Creating a Global Information Commons for Public Science

Salvador, Brazil
21 September 2005

by
Paul F. Uhlir
Director, Office of International Scientific and Technical Information Programs
The National Academies
Washington, DC
USA
puhlir@nas.edu

Creating a Global Information Commons for Public Science

Key aspects of scientific research using global digital networks:

- **Data-driven science through collection and creation of ever-increasing amounts and types of raw data;**
- **Interpreting and transforming the data into unlimited new configurations of information and applications;**
- **Collapsing the space and time in which data and information can be made available and used to advance science;**
- **Making the marginal cost of instantaneous and global dissemination of information near zero; and**
- **Facilitating entirely new forms of distributed research collaboration and information production.**

Creating a Global Information Commons for Public Science

Comparison of some key characteristics of the print 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, with time/space collapsed
limited content and types	unlimited contents and multimedia
distribution difficult, slow	effortless, immediate availability
copying cumbersome, not perfect	copying easy and identical
significant marginal distribution cost	zero marginal distribution cost
single user (or small group)	multiple, concurrent users/producers

Creating a Global Information Commons for Public Science

New forms of distributed collaborative research and information production on the Internet:

- **Open-source software movement (e.g., Linux and 10Ks of other programs worldwide, many of which originated in academia);**
- **Distributed Grid computing (e.g., SETI@Home, LHC@home);**
- **Community-based open peer review (e.g., Journal of Atmospheric Chemistry and Physics); and**
- **Collaborative research Web sites and portals (e.g., NASA Clickworkers, Wikipedia, Project Gutenberg).**

Creating a Global Information Commons for Public Science

New forms of open dissemination online:

- **Open data centers and archives (e.g., GenBank, Chinese Meteorological Agency, space science data centers);**
- **Federated open data networks (e.g., World Data Centers, Global Biodiversity Information Facility; ILTER);**
- **Virtual observatories (e.g., the International Virtual Observatory for astronomy, Digital Earth);**
- **Open access journals (e.g., BioMed Central, Public Library of Science, + > 1500 scholarly journals, many in developing world);**
- **Open institutional repositories for that institution's scholarly works (e.g., the Indian Institute for Science, + > 400 globally);**
- **Open institutional repositories for publications in a specific subject area (e.g., PubMedCentral, the physics arXiv); and**
- **Free university curricula online (e.g., the MIT OpenCourseWare).**

Creating a Global Information Commons for Public Science

Compelling reasons for placing government-generated data and information in the public domain and under open access conditions:

- **A government entity needs no legal incentives from exclusive property rights to create information. Both the activities that the government undertakes and the information produced by it in the course of those activities are a [global] public good.**
- **The taxpayer has already paid for the production of the information. Burden of additional access fees falls diisproportionally on the individuals least able to pay.**
- **Transparency of governance is undermined by restricting citizens from access to and use of public data and information. Rights of freedom of expression are compromised by restrictions on re-dissemination of public information, particularly of factual data.**
- **Numerous economic and non-economic positive externalities—especially through network effects—can be realized on an exponential basis through the open dissemination of public-domain data and information on the Internet. Conversely, the commercialization of public data and information on an exclusive basis produces de facto public monopolies that have inherent economic inefficiencies and are contrary to the public interest on other social, ethical, and good governance grounds.**

Economic Comparison of U.S. and European Public Sector Information
PIRA International, 2000

	EU	US
Investment Value in PSI	**9.5 billion Euro/year**	**19 billion Euro/year**
Economic Value	**68 billion Euro/year**	**750 billion Euro/year**

Creating a Global Information Commons for Public Science

Free and open access to publicly-funded scientific data and information online:

- Facilitates transfer of information from more developed to developing countries,
- Promotes capacity building in developing countries,
- Promotes interdisciplinary, inter-sectoral, and international research,
- Avoids duplication of research and 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,
- Enables the exploration of topics not envisioned by the initial investigators,
- Permits the creation of new data sets when data from multiple sources are combined, and
- Generally helps to maximize the research potential of new digital technologies and networks, thereby providing greater returns from the public investment in research.

Creating a Global Information Commons for Public Science

Countervailing polices and practices that may limit the free and unrestricted access to and use of government information:

- **1. Statutory exemptions to public-domain access and use based on national security concerns, the need to protect personal privacy, and to respect confidential information (including limited periods of exclusive use by researchers prior to publication of results), plus other exemptions to Freedom of Information Laws.**
- **2. Government-generated data are not necessarily provided free, even if there are no restrictions on reuse.**
- **3. Government agencies should not directly compete with the private sector in providing information products and services.**
- **4. Government agencies should respect the proprietary rights in data and information originating from the private sector that are made available for government use, unless expressly exempted.**

Creating a Global Information Commons for Public Science

Government-funded databases and research articles in academia not under direct control of government sponsors. However:

- **Motivations for public-interest research of academic and not-for-profit scientists generally are not market-driven and are rooted predominantly in intellectual curiosity, the desire to create new knowledge, peer recognition and career advancement, and the promotion of the public interest.**
- **These values and goals are best served by the maximum availability and distribution of the public research data and information, at the lowest possible cost, with the fewest restrictions on use, and the promotion of the reuse and integration of the fruits of existing results in new research. The dissemination of scientific data and information under a policy of full and open access by the government and academic sectors, reflects these values and serves these goals.**

Creating a Global Information Commons for Public Science

Pressures on the public domain status and on open access to publicly funded scientific data and information

Economic

- **Privatization of government research and data collection activities**
- **Commercialization of academic research (and of government data outside U.S.)**
- **Long-term sustainability of data dissemination/archiving activities**

Legal—IP laws have become broader, deeper, longer

- **Copyright**
- **Restrictive licenses**
- **Statutory database protection**
- **Other restrictive laws, especially those based on national security**

Sociological—competitive behavior, traditional discipline cultures/norms

Creating a Global Information Commons for Public Science

Broad implications of excessive restrictions on access to and reuse of data and information from public research:

3) **Disadvantage and marginalization of developing country researchers.**
4) **Significant lost opportunity costs, and the related failure to capture maximum value from public investment in public research activities.**
5) **Sole-source problems exacerbated in scientific database and journal markets, both public and private.**
6) **Higher transaction costs (not just cost of access).**
7) **Less effective international, inter-institutional, and interdisciplinary cooperation using digital networks.**

Creating a Global Information Commons for Public Science

Create a legal and policy regime that supports the open availability and unfettered use of publicly funded scientific data and information produced by government and government-funded sources by:

- **Expressly prohibiting intellectual property protection of information produced directly by government;**
- **Contractually reinforcing the sharing norms of science through open data terms in government research grants and contracts, and deposit of scientific literature in open repositories;**
- **Maintaining a large and robust public domain for non-copyrightable data, as well as other immunities and exceptions under IP law favoring science and education;**
- **Treating data/info resources as a major research infrastructure investment, with appropriate institutional mechanisms/support**
- **Disseminating data/info freely online.**

Creating a Global Information Commons for Public Science

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

- ***Bits of Power: Issues in Global Access to Scientific Data (NAS, 1997)***
- ***A Question of Balance: Private Rights and the Public Interest in Scientific and Technical (S&T) Databases (NAS, 1999)***
- ***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)***

All NAS reports available at www.nap.edu