Revolution and Evolution in Scientific Information: Moving from Restrictive Dissemination of Publicly-Funded Knowledge to Open Knowledge Environments

COMMUNIA Conference
Torino, Italy
29 June 2009

by
Paul F. Uhlir
Director, Board on Research Data and Information
National Research Council, U.S.
Washington, DC
puhlir@nas.edu

Revolution and Evolution in Scientific Information

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

Revolution and Evolution in Scientific Information

Principles for deconstruction of institutional mechanisms for scholarly communication and reconstruction in networked context:

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
6. Maintain 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, long-term preservation and sustainability)

Conclusion: Open access and unrestricted reuse of research data and information produced from public funding online is in most cases far superior to proprietary and restricted dissemination, which maximizes value for the disseminating organizations rather than for the content producer and user community.

Revolution and Evolution in Scientific Information

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

- **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 data sets when data from multiple sources are combined;
- **Promotes capacity building in developing countries and global research;**
- **Supports economic growth and social welfare; and**
- **Generally provides greater returns from public investments in research.**

Broad implications of excessive restrictions (economic, legal, technical) on access to and reuse of data and information from public and academic sources:

1) **Higher research costs (monopolization of public goods, transaction costs)**
2) **Lost opportunity costs (automated knowledge discovery, failure to capture full benefits of public investments)**
3) **Barriers to innovation (new uses and serendipity limited)**
4) **Less effective scientific cooperation and education**
5) **Widening gap between OECD and developing countries**

Openness **thus should be the default rule, subject only to legitimate and well-justified exceptions. But how to get there?**

Revolution and Evolution in Scientific Information

Existing digital commons models and emerging open knowledge environments:

- **Open-source software movement (e.g., Linux and 10Ks of other programs worldwide, many of which originated in academia for research applications);**
- **Distributed Grid computing or *e*-science (e.g., LHC@Home);**
- **Open data centers and archives (e.g., GenBank, EROS Data Center);**
- **Federated open data networks (e.g., World Data Center System, Global Biodiversity Information Facility, NASA DAACs);**
- **Open access journals (e.g., PLOS + > 4000 scholarly journals, many in developing world—SciELO, Bioline International, Hindawi Press);**
- **Open repositories for an institution's scholarly works (+ > 300 formally registered globally on Open DOAR, plus 1000s more not registered)**
- **Open repositories for publications in a specific subject area (e.g., the physics arXiv, CogPrints, PubMedCentral in US and UK);**
- **Free university curricula and lectures online (e.g., the MIT OpenCourseWare);**
- **Emerging discipline or applications commons, peer production of info, and integrated thematic open knowledge environments (e.g., virtual observatories, wiki encyclopedias, subdiscipline OKEs).**

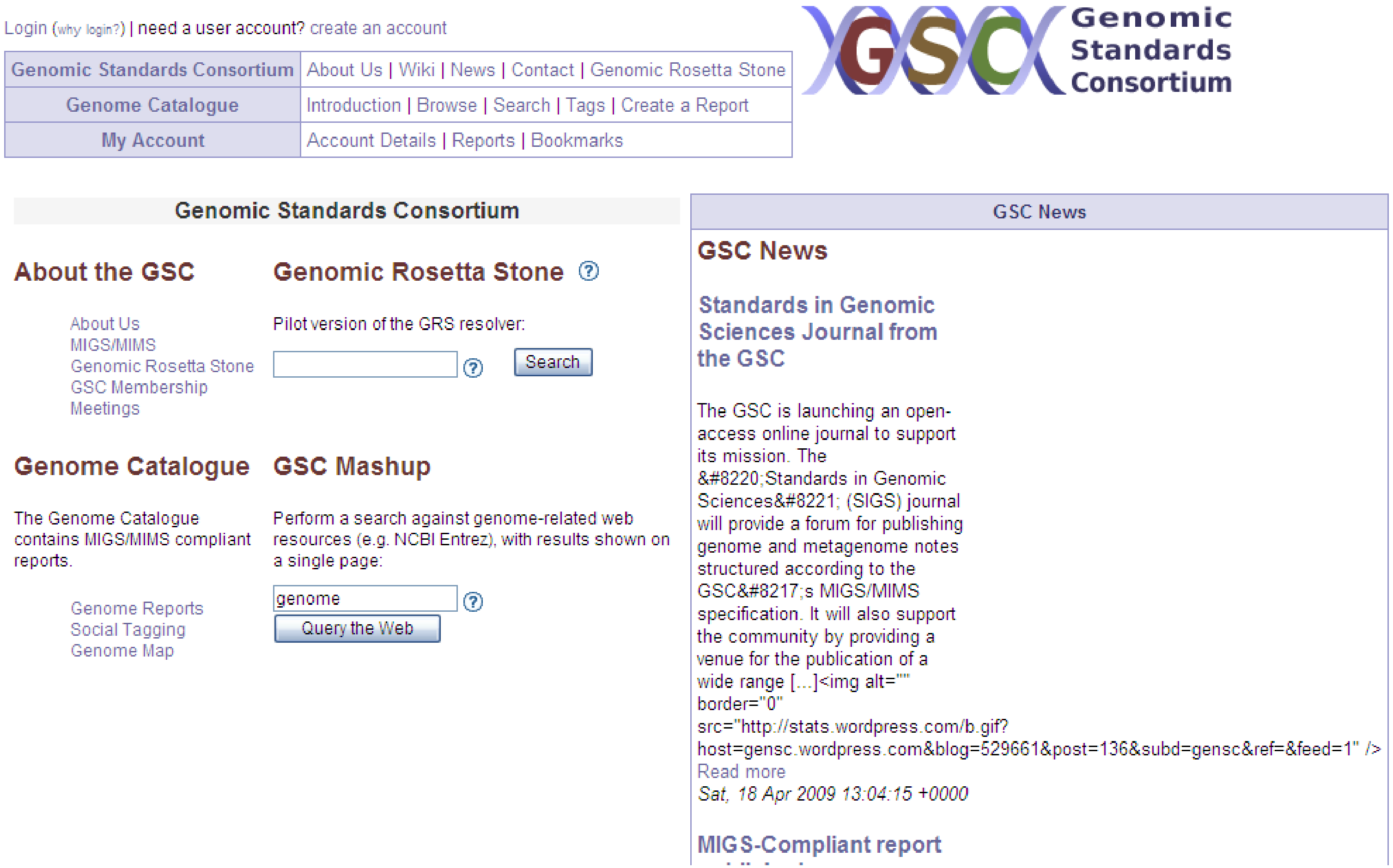
Revolution and Evolution in Scientific Information

Vision for open knowledge environments (OKEs) at universities

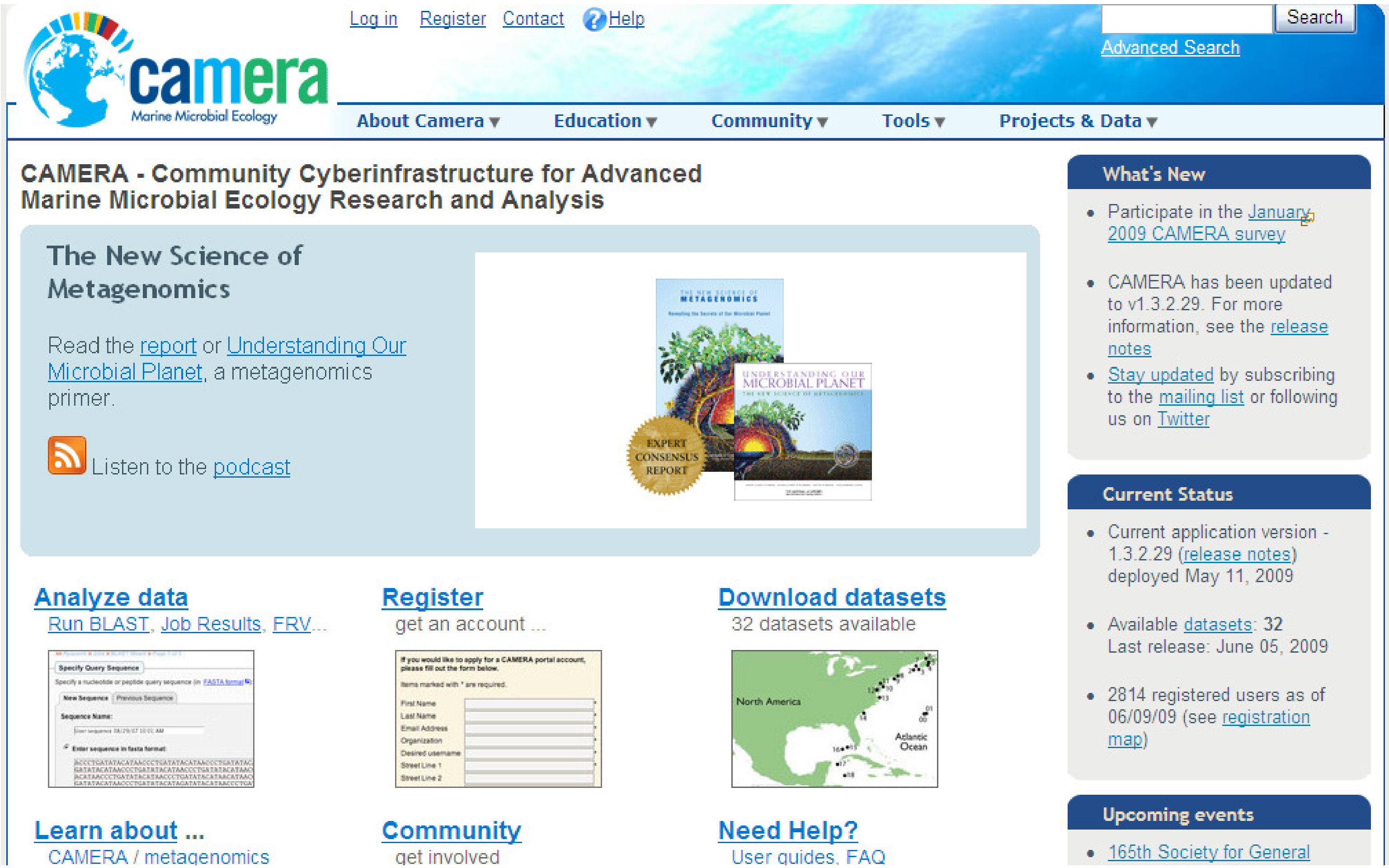
The restructuring of the print paradigm journal system through the formation of thematic OKEs in all universities:

- Organized around OA journals, gray literature, databases, OSS, and peer production of information in a focused thematic area.
- Supporting and integrating the university mission of public knowledge creation, dissemination and use, and of education (enhanced U.S. law school prototype).
- Common use licensing of content and tools (e.g., CC, GNU), and technically optimized (semantic web) for broad access and reuse.
- In-house and external OA content augmented by interactive collaboration tools in OKE, coupled with effective social networking and outreach.
- Managed by academic departments that integrate domain discipline(s), computer engineers, information scientists, libraries, and other collaborating departments at one or more universities (a consortium).
- Involving professors, students, and possibly external consultants and services (e.g., STM publishers, but that do not capture the content).

Revolution and Evolution in Scientific Information



Revolution and Evolution in Scientific Information



Limitations on creating OKEs at universities:

- Implementation and acceptance of new policy and institutional frameworks, frequently with conservative management and socio-cultural milieu.
- Development of adequate incentives for participation in OKE formation and use at the individual, community, institutional, and governmental levels.
- Long-term financial sustainability of different OKE models (university OKEs should have low cost, high positive externalities).
- Overcome pressures in universities to commercialize the OKE (e.g., by University Presses).
- In all cases, must balance with legitimate countervailing values and legal restrictions (protection of national security, privacy, confidentiality, and IPRs in bona fide commercial opportunities).

Additional works by the author on this topic (all available freely online):

- ***Bits of Power: Issues in Global Access to Scientific Data (NAS, 1997)***
- ***The Role of S&T Data and Information in the Public Domain (NAS, 2003)***
- ***Reichman, J.H. and Paul F. Uhlir, "A Contractually Reconstructed Research Commons for Scientific Data in a Highly Protectionist Intellectual Property Environment, 66 Law & Contemporary Problems 315-462 (2003)***
- ***UNESCO Policy Guidelines for the Development and Promotion of Governmental Public Domain Information (2004)***
- ***Open Access and the Public Domain in Digital Data and Information for Science (NAS, 2004)***
- ***Strategies for Open Access to and Preservation of Scientific Data in China (NAS, 2006)***
- ***Uhlir & Schröder, "Open Data for Global Science", Data Science Journal, CODATA, (2007).***
- ***Reichman, J.H., Tom Dedeurwaerdere, Paul F. Uhlir, "Designing the Microbial Research Commons: New Strategies for Accessing, Managing, and Using Essential Public Knowledge Assets" (forthcoming).***