, which makes literature in PubMed available in other countries. Last year on HighWire Press alone, there were 18.6 million PDFs downloaded. To me, that spells access.

About 40 percent of the manuscripts submitted to our peer-review journals are accepted. We could make our journals more "boutique" by lowering the acceptance rate, but we feel that a scientific society should encompass the broadness of the field, i.e., covering the field of microbiology. For us, this is a very important reason for our being.

More than 50 percent of our publications now have foreign authors, a figure that has increased steadily over the past decade and a half. Fifteen years ago, approximately 30 percent of accepted publications were from foreign authors.

The journals are important to scientists in various ways. In order to get grants, for instance, you need publications in good, peer-review journals. The quality of manuscripts and where they have been published also affects career advancement.

This is not free, of course. Those 55,000 pages—print, as well as digital—cost about $20 million yearly. It is a big operation, and it costs a lot of money to get this kind of quality. It is impossible to imagine such an operation without the effort and structure that is put into it.

If we were to stop printing paper copies, the cost would drop to about $13 million yearly. I tried very hard to get rid of print journals for the ASM before I stepped down as chair of the Publications Board. However, it was felt we could not drop print even though getting rid of that format would have allowed us to perform a variety of other activities. I should note, however, that the digital form of the journal, not the print journal, is now the official copy of record.

How do the ASM journals fit into the microbial commons? We have 144 editors and a stable of 28,000 reviewers, of which 13,000 were actually used in the past year. So if you are thinking about supplanting the established journal process, you should think carefully about the magnitude of that undertaking and how this could be accomplished.

Thanks to the National Library of Medicine, all of our journals are digitally archived and available, beginning with the first publication of the *Journal of Bacteriology* in 1916. We have standardized the literature across the field. Researchers, who communicate, need to have a sense of how to talk to one another in a field, so all of our journals try to standardize that process. This is not always successful, but there is a reason for having print journals or, if not print, at least a recognized subscription and editorial process.

---

[50] Presentation slides available at *http://sites.nationalacademies.org/xpedio/idcplg?IdcService=GET_FILE&dDocName=PGA_053737&RevisionSelectionMethod=Latest.*

The journals also provide quality control and readability. We have over 60 people on staff, a large number of whom are involved in copyediting. Although it costs a lot of money to copyedit journals, it is important. Moreover, the fact that more than half of our articles are now authored by foreign nationals makes the issue of copyediting and readability in English much more important across the field.

Furthermore, the journals provide control over the format and presentation of figures and tables. Despite the many pages of information we provide concerning how to submit a paper and what format the figures and tables should take, they come in as a hodgepodge. So unless somebody does the work to put them all into a consistent format, readability would be impossible.

Journals also digitize content for broad distribution. Soon, perhaps, the print version will be dropped altogether in favor of the digital format because this would make it possible to be involved in more areas, such as the gray literature, data mining, and other functions that can only be done online and that are not available through the print journal.

Journals also help with nomenclature in microbiology. This is vital in dealing with the coming avalanche of digital information. There are 40 terabytes of genes coming down the road, and you have to call them something. It is no good simply to have a million genes and wish to communicate about them. They have to have names. There has to be some way of recognizing these genes, which means that nomenclature is a critical issue, but it does not arise *de novo*, nor is it maintained within a vacuum.

The ASM journals try very hard to standardize nomenclature according to the old Demerec's rules, but it is simply impossible. Last year there were 540 manuscripts published where nomenclature was a serious issue in terms of trying to standardize what people call genes. In establishing a microbial commons, this is something that should be addressed up front. What do you call genes? You can have digital identifiers which are a best effort to identify a gene, but within such a context there are some genes that are known with 100 percent certainty, there are many genes where the probability of their function is less than 60 percent, and there are many, many genes whose functions are unknown at any level of probability. This is a real problem and only getting worse. Without rigid structural context, we only will exacerbate this problem.

Finally, as has been mentioned several times in this meeting, journals play a major role in making sure that biological materials are shared. We have a strict rule in all of our ASM journals: If you do not provide the materials to the community that you published about, you are not going to publish in an ASM journal again. Indeed, we very often get complaints from researchers who tell us that they have had difficulty getting the biological material from a cited author. If the report was published in an ASM journal, we will write to the author and remind the author of our rule. We cannot take the author to court, but we try to use whatever powers we have (denial of future publication in an ASM journal) to ensure the sharing of biological materials.

We also try to use the power that we have to make sure that the materials and methods are completely described. We have refused to publish papers—even otherwise scientifically acceptable papers—where authors have not provided enough detail concerning the materials and methods.

What should journals do to further the advance of science? Here is my personal wish list assuming that all were digital and we did not have the constraints of print copy.

I would start with free and immediate availability. If you are a member of the ASM, you can get immediate digital access to all 55,000 pages for $235 a year. That may seem like a lot of money, but an iPod costs $200, and everyone seems to have one.

In addition, everything we publish is put on PubMed Central. When I took over as chair of the ASM Publications Board, all our journal content was made freely available after one year through PubMed Central. We reduced that period to six months, and then to four months free, during my tenure. When I left the Publications Board, they increased it back up to six months.

I also would like to remove all copyright restrictions. The only time that I know of that the ASM journals enforced any copyright restriction has been when a commercial publisher has used some of our material without the appropriate attribution. We did not seek to charge a fee, but only to require the appropriate attribution.

I would like to be able to crosslink our content with that of all the other microbiology journals. I would like to be able to link references and tables of contents. I would like to incorporate unsolicited peer review and the gray literature. You can only do this, however, if everything is digital.

I would certainly like to encourage data mining, and I would like to arrange for the development of "critical tables". This is my pet peeve. Physicists and chemists have critical tables: You can find the boiling point of anything; you can find the molecular weight of anything or the refractive index. For the genome projects, I would like to see a series of "genomic critical tables". It is already possible for me to click on *E. coli* and see what pathways it has and so on, but I would like to be able to do more. I would like, for example, to start with something like the enzyme lysine decarboxylase, then click on that and find all the organisms that have lysine decarboxylase in order to find out the map position of its cognate gene. I would like to be able to find out many things about that enzyme and to find out its nomenclature as well as its molecular size, pH and temperature optima, as well as regulatory elements and so much more.

Returning from this ideal world to the real world, I am going to touch on several issues facing scientific publishing, and I will begin with the cost of content.

When print was the only option, the cost of content or of getting access to content was based upon the "three-legged stool" model. The three legs were subscriptions to libraries or institutions, subscriptions to members, and page charges to authors. Now, institutions are rebelling, and everyone is saying the literature should be free. Even members of the ASM find it pricey to pay $235 to get all this content. So, with the advent of open access, the question has been raised, Why not have the author pay the full cost? Here is the problem. Recently *Nature* announced that it will cost $3,000 to have an article published under open access conditions and, in general, the going price seems to be somewhere between $2,800 and $3,500 to have an open access article published. If you as a researcher publish 10 articles a year, that is $28,000 to $35,000, but even three per year would be costly. No study section at the NIH that I know of is going to give you $35,000 in your grant for publication costs. The NSF is even worse.

When we raised the page charges of certain high-impact ASM journals from $55 to $75 we got a flood of complaints from prospective authors saying it was outrageous and they would never publish in those journals again. Given that the total price might have been around $800, imagine what they would have said if it was $3,000?

So, we are going to have to address this issue. One way or the other, it costs money to publish. So who is going to pay if the authors cannot or will not do it?

Protection against misuse is going to be another major issue. This is something that the organized journals do very well. If, for example, an author ever came to us and said, "I just saw my figure published in another paper," we would approach the editor or the publishers of that paper and investigate the situation.

The detection of falsification and plagiarism is another job of the scientific journals. This is a real problem. Believe it or not, it happens a lot. It takes a concerted effort, costing time and money, to perform the necessary sorts of policing activity, and it cannot be done in a lackadaisical fashion.

Dual use refers to situations where a given scientific finding can be used for good—the purpose of its publication—or it can be used for evil, say, to make some bioweapon. No one has mentioned such dual use regarding the microbial commons and open access, mainly because it has not been on people's radar screens recently, but if there is another anthrax incident or something similar, it will quickly get attention. As soon as you were to start releasing all of this information free without any oversight—which, by the way, is what I believe should be done—then the issue of the "bad guys", whoever they are, using that information for evil purposes becomes a concern. Thus dual use poses a real threat to open access.

From my perspective, one of the most important issues is journals' income and the role of professional societies in promoting science. Professional societies do many things beyond publishing journals and holding meetings. There is a collective membership, for example. The ASM has about 40,000 members, but in a sense that membership extends out to all the people who publish in ASM journals or attend ASM meetings, which by the way, are both populated by a minor fraction of members of the ASM.

The ASM has educational programs that range from kindergarten to high school, and we spend a lot of money on those programs. We serve on government panels, for example. When the government asks someone from the ASM or its leadership to serve on a panel or committee, or to represent the government overseas on some issue, the government does not pay for it. The travel expenses come out of ASM money. This is part of what we do to promote the field of microbiology.

We interact and cooperate with other societies, as well. We hold joint meetings and have joint publications with other professional societies. We develop standards for quality control. We lobby the government on behalf of microbiology and of biology in general. We spend considerable sums to enhance the profession in terms of CMEs (educational credits required to enhance the professionalism of scientists and technicians) and other things. We provide scientific information to the public in many forms, such as the *Microbe Minute* on National Public Radio. The *Wash Your Hands Program*, which began 8 or 10 years ago, was an ASM-sponsored program.

The journal of tomorrow, as I see it, will be one that is fully digital and that is interactive at all levels with the community at large.

When considering the establishment of a microbial commons, there are a number of questions we should be asking: Who contributes to this commons, and how do they contribute? What will the content be? Will recognition be attributed, and by whom, from whom, and to whom? Who will pay for the cost of the commons? Who will provide the upkeep? Who will validate, vet, verify, and provide access to the commons? Who will maintain the commons over time and so much more, as described above?

24. Microbial Commons: Overview of the Governance Considerations—A Framework for Discussion – Tom Dedeurwaerdere[51]

Université Catholique de Louvain, Belgium

Exchanges of microorganisms among culture collections, laboratories, and researchers worldwide have historically occurred in an informal way. These informal exchanges have facilitated research activities, and, as a consequence, science and exploitation of microbial resources have advanced rapidly. During the last decades of the twentieth century, this situation has changed. Major drivers of this transformation are the increasing commercial pressures from biotechnology firms active in microbiology and the introduction of new legislation on the use of and access to biological resources. As a result, the access and distribution of genetic resources are now more strictly regulated and, therefore, exchanges are becoming more and more formalized.[52]

Before addressing these issues, let me make a brief note about terminology and, in particular, about the meaning of "commons" because the term is used very differently by legal scholars and by economic scholars. To clarify, I use the consensus definition that came out of a workshop organized by the Center for the Study of the Public Domain at Duke Law School. In that workshop there was a great deal of discussion about the definition because the term comes from natural resource management, but has now moved into the field of the Internet and the science commons. Therefore, an approach was needed that covers both shared resources that are depleted upon use, which are designated as common pool resources, and shared resources which are not, such as ideas, which are pure public goods. The workshop came up with a very simple standard definition. A commons is a resource shared and managed in common by a group of people. The group of people can be very small, like a club in the sense discussed by Minna Allarakhia[53] this morning, or it can be at the level of a community or even multiple countries.

The concept, as defined, includes both the semicommons, which is a partially restricted area of exchanging resources and digital data, and a fully open commons. Those distinctions can be evolving for the same resources; materials may remain in a semicommons for six months or a year and then come into the full commons.

With that in mind, I would like to give some examples of the benefits of global and regional exchange of microorganisms in a commons, provide an analysis of the patterns and norms of exchange, and then examine the institutional design implications for the development of research friendly formal institutional arrangements.

I will begin by examining some existing practices where broad worldwide sharing of materials and information provides key benefits to both public and private actors. Well-known examples of the worldwide sharing of biological resources involve microbial materials in the field of food and agriculture. In the early 1950s, stem rust (race 15B)

---

[51] Presentations slides available at *http://sites.nationalacademies.org/xpedio/idcplg?IdcService=GET_FILE&dDocName=PGA_053743&RevisionSelectionMethod=Latest.*

[52] See also, T. Dedeurwaerdere. Global microbial commons: institutional challenges for the global exchange and distribution of microorganisms in the life sciences. Research in Microbiology 2010;161(6):414-421.

[53] See Chapter 20 within this publication.

devastated the US and Canadian wheat crops, leading to estimated losses of around US 3 billion (in 2007 dollars). This disease prompted the organization of the first international nursery trial to test wheat lines for resistance in seven countries. As a result of this international breeding program, stem rust was brought under control by the mid-1950s.

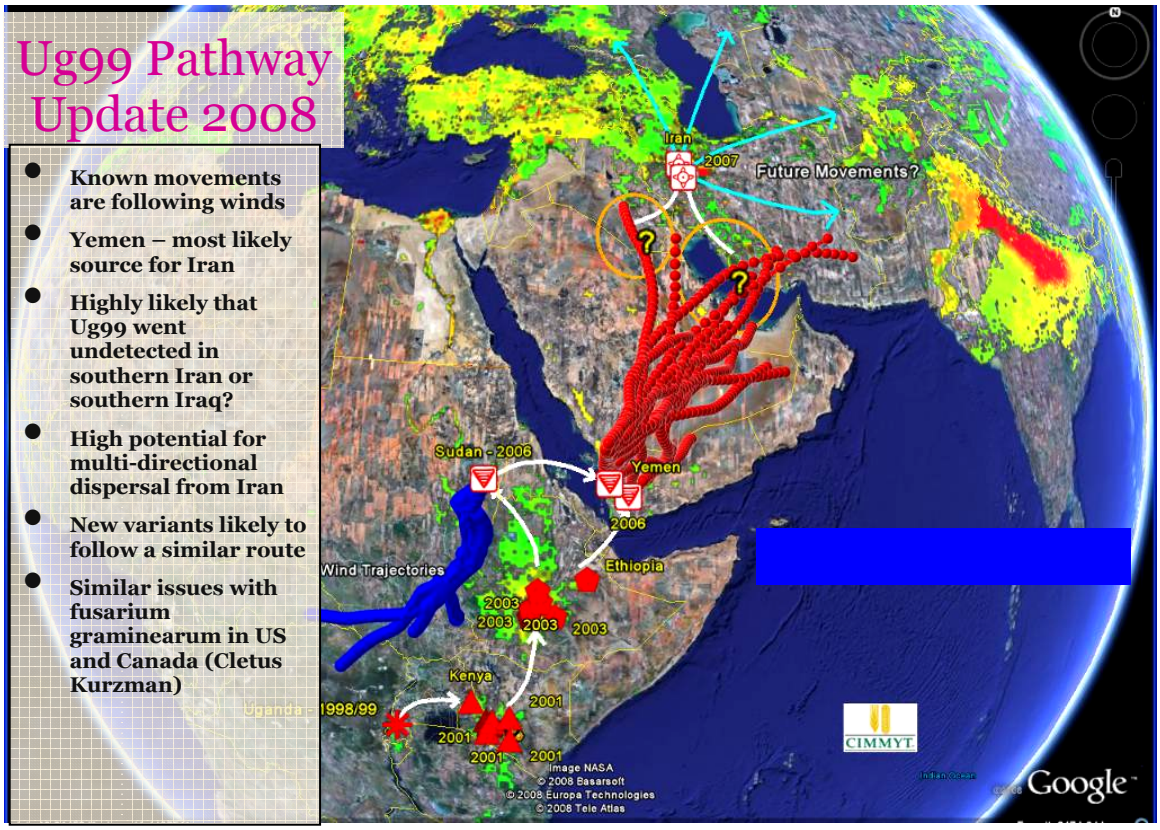

FIGURE 24–1 Path of a disease outbreak.
SOURCE: Amri Ahmed, 2nd microbial commons expert workshop, Brussels, March 25, 2009.

This success was one of the motivating factors to establish the Consultative Group on International Agricultural Research (CGIAR), which coordinates international breeding programs for the main food crops based. This in turn led to the Green Revolution, golden rice, and other agricultural breakthroughs.

What is interesting is that the disease is back again. Figure 24–1 shows the path of a recent epidemic of stem rust spreading from Uganda since 1999. So we are faced here with an evolving biological reality—there is a race going on in which microbial pathogens evolve and make previous crop improvement programs obsolete. The microbial pathogen population of Puccinia graminis has been evolving and a new program of international collaborative research is needed to bring stem rust under control in East Africa and the Middle East.

An important lesson from these programs is that it is only possible to derive the benefits for disease diagnostics and crop improvement if one can gather and identify the microbial materials from all the places where the diseases are, which requires collaborative DNA sequencing arrangements of the main materials, and track down all

the existing scientific information required to identify the mutations of the pathogens that are responsible for the disease.

There are many other examples of research with microbial resources where the end results require not just one local study but a large collaborative project. One such case is the use of microbial ferments in food processing, such as yoghurt.

By gathering all the strains of yogurt of a great number of the farmers' markets in Georgia, public sector researchers were able to design uniform starter cultures for the dairy industry in Georgia, whose milk products are known to possess special pro-biotic properties. They were able to develop a standardized yogurt product by collecting all the strains, analyzing the genetic sequences, and then selecting the strain that was most useful. Again, this was a major collaborative project with direct social benefits.

In the commercial context, let us briefly examine bio-prospecting outside national jurisdiction, which can still be done in places like Antarctica or the open seas. These are areas where any company can go and harvest valuable microbials.

For example, companies active in Antarctica have collaborated on a website (*www.bioprospector.org*), which shows innovations based on microbials from Antarctica. All the ice cream companies are there because they are interested in microbes that can affect the freezing point. The bioportal shows an incomplete list, but it already does demonstrate that the private sector has a big interest in accessing microbials throughout the world. It is a list of examples where benefits have been transferred from a public commons, with shared use of a whole range of resources and digital information, to private sector innovation.

These three examples aim to illustrate how emerging groups and cooperative networks, both from the public and private sector, are trying to produce benefits from global and regional exchange to the broader society. This symposium has highlighted a number of other interesting initiatives involving emerging cooperative networks trying to provide such benefits, but overall, these initiatives are still disjointed and involve many ad hoc arrangements. As yet, no overall vision of an integrated infrastructure has emerged.

For the design of a worldwide microbial commons, a more systematic approach that is based on a set of agreed rules between the collections, the users and the provider countries, is needed. The main issue that has to be addressed in this context is the creation of a better fit between the formal institutional arrangements required for building a global science infrastructure and the norms and goals of the microbial science communities. In particular, to foster wide acceptance and thereby accelerate scientific progress, any formal arrangement needs to be committed to facilitate the exchange of materials and need to be easy to implement by regulatory bodies, as well as both parties involved in the exchange (providers and recipients). This raises a double set of problems. On the one hand, institutional frameworks that rely excessively on monetary incentives or formal control can crowd out the social norms of communalism and the intrinsic values that drive scientific communities. This is especially relevant for the bulk of microbial resources that are exchanged for public research purposes. On the other hand, without a formal arrangement of some kind for regulating the exchanges, the benefits of the infrastructure might be restricted to the most advanced researchers, who organize exchanges on the basis of networks of personal relationships. The goal of further harmonization of the institutional frameworks should therefore be to provide the broadest access possible to essential research materials—within the constraints set by biosecurity

and quality management requirements, while preserving the community norms which motivated the practice of exchange to start with.

Let us discuss some of the available institutional options. One option is the adoption, on an international level, of a set of legally binding rules to govern transactions involving microbial resources. This would potentially alleviate many of the problems caused by the lack of standardization and agreed formal rules which characterize the current system of exchange. The development of a fully fledged international regime takes time however, and, in the light of the threats to the commons and the public good benefits that may potentially be lost, it is urgent to work on interim solutions for putting the global microbial commons on a sound legal basis.

In the area of the microbial commons, there are some emergent examples of such interim solutions which might lead to the building of a global commons. It is interesting to see that those are appearing throughout the world. It is not a north-south divide. For microbiology collections in Thailand, for example, the institutional arrangement is based on two Material Transfer Agreements (MTAs): one for regular distributions, and one for what they call legal or legitimate exchange among culture collections. In the case of the latter, as long as the strains circulate among the pool of more than 600 culture collections who are members of the World Federation of Culture Collections, anyone can redistribute it within that same network.

Russia is adopting the same approach, as is the European Culture Collection Organization. One could envision extending this approach beyond just exchanges among culture collections, to encompass also type strains and reference strains held by qualified research collections, because those are basic research materials that everybody needs, whether they are in the commercial or the public sector.

A third model is the clearinghouse model, where only information is shared and not the materials. In an information clearinghouse, all the information on available strains is put on a common bioportal, and people get the materials from the most nearby places or where the license conditions are the most open. One can go onto each website and see the different conditions the culture collection imposes. This clearinghouse model has been developed for research into Huntington's disease by Science Commons, for example (cf. Science Commons MTA project, *http://sciencecommons.org*).

Finally, you have the public domain, as in the case of bioprospecting in Antarctica and the high seas. That is also a commons. But there, you do not need contracts. It is unregulated and in the open.

How should principles of governance be designed for this whole galaxy of projects and emerging initiatives? We need to move from a disjointed set of bottom-up initiatives towards an integrated, but still distributed, infrastructure. The research question that needs to be answered is: How can we create the best possible fit between the governance of scientific infrastructures, on the one hand, and the normative practices and needs of the microbial research commons, on the other?

In short, there are trade-offs and many complex social mechanisms to be considered in designing the governance rules. It has to be done in a way that keeps and reinforces the existing normative practices, but that also adds new mechanisms for coordination wherever needed. That is the challenge that we have in front of us. What are those norms and collaborative practices that must be taken into account when thinking about governance?

Currently, more than half a million microbial samples, which have been collected in various countries, are distributed throughout the world every year by the public *ex-situ*

collections that are members of the World Federation of Culture Collections alone, mostly for the marginal costs of distribution. Each of these collections contains a very substantial set of unique materials. An average of 40 percent of the strains in the WFCC that are referenced on StrainInfo (*www.straininfo.net*) are unique. Intense collaboration and exchange amongst culture collections is a necessary consequence of this situation. It is difficult to estimate how many *ex-situ* materials are exchanged between research collections outside the WFCC collections on an informal basis, but it is fair to say that the volume of materials exchanged between these collections is probably even greater.[54]

In order to get a better picture of the institutional arrangements within the microbial commons, a set of original surveys and interviews were conducted in 2005 and 2009. In 2005 Stromberg et al.[55] surveyed the 499 public collections that were members of the WFCC at that time (119 completed survey forms). In 2009 Dedeurwaerdere et al.[56], undertook a quantitative assessment of the entire accession database of a geographically representative set of 9 major collections over 3 years (2005, 2006, 2007 : totalling more than 15,000 single accessions), conducted semi-structured interviews with administrators of these collections, organized a short complementary email survey on access and benefit-sharing measures with 238 WFCC collections (43 completed questionnaires), and completed 16 in-depth phone interviews with scientists from both public and laboratory culture collections.

The quantitative assessment of the databases of the 8 major collections showed that for 6 collections more than 98 percent of all the deposits of 2005 to 2007 came in without restrictions, in spite of the use of formal deposit forms by these collections. For the other 2 collections around 85 percent came without restrictions. The overall experience was that after formalizing the process, the vast majority of deposits were still done without restrictions. The lesson is that people not only operate on the presumption that their work is part of a global research infrastructure, but when you ask them to sign a form, they are willing to give up their proprietary interests as well, in exchange for the benefits from making the microbial material available for follow on research and publication purposes. The collections we examined were selected from a set of major collections throughout the world, some in the United States, some in Europe, and some in Asia and South America. We did not see any exceptions to this behavior.

It is also worth noting that the depositors to those culture collections often come from other countries. Researchers in India or Brazil are regularly depositing to collections in the United States, Europe, and Japan. Even for the deposits done by national researchers, if you look at the country of origin of the material that is deposited, 60 to 75 percent of those materials come from other countries. So the people are collecting and depositing throughout the world.

---

[54] Dedeurwaerdere T. Institutionalizing Global Genetic Resource Commons: Towards Alternative Models for Facilitating Access in the Global Biodiversity Regime. 2010

[55] Stromberg, P., Dedeurwaerdere, T., Pascual, U., 2007. An empirical analysis of ex-situ conservation of microbial diversity. Presented at the 9th International BIOECON Conference on "Economics and Institutions for Biodiversity Conservation", Kings College Cambridge , 19-21 September 2007. Available on line at *http://www.bioecon.ucl.ac.uk/10chapters9.htm*.

[56] Dedeurwaerdere, T., Iglesias, M., Weiland, S., Halewood, M., 2009. Use and exchange of microbial genetic resources relevant for food and agriculture. CGRFA Background Study Chapter No. 46. Commission on Genetic Resources for Food and Agriculture, Rome. Available on line at *http://www.fao.org/nr/cgrfa/cgrfa-back/*.

There is a great deal of material deposited in the World Federation of Culture Collections (WFCC) collections that comes from in house laboratory and university research collections. It is expensive to maintain the strains. The informal research collections do not have the money to do that, so typically upon publication they will put the materials in a WFCC collection. However, there is also a lot of exchange between WFCC collections. The survey found that on average 20 percent of the strains acquired in 2005 by 119 WFCC culture collections came from other WFCC culture collections, and 10 percent of distributed materials went from WFCC collections to other WFCC culture collections. Those are the high-value and unique research materials. This is how a collection fills in its own gaps in order to have all the type strains and remain up to date. This "conditional reciprocity" between the culture collections—I can order unique materials at other WFCC collections because I also distribute my own unique materials under open access conditions—is very strong, and that is an important fact to take into account.

Regarding digital information, our study of open access publishing in the field of microbiology found that about 30 percent of the academic literature is in full open access journals[57]. However, that figure includes hybrid access, which means that the information is available both through purchased open access and by subscription. The prices to purchase open access can be quite high, so full open access is still not prevalent in this field.

What are the implications for governance principles that one can draw from these surveys? I offer three.

The first is that you will need a governance framework driven at least in part by the scientific community. It should not be driven by a government entity only. The main reason for this is that most decisions on governance require deep knowledge of the technical specifics of the field. Regarding issues like prior informed consent or quality management, although these do have a regulatory component, the decisions require thorough knowledge of the scientific aspects.

The second principle is the need for multi-level governance. This requirement arises primarily because of the extreme heterogeneity between the collections and the various research environments. Some of the collections produce international public goods. The World Health Organization's network of microbial laboratories supporting research on H1N1 is a case of a global public good. Others produce just regional or trans-regional public goods, that is, they operate as knowledge hubs that are very strongly integrated with a local or regional microeconomic environment. We saw an example of this in the dairy industry in Georgia, where the goal was to have new starter cultures for the national yogurt industry. They did the genetic sequencing locally and then looked in GenBank to compare the sequences. So they accessed the global infrastructure, but overall the sharing of microbial material occurred on a regional basis. That is quite common in this field.

The third principle is the need for specialization and cooperation. To understand this need, it is sufficient to recall that, on average, 40 percent of the strains in each WFCC collection are unique. Thus if you want to solve the stem rust disease discussed above,

[57] Reichman, J.H., T. Dedeurwaerdere, and P. F. Uhlir. Designing the microbial research commons: Global intellectual property strategies for accessing and using essential public knowledge assets (Cambridge Univ. Press, forthcoming 2013).

you probably will need access to pathogenic strains from many places, and some unique strains will probably be in collections far away. Thus we need specialization and cooperation.

The implementation of these principles needs to be articulated to the regulatory frameworks developed at the international level. Therefore, there are some fora where representation of the microbial science community is a key issue. For instance, the discussion on access to knowledge going on at the World Intellectual Property Organization will be crucial for building common ground between developing and developed nations on open-access infrastructures. There are also the discussions concerning access and benefit sharing for microbial materials taking place in the Convention on Biological Diversity and in the FAO's Commission on Genetic Resources for Food and Agriculture, which have been addressed in other presentations at this symposium.

Here the key message is that there are limits to a voluntary scheme such as a microbial commons for an international access and benefit-sharing regime. Even if the microbial research commons would contribute to access and benefit sharing through a standard material transfer agreement and a compensatory liability scheme, it remains a voluntary regime, so people can always decide whether to join the research commons or, if they have microbial materials with a very high commercial potential, to go in a different direction and step out of the commons. Major contributions for addressing these problems can be expected from international agreements between competent science ministries that oversee the collection on measures that provide for a standardized solution to benefit-sharing with the original providers of the strains to culture collections and support further standardization of the license conditions used in the various MTAs.

These and other considerations lead us to think about the possibility of further formalizing the informal exchange practices in the microbial commons and developing a sound legal and institutional framework for the operations of the collections. This in turn can support the further development of a fully digitally integrated research infrastructure building upon and extending the emerging global initiatives in the microbial commons.

25. Institutional Design and Governance in the Microbial Research Commons – Charlotte Hess[58]

Syracuse University

I was invited to talk about institutional and governance issues in microbiological research commons. While the microbial commons is a new type of commons and a part of the larger knowledge commons sector, I would first like to situate it within the study of the traditional national-resource commons. The challenges in constructing a viable commons are:

1. Understanding the nature of the resource and the users;
2. Dealing with the complexity of new—and especially global—commons; and
3. Managing possible fragility and threats.

The microbial commons is an example of a dynamic, international, *new commons*. While this commons encompasses *in vivo*, *in vitro,* and *in silico* resources, the focus here will be on the latter: the microbial commons in digital format.

Commons are about the relationship between a resource and human institutions or rules. The salient question in the governance of any commons is: how can fallible, heterogeneous individuals come together with incomplete information to make rules and decisions in order to effectively manage and sustain a resource? The analysis of this question about effective sharing requires an interdisciplinary approach that combines law with biological, economic, and other social sciences.

The first literature that resembles the approach that commons scholars take today was produced in the 1950s and applied the field of economics to the fields of biology and fisheries[59]. The concerted study of commons, however, did not really take off until the late 1980s with the organization of the International Association for the Study of Common Property (IASCP).[60] So you can see that this is a relatively new area of study.

Many of the early IASCP studies were focused on either demonstrating the accuracy or refuting Hardin's thesis of the tragedy of the commons.[61] Other predominant foci continue to be:

1. The threat of enclosure and the lessons of the historical enclosure movements in Europe;
2. The relationship between formal and informal property rights and the health of the resource; and
3. How different types of social dilemmas, such as free riding or non-compliance, lack of trust, competition, or secrecy, affect the outcomes.

---

[58] Presentation slides available at: *http://sites.nationalacademies.org/xpedio/idcplg?IdcService=GET_FILE&dDocName=PGA_054555&RevisionSelectionMethod=Latest.*

[59] See Gordon, H. Scott. 1954. "The Economic Theory of a Common-Property Resource: The Fishery*,"* 62 *Journal of Political Economy*; and Anthony D. Scott. 1955 "The Fishery: The Objectives of Sole Ownership*."* 63 *Journal of Political Economy*.

[60] The name was changed to the International Association for the Study of the Commons (IASC) in 2006. See *http://www.iasc-commons.org/*.

[61] Hardin, G. 1968. "The Tragedy of the Commons." *Science,* Dec. 13, at 1243.

The growing international commons literature on traditional, natural-resource commons allowed scholars to more deeply analyze how commons work and to better understand why they fail. For Elinor Ostrom, it led to her seminal book, *Governing the Commons* in1990. Ostrom applied a complex set of instruments to eighty-six case studies to commons of different sectors and varying geographical regions. From her analysis, she was able to determine eight design principles that long-enduring, robust commons shared.

The Ostrom design principles[62] are:

1. Group boundaries are clearly defined;
2. Rules governing the use of collective goods are well matched to local needs and conditions;
3. Most individuals affected by these rules can participate in modifying the rules;
4. The rights of community members to devise their own rules is respected by external authorities;
5. A system for monitoring member's behavior exists; the community members themselves undertake this monitoring;
6. A graduated system of sanctions is used;
7. Community members have access to low-cost conflict resolution mechanisms;
8. For CPRs that are parts of larger systems: appropriation, provision, monitoring, enforcement, conflict resolution, and governance activities are organized in multiple layers of nested enterprises.

The Ostrom design principles are considered today as useful tools by many scholars in commons study. All the commons in the study, however, were managed by relatively small, homogenous groups. We do not know if these principles scale up nor do we know how the design principles would apply to the microbial commons. We might be able to use some of the principles as a place to start, although certain principles—such as group boundaries being clearly defined—may be harder to apply. Other principles, such as the importance of monitoring mechanisms, may take on even greater importance.

Interest in *new commons*, for the most part, emerged after the World Wide Web had gained ubiquity in the mid-1990s. They tend to have several characteristics that distinguish them from traditional natural-resource commons. Many are human-made resources, such as open source software, the Internet, and scientific research commons. Or they are resources that have been newly recognized as commons, such as urban landscapes, parking spaces, parks, and even garbage dumps. Many new commons have arisen out of the development of new technologies or the growth of new communities. Unlike traditional natural resource commons, new commons tend to be dynamic, quite complex, and heterogeneous. Many are global in scale and have fuzzy boundaries. There is a great deal that we do not yet know about new commons, particularly how they work and if they can be sustained.

Figure 25–1 is a map of new commons[63] based on the emerging literature of new commons sectors. As one can see, the knowledge commons is quite dominant and takes on many forms.

---

[62] See Ostrom, E. 1990. *Governing the Commons: The Evolution of Collective Action*. Cambridge University Press.

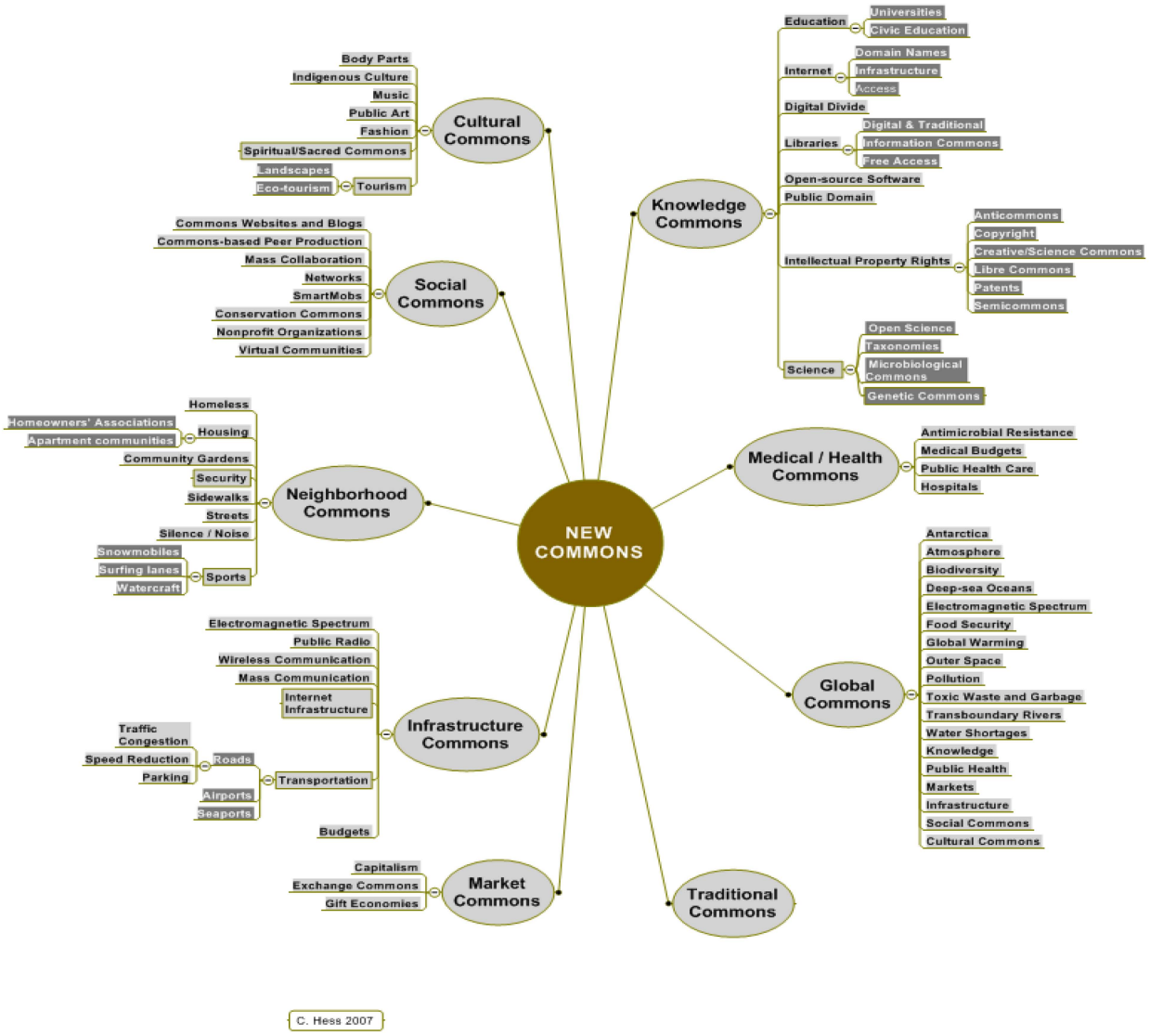

FIGURE 25–1 A map of identified new commons and new commons issues. SOURCE: Ostrom, E. 1990. *Governing the Commons: The Evolution of Collective Action*. Cambridge University Press.

Not only have the number and types of commons expanded over the past 25 years, but the way that we think about commons has changed considerably as we have learned more about them. In the 1980s, the term "common property" was the preferred term applied to commons institutions. After many case studies and a literature began to be built, researchers found that commons could exist in all kinds of property regimes. In Africa, for instance, there are many community forests or commons that are managed and used collectively, but privately owned.

Elinor Ostrom argued the importance of distinguishing between the resource and the regime—the regime being the property rights, and the resource being a type of

[63]In Hess, C. 2008. "Mapping New Commons." *http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1356835* .

economic good: a common-pool resource. The oft-used schema of the four types of goods: public goods, private goods, toll goods and common-pool resources was, in fact, developed by Ostrom and Ostrom in the 1977 paper "Public Goods and Public Choices."[64]

With the rise of new commons, there are now a number of legitimate scholars who are using the word "commons" without defining it as either a property regime or as an economic good. In fact, whether in new or traditional commons literature, the word "commons" is rarely ever defined. In the new commons literature, the word "commons" is almost always the preferred term (rather than common-pool resources or common property). As an emerging area of study, much of the literature is aimed at identifying a particular resource as a type of commons. Only in the legal literature do we find rich and multilayered studies, particularly of the knowledge and cultural commons[65]. Ostrom and I discussed the lack of any clear definition of "commons" at length while working on our book *Understanding Knowledge as a Commons.* (2007, MIT Press). We decided to attempt a definition and settled on the following: *A commons is a resource that is shared by a group of people that is subject to social dilemmas.* In further study, I have found that new commons almost always carry with them an element of vulnerability. Resources shared in commons are vulnerable to threats of various types of enclosure and capture (Hess fn. 6). And this vulnerability creates an ever-present need for monitoring and protection.

We do not know a great deal about new commons nor how to govern them, and we know much less about global commons. One important thing we do know about most global commons is that they are also local, either in creation or in implementation. Unlike with traditional commons, we can much more easily study how new commons come into existence. We have learned, for instance, that the considerable attention to the knowledge commons has arisen because of the collective witnessing of enclosure or threats of enclosure of open knowledge as a public good. New capabilities of information technology allow the capture (enclosure, privatization) of data and information that was previously "uncapturable." In other words, the commons is created by the enclosure or threat of enclosure of a public good.

Ostrom has more than once pointed out that the governance of a commons is really hard work. Since they tend to be self-governing and participatory, some types of a social dilemma are inevitable. Commons governance, therefore, requires ongoing attention, persistent effort, mindful adjustment of rules, and adaptation to new situations. As Vincent Ostrom has often noted, members of a commons are "artisans" who "craft" appropriate institutions. Finally, communication is essential in order to build trust and reciprocity.

Figure 25–2 shows a useful framework that facilitates a better understanding of the commons concept.

---

[64]Ostrom, Vincent, and Elinor Ostrom 1977. "Public Goods and Public Choices." In *Alternatives for Delivering Public Services; Toward Improved Performance*. E. S. Savas, ed. Boulder, CO: Westview. Reprinted 1999 in *Polycentricity and Local Public Economies: Readings from the Workshop in Political Theory and Policy Analysis*. M. D. McGinnis, ed. Ann Arbor: University of Michigan Press. (Institutional Analysis). Online (on 4-1-11) at *http://theworldbuilders.witesman.com/v372/Ostrom%20public%20goods%20and%20public%20choices.pdf*

[65] I am referring to the work of Yochai Benkler, James Boyle, Brigham Daniels, Brett Frischmann, Michael Heller, Mark Lemley, Lawrence Lessig, Jerome Reichman, Carol Rose, and others.

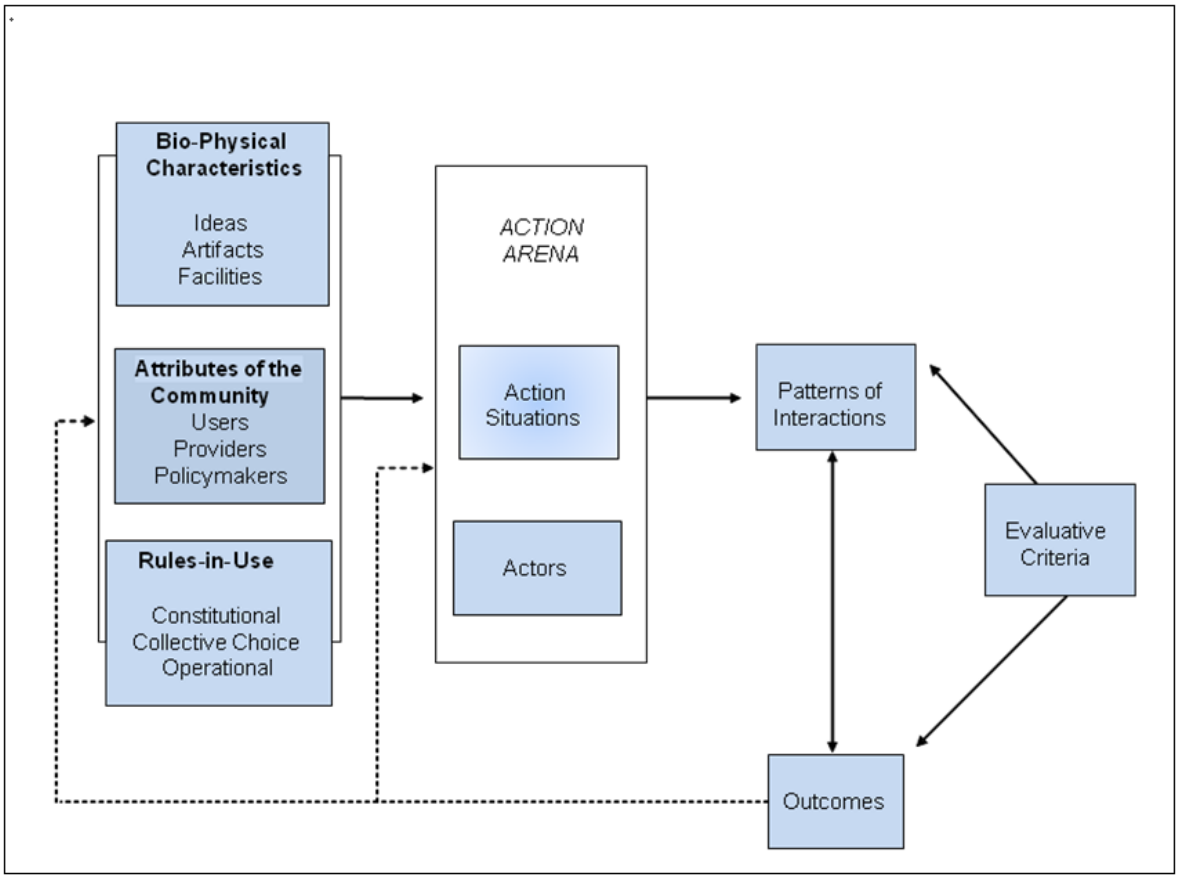

FIGURE 25–2 Institutional Analysis and Development (IAD) framework.
SOURCE: Adapted from Ostrom, Elinor, Roy Gardner, and James M. Walker 1994. *Rules, Games, and Common-Pool Resources*. University of Michigan Press.

The left part contains various exogenous characteristics, including the biophysical characteristics, the attributes of community, and the rules in use. In a microbial commons those are important for everybody to understand clearly: What exactly is the (common) resource? Is there really one microbial commons, or are there many? The culture collections will have different attributes than the digital information databases, and those will have different attributes than proprietary literature databases, which will have different attributes than the secondary literature. Who are the information users and who are the providers? What kind of rules (including laws, norms, etc.) are in place? Do they work? Are they appropriate?

The middle section, the action arena, concerns specific actions taken, how people interact and what they do. This is the area on which game theorists and modelers are usually focused.

On the right side, the *patterns of interaction* are the institutional reactions of the action arena. The *outcome* is the current state of the resource and/or the resource users. It is the area of analysis that many researchers start with. For instance, why is this land constantly degraded while a parcel of land 50 miles away is doing really well? Why do some science collaboratories thrive while others run into problems of conflict and noncompliance? The researchers start with an outcome and trace it back through the framework.

Figure 25–3 shows an adaptation of new framework developed by Ostrom and a group at Arizona State University to analyze the robustness of complex social-ecological systems. It focuses on the institutional configurations that affect the interactions of resource users and resource systems. One of the purposes of this framework is to help researchers look closely at the individual system at hand and not defer to "blueprint solutions."

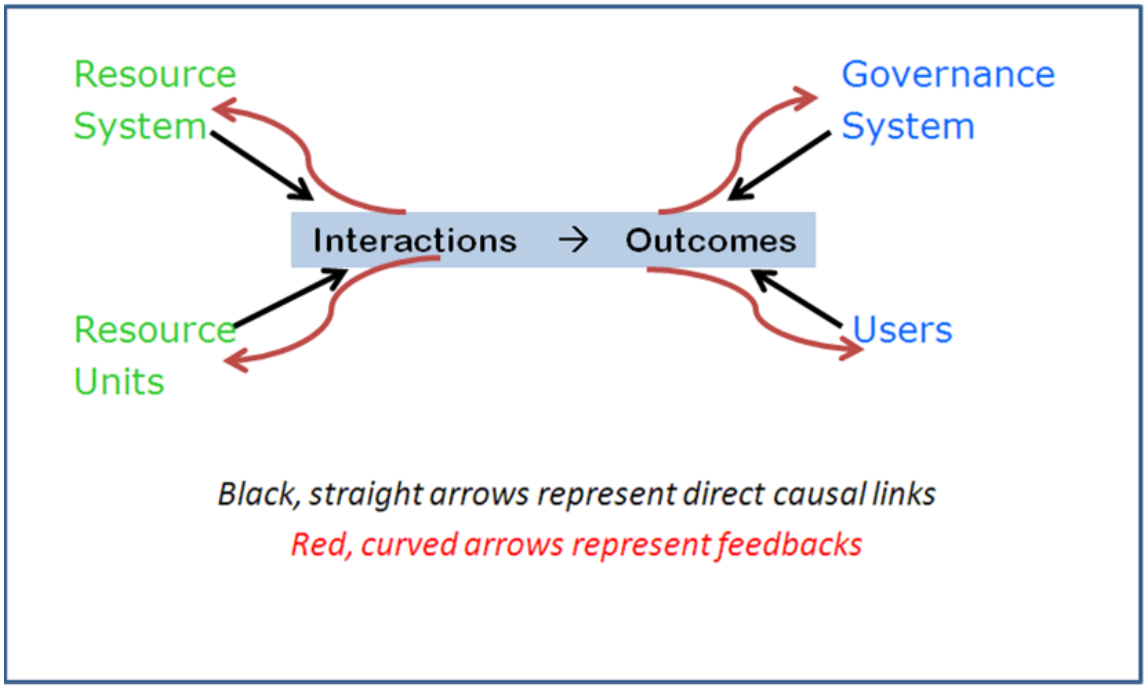

FIGURE 25–3 Diagnostic tool for analyzing a social-ecological system (SES)[66]

In either case, it is always essential to identify and understand the physical nature of the resource at hand. From that perspective the microbial commons shares many characteristics with other digital scholarly or scientific information. It is still surprising to many how extremely fragile digital information is. Important *e*-information is being lost every day in many ways. Scientific information has been withdrawn by the U.S. Government under the Patriot Act. Where publishers have gone out of business, access to once available files can be closed off. Primary genomic data is being rapidly patented and therefore cordoned off to most future research. Digital information is also being lost due to inattention, lack of robust preservation strategies, underfunding, or people simply do not know what to do with it obsolete formats. Until the recent NSF and NIH mandates, many universities were not taking responsibility in the storage of massive datasets developed by their scholarly community. In the global south, there is also unequal access to digital information because of a lack of technology, a lack of infrastructure, a lack of

[66] SOURCE: Ostrom, E. 2007. "Sustainable social-ecological systems: An impossibility?" *http://www.indiana.edu/~workshop/publications/materials/conference_papers/W07-2_Ostrom_DLC.pdf*. and Anderies, John M., Marco A. Janssen, and Elinor Ostrom 2004. "A Framework to Analyze the Robustness of Social-Ecological Systems from an Institutional Perspective." *Ecology and Society* 9(1). *http://www.ecologyandsociety.org/vol9/iss1/art18/.*

electricity, or a variety of other reasons. In other words, scientific information in general, is still quite siloed in the digital era.

As we have heard from others today, the amount of scientific information is growing exponentially, and it is getting harder to collect it, to preserve it, and to store it. I say this not as a scientist, but as an information professional who is trying to keep track of the scholarly and scientific information that a single university is generating. It is very challenging. Many predict that there will be dramatic changes to how science is done and that few traditional processes will survive in their current form by 2020.

Considering the microbial commons in the academy, we are all aware of the conflict between the desire to open up information and make it accessible and the increasing mandates on the university to monetize or commoditize information, coupled with the growing influence and power of the universities' technology transfer offices.

Another problem is the high transaction costs and lack of strong incentives for university scientists to annotate an organism's genome in collaborative information repositories.[67] The prevailing system, described by Syracuse University faculty, is outmoded, inefficient, circuitous, and does not count toward tenure. There is a clear time lag and disconnect between current practices in digital scholarship, whether it be genomic annotation or experiments with new media. Official recognition and clear rewards need to be built into university tenure and promotion structures.

Some interesting work is being done by forest researcher Charles Schweik at UMass-Amherst on cooperation in open source communities. He unpacks the traditional theories of collective action to show how people will cooperate online in nontraditional and unprecedented ways. He has found that in environmental commons norms, rules, and governance structures often help to overcome tragedies. His research suggests that too much governance structure and rules may get in the way of collaboration.[68]

With regard to scholarly communication, one thing that has not been mentioned here is the changing role of the university library (which is relevant to how we access scientific research). Traditionally, the mission of academic libraries was to collect, organize, disseminate, and preserve the cultural and scholarly record. However, with each passing year in the digital environment, libraries are moving farther away from that mission because the massive amount of digital scholarly in multiple formats that is being generated on campuses today is not what is being collected by research libraries. Libraries are still focusing only on the published record which, because of ever-declining budgets, is a decreasing percentage of the whole. It is a huge problem, and in addressing it is important to think about preexisting infrastructures. One should, for instance, use library expertise when it makes sense in terms of organization and archiving and build library–academic departmental collaboration for funding open access. I would also suggest making sure that the tenure and promotion process take reputation, global networks and research distribution into account. Some universities, such as Harvard and MIT, have already passed open access mandates, but most universities are not yet addressing that. The bottom line here is that academic libraries face huge challenges and need to work more fluidly with researchers. University scientists should support them

[67] See Welch, R. and L. Welch. 2006. "If You Build, They May Come." *Nature Reviews Microbiology* 7, 90 (February 2009) | doi:10.1038/nrmicro2086, examining why researchers seem reluctant to be more directly involved in the annotation of microbial genomes.

[68] See Schweik Open Source Project with links to articles at *http://www.umass.edu/digitalcenter/ossuccess/*.

with their departments because I do not know if they are going to be sustainable otherwise. This is a very critical issue.

What do we know about what works? There are several different models that are working in the new commons environment. We need to study them. If nothing else, we can use the time-worn measures of equity, efficiency, and sustainability in guiding what we are building.

In conclusion, we need to do more outreach to build greater awareness of the commons and of open access. We need to document the steps that we are taking and the lessons we are learning in the course of building the microbial commons, because not much is known about this process. We need to translate our knowledge to a wider, nonscientific community.

I would like to encourage you to join the International Association for the Study of the Commons, form a panel, and present your work at the upcoming 2011 conference in Hyderabad, India.

Question and Answer Session

PARTICIPANT: There were many rich points in your presentation, but I was particularly interested in your plea at the very end for the integration of libraries in the knowledge production process. This is a point we are trying to make in developing the open knowledge environments—that there is a fantastic role for libraries there. Do I understand you in supporting that?

DR. HESS: Absolutely.

PARTICIPANT: That was one of our objectives, and that is why we are thinking that there is a large role for the universities. It has so many opportunities for integrating knowledge that cannot be done anywhere else, so it would seem to be a way of transforming the way universities organize the production of knowledge.

DR. HESS: I fully agree. It is not just going to happen on the academic side, but also on the side of the libraries. Some of these are really not in tune with what you are doing and not really focusing on gray literature or on databases that are not in their library. There are so many aspects where that coordination needs to happen on both ends. So absolutely, you are on target as far as that goes.

26. International Developments: A Context for the Creation of a Microbiology Commons – Anita Eisenstadt[69]

Department of State

I will first lay out a couple of key principles for engaging in international cooperation. Then I will talk about some of the challenges that are particularly associated with biological data. And finally, I will describe the past and present work at the Organisation for Economic Co-operation and Development (OECD). This is an international organization where it may be possible to carry out some of the work we have been discussing.

Let me begin with some basic observations. Science, technology, and innovation are accelerated by international cooperation. Science today is absolutely global. I recently attended an event at the Finnish Embassy here in Washington, DC. Finland has developed a new technology award, the Millennium Technology Award, which is their attempt to do the same thing for technology that the Nobel prizes from Sweden do for different areas of science. Some amazing people have won the awards—the researcher who developed the World Wide Web, the scientist who developed the technology for light-emitting diodes, and so forth. The Finnish Embassy had some really amazing speakers at this event.

The message that I brought home is that science is global. We have key global challenges that we must address today and together internationally. No one country has the resources or the solutions by itself. Science, technology, and innovation are all key to us resolving these global challenges. I think we all agree that the life sciences will play a major role in resolving these global challenges and that they are a really important future research area.

The Obama Administration has been incredibly supportive of science, and one of the early speeches that President Obama gave was here at the National Academy of Sciences. The fact that he chose this place spoke volumes. In his speech he emphasized the importance that science plays in addressing issues in society, such as our economic well-being, and stressed how global and international science needs to be. His administration is also dedicated to ensuring openness and transparency in government, including access to scientific data and information.

The U.S. government has a long history of promoting access to federally funded research. The Office of Management and Budget Circular A130 and the Paperwork Reduction Act have been part of the legal and policy foundations for that policy. A study came out in January 2009 called *Harnessing the Power of Digital Data for Science and Society* which was put together by the National Science and Technology Council's Interagency Working Group on Digital Data. I am the State Department representative on that Working Group, which represents a large number of science agencies. That report contains this quote:

[69] Presentation slides available at *http://sites.nationalacademies.org/xpedio/idcplg?IdcService=GET_FILE&dDocName=PGA_053674&RevisionSelectionMethod=Latest.*

> The ability to achieve innovation in a competitive global information society hinges on the capability to swiftly and reliably find, understand, share, and apply complex information from widely distributed sources for discovery, progress, and productivity. Limits on information access translate into limits on all other aspects of competitiveness. Thus, digital information preservation and access capability are critical to the progress of individuals, nations, science and society.

This states very succinctly the issues that we are dealing with here at this symposium and captures how incredibly important it is to have access to digital information.

There are a variety of challenges associated with biological (and other types of) data. First, we have unprecedented amounts of such data. Given all this raw data, how do we enhance access to it? How do we organize it so that researchers working in one discipline can have access to data from a totally different discipline in a different lab? Can we find a way to accelerate understanding and knowledge to facilitate advances in important fields, such as biotechnology, health, agriculture, environmental remediation, and sustainable biofuels?

We need a framework with which to compare and combine experimental data collected in different labs so that we can get a fuller understanding of the identity, structure, and biological functions. Science is moving away from looking at small, individual items to looking at the larger system—the systems approach to science.

Someone who has collected data for one reason may be shocked at how that data can be used by someone else in a totally different field. The founder of the World Wide Web was at the Finnish Embassy this week. His key message was that it is astounding what other people will do with your data, based on just a few days of you putting the data out there. He urged people not to be so cautious about trying to make their data available in a perfect form on a fancy Web page; even if you do not have time to format the data in a certain way, once you get the data out somebody else might be able to take that next step.

One problem with this, of course, is that if you get the data out there, but there is no way to organize it or access it, and there is no shared vocabulary, people will find it difficult to use the data. To address these issues, we probably need some enhanced analysis methods for large databases and we need to pay more attention to data interoperability and compatibility.

There are also a variety of legal implications to sharing data, as has been already discussed here. These tend to be more complex when human data are being shared, and they raise questions in the areas of intellectual property, privacy, and dual use.

For the rest of my talk I will focus on the OECD. It is headquartered in Paris, has 30 member nations, and was established in 1961. It came out of the Marshall Plan at the end of World War II, and the United States was one of founding members. The OECD brings members together to support sustainable economic growth and maintain financial stability. The idea behind the OECD is to provide a forum in which to compare policy experiences, identify good practices and guidelines, and coordinate national and international policies.

It is a very large organization and it would thus be impossible for any one person to keep track of everything that goes on there. Much of what goes on at the OECD is trade-related, or finance-related, but there is a part of the OECD that deals with research

issues and science policy. This part includes the Committee on Science and Technology Policy and also a Working Party on Biotechnology.

A number of years ago at the OECD, we decided it would be helpful to develop guidelines to promote access to publicly funded research data. Paul Uhlir was one of our experts on that working group. A large part of the goal of our working group was to try to get other countries to adopt the U.S. approach, because the United States has one of the strongest records of providing access to federal and federally funded research information.

The European Union has a very different approach from ours because of the Directive on the legal protection of databases that they have. Various other countries sometimes have very different approaches as well. So we were pleased to come up with a set of recommended guidelines, released in December 2006,[70] which promoted access to publicly funded research data at little or no cost.

A second OECD recommendation, this one released in 2008 in preparation for a ministerial on the future of the Internet, offered ways to enhance access to public sector information.[71] If you are looking for any kind of guidelines in these areas, you might find these two documents of help.

Why did we decide to develop these research guidelines? Because we really believe that the exchange of data, knowledge, and ideas are fundamental to progress. This is particularly true now because the Internet has opened up new applications for research data that were never available before, and you must access to the research data in order to take full advantage of the opportunities afforded by the Internet. Another reason was that access to data increases the return from public investment and reinforces open scientific inquiry.

We also wanted to encourage governments around the world to address the issues underlying access to data in their national policies, because some of them had not yet done so. We did not find in other countries many of the things that the United States had done concerning grants, such as the guidelines from the National Science Foundation and the National Institutes of Health that encourage sharing of data. They were hoping that those ideas would be shared and that maybe some of them would be adopted. And we were hoping to enhance international data sharing.

A second project at the OECD—and one that is closer to what we are talking about here today—was the development of a set of best practices for biological research centers. These centers are repositories and providers of high-quality biological materials, and there was a concern that there was not enough sharing of these biological materials as well as a concern that there was insufficient control over quality.

Some of the countries were interested in having a set of best practices so that when people obtained a sample, there could be some assurances that it had been collected and maintained in accordance with those established practices. The definition for biological resources included living organisms and all the other materials needed for the advancement of biotechnology and human health research, including microbiological materials.

---

[70] Organisation for Economic Co-operation and Development. OECD Principles and Guidelines for Access to Research Data from Public Funding. 2007; *http://www.oecd.org/dataoecd/9/61/38500813.pdf.*

[71] Organisation for Economic Co-operation and Development. OECD RECOMMENDATION OF THE COUNCIL FOR ENHANCED ACCESS AND MORE EFFECTIVE USE OF PUBLIC SECTOR INFORMATION. 2008; *http://www.oecd.org/dataoecd/0/27/40826024.pdf* .

We came up with four sets of best practices. The first one focused on generic quality aspects that could apply to labs that were collecting and maintaining any type of biological materials. The second set of guidelines concerned biosecurity practices. We understand that these have been adopted by a number of countries since they were developed. These first two sets apply to all types of biological resource centers (BRCs). The third set was specific to BRCs holding and supplying microorganisms, and it is probably the one most relevant to your discussions here. Finally the fourth set was for BRCs holding and supplying human-derived materials.

The guidelines set forth methodologies for preservation, replication, and quality control, and they attempted to provide internationally unified quality control. The recommendations include such things as minimum data sets—that is, the minimum amount of information that should be collected for the samples—and various other practices aimed at insuring good quality.

One thing that came out of this discussion was that a number of countries were very interested in forming a global biological research centers network, or GBRCN. The idea was to develop an international mechanism for cooperating and linking these various biological centers and for sharing data. Although there has been no consensus at the OECD on a specific recommendation, the GBRCN is now a pilot project.

The GBRCN is headquartered in Germany. It has 15 countries participating in it, and the World Federation of Culture Collections is a major player as well. The goal of this pilot project is to look at whether it is possible to create networks that are more valuable than the status quo. If so, this pilot may end up being a model for a broader network.

Another issue that has been discussed is certification of BRCs. There are differing views on this. Some countries want the OECD to put in place a certification process that would review the practices of a BRC and provide a certificate if it met the requirements of the best practices. The Unites States was not comfortable with that approach, and so at this time it is optional whether a country chooses to have a national certification process for BRCs, as opposed to having BRCs accredited by private third parties. France and Germany, by the way, were quite interested in national accreditations systems.

We heard some discussion this morning about knowledge markets. In October 2008 the OECD Working Party on Biotechnology held a workshop here at the National Academies to examine the different practices for knowledge market mechanisms and to try to identify the knowledge markets in the biomedical sector.

Knowledge markets can include both proprietary intellectual property and goods in the public domain. They are intended to facilitate the sharing of intellectual assets, including data, materials, expertise, and services. The report that came out of the workshop is very useful. It provides details about knowledge markets and offers examples of where they have worked.

One of the key things we are looking at now at the OECD is putting together a topology of knowledge markets. We are also thinking about doing further work in this area. One thing that I would be most grateful for is if any of you have suggestions on what we might do to move forward in the area of knowledge markets related to biotechnology.

At the moment, OECD is putting great effort into understanding exactly what innovation is and how countries can best promote it. The work we are doing on knowledge markets will be incorporated into next year's ministerial on an innovation

strategy, so this work could get quite a bit of attention if we do some sound work over the next year.

My final example is a current effort by the OECD Committee on Science and Technology Policy to examine multilateral scientific cooperation and look at new approaches and governance mechanisms for multilateral cooperation in international science. We have a steering group meeting later this month. It will be a two-year project, and we will be looking for proposals on ways to set forth better frameworks for promoting international science and technology cooperation. This is another area in which, if the participants in this symposium have ideas you would like me to carry forward, I would be really open to hearing your suggestions. We are at a pivotal point right now in steering that group in directions that we think would be useful.

Question and Answer Session

PARTICIPANT: There have been a number of countries that did not have any data policies, which have sought to implement national policies since the OECD guidelines were written—notably South Africa, Chile, and some other countries that recently joined OECD. So these things do matter.

MS. EISENSTADT: Yes, and the OECD is doing a lot of work now with developing countries as well. So some of the guidelines are not only going to be used by member countries but will be very useful examples for developing countries as well.

PARTICIPANT: In addition to the groups that you mention in OECD, there is also a working group on harmonization of regulation of biotechnology and a microorganisms sub-working group. One of the projects they have been dealing with is the development of a unique identifier for certain kinds of microorganisms used in biotechnology. OECD is looking for feedback on the kind of proposal it had in mind.

PARTICIPANT: My experience in directing an OECD effort for neuroinformatics led me to understand that OECD basically anoints or allows a group to commission an effort—a community group—and most often steps back and just encourages member nations to participate in that community effort. I am glad to hear about the kind of thing you spoke about because it says that the OECD is actually getting involved in helping to promote standards that allow for interoperability, which it particularly important.

Our experience with 15 member nations was that there was, of course, reluctance to participate. Even though it was something that they were encouraged to do through their country's membership, they did not actually understand how to participate because there were few guidelines about data sharing. Even if they wanted to share data—this was in brain research, which spans the range from molecular and genomic on up to information that is clinically relevant—they did not know how to operate with regard to identifying or making useful the information that was put away in data collections.

So when you talk about the Berners-Lee World Wide Web activity and being able to find things on the Web, I think what people are missing is the real challenge of what we call the long tail of small data. That is the data that we get together and figure out that we can put in collections, whether they are physical collections or data collections. How do we make sure they are well-curated, so they can be more easily discovered? We are really dealing with a lot of data that are not discoverable because people do not organize

them in any way that makes them easy to find. It would be great if the OECD would organize an effort with the information technologists—the knowledge management engineers, if you like—to help make that easier for the inexperienced scientists who are dealing with this problem.

27. Options for Governing the Microbial Commons – Michael Halewood[72]

Bioversity International, Italy

Bioversity International is one of 15 International Agricultural Research Centers supported by the Consultative Group on International Agriculture Research, often referred to as the CGIAR Centers. Most of my work involves assessing the impact of policies and policy-making processes on the use and conservation of genetic resources for food agriculture. In recent years, I have also dedicated a considerable amount of time to coordinating the representation of the CGIAR centers, as observers, in the negotiations of international access and benefit sharing agreements. The CGIAR centers' main concern has been with plant genetic resources, though lately we have become increasingly interested in contributing to the development of access and benefit sharing norms that support the use and conservation of other genetic resources for food and agriculture, including microbial genetic resources used in agriculture production systems and plant pathology research. In particular, we are interested in promoting the development of an internationally coordinated system for the common pooling, management and use of agricultural microbial genetic resources. We see such a system (or systems) as an essential supportive component of the agricultural research and development continuum. It is for this reason that we are very pleased to participate in this meeting focusing on the development of the microbial commons.

In this presentation, I am going to focus on challenges associated with populating the microbial commons, on an ongoing, dynamic basis, with previously unavailable, microbial genetic diversity. I am going to focus in particular on challenges associated with access and benefit sharing, and the impact that the Convention on Biological Diversity (CBD) has (or has not) had to date with respect to those challenges. Ultimately, I will argue that intergovernmental participation in developing international norms and administrative mechanisms to support the microbial commons will be essential to overcoming these challenges. I will also identify opportunities for champions of the microbial commons to engage in ongoing international policy-making processes in pursuit of the necessary policy support.

Jerome Reichman already highlighted a number of the most essential components of the microbial commons. Among other things, he stressed the importance of establishing, up-front, access and benefit-sharing terms that are reflected in a single, standardized material transfer agreement that would accompany all transfers of materials in the commons. This is an essential aspect of the commons that would contribute to lowering transaction costs associated with using microbial genetic resources in agricultural research. Furthermore, the benefit sharing terms that are ultimately agreed-upon could encourage would-be depositors to overcome reservations they may have about depositing materials in publically-accessible culture collections. My presentation addresses access and benefit sharing related issues that will need to be resolved before it is reasonable to expect wide-spread adoption and use of such a standardized Materials Transfer Agreement (MTA) in the context of the microbial commons.

[72] Presentation slides available at *http://sites.nationalacademies.org/xpedio/idcplg?IdcService=GET_FILE&dDocName=PGA_054720&RevisionSelectionMethod=Latest.*

The microbial commons will need to encompass both genetic resources in culture collections, and materials "in the field" that have never been collected (much less identified). Today, I will focus primarily on materials in collections. Partly because shortness of time limits my ability to focus on all aspects of the microbial commons, and because the culture collections are already so wide-spread and functioning as essential elements of nationally and internationally supported research and development systems. They will clearly play a key role in the commons, authenticating, maintaining and distributing strains. There are approximately 550 culture collections listed under the World Data Center for Micro-organisms, most of them hosted by public and semi-public organizations and universities. They hold approximately 1.5 million strains. Each year, these listed collections distribute approximately 500,000 isolates. Tom Dedeurwaerdere estimates that a considerably higher number of isolates are exchanged each year informally, without legal agreements, as part of informal networks through peer-to-peer exchanges. Tom has also estimated, based on a survey of a number of genebanks, that approximately 50 percent of the materials in culture collections were acquired by those collections prior to 1993, the year the Convention on Biological Diversity came into force.

The year 1993 and the coming into force of the CBD is a very significant date to bear in mind when thinking about promoting a microbial commons, especially when thinking about the status or role of material in microbial collections. Stated bluntly, a culture collection (or a country hosting a culture collection) is free to determine what it wants to do with microbial genetic resources it has acquired from other countries prior to the CBD being implemented in those countries. It can decide to adopt and use an MTA like that described this morning by Jerome Reichman when distributing microbial genetic resources, and in so doing, voluntarily subscribe to the "rules of the game" for the microbial commons. The fact that collections have the possibility treating up to 50 percent of the materials they hold in this way is good news for the microbial commons; it provides a substantial basis upon which to found the commons.

The situation is very different for materials acquired after the CBD came into force (and after it has been implemented by countries). The practical consequence of the implementation of the CBD's Article 15 is that since 1993—assuming countries have implemented the CBD—acquiring new genetic material requires first getting prior informed consent on mutually agreed-upon terms from competent authorities in the countries of origin of that material. If prior informed consent from the competent national authority is not obtained, the collector of the culture collection cannot take the material out of the country. Nor can a research scientist voluntarily deposit such material in a collection outside the country concerned without the requisite permission.

Consider the implications of this with respect to materials acquired by culture collections between 1993 and the present. The culture collections concerned would have to have the prior informed consent from the competent authority of the country of origin to distribute such material using the microbial commons-inspired MTA. Unless it was mutually agreed when the material was deposited that that the culture collection had the right to later change the MTA it uses to distribute material, the collection would have to go back to the competent authority from the depositing country to obtain permission to use the new instrument.

As far as future acquisitions by culture collections are concerned, it will be necessary to obtain prior informed consent from competent authorities in the countries of origin of microbial genetic resources to redistribute those resources using the commons-

inspired MTA. This is a particularly important consideration given that 99 percent of microbial diversity currently exists in *in-situ* conditions behind national borders. Over time, to maintain its relevance, an increasing proportion of the material in the microbial commons will have to be materials accessed after 1993, from *in-situ* conditions.

Before proceeding further, I would like to comment on a closely related issue. There seems to be confusion in some of the literature written about the microbial commons concerning CBD-related obligations. Some commentators appear to suggest that a culture collection could extinguish its prior informed consent-related obligations by voluntarily including benefit sharing conditions, in an MTA, that promise to share a percentage of royalties to the country of origin, in the event of commercial exploitation of the resources. While such a clause might well be appreciated by the country of origin, it does not satisfy CBD standards. The CBD is clear that the country of origin has to consent to access for any purpose, whether it is commercial or not.

A newcomer to the field could be forgiven for thinking that these obligations rooted in the CBD should not create significant impediments for microbial genetic resources continuing to flow into the microbial commons, and that the pre-1993/post-1993 divide could be bridged in ways that made it attractive for depositors to proactively place new microbial genetic diversity in the commons. However, the evidence that has been slowly accumulating since the mid-1990s reveals that providers have in fact become increasingly reluctant to make new genetic diversity available to the agricultural research community, and that this reluctance is fueled in part by access and benefit sharing issues. The combined effects of *a)* high levels of geo-politicized controversy about access and benefit sharing equity, and *b)* low levels of legal certainty about the conditions under which national authorities can provide access, has contributed to significantly increased transaction costs for research that requires access to agricultural genetic resources, with research having to be terminated or not started, in some cases.

To illustrate this phenomenon, I will focus for a moment on the experiences of the CGIAR centers and their efforts to attract new deposits of plant genetic resources to international crop and forage collections which they host. Those collections were originally assembled over the 1970s and 1980s. The centers hold the collections "in trust" for the international community, which means that they agree to maintain and distribute materials for agricultural research and breeding purposes to anyone who requests samples, anywhere in the world. Cumulatively, there are approximately 650,000 accessions of plant genetic crop and forage genetic resources in those collections. Rates of global acquisition and distribution of materials to and from those collections were relatively stable until the mid 1990s, after which time countries have been increasingly unwilling to deposit new materials. Between about 1995 and 2004, the rates of new deposits dropped approximately 80 percent. Part of the reason was the legal and political insecurities about access and benefit sharing associated with the CBD.

I want to share a few illustrative examples of how this plays out "in the field". The CGIAR centers have encountered situations where it was not clear who within the countries-concerned has authority to give prior informed consent to provide access to genetic resources. In light of that uncertainty, no one was willing to take responsibility for agreeing to deposit new material into the Centers' international collections.

If a genetic resource turns out to lead to a commercial success, no one wants to have been responsible for having agreed to make it globally available through the Centers' genebank—At least not without well defined legal authority to fall back on in their own defense. The problem is, most countries that ratified the CBD still have not

managed to put access and benefit sharing laws in place. And even the countries that have implemented national laws still do not have all the supportive mechanisms in place to make the laws actually work.

In numerous such cases, technical level partners in national agricultural research programs have clearly expressed an appreciation for the importance of conserving and making such materials from their country available for research through the international collections concerned. However, they were unable to "get to the end" of the consent-granting processes in their own countries, and the resources were ultimately not made available after extremely long delays.

In one instance when Bioversity wanted to coordinate collecting of papaya with a national research organization, we were informed that according to national law, it was necessary to get prior informed consent from local communities. Unfortunately, the national government said it did not know who we should contact in those communities to get prior informed consent. Instead, we were advised that we should contact the communities ourselves and establish to establish who had the right to provide or withhold approval and then inform the competent national authority. In light of the likely complications we did not purse the related collection and research project.

Repeated experiences of this nature lead some of the CGIAR Centers to adopt informal policies to stop approaching governments with proposals to organize new collecting missions. It was too complicated, the transaction costs were too high, and the negotiations attracted too much negative political attention. Instead, the centers put their hopes in the idea that the International Treaty on Plant Genetic Resources for Food and Agriculture—the treaty Shakeel Bhatti[73] described this morning—would provide a response to their access and benefit sharing related challenges. In the meantime, the end result has been a gradual tapering-off of the levels of new plant genetic diversity that is being made available to the plant research and breeding community through the international public genenbanks.

I have been focusing on how access and benefit sharing-related challenges have created disincentives for potential providers of plant genetic resources to make new diversity available. The situation appears to be similar with respect to microbial genetic resources for food and agriculture. This morning, Flora Katz[74] highlighted the difficulties associated with making arrangements to get access to microbial resources in a post-CBD world, stating that it takes an average of two years to get access to materials for the project she has been working on. Others have shared similar accounts. [75]

In this context, it is interesting to revisit the statistic mentioned earlier this today that up to approximately 60 percent of the isolates that are transferred are transferred informally. Why? Presumably one of the incentives for continuing to use informal mechanisms is that the formal procedures are considered to be too onerous with too high transaction costs.

What would happen to those exchanges if the transferors opted to bring those exchanges to the attention of relevant competent national authorities of the countries of origin of the materials for case-by-case adjudication? Presumably, the speed and volume

[73] See Chapter 19 within this publication.
[74] See Chapter 18, Proposal for a Microbial Semi-Commons: Perspectives from the International Cooperative Biodiversity Groups, Flora Katz, Fogarty International Center, National Institutes of Health, within this document.
[75] See Chapter 24. Microbial Commons: Overview of the Governance Considerations—A Framework for Discussion Tom Dedeurwaerdere, Université Catholique de Louvain, Belgium, within this publication.

of those exchanges would be negatively affected, possibly brought to a halt. At the same time, it is clearly unacceptable that exchanges that should be subject to formal legal approval are continuing "under the radar". Ultimately, the scene appears to be set for considerably more controversy, with negative impacts on the management and volume of material included in the hoped-for microbial commons.

What options exist to address the increasing reluctance of would-be depositors to make more microbial genetic resources publicly available, and in particular to make them available using a standardized MTA that reflects the basic tenets of the microbial commons? How is it possible to bridge the "1993 CBD divide," so that the commons is not limited to materials in culture collections acquired before 1993? It seems to me that one of the potentially most effective means to address these challenges in effective, long-lasting way is for national governments to meet in international, intergovernmental fora to create internationally harmonized standards, mechanisms and tools to support the collective pooling and management of microbial genetic resources.

Intergovernmental action and support could come in number of different forms. Perhaps the highest-level form of intergovernmental intervention would be to agree to a system whereby governments would agree to a tax on commercial sales of microbial genetic resources products, the sum total of which would be directed to an international benefit sharing fund. In return, all countries who are part of the agreement would provide facilitated access to one another, upon request for microbial genetic resources.

This was a model of horizontally constructed multilateral access and benefit sharing that many people once hoped-for in the early stages of the seven year negotiations of the International Treaty on Plant Genetic Resources for Food and Agriculture. It quickly became apparent that many countries would not go along with such an approach. However, with 20 years of experience now under the CBD as proof of the inherent difficulties of making bilaterally oriented ABS systems to work, perhaps the international community could possibly consider such an approach again, this time focusing on microbial genetic resources for food and agriculture. Quite frankly, to me, it still seems like an unlikely scenario, no matter how practical it may be, given many developed countries' aversion to such schemes.

Another, equally high-level form of intergovernmental intervention would be the creation a new legally binding international treaty on microbial genetic resources, establishing standard conditions for access and benefit sharing, a standard material transfer agreement, a common information-sharing platform, reporting schedules, tracking mechanisms, etc.

All participating organizations and individuals in contracting parties would use the MTA adopted under the treaty. This is clearly an approach inspired by, and similar to, the International Treaty on Plant Genetic Resources for Food and Agriculture, but focused on microbial genetic resources. Obviously, a treaty of this nature would have the benefit of creating legal certainty and would, presumably, create the possibility of high-levels of political level buy-in and commitment. It could also—like the International Treaty on PGRFA—bridge the "1993 CBD divide" in a very interesting way.

The International Treaty states that PGRFA that are "in the management and control" of state parties "and in the public domain" are automatically included in the Treaty's multilateral system of access and benefit sharing, regardless of when they were collected.

Even though the Treaty negotiators initial mandate was to consider access and benefit sharing conditions for *ex-situ* collections collected before the coming into force of

the CBD, they developed this formula that bridges the "1993 CBD divide" very artfully. However, negotiating such a treaty, and putting in place the supporting mechanisms and processes for its implementation is an extremely process-heavy, time-consuming procedure. Other, less resource-demanding mechanisms are to be preferred, if they can be effective.

One such possibility would be for an intergovernmental body to develop generic agreements that could be entered into, on a voluntary basis, between culture collections (or the governments of the countries in which the collections are located) and an intergovernmental body.

Those agreements would reflect the main characteristics of the microbial commons, for example, establishing that that culture collections would hold and make materials available for purposes X, Y or Z, the conditions under which materials held by the culture collections would be distributed, how benefits derived from commercial use will be shared, the MTA to be used for distributing materials, where information about materials could be publicly posted and shared, and so on.

Indeed this was the approach taken by the FAO Commission on Plant Genetic Resources for Food and Agriculture of the Food in its efforts, from the mid 1980s to the mid 1990s, to develop the "international network of *ex-situ* collections of plant genetic resources."

The commission developed model agreements that could be signed by organizations hosting PGRFA collections and or the country where those collections were located. In the end, this approach to developing an international PGRFA commons was overtaken by the negotiations of the International Treaty on Plant Genetic Resources for Food and Agriculture.

However, the enterprise did show some promise. In 1992, before efforts to develop the system were halted, nearly 30 countries indicated that they would sign such agreements to make their national collections available.

This incremental-federated approach has the advantage of being lighter weight (at least potentially) than a full blown treaty. Most important, the genetic agreements and related MTA would be very useful tools in the hand of national champions of an international microbial commons, allowing them to present hesitant national competent authorities with concrete, constructive options for how to administer their access and benefit sharing responsibilities.

At least for an important subset of the countries genetic resources, the fact that the general policy approach (in support of international harmonized standards for pooling and facilitated access and benefit sharing) and MTA have been endorsed by an intergovernmental forum in which the competent authority's own government has participated will give the option added credibility.

Such an initiative could complement, and build upon, the kind of coordination that has been promoted to date by the World Federation of Culture Collections, albeit without formal intergovernmental support and without all of the commons-related focus that is the subject of this meeting, for example.

The downside of this incremental federated approach is that it could potentially take a long time for a critical mass of collections and or competent national authorities to voluntarily decide to sign such agreements. So it could be a while before there is a significant amount of microbial genetic resources pooled and available under standardized terms and conditions. Another potential downside is that the approach is largely geared towards collections, and not so appropriate for *in-situ* materials.

Another, still lighter-weight approach which nonetheless exploits some of the goodwill that can potentially be purchased by engaging intergovernmental bodies would be to develop non-legally binding guidelines or codes of conduct. Such codes or guidelines could recommend following commons-informed principles and approaches, and recognize the advantage of terms and conditions that encourage the common pooling of microbial genetic resources and adopting low-transaction approaches to access and benefit-sharing. They could go so far as including model MTAs and recommending their use under various circumstances, by certain classes of users.

Again, such guidelines or codes would be useful tools in the hands of champions of a microbial commons. As instruments endorsed by an intergovernmental body, they could help to move competent national authorities in the direction of participating in the microbial commons.

The very light-weight form of intervention by an intergovernmental body would be some form of endorsement of the projects that subscribe to microbial commons principles, and that are dedicated to expanding the coverage of voluntarily adhered to practices by natural and legal persons. Such project could be presented to the next meeting of the CBD or the Commission on Genetic Resources for Food and Agriculture, and those bodies could make explicit statement recognizing the value of such projects and their objectives.

One advantage common to all of these ways of engaging intergovernmental fora is that they stimulate discussion at a very high policy level, with the potential to lift individual organizations and competent authorities above their national contexts, allowing them to investigate more broadly-conceived options for using and conserving their genetic resources. Intergovernmental processes would facilitate national competent authorities and other stakeholders to engage with their peers from other countries in a constructive, goal oriented context, setting the stage for the pursuit of common objectives through means that transcend purely national competencies.

As it turns out, there are currently opportunities for engaging in such internationally sponsored discussions and for promoting ABS norms supportive of the microbial commons in at least two ongoing intergovernmental policy-making processes. One of these processes is the ongoing negotiation, under the aegis of the CBD, of an international regime on access and benefit sharing. These negotiations have been ongoing since 2004, and there has been little substantive progress.

As one way of moving forward, a growing number of delegations are advocating sectoral approaches to the development of ABS norms under the international regime. Since there is no time to actually work out what the appropriate norms would be for each sector, the international regime would create flexibility, or even create a mandate, for considering such norms sometime in the future, after the framework of the regime is adopted.

To date, however, there has been practically no discussion at these negotiating meetings of what sectors might deserve special rules or what those rules may look like. The CGIAR centers are organizing a side-event at the next negotiating meeting (in November in Montreal) to sharing information with delegates about the (food security and economic development) benefits to be gained through internationally coordinated strategies to pool and use microbial genetic resources in support of agriculture research and development. We will also highlight possible options related to access and benefit sharing that the delegates could advance to support the microbial commons, in the course of the negotiations of the regime.

Another intergovernmental forum that is currently open to technical inputs concerning access and benefit sharing issues is the Commission on Genetic Plant Resources for Food and Agriculture. The commission adopted a multiyear plan of work in 2007 which includes examination of "policies and arrangements for access and benefit sharing for genetic resources for food and agriculture". The Commission secretariat is supporting the development of a study of on the use and exchange of microbial genetic resources, and another on the impact of climate change on countries' interdependence on microbial genetic resources.

These studies are intended to inform the discussion at the Commission level of what might be appropriate ways of regulating access to microbial genetic resources used for food and agriculture. Similarly, the Commission has a scoping study on microbial organisms and invertebrates as an input into its upcoming session. Then two or four years later, the Commission will review key issues related to microorganisms and invertebrates used in food and agriculture.

This is all fairly new—until very recently, the Commission has generally focused almost exclusively on plant genetic resources, and more recently on animal genetic resources. The fact that the Commission is widening its scope of enquiry, and possibly norm-setting, represents potentially rich opportunities for introducing consideration of access and benefit sharing norms to support the microbial commons.

To summarize, I have argued that access and benefit sharing related challenges have the potential to undermine the development of a vibrant and active international microbial commons. Intergovernmental participation in the development of access and benefit sharing norms (and related instruments) will be essential to overcome these challenges. There are currently opportunities in ongoing international policy-making processes to introduce consideration of access and benefit sharing norms that would support the microbial commons.

Question and Answer Session

PARTICIPANT: I take it that you read the convention as definitively eliminating the problem of pre-1993 claims, because the contrary reading would be the International Declaration on Sovereignty over Natural Resources of 1967. Are you confident that it is actually decided, or is it just there for future argument?

The second thing is more important. I take it that your third category would be analogous to the soft law approach that preceded the treaty that Shakeel Bhatti talked about—that it grew out of the soft law approach. I would hope that one would try to make it more successful than the preceding soft law approach, although you could say that it succeeded in the treaty. Is that where you are going—that it was a true soft law approach on which you could act?

MR. HALEWOOD: Yes. I think that the soft law approach could potentially work better now than it did before—partly because the International Treaty on Plant Genetic Resources for Food and Agriculture now exists. And partly because, in the context of the ongoing negotiations of the CBD's international regime, there are potential opportunities now for quick adoption of soft laws so that that the international community can demonstrate it is making some progress. Perhaps these options will look increasingly attractive the slower things go.

PARTICIPANT: So, you are implying that they might find that attractive in order to avoid another seven-year negotiation and just profit from work already done?

MR. HALEWOOD: I am not sure. It is possible that some of the steam is coming out of the geopoliticization of the access and benefit-sharing of genetic resources. When you see the way Brazil, China, and India are conducting themselves in the negations of the International Regime, it may be that there is room now for a less geopoliticized discussion and more open consideration of some of these soft-law possibilities.

28. Access and Benefit Sharing under the CBD and Access to Materials for Research – Stefan Jungcurt[76]

International Institute for Sustainable Development, Canada

After having heard so much about the tsunami of information that researchers have to deal with, I am going to talk about the storm on the horizon, which might turn into a thunderstorm or it might turn into a hurricane. I work for the International Institute for Sustainable Development, which is a green policy think tank in Canada. We have a branch called reporting services, and the people in that branch monitor all multilateral environmental negotiations on the international level and write daily reports, called the *Earth Negotiations Bulletin,* of what is happening in these negotiations. I have been performing that function regarding the negotiations on an international regime on access and benefit sharing (ABS) under the Convention on Biological Diversity (CBD) for the past six years, and now I am like everybody else: I need to figure out how this works.

In essence, the access and benefit sharing regime will add another element of demands or conditions on the international level that will affect the management of microbial resources. In other words, it will add a layer of complexity to something that is already quite demanding to deal with.

When the Convention on Biological Diversity was adopted in 1992, it included a provision on access and benefit sharing in Article 15. The main principles are countries have national sovereignty over their genetic resources, but that there should be facilitated access and benefit sharing under mutually agreed terms and with prior informed consent.

In 2002, the first major effort was made to implement this provision, which was the Bonn Guidelines for access and benefit sharing. These were basically a collection of items that should be taken into consideration when a national access and benefit sharing law is implemented. The guidelines were voluntary.

At around the same time, at the World Summit for Sustainable Development, the G77 countries which act mostly as suppliers of genetic resources approved an initiative to negotiate, within the framework of the CBD, an international regime for benefit sharing. In essence, this conference more or less told the CBD Secretariat, "You have been too slow in implementing Article 15, so please take action and implement it soon."

The CBD Secretariat reacted with a decision in 2004 to give a formal mandate to the ABS working group discussing the benefit-sharing regime. The user countries insisted on putting access back into the mandate, so the discussions focused on a regime for both access and benefit sharing and not just benefit sharing.

In 2006, the eighth conference of the parties (COP) established a deadline and asked that this regime be negotiated at the earliest possible time before the tenth COP. COP 10 will be held in October 2010, barely 12 months from now. The conference required that the negotiations be finished by then, but it is questionable whether this will actually happen, and even the delegates within the process have their doubts.

After the last meeting in Paris, one of the delegates who had been involved in this process from the beginning told me that this process now is where climate change was at in the 1970s. He said that the scientific basis for what we are trying to do here is so slow

[76] Presentation slides available at *http://sites.nationalacademies.org/xpedio/idcplg?IdcService=GET_FILE&dDocName=PGA_053742&RevisionSelectionMethod=Latest.*

in penetrating the process that it will take us many years of negotiation and learning to find out how we will realize this access and benefit sharing provision in the framework of the CBD.

That delegate was frustrated after a meeting that had not been very successful, so he may have been too pessimistic, but, nonetheless, the negotiations are certainly progressing slowly. We are now just a little more than 10 negotiation days away from the day when this should be adopted, and about half of the subject matter that should be covered by the regime has been consolidated into one text, but the text contains an enormous amount of the square brackets that indicate instances of disagreement. We have about 2,000 brackets or at least 1,000 cases where countries cannot agree on what the text should say concerning this regime.

At the same time, the pressure to adopt something in 2010 is very strong. For the CBD, 2010 will be the International Year of Biodiversity, and the CBD has an objective to significantly reduce the rate of biodiversity loss, so it might happen that a very broad declaration or framework agreement may be adopted, but there is sure to be an ongoing process afterwards. That means that there will be opportunities to continue engaging with this process after 2010, and the important thing will be to keep the door open at the tenth Conference of Parties in Japan in 2010.

Box 28–1 contains the first paragraph of the article on scope. Everything inside the brackets must still be negotiated. It demonstrates just how confused the question of what will be covered by the regime is right now.

BOX 28–1

Text on Scope

The International Regime on Access and Benefit-sharing applies to **[all] [biological resources,] genetic resources, [including viruses and other pathogenic , [as well as potentially pathogenic] organisms and genetic sequences regardless of their origin] [derivatives,] [products] [benefits arising from commercial and other utilization]** as well as [to their] [associated] traditional knowledge, innovations and practices [covered by the Convention on Biological Diversity] [in accordance with Article 8(j)] [within national jurisdiction and of a transboundary nature] [in accordance with the relevant provisions of the Convention on Biological Diversity] [subject [and mutually supportive] to other [relevant] international obligations] [and without prejudice to other international obligations]. [The International Regime will also apply to genetic resources of migratory species that for natural reasons are found on the territories of the Parties.]

The emphasized part of the text in the box is where a definition of the genetic material that is covered will eventually go. The only part of the definition that is not in brackets here is "genetic resources." The reason is because this is covered in the convention, and it cannot be renegotiated the delegates have not yet started to renegotiate the convention. All the other terms in the text are highly diffused, overlapping, and sometimes conflicting. The one big source of frustration at the last meeting was that—for some reason that I do not understand—the European Union insisted on inserting a

reference to pathogens as a very specific subset of microorganisms. Everyone else's reaction was, "Why do you do that?" Because the process is so politicized, there is a lot of suspicion about what it means.

It is possible that the background of that controversy can be found in some other recent discussions. Specifically, in the World Health Organization, Indonesia has been refusing to give access to virus strains for Avian influenza—not H1N1, but the one before that. The Indonesians were arguing that this should be regulated by the CBD first because they want access to the results of the vaccination and the medical treatments coming out of it.

The big question now is: Will all microbial materials be affected and how will they covered? It is not clear right now.

To offer some insight into the answers, I will start with a quick overview of the politics surrounding this process. When the CBD was negotiated, access and benefit sharing, or ABS, were the result of a "grand bargain." As the negotiations drew to a close, there was a large fraction of developing countries that said, "Why should we join this regime? It is just going to increase costs for biodiversity conservation, and the question of how fairly this is going to be financed is not clear." This led to the idea that access and benefit sharing could be a source of financing for biodiversity conversation. So, in essence, ABS was a promise that there would be, at some point down the road, markets that would generate revenue to cover the costs of implementing this convention. This promise was based on commoditization logic, which we have now discovered is not very adequate when working with genetic resources and information resources.

Thus, the ABS process faces the conundrum of coming to terms with this promise, which many developing countries are very adamant about. They say that this is a bill that you have not yet paid, so give us the access and benefit sharing or come up with something very big on financing. The latter option is almost impossible to expect from the developed countries right now, of course, so naturally the developing countries are expecting some sort of monetary benefits. For them, the idea has been that this resource will become part of our national income, and increasingly, as the intellectual property rights agenda has developed under the TRIPS agreement, the developing countries have seen it as a way to counterbalance what is sometimes called the "voracious appetites of intellectual property rights (IPRs)"—the trend to protect products of genetic resources through patents, which restricts access to those products, and sometimes even restricts traditional uses of genetic resources, for example as traditional medicines. Developing countries fear that IPRs will restrict them from benefiting from their own resources because of illegal or unauthorized access to genetic materials, or bio-piracy.

For a long time the rhetoric has been dominated by this idea that bio-piracy is a modern form of colonialism, that bio-prospecting is a criminal activity, and so on. It is getting better now. There is not such a strong divide in the process any more, but the mistrust still complicates the negotiations enormously.

On the other hand, user countries are increasingly offering other types of benefits. They have always been focused on making sure that access is not too restricted, so they have tried mostly not to get involved in IPRs. The user countries, however, have been coming forward with offers to share other types of benefits in areas such as technology transfer.

To summarize, we have a situation in which conflict about the distribution of potential benefits dominates the discussion of how we can safeguard the creation of those benefits. Unfortunately, this is a phenomenon that we see very often in international

negotiations. The issue of dealing with climate change faces a similar conflict, as the fight over the costs of dealing with it overshadows the considerations of how you can actually do it. It also resembles very clearly the blockbuster phenomenon—that the developing countries expect that they have something extremely valuable in their biological materials, and they do not want to miss out on benefiting from that. The potential value of such materials is overestimated.

It is worth noting that until the next-to-last meeting, of the sixth meeting, of the ABS working group, there was very little involvement of the scientific community. There have been a couple of position papers prepared. At some point, for instance, the WFCC submitted a position paper which is referred to quite a lot. There has been no active engagement of scientists on the delegations, however, and there has been little participation by nongovernmental organizations or other observers, although that has increased somewhat recently.

In general, there is some reason for optimism that at some point it will be possible to negotiate an agreement because there is an increasing recognition of the fact that no country is self-sufficient in genetic resources so that restricting access does not make much sense as a general principle. There is also greater recognition that any new ABS regime will have to address the link between access and use—that is, in order to have benefits to share; you must first be able to use the resource. Furthermore, there are several countries that are developing the capacity to use the resources themselves, and that affects their interests in the process. They are starting to explore on an informal level with the countries of the Organisation for Economic Co-operation and Development what the alternatives could be. What might be in there for them if they backed off of this extreme position that they have been taking so far?

The delegation of Brazil, which is obviously very large in these meetings, offers an interesting example. Their head negotiator was leading the group of like-minded mega-diverse countries - a coalition of biodiversity-hotspot countries. However, some people from the Brazilian delegation were talking to the European Union and asking questions like, “What do you mean when you talk about noncommercial research? What do you mean when you talk about collaborative projects?”

There is also an increasing recognition that the International Treaty on Plant Genetic Resources for Food and Agriculture (ITPGRFA) actually seems to be a good model. It has been cited in the process.

There have been a number of recent activities intended to achieve a broader involvement of experts and user groups. For instance, there has been a total of four technical experts groups: one on certificates; one on concepts, terms and working definitions; one on compliance; and one on traditional knowledge. These groups were supposed to have expert participation—scientists and other users of biodiversity. Unfortunately, most of the time the countries just nominated their normal negotiators as their delegates to these groups. Nonetheless, the discussions at least have raised awareness of some of the kinds of problems that this process has to address, although you cannot yet see the effects in the negotiating text. The positions have not changed much so far.

There have been a couple of other, more informal initiatives. One was a workshop on noncommercial research, which was led by the Consortium for the Barcode of Life. Another was a workshop on sectoral linkages, led by the United Nations University. There was also a workshop on traditional knowledge. One of the big issues is always the traditional knowledge tree of genetic resources. If an organism or genetic resource that

has been used traditionally leads to the development of some treatment or product, the idea is that the community should also share in the benefits.

If we judge these activities intended to involve the experts and user groups on the basis of success in meeting the 2010 deadline, it has been too little involvement and far too late in the process. If you look at this as something that might continue in an ongoing fashion, however, it could be a good starting point.

It is possible to discern some emerging trends from the activities that have taken place so far. However, keep in mind the usual caveat that is quoted in international negotiations: Nothing is agreed until everything is agreed. Still, some things seem to be coming into focus.

First, there seems now to be a broad acceptance that the regime will be legally binding. For a long time that was a point of contention. There was some idea of making a hybrid with some elements binding and some elements nonbinding. That is mostly gone.

Second, the process has been proceeding on the assumption that the outcome will be a protocol to the CBD. So this will be similar to the Kyoto Protocol—a separate part that only members of the convention can sign. This would mean that the United States would not be able to ratify this protocol. It would have to ratify the CBD first.

This implies that there will probably be some kind of a minimum participation clause. For the ITPGRFA it was 40 countries, but it might be higher or lower for the ABS protocol. This implies that once the regime has been adopted, there will be a period before it enters into force after the minimum number of countries have ratified. This could take several years.

There also seems to be an assumption emerging that the objective of the regime will be to support implementation of national ABS legislation. There would be a series of international minimum standards, both for access and for benefit sharing. There might be model domestic legislation and model clauses for material transfer agreements. A very important part will be certificates that will be used as a tool to monitor compliance.

One interesting thing about terminology is that they talk about minimum standards for access, but they are wary of talking about anything that goes into harmonization of access law. That is a result of the mistrust I spoke of earlier, or perhaps it is just that most of the supplier countries want to keep control over how this works. Unfortunately, the implication of that is that there will be a multitude of regulations and procedures in domestic legislation and different conditions for access, so the transaction costs will remain high and the situation Michael Halewood described will mostly prevail even after the regime has been adopted. Some kind of agreement therefore is needed to address the clarity and transparency of these laws. There needs to be legal certainty, but it is not clear now just how to achieve it.

Concerning benefit sharing, there are a couple of interesting ideas currently in the text. These are not yet agreed to, but they are there. One is that there might be something like a research exemption, both in terms of sharing the results of research and in terms of access. It is clear that this will not just be about access to journal articles or that type of information, but there would also be a type of technology transfer if the results of research are other products that can be used as basis for further research.

There is a clause concerning participation in research activities and joint activities, which is fairly uncontested at the moment. The idea of minimum conditions and standards is in there. And at the last meeting delegates inserted two things inspired by the ITPGRFA Treaty. One would be the option for multilateral sharing of benefits when

the origin is unclear or when resources exist in several countries. The other idea is that there might be trust funds for realizing the benefit sharing in those cases.

Compliance is always a big issue, but it is probably the least certain at this moment because compliance is usually negotiated after everything else has been agreed. The wording now has hints that there might be codes of conduct for important user groups, and I think there are explicit references to the Food and Agriculture Organization codes of conduct as well.

Another idea is that research funding agencies should specify what ABS legislation must be followed when the projects are funded. A very big issue has always been the question of disclosure requirements, both in material transfer agreements and in patent applications on downstream innovations.

The take-home message is that some of the things being discussed are already done. A lot of the information being tracked—the idea for developing unit identifiers, for instance—are reasonable ideas that have made it into the text and will probably stay there.

As for possible implications, as I was saying, it looks like the regime will not solve the problems we have been talking about and it might actually aggravate them. An important point to note is that there has been no recognition of the diversity of user practice. This workshop has made it clear that microbial research is not one monolithic process, but that it includes many different and very specific practices. If the regime is not to have a very negative impact, this range of practices will have to be reflected somehow in the regime. There must be some flexibility built in.

A couple of things will probably be very hard to negotiate. One will be the disclosure requirements and how far they go. The basic idea is to achieve transparency so that you can trace back to the country of origin, which is, of course, impossible for all the materials already in collections for which this was not documented. Disclosure requirements might also be difficult to decide in the case of resources collected in multiple countries or when one does not know which country should be associated with a resource.

How can the scientific community engage more in the process? The experience of the ITPGRFA is very instructive in the sense that there should be a concrete proposal brought forward by a large group announcing what it sees as the best way of achieving the objectives. One way that this could be done would be to use existing institutions and networks to give clout to a proposal like that and to promote it at several points. These institutions should not just focus on the process itself, but should also go to national governments and tell them that if research capacity is not to be destroyed, then that needs to be considered in the negotiations.

I think it is important that the COP 10 decision does not close the door to these activities. Hoping for anything else is probably unrealistic, but the take-home message should be not to close the door.

Question and Answer Session

PARTICIPANT: Can you define bio-piracy precisely?

DR. JUNGCURT: There is no precise definition of bio-piracy, but it is on the list of things that the developing countries want to control. They want to have an "internationally agreed-upon understanding of misappropriation and misuse," which is the politically accepted terminology for what bio-piracy would stand for. The word has been used by nongovernmental organizations (NGOs) basically for any kind of bio-prospecting activity that did not comply with the CBD, either deliberately or inadvertently. Most of these activities have been called bio-piracy after the fact, and there have been some studies that have concluded that those companies who wanted to negotiate prior informed consent were the ones identified by the NGOs and called bio-pirates. These companies actually had the good will but did not know how to negotiate or fell into the trap that there was no national authority defined, or else they negotiated with the wrong party, and they were then called bio-pirates. The situation results in punishing the compliant, which, of course, has driven every other private-sector player away from even trying to comply.

PARTICIPANT: This may be naïve, or maybe inflammatory, but the whole discussion has been in terms of benefits, but you mentioned pathogens. At what point do lawsuits among countries begin for not controlling infections?

DR. JUNGCURT: It is part of the frustration that some felt during the process when this suggestion came up. Many delegates did not understand why this was specifically mentioned by the EU delegation. Most delegates seemed to agree that pathogens, as a type of resource, should be covered by a broader definition.

You can think of benefits in different ways. If a cure to a disease has been based on accessing a certain virus strain or a certain collection, there could be a way of doing that, but this is the first time this has actually been raised in the discussion.

29. Closing Observations – Cathy Wu

University of Delaware

I would like to congratulate all the participants for this very successful and thoughtful symposium. I do not think I can do it justice by trying to recap the discussion from the symposium—the vast issues being explored and the different opinions being expressed. As mentioned by Dan Drell, our sponsor from the Department of Energy, this symposium is quite critical and timely, in light of the tsunami of data and information we are witnessing.

We certainly do need to address all these issues concerning the barriers and challenges in the different scientific, technical, institutional, legal, economic, and socio-cultural areas due to the increasing rate of the data flood. We also have heard many presentations about the opportunities for new research and discoveries because of this wealth of data and digital resources.

Some of the issues brought forth in this symposium may lead to additional studies. For example, how do we characterize knowledge? How do we better understand knowledge structures? How do we study the perceived value of knowledge? These and other questions have legal, economic, and governance implications.

What about training? There was some discussion about training data scientists. And what about open source software in this integrated research commons with open data, publications, and materials? What is the role of the software tools?

Many of these discussions are relevant to scientific disciplines beyond biology, and the issues discussed at this symposium will be shared with the Board on Research Data and Information. They will also no doubt be considered within your respective communities. I think this symposium has really planted seeds for many more interesting and useful things to come and will inspire new discussions and approaches.

Appendix A – Microbial Commons Symposium Agenda

Designing the Microbial Research Commons: An International Symposium

Board on Research Data and Information
Policy and Global Affairs Division
National Academy of Sciences
In collaboration with
Board on Life Sciences
and
Board on International Scientific Organizations
National Academy of Sciences

8-9 October 2009

The Lecture Room
National Academy of Sciences
2100 C Street NW, Washington, DC

AGENDA

THURSDAY, 8 OCTOBER 2009

SESSION 1: Statement of the problem from the research perspective—identification of opportunities and barriers
Chair: **Cathy Wu, University of Delaware**

8:45	Welcoming remarks and overview of the symposium	Cathy Wu, University of Delaware
9:00	Microbiology in the 21st Century	Joan Bennett, Rutgers University
9:40	Digital science perspective - From Brains to Microbes	Mark Ellisman, UC, San Diego, CA
10:10	Coffee Break	
10:40	Industrial perspective: Development of an MTA harmonious with a Microbial Research Commons	Stephen McCormack, Exela, Inc., Claremont, CA
11:10	Developing country perspective: Microbial Research Commons Including Viruses	Ashok Kolaskar, University of Pune, India
11:40	Panel discussion of Session 1 speakers with other Symposium participants regarding the commonalities and potential conflicts among different groups/sectors	

12:15 Lunch at the Academy

SESSION 2—Promoting access to and reuse of microbial materials
Chair: **James Staley, University of Washington**

13:15	Designing a semicommons for materials in microbiology	Jerome Reichman, Duke University Law School

13:45 Comments from different perspectives and panel discussion

- Federal Government Culture Collection - The Agriculture Research Service Culture Collection (NRRL): Germplasm Accessions and Research Programs
  Cletus P. Kurtzman, National Center for Agricultural Utilization Research, USDA
- Not-for-profit Culture Collection - ATCC: A Model for Biological Materials Resource Management
  Frank Simione, American Type Culture Collection
- Legal - Contracting to Preserve Open Science: Lessons for a Microbial Research Commons
  Peter Lee, UC Davis School of Law
- Economic - The Impact of Open Access Institutions on Life Sciences Research: Lessons from BRCs and Beyond
  Scott Stern, Northwestern University

15:30 Break

SESSION 3—Promoting access to and reuse of digital knowledge resources
Chair: **Michael Carroll, American University, Washington College of Law**

16:00	Designing the digital commons in microbiology - Moving from Restrictive Dissemination of Publicly-Funded Knowledge to Open Knowledge Environments: A Case Study in Microbiology	Paul Uhlir, National Research Council

16:30 Comments from different perspectives and panel discussion

- Web applications - The Web-Enabled Research Commons: Applications, Goals, and Trends
  Thinh Nguyen, Creative Commons and Science Commons

- Legal - Comments on Designing the Microbial Research Commons: Digital Knowledge Resources
  Katherine Strandburg, New York University Law School
- Federal information policy - Toward a biomedical research commons: A view from NLM-NIH
  Jerry Sheehan, National Library of Medicine, NIH

- Academic publications
  Frederick Rainey, Louisiana State University
- Web information services - StrainInfo.net: Reducing Microbial Data Entropy
  Peter Dawyndt, Ghent University, Belgium

18:30 Dinner for the speakers in the Members' Room

FRIDAY, 9 OCTOBER 2009

SESSION 4: Thematic focus on microbiology research and applications in energy and environment

8:30 **The Materials Semicommons in Microbial Energy and Environmental Research and Applications**
Chair: **Stephen McCormack, Excela, Inc.**

- Research funder- The Department of Energy: Genome Sciences
  Daniel Drell, Department of Energy
- Researcher - Large Scale Microbial Ecology Cyberinfrastructure (CAMERA)
  Paul Gilna, UC San Diego
- International cooperation - Proposal for a microbial semi-commons: Perspectives from the International Cooperative Biodiversity Groups
  Flora Katz, Fogarty International Center, National Institutes of Health
- Intergovernmental organization - The International Treaty on Plant Genetic Resources
  Shakeel Bhatti, Food and Agriculture Organization
- Institutional design - Microbial Commons: Governing Complex Knowledge Assets
  Minna Allarakhia, University of Waterloo, Canada

10:15 Coffee Break

10:45 **The Digital Commons in Microbial Energy and Environmental Research and Applications**
Chair: **Micah Krichevsky, Bionomics International**

- Digital research - Microbial Genomics
  Nikos Kyrpides, Lawrence Berkeley National Lab
- Digital user - Accessing Microbiological Data: A User's Perspective
  Mark Segal, Environmental Protection Agency
- Academic journals - The Microbial Commons: Journals and Professional Societies
  Samuel Kaplan, UT Houston Medical School, and former Chair of the American Society for Microbiology's Publications Board
- Economic and institutional - *Mitigating Anti-commons* Constraints on

Global Scientific Research: A "bottom up" approach to instutitional reforms
Paul David, Stanford University

12:30 Lunch at the Academy

SESSION 5— Governance of the integrated microbiology commons
Chair: **Paul Gilna, UC San Diego**

13:30	Microbial Commons: Overview of the Governance Considerations - A Framework for Discussion	Tom Dedeurwaerdere, UC Louvain, Belgium

14:00 Comments from different perspectives and panel discussion

- Institutional design and Governance in Microbial Research Commons
  Charlotte Hess, Syracuse University
- U.S. foreign policy - International Developments: A Context for the Creation of a Microbiology Commons
  Anita Eisenstadt, Department of State
- International food and agriculture - Options for governing the microbial commons informed by the need to bridge the 1993 CBD divide
  Michael Halewood, Bioversity International, Italy
- Access and Benefit Sharing under the CBD and access to materials for research
  Stefan Jungcurt, International Institute for Sustainable Development, Canada

15:45	Concluding remarks by the Symposium Chair	Cathy Wu, University of Delaware

16:00 End of meeting

Appendix B – Microbial Commons Symposium Participants

SPEAKERS:

Allarakhia, Minna
Bennett, Joan
Bhatti, Shakeel
Carroll, Michael
David, Paul
Dawyndt, Peter
Dedeurwaerdere, Tom
Drell, Daniel
Eisenstadt, Anita
Ellisman, Mark
Gilna, Paul
Halewood, Michael
Hess, Charlotte
Jungcurt, Stefan
Kaplan, Samuel
Katz, Flora
Kolaskar, Ashok
Krichevsky, Micah
Kurtzman, Cletus
Kyrpides, Nikos
Lee, Peter
McCormack, Stephen
Nguyen, Thinh
Rainey, Frederick
Reichman, Jerome
Segal, Mark
Sheehan, Jerry
Simione, Frank
Staley, James
Stern, Scott
Strandburg, Katherine
Wilbanks, John
Wu, Cathy

PARTICIPANTS:

Berger, Kavita
Bowden, Robert
Bowman, Katie
Chang, Richard
Chen, Yu-Fen
Chong, Lisa
Contreras, Jorge
Epstein, Gerald
Garges, Susan
George, Carol
Gregurick, Susan
Heaney, Chris
Kapustij, Cristina
Madhavan, Guruprasad
McCluskey, Kevin
McCreight, Robert
Nelson, Karen
Palm, Mary
Santos, Ana
Seto, Belinda
Siebenga, Joukje
Tyler, Brett
Wheeler, Terrie

STAFF:

Cohen, Daniel
Kuvelker, Subhash
Levey, Cheryl
Uhlir, Paul