

*Designing microbial research commons:
Learning from institutional policies and collaborative network technologies*

Centre for the Philosophy of Law, Université catholique de Louvain
Laboratory of Microbiology, Ghent University
U.S. National Committee for CODATA
U.S. National Committee for Biology
Genomic Standards Consortium
Bioversity International
Science Commons

In collaboration with
World Federation on Culture Collections
International Union of Microbiological Societies
Committee on Data for Science and Technology
Research Unit Knowledge-based Systems, Ghent University
Department of Applied Mathematics and Computer Science, Ghent University

SUMMARY

Within the life sciences there is a growing need for integration of biological information across geographical and disciplinary boundaries to address complex global problems. Efficient information fusion is facilitated through advances in information and communication technologies. It also requires the involvement of various communities, institutions, and government administrations in ensuring that the available information is effectively used and integrated, and that the results are distributed and applied. The objective of this project is **to gain a better understanding of the impact of policies on access and reuse that are internal and external to the microbial community, to evaluate the role of public domain and contractually reconstructed “commons” in microbial data, information, and materials, and to explore opportunities opened up by the collaborative nature of a new generation of Internet applications**, which allow users to become active participants in the creation, distribution, annotation, and change of the information content itself. As an application domain this project will primarily focus on the ongoing efforts to create a worldwide microbial research commons. This is a major undertaking that has strong connections to other similar initiatives in genomics and open access publishing, and potentially has big payoffs for the microbial community that will draw upon some of those other initiatives. Starting from existing approaches in the broader science commons, such as common-use licensing, web 2.0 technologies, and collaborative institutional policies, a major goal is to organize a structured learning process between communities and policy makers that can be useful for other problem situations and actor settings.

Specifically, the project will be focused on answering the following questions:

1. What are the research and applications opportunities from improved integration of microbial data, information, and materials and from enhanced collaboration within the global microbial community?
2. What are the challenges and barriers—the scientific, technical, institutional, legal, economic, and sociocultural—that hinder the integration of microbial resources and the collaborative practice of scientific communities in the microbial commons?

3. What are 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. What are the contributions of new 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 commons.

An initial two-day international workshop with high-level invited speakers will be held to address the tasks above. The results will be published in international peer-reviewed journals, disseminated through the collaborating organisations' websites, and further developed through their international and national initiatives. In a second phase of this project, another workshop will be organized approximately 9 months after the first one, which will focus on applying and implementing the outcomes from the first workshop in the collaborating organisations' national and international initiatives.

Intellectual merit of the proposed activity

The introduction of new standards for the protection of intellectual property during the last twenty years has had a profound impact on the practice of sharing information and resources in the life sciences. Even if there is no clear evidence of an anti-commons based on patent thickets, there is considerable evidence of related anti-commons effects inhibiting the development of real information integration. Such evidence includes decreased confidence among providers of biological information (Roa-Rodríguez and van Dooren 2007) and massive under-use of information held in private databases within research laboratories (Esanu and Uhler 2003).

For this reason appropriate governance mechanisms are needed to enforce the norms and practices of information sharing in the institutional and organisational environment of public-sector e-science. For instance, in the field of public e-science, a report of the Oxford Internet Institute proposes the establishment of an independent advisory board for collaborative arrangements which would develop a set of contractual prototypes ('model contractual clauses'), each treating some specific problem, which could be readily assembled into a variety of contractual arrangements (David and Spence 2003). Another proposition, more directly articulated for the field of scholarly information, is based on the development of standard contracts for information sharing by leading university institutions as a model on which other institutions could draw (Reichman and Uhler 2003). When the end-products of information integration are subject to intellectual rights, a commons could be reconstructed through the use of common-use licenses for non-commercial research purposes (David and Spence 2003, pp. 62–64).

The same trends of erosion of norms and sharing practices can be observed with the exchange of biological materials. For instance, in the field of non-human biological materials, scientists usually shared materials in an informal way for research purposes (Nguyen, 2008; Stromberg, Pascual and Dedeurwaerdere, 2006). This practice has been based on a sense of reciprocity and a common research ethic that integrates the values of quality management and curation of

materials held in a world-wide network of biological resource centres. However, increasingly this system faces the pressures of different national interests¹ and competition amongst scientists for prior publication and access to project research funding². A combination of new legal approaches, institutional tools, and normative guidelines will be required to preserve the access to basic biological materials and research tools (Dedeurwaerdere, 2008). Examples of initiatives in this direction are the proposition of a new standard Material Transfer Agreement, such as recently by the European Culture Collections Organisation at its October 2007 meeting³, appropriate enforcement of the Bacterial Code governing the exchange of type strains (Lapage, S.P., *et al.*)⁴, and the negotiation of niche-specific Material Transfer Agreements establishing new commons regimes, such as in the Science Commons neurocommons project⁵.

To provide a real world context for the analysis of these trends and proposed remedies, we will primarily focus on the ongoing efforts to create a worldwide microbial research commons. This is a major undertaking that has strong connections to other initiatives in genomics and open access publishing and potentially big payoffs that draws from all the other initiatives. It has an ambitious scope, and substantial organizational and governance challenges. Consistent with the tensions mentioned above, this project could produce a microcosm for a federated networked model that others could use, but only if it can overcome some of the negative trends that are rapidly encroaching on the sharing ethos.

What makes this project important for both theoretically and practice is its goal to expand the design principles of collaborative networking from a relatively homogeneous community, organized about a common purpose to a heterogeneous community, which could use the research tools for a variety of applications from pure data-driven research to commercial products. What also makes this project interesting is that it attempts to combine materials, data and information into an integrated commons – a mechanism for lowering transaction costs, speeding up research, promoting use of raw materials as well as research results, both published and private.

Broader impacts of the proposed activity

Historically, microbiological data, information, and materials have been the main promoter of open access in the life sciences, through the organisation of public and non-profit ex situ conservation facilities. In the last decade, these pre-genomics ex situ facilities have progressively developed into multi-service facilities called Biological Resource Centres (BRCs, for a discussion of the concept, see OECD 2001), that manage the collection, organisation, curation, and exchange of biological resources and their associated data and information. Numerous examples exist of cases where microbiological resources held in open access (in nature or in ex situ facilities) have played a key role in new developments in the life

¹ For example, according to US legislation, a US laboratory has to buy its biological material in US culture collections, if available, even if it would be available at better (less restrictive) conditions in another country.

² For example it is current practice for a researcher to ask that a deposited strain of biological material be kept secret until his or her publication on that strain is published. This delay in allowing open access to the strain is often informally agreed, and can mean a delay of months or even years.

³ In Goslar, Germany, 12th of October 2007.

⁴ Due to the problems associated with the availability of type material that has only been deposited for the purpose of patent deposit, the International Code of Nomenclature of Bacteria or “Bacterial Code”, has decided that such strains cannot / can no longer serve as type strains. Cf. also *infra* footnote 11.

⁵ Cf. <http://sciencecommons.org/projects/data/>

sciences. The use of natural enzymes (to catalyse chemical reactions), for example, is widespread throughout industry, which continues to rely upon the provision of purified strains from natural specimen of bacteria kept in BRCs. In health research, it is estimated that 25 percent of all new drugs are based directly or indirectly on natural plants or organisms⁶. A similar situation prevails in related field of agricultural *ex situ* collections. A survey amongst major firms in the seed/agriculture plant breeding industry showed that agricultural R&D depends on a yearly 8% input of new material to deal with increasing resistance of existing commercial varieties, of which approximately 60% comes from *ex situ* facilities (Swanson 1997, pp. 73-75). These diagnostics are probably valid also in the case of applying microorganisms in agriculture, where bacteria are extensively used for enhancing productivity, representing a world market measured in billions of US\$⁷.

An important further step in the domain of microbiology is the extension of these worldwide networks for conservation and innovation in microbial diversity to world-wide information facilities that can be accessed through the Internet. These information facilities operate on a global scale, such as the initiative for Common Access to Biological Resources and Information (CABRI) and the Straininfo.net bioportal, both aiming at the integration of information on the biological material held in BRCs belonging to the World Federation of Culture Collections (WFCC), and more focused issue networks such as the European Human Frozen Tumour Tissue Databank (TuBaFrost).

By analyzing real-world examples of technical and institutional design elements in the microbial commons, such as common-use licensing, web 2.0 technologies and collaborative institutional policies, it will be possible to improve and apply these tools to other problem situations and actor settings.

⁶ A study made in 1989 in the US estimated that 25 percent of drugs' active ingredients were extracted or derived from plants. Another study carried out in 1993 estimated that in the US 57 percent of the prescriptions contained at least one major active compound now or once derived after compounds derived from biodiversity, cf. Brahy 2006, p. 266 ; Principe, 1989 ; Grifo and Downes, 1996.

⁷ A case in point are the soybean root nodule bacteria (RNB). These bacteria are mixed with soybean seeds for enhancing productivity, which represents a world market measured in billions of US\$. The RNB strains themselves are freely exchanged amongst collections of developed and developing countries both for custody of reference strains (so that each country can provide them to its local industry) and for research.

PROJECT DESCRIPTION

Policy Context

Within the field of the life sciences there is a growing need for integration of biological information across geographical and disciplinary boundaries. Increasingly complex global problems such as the effects of climate change on biodiversity, the development and geographical patterns of drug resistance, and the potential impact of genetically modified organisms on the natural environment, require the integration of different data sources and other available information.

Efficient integration of information is not only facilitated through advancements in collaborative networking technologies. It additionally requires the involvement of various communities in ensuring that the available information is effectively used and integrated, and that the results are distributed and discussed. The design of appropriate organisational and institutional rules for coordination and cooperation among the various communities is therefore crucial in building a sustainable information infrastructure.

In the face of the considerable uncertainty and great speed of development of the cyber-infrastructure, a flexible and adaptive social and institutional infrastructure is needed. The sustainability of the institutional architecture will depend on its capacity to draw upon diverse institutional resources. Information sharing in the life sciences must draw upon a variety of existing institutional arrangements and types of ownership, including government, private and community ownership of information resources.

Recently, several attempts have been undertaken to disentangle the issue of institutional choice in the development of e-science. In this context, an OECD report on research data recommends, as a general design principle, that publicly funded research data should be openly available to the maximum extent possible (Arzberger *et al.* 2004). As proposed in the report, ‘improving and expanding the open availability of public research data will generate wealth through the downstream commercialisation of outputs, provide decision-makers with the necessary facts to address complex, often trans-national problems, and offer individuals the opportunity to better understand the social and physical world in which we all live’ (Arzberger *et al.* 2004, p. 139). On the other hand, when the end-products of information integration are subject to intellectual rights, the respective rights of the participants in the public research system can be mediated most effectively through the use of contracts at the individual researcher, institutional, and governmental levels. Common-use licensing approaches that promote broad access and reuse rather than restrict it, can preserve essential ownership rights while maximizing the social benefits and returns on the public investments in research.

Open access⁸ to publicly funded research is also an important general design principle of the institutional architecture for information sharing in the life sciences. However, this project

⁸ There can be variants of many terms marching under the banner of open science or public research (Cook-Deegan and Dedeurwaerdere 2006, p. 300). “Open access”, for example, can mean free access to view information, but not necessarily freedom to use it in all ways without restriction. To some, open science means no one can fence it in. Access to information, say through “viral” licensing or copyleft, may be conditioned on agreeing *not* to restrict subsequent users. Information may also simply be put into the public domain, by deposit at a freely available public database for example, for any and all subsequent uses, both proprietary and open. In the latter case we will refer to it as “public domain”, in the former cases to it as “contractually reconstructed commons”.

will go beyond the question of data availability and accessibility to address the challenges related to the process of information integration, which build upon the open-access infrastructure of publicly funded research⁹ and produce new knowledge by combining, re-using and extrapolating the available information.

An incomplete, informal commons for microbial data, information, and materials

The WFCC is a multidisciplinary commission of the International Union of Biological Sciences (IUBS) and a federation within the International Union of Microbiological Societies (IUMS). As an organisation it is concerned with the collection, authentication, maintenance and distribution of cultures of microorganisms and cultured cells. It aims to promote and support the establishment of culture collections and related services, to provide liaison and set up an information network between the collections and their users, to organise workshops and conferences, publications and newsletters and work to ensure the long term perpetuation of important collections.

There are over 525 WFCC member organizations in 67 countries. Members include public and not-for-profit entities—government agencies, universities, NGO research organizations. They have been collaborating with exchanges at marginal cost of dissemination (small fee per sample and free online access to limited data and information).

One of the missions of the WFCC is to share data about its members' microbial culture collections by providing links to their web sites. Recently, however, the StrainInfo.net bioportal (www.StrainInfo.net) was created to provide “a technology platform that can stimulate the movement towards using multi-perspective integrated information in a broadened biological and clinical context”. The bioportal thus seeks to serve as an integrator of diverse public and private information resources to serve the microbial community.

Information sharing: barriers and opportunities

Within the field of the life sciences, initiatives have emerged for sharing microbial information through networking of distributed databases and related information. From a governance perspective, these networks face increasing pressure from the development of global markets, which has led to competition for the ownership of previously shared resources. At the same time, as economy shifts more towards information-based production, the prevalence of public-good and informational concerns looms larger (Stiglitz *et al.* 2000).

Two of the most influential policy changes in this debate have been the 1980 Bayh-Dole Act in the US (Rai and Eisenberg 2003) and, more recently, the 1996 EU Database Directive 96/9/EC (Reichman and Uhler 1999). The Bayh-Dole Act explicitly gave universities the right to seek patent protection on the results of government-sponsored research, and to retain patent ownership. As a consequence, in the period from 1980 to 1992, the number of patents granted per year to universities in the US increased from fewer than 250 to almost 2700 (Rai 1999, p. 109). The EC Database Directive 96/9/EC was a landmark decision that lowered the standards of eligibility to database protection. Indeed the Database Directive offers copyright protection to databases that are original in the selection or arrangement of their contents, but also to non-original databases if it can be shown that there is a substantial investment either in obtaining,

⁹ Information integration can also use the information made openly available by private-sector providers. Indeed the latter also have a major interest in a broad and open information infrastructure in basic research tools. This is, for instance, clearly the case in the field of genomics (Cook-Deegan and Dedeurwaerdere 2006).

verifying, or the presenting their contents. This extended protection to library catalogues, for instance, but potentially also to biological information facilities that network existing databases.

These rulings have to be situated within the wider scope of globalisation of intellectual property rights, a phenomenon that has accompanied the genomic revolution in the life sciences and the digital revolution in information technologies. This new institutional environment has played a key role in stimulating innovation and new market developments in the life sciences. However, it is also posing a challenge to life science research for public purposes, as the research communities have to adapt their strategies and design new institutional arrangements to allow them to provide services of general interest in an increasingly competitive and international environment.

Federated databases such as the Straininfo.net bioportal and the Global Biodiversity Information Facility (GBIF) were created to provide a technology platform to stimulate the use of multi-perspective integrated information in a broad biological and clinical context. In the case of the Straininfo.net bioportal, advanced software tools are used to automatically assign persistent and globally unique identifiers to cultured samples of micro-organisms and to integrate information about the organisms held in a global network of biological resource centers with relevant downstream information provided by third party information providers (e.g. literature references from PubMedCentral and CrossRef or genome sequence records from the International Nucleotide Sequence Database Consortium (INSDC) and SILVA). From a governance perspective, these federated infrastructures may be characterised as an informal commons. The exchange of information draws upon the professional practices and well understood needs of the scientific communities. The main benefit is the multiplier effect of the multilateral system: researchers provide access to their own limited resources and information and in return they gain access to resources and information of all other member organisations. However, for their further development they will have to cope with the economic pressures and the threats arising from the commercialisation of resources. The analysis of more formal governance models based on common-use licensing will provide design elements for rising to this challenge.

Policy response: establish an integrated microbial commons for the exchange of biological material and information

An integrated infrastructure would provide basic common use principles for access to both materials and information. It needs to provide at least:

- new resources with a focus on existing barriers, governing structure, access to a whole that is greater than parts, and an evolving structure that facilitates collective research and applications and grows over time.
- a more formal role for international coordinating bodies, such as the World Federation of Culture Collections, the European Culture Collections Organization or the World Data Centre for Microorganisms.
- further benefits: conservation of ex-situ microbial biodiversity¹⁰. Information integration through the Straininfo.net bioportal allows the mapping of

¹⁰ We focus here on the “informational” benefit that comes from better knowledge and understanding of the ex-situ biodiversity. In the field of plant diversity for agriculture, one can establish however also a direct contribution to biodiversity. For instance, a summary of scientific findings concludes that genealogical diversity and molecular genetic diversity have increased over time in germplasm held at the International Maize and

endangered bacterial and archaeal type strains – these are type strains (the reference strains for a particular species) with no known culture in a biological resource center – or to map whole-genome sequences that have no known publicly available culture of the sequenced organism. Enhanced knowledge about endangered strains might help to enforce the bacterial code¹¹ or to incite researchers in genomics to deposit organisms in public culture collections.

An analogous movement for building the biological material commons: the CGIAR network and the ITPGRFA treaty

Historically, the Consultative Group on International Agricultural Research has played a leading role in promoting open access to biological resources through the organisation of a network of specialized ex-situ conservation facilities throughout the world. As stated in the 2003 CGIAR policy guidelines:

“The germplasm¹² designated by the Centres is held in trust for the world community in accordance with the agreements signed with FAO [...]. Based on the conviction that their research will continue to be supported by public funds, the Centres regard the results of their work as international public goods. Hence full disclosure of research results and products in the public domain is the preferred strategy for preventing misappropriation by others”.

An interesting example of the application of this open access policy is the case of the World Fish Centre (Greer and Harvey 2004, pp. 135–141). In the World Fish Centre, new breeds of tilapia, a tropical river fish originally found in Central and East Africa, were developed for use in aquaculture. In 1999, a corporate partner, GenoMar, received the exclusive licence to market a specific variety of a new breed, called “super-tilapia”, while the property rights on all the other varieties of new breeds remained with the World Fish Centre. Those varieties remain freely available for direct use in developing countries¹³. Both the CGIAR intellectual property policy guidelines and the ITPGRFA treaty reflect this open access strategy, even if the new treaty raises some issues of harmonization between the ITPGRFA standard material transfer agreements (MTAs) and the CGIAR standard MTAs¹⁴.

An analogous movement for building information commons: the International Nucleotide Sequence Database Collaboration

Wheat Improvement Centre in Mexico (CIMMYT). CIMMYT-related varieties represent a vast array of germplasm constituted by genetic recombination of diverse sources of materials throughout the wheat-growing world (Fowler *et al.* 2001 : 194, note 8).

¹¹ The Bacteriological code is a product of the International Committee on Systematics of Prokaryotes (ICSP), a IUMS (International Union of Microbiological Societies) ComCoF. The Bacterial Code requires the deposit of new type strains in at least two culture collections in two different countries. No such code exists however for yeast, fungi, plasmids or algae.

¹² Technically, one speaks of collections of “germplasm”, referring to “seeds, plants or plant parts that are useful in crop breeding, research or conservation because of their genetic attributes (Fowler *et al.* 2001: 182).

¹³ Both public and private research has made extensive use of the materials held in the CGIAR collections. For instance, a quantitative analysis of 15 years of exchange of maize germplasm between the International Maize and Wheat Improvement Centre in Mexico (CIMMYT) and 15 other developing countries shows that the recipient countries received 4 times the amount of specimen that they have contributed to the international CGIAR repository (Fowler *et al.* 2001). Moreover, an estimated 75% of all seeds sold by private companies in Latin America in 1996 contained CIMMYT-derived germplasm (*Ibid.*, p. 194).

¹⁴ The CGIAR intellectual property policy clearly reflects this open access strategy: “Consequently, the Centres will not assert intellectual property control over derivatives except in those rare cases when this is needed to facilitate technology transfer or otherwise protect the interests of developing nations” and “In the event that a Centre secures financial returns as a result of the commercialization by others of its protected property, appropriate means will be used to ensure that such funds are used for furthering the mandate of the Centre and the objectives of the CGIAR” (CGIAR 2003, p. 31-32).

DNA sequence data of the International Nucleotide Sequence Database Collaboration (INSDC) is primarily collected by a trio of databases in the United States (GenBank), Europe (EMBL) and Japan (DDBJ), who share data among themselves. In the case of INSDC, intellectual property on the gene sequences that are published depends on the data provider¹⁵. Some of them are related to patented material, but most are in the public domain. However, both for patented and public domain sequences, important incentives exist to publish rapidly and deposit data in the INSDC databases. Having one's sequence published in the INSDC databases or any alternative recognized international e-repository is a requirement for publishing scientific research on a new gene sequence in an international journal. The provision of gene sequences to this international science commons is thus assured through its connection with the collaborative effort of the basic scientific research in the life sciences¹⁶.

Genomics provides several examples of the value of public information (Cook-Deegan and Dedeurwaerdere 2006). For instance, the 2002 report from the World Health Organisation, *Genomics and World Health* gave the example of fosmidomycin (WHO 2002, p. 49). This drug is currently tested for the treatment of malaria in Africa (Missinou *et al.* 2002). Its use came to light as a consequence of sequencing the complete genome of the malaria parasite and noticing a metabolic pathway not previously known to exist. Another case study is SARS. Strains of the corona virus that causes SARS were identified and sequenced within a month by at least three laboratories in Singapore, Canada and the United States. Making progress with such speed required strong norms of open science, with obvious social benefits (Cook-Deegan, p. 304).

Design elements for implementing the integrated microbial commons

Design elements for implementing the integrated microbial commons include appropriate regulation, institutional policies, standardized contractual templates for exchange of materials, collaborative network technologies and access to research results published in the literature. In this project we will focus more precisely on some real world examples in the microbial commons where these design elements are already being implemented.

Impact of institutional policies

Biological Resource Centres (BRCs) are an essential infrastructure for life science research. If, for financial reasons, BRCs are unable to perform their tasks under conditions that meet the demands of scientific research and the requirements of industry, countries will inevitably see high value-added products being transferred into a strictly commercial environment with at least two consequences (OECD 2007, p. 17):

- blockage of access to these products or requiring payment of a high price (which may not take account of the initial public investment required to develop them).

¹⁵ Cf. for an introduction to this <http://www.ncbi.nlm.nih.gov/Genbank/> (accessed 6th of April 2006).

¹⁶ One important issue for collective action however is the quality management of the information in GenBank. Because of the pressure to publish rapidly, the information submitted is often incomplete and poorly verified and as a result the databases contain many errors (Pennisi 1999; Gilks, Audit *et al.* 2005). The best results in data curation in genomics have been obtained where there are appropriate incentives for providers and data managers. Measures such as appropriate citation of the data provider by the user, privileged data access for providers, data-deposit practices enforced by publishers and appropriate career structures and rewards for data managers have been proven to be effective tools in the establishment of data-sharing regimes (The Digital Archiving Consultancy, The Bioinformatics Research Centre, The National e-Science Centre, 2005).

- definitive loss of products and elimination of technology transfer of those products for the foreseeable future.

In order to secure this essential infrastructure, the OECD guidelines for Biological Resource Centres (OECD 2007) state some of the actions national governments should undertake in concert with the international scientific community:

- selectively seek to strengthen existing ex-situ collections of biological data and materials, create collections of new resources, including in non-OECD countries, and elevate those collections to the quality required for accreditation as national BRCs.
- support the development of an accreditation system for BRCs based upon scientifically acceptable objective international criteria for quality, expertise and financial stability.
- facilitate international co-ordination among national BRCs by creating an agreed system of linkage. This should be based upon modern informatics systems that link biological data to biological materials across national BRCs and upon common technological frameworks.
- take into account the objectives and functioning of BRCs when establishing and harmonising national or international rules and regulations.
- develop policies to harmonise the operational parameters under which BRCs function, including those governing access to biological resources as well as their exchange and distribution, taking into account relevant national and international laws and agreements.
- support the establishment of a global BRC network that would enhance access to BRCs and foster international co-operation and economic development”.

In this project we will evaluate the impact of the OECD guidelines for BRCs on information integration in the microbial commons and compare the guidelines (and their impact) to the CGIAR guidelines in agricultural research, the Bermuda Rules in genomics and the principles used by the Genomic Standards Consortium.

Contractually reconstructed commons

The past quarter century has seen the emergence of a pronounced world wide trend towards commoditization of publicly funded research outputs, including the underlying data and information resources (David 2005). This tendency has gained impetus from the intensification of global economic competition and the continuing fiscal pressures on governments, with a concomitant commercialization and privatization of functions previously conducted by public agencies, including research and the dissemination of government data and information.

The “public goods” properties of data and information, however, permit concurrent use and reuse at negligible incremental costs by an unlimited number of users whose access to and use of the content leaves it undepleted. Given the expansible nature of information it is unreasonable to ignore efficiency losses on the functioning of the research system by the enforcement of intellectual property rights in digital scientific resources. Based on this economic *rationale*, it has become increasingly apparent that there is another and rather different approach whose practical aspects merit wide attention and support to its further development. An example of this approach is *Global Information Commons for Science Initiative*, which is an international initiative of the Committee on Data for Science and Technology (CODATA)¹⁷. The proposed approach of the science commons consists of the

¹⁷ The original ideas for the Global Information Commons for Science Initiative were presented in a series of reports published at the U.S. National Academies as well as in a seminal article by Reichman & Uhler 2003, note 3, and in David and Spence 2003. These ideas were more fully fleshed out following an international workshop at UNESCO Headquarters in Paris on 1-2 September 2005 on the theme “Creating the Information Commons for

voluntary use of the rights held by intellectual property owners, which allow them to construct by means of licensing contracts conditions of “common-use” that emulate the key features of the public domain that are most beneficial for collaborative research in all its forms. The intention is to form legal coalitions or “clubs” for the cooperative use of scientific data, information, materials and research tools that actually are not in the public domain, and whose licensed use is therefore legally protected by an intellectual property regime. Such an undertaking may be properly described as creating a network of “global information commons for science”, inasmuch as each “common” constitutes a collectively held and managed bundle of resources to which access by cooperating parties is rendered open (though perhaps limited in its extent or use) under minimal transactions cost conditions.

The respective rights of the participants in the public research system can be mediated effectively through the use of contracts at the individual researcher, institutional, and governmental levels. Common-use licensing approaches that promote broad access and reuse rather than restrict it, such as those being developed by the new Science Commons under the Creative Commons (see <http://science.creativecommons.org>), can preserve essential ownership rights while maximizing the social benefits and returns on the public investments in research. They can help to achieve a productive balance between the domains of proprietary R&D and publicly funded open science, particularly in a highly protectionist intellectual property environment.

Contribution of new ICT tools

BRCs are important primary providers of relevant information about the microorganisms that are publicly available through a global network of biological resource centers, but they are definitely not the only ones out there. Much more downstream information resides scattered over a veritable cottage industry of databases across the Web¹⁸. These information providers form the “knuckles” and “nodes” in a holistic approach to integration proposed by Lincoln Stein (Stein 2003). This knuckles-and-nodes strategy prescribes integration to take place both at the level of the primary information sources (the nodes) and some special-purpose lightweight secondary information sources (the knuckles), thereby creating a complex semantic information network that remains both scalable and flexible. Among other tasks, the knuckles may serve as transparent hubs to primary information sources that provide data on the same type of objects and coordinate mappings between different objects in the information network. More detailed levels of information about a given object are then provided by one or more of the nodes.

Following the knuckles-and-nodes approach, the StrainInfo.net bioportal plays a key role in compiling a BRC hub as described above and establishing some of the important mappings wherein biological resources are involved (Van Brabant *et al.* 2006). The integration efforts underlying the StrainInfo.net bioportal, both with respect to the establishment of the biological resource knuckle and the diverse mappings to third-party nodes and knuckles, have proven to be very useful for its user community. Most notable is the improved discoverability

Science: Toward Institutional Policies and Guidelines for Action” [details of the Workshop rationale and proceedings, are available at: <http://www.codataweb.org/UNESCOmtg/index.html>]. The response from the participants in that Workshop (which was organized by CODATA with the joint sponsorship of ICSU, ICSTI, INASP, UNESCO, and TWAS, with the collaboration of the OECD) has led to the development of this Initiative.

¹⁸ See the ExpASy Life Science Directory (us.expasy.org/alinks.html) and the Organism-Specific Genome Databases (mbcf.dfci.harvard.edu/cmsmbr/biotools/biotools10.html) to get an idea of the scale of the landscape.

yielded by the bioportal. Several roadblocks however lie in the way of more advanced integration scenarios and improved quality of the integration results.

Synchronisation between the StrainInfo.net bioportal and the individual BRCs is currently mediated through screen scraping. This somewhat mediaeval approach is justified by the fact that it does not put any requirements on the culture collections, other than having an online catalogue that can be smoothly parsed. As screen scraping is simply the only possible solution that allowed for an instantaneous implementation, it definitely lacks the scalability required to process the more than 500 BRCs that are currently registered with the World Data Centre for Microorganisms (WDCM) at an acceptable rate. What is clearly missing here is a standardised data exchange format to swiftly transfer information between the culture collections and the bioportal. The CABRI standard that was recently adopted by the OECD as a guideline for BRCs (OECD 2007) lacks the expressiveness of present-day ontologies to be considered as a mature candidate.

Another considerable impediment to integration of the biological resource knuckle, as well as to mapping it to external knuckles and nodes, is the presence of inconsistencies and missing data in the information produced by primary information providers. Although the StrainInfo.net bioportal was carefully designed to tackle this kind of problems by incorporating clever inference techniques and novel error detection/correction algorithms (Dawyndt *et al.* 2005), it would be naive to assume that its integration efforts were free from false positives and false negatives. A collaborative solution to this problem that is rapidly gaining popularity in the Internet community is to give the power to the user community to manually correct (or at least report) inconsistencies whenever they are encountered, and to fill in missing information whenever they have it at their disposal.

This form of collaborative action is definitely finding its way into a new generation of Internet applications, an evolution that is nicknamed Web 2.0. One can see early signs of this countervailing trend in open data projects such as Wikipedia, the Creative Commons, and in software projects like Greasemonkey, which allow users to take control of how data is displayed on their computer. This is the idea of the participative economy rather than the consumer economy, and of user-created content rather than mere end use. The collaborative nature of Web 2.0 is a revolution that the StrainInfo.net bioportal is ready to address. As the bioportal further unfolds itself as a helpful add-on to the microbiologists' toolbox, we therefore hope to gather around it a growing community that might instead become active contributors to both its content and implementation.

URL references

ExpASy Life Science Directory, <http://us.expasy.org/alinks.html>, an extensive list containing pointers to information sources for life scientists.

Genomes OnLine Database (GOLD), <http://www.genomesonline.org>, the reference resource for comprehensive access to information regarding completed and ongoing genome projects around the world.

International Nucleotide Sequence Database Collaboration (INSDC), <http://www.insdc.org>, has developed and maintained the International Nucleotide Sequence Databases for over 18 years as a collaborative effort among DDBJ, EMBL and GenBank.

Organism-Specific Genome Databases, <http://mbcf.dfci.harvard.edu/cmsmbr/biotools/biotools10.html>, is a compendium of Internet-accessible organism-specific genome databases and gene map/linkage analysis tools.

StrainInfo.net bioportal, <http://www.StrainInfo.net>, a one-stop-shop to navigate the microbial information landscape.

References

- Arzberger, P., Schroeder, P., Beaulieu, A., Bowker, G., Casey, K., Laaksonen, L., Moorman, D., Uhler, P., Wouters, P. (2004). Promoting Access to Public Research Data for Scientific, Economic, and Social Development. *Data Science Journal* 3, 135-152.
- Brahy, N. (2006). Databases and protecting traditional knowledge. *International Social Science Journal*, Vol. 188, June 2006, pp. 259-282.
- CGIAR (2003). *Booklet of CGIAR Centre Policy Instruments, Guidelines and Statements on Genetic Resources, Biotechnology and Intellectual Property Rights, Version II*, produced by the System-wide Genetic Resources Programme (SGRP) with the CGIAR Genetic Resources Policy Committee, Rome, July 2003, 37pp.
- Cook-Deegan, R., Dedeurwaerdere, T. (2006). The science commons in life science research: structure, function and value of access to genetic diversity. *The International Social Science Journal* 188 (June), 299-318.
- David P.A. (2005). Koyaanisqatsi in Cyberspace: The Economics of an “Out-of-Balance” Regime of Private Property Rights in Data and Information. In K.E. Maskus and J.H. Reichman, *International Public Goods and Transfer of Technology*, Cambridge University Press, pp. 81-120.
- David, P.A., Spence, M. (2003). Towards institutional infrastructures for e-Science : The scope of the challenge. Oxford Internet Institute, Research Report N°2, September 2003, 98 pp.
- 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.
- Esanu, J.M., Uhler, P.F., eds. (2003). *The Role of Scientific and Technical Data and Information in the Public Domain*. Washington, DC: National Academies Press.
- Fowler C., Smale M., Gaiji Samy (2001). Unequal Exchange? Recent Transfers of Agricultural Resources and their Implications for Developing Countries. In *Development Policy Review*, 19(2), pp. 181-204.
- Gilks, W.R., Audit B., de Angelis D., Tsoka S. and Ouzounis Ch. A. (2005), “Percolation of annotation errors through hierarchically structured protein sequence databases”, *Mathematical Biosciences* 193 (Feb): 223-34.
- Greer, D. and Brian Harvey (2004), *Blue genes. Sharing and conserving the world’s aquatic biodiversity* (London: Earthscan).
- Grifo, F. T. and D. R. Downes (1996), Agreement to Collect Biodiversity for Pharmaceutical Resource: Major Issues and Proposed Principles, in Brush, S. and D. Stibansky, Eds, *Valuing Local Knowledge*, Washington D.C. Island Press.
- Hess, Charlotte and Elinor Ostrom (2006), A framework for analysing the microbiological commons, *International Social Science journal*, 335-349.
- OECD (2007). *OECD best practice guidelines for biological resource centres*. Organisation for Economic Co-operation and Development.
- Pennissi, E. (1999). Keeping Genome Databases Clean and Up to Date *Science* 286 (15 Oct): 447-450.
- Principe, P.P. (1989) The Economic Significance of Plants and Their Constituents as drug, in H. Wagner, H. Hikino and N.R. Farnsworth (eds.), *Economic and Medicinal Plant Research*, Volume 3, pp. 1-17, Academic Press, London, U.K.
- Rai, A.K. (1999). Regulating Scientific Research: Intellectual Property Rights and The Norms of Science. *Northwest University Law Review* 77, 77–152.

- Rai, A.K., Eisenberg, R.S. (2003). Bayh–Dole Reform and the Progress of Biomedicine. *Law and Contemporary Problems* 66 (1/2), 289–314.
- Reichman, J.H., Uhler, P.F. (1999). Database Protection at the Crossroads: Recent Developments and Their Impact on Science and Technology. *Berkeley Technology Law Journal*, 14 (2), 793, 799–821.
- Reichman, J.H., Uhler, P.F. (2003). A Contractually Reconstructed Research Commons for Scientific Data in a Highly Protectionist Intellectual Property Environment. *Law and Contemporary Problems* 66 (1-2), 315–440.
- Roa-Rodríguez, C. and T. van Dooren (2007). The Shifting Common Spaces of Plant Genetic Resources in the International Regulation of Property. In Preparation (Manuscript with authors).
- Stein LD (2003). Integrating biological databases. *Nature Reviews Genetics* 4(5), 337-45.
- Stiglitz, J., Orszag, P.R., Orszag, J.M. (2000). *The Role of Government in a Digital Age*. Report for the Computer and Communications Industry Association, Washington DC.
- Swanson, T. (1997). *Global action for biodiversity: an international framework for implementing the convention on biological diversity*, London : Earthscan .
- The Digital Archiving Consultancy, The Bioinformatics Research Centre, The National e-Science Centre (2005). *Large-scale data sharing in the life sciences*. August 2005, 75 pp.
- Van Brabant B, Dawyndt P, De Baets B, De Vos P (2006). A Knuckles-and-nodes approach to the integration of microbiological resource data. *Lecture Notes in Computer Science*, 4277, 740-750.