Open Knowledge Environments

Academia Sinica
Taipei
2017 December 11

by
Paul F. Uhlir, J.D.
Consultant, Information Policy and Management
New York, USA
www.paulfuhlir.com

CC BY

Open Knowledge Environments

Focus of my remarks:

- Changing context for scholarly communication
- Guiding principles for the (relatively) new context
- The value of openness
- Legitimate and non-legitimate limitations on openness
- Established and emerging models
- The Open Knowledge Environment (OKE) model

CC BY

Open Knowledge Environments

What is "**scholarly communication**"?

The transfer of knowledge among scientists and scholars, and between them and the general public, increasingly machine intermediated.

"Fixed" mechanisms:

- **Scholarly literature**, both peer reviewed and not—journals, books, monographs, conference presentations, reports, etc.
- **Datasets and databases**—factual compilations, both centralized and decentralized, made useable with metadata.
- **Software tools**—programs for managing research information.
- **Multimedia tools**—visualization, animation, web sites, video, music, social media, etc.

CC BY

Open Knowledge Environments

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	nearzero marginal distribution cost
single user (or small group)	multiple, concurrent users
centralized production	distributed production
slow knowledge diffusion	accelerated knowledge diffusion

CC BY

Open Knowledge Environments

Principles for deconstruction of print-paradigm models for scholarly communication and reconstruction in the digitally networked context:

Economic Principle 1

Maximize the (global) public good aspects of
publicly funded research data, literature,
and other information products

Public good = cannot be depleted and excluded

Quasi public good = cannot be depleted, but inefficient to exclude

CC BY

Open Knowledge Environments

Economic Principle 2

Take advantage of the near-zero marginal cost for (global) dissemination in the public interest and in the researcher's own interest

CC BY

Open Knowledge Environments

Economic Principle 3

Bridging the “digital divide” must include reducing the “content divide”

CC BY

Open Knowledge Environments

Legal and Ethical Principle 1

Protect only legitimate and required interests.

CC BY

Open Knowledge Environments

Legal and Ethical Principle 2

Keep the intellectual property rights with the author/producer
and not with the distributor
(promote a service, not a captured product)

CC BY

Open Knowledge Environments

Political Principle

Transparency of governance is undermined by restricting citizens from access to and use of public information, especially factual data

CC BY

Open Knowledge Environments

Scientific Management Principle

Maintain and promote characteristics that are essential to the research community and the progress of science:

- promote freedom of inquiry
- maintain rigorous quality control through peer review/editing
- enhance research impact and reputational benefits
- promote speed of publication
- place premium on ease of access and dissemination
- ensure long-term preservation, if needed

CC BY

Open Knowledge Environments

There are many scientific advantages of open access to and unrestricted reuse of publicly generated or funded research data and information on digital networks:

- **Promotes interdisciplinary, inter-institutional, and international research;**
- **Enables automated knowledge discovery;**
- **Avoids inefficiencies, including duplication of data collection and 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.**

CC BY

Open Knowledge Environments

Legitimate restrictions on public access to or use of government or government funded data and information for protection of:

- National security and public safety
- **Personal privacy**
- Confidentiality
- Proprietary rights of private-sector parties or commercial potential of research
- Exclusive use of PI data prior to publication-embargo periods
- Restrictions for specific reasons (e.g., endangered species, archeological digs, specific indigenous cultural rights)

CC BY

Open Knowledge Environments

Other legitimate practical concerns of researchers for openly sharing research data include:

- Unfunded mandates to make the data useful to others (i.e., a lack of money and resources).
- Lack of technical means to share the data.
- Lack of recognition and rewards from home institution.
- The potential for being "scooped" by other researchers if the data are released *before* publication.
- Lack of recognition and proper attribution by the users of the dataset.
- Concerns about the misuse of the data.
- General lack of a data sharing ethos in the immediate community or discipline.

CC BY

Open Knowledge Environments

Restrictions that are not legitimate:

- Hoarding data that are not legally restricted (not releasing them *after* publication, even when asked).
- Hiding bad data or erroneous research results based on the data.
- Fear of "plagiarism".
- Socio-cultural attitudes against collaboration and data sharing, even when the research demands it (i.e., absence of a data sharing ethos).
- Spurious reasons prohibiting open data sharing by the researcher's employer or government.
- Use of "other peoples' data".

CC BY

Open Knowledge Environments

Broad implications of excessive restrictions (economic, legal, political, scientific & technical) on scholarly communication using 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?**

CC BY

Open Knowledge Environments

What is a digital commons?

Digital data and information originating principally from government or publicly-funded sources;

- Made freely and openly available for broad, common use online;
- Without reuse restrictions, with the material in the public domain, or with only some rights reserved (using waivers of rights or common-use licenses, such as Creative Commons); and
- Typically organized thematically.

CC BY

Open Knowledge Environments

Existing digital commons models for different scholarly information types:

- **Open-source software movement** (e.g., Linux and 10Ks of other programs worldwide, many of which originated in academia for research applications);
- **Open data centers and archives** (e.g., GenBank, Hubble Telescope archive);
- **Federated open data networks** (e.g., World Data System, Global Biodiversity Information Facility, Group on Earth Observations);
- **Open access journals** (e.g., > 10,500 scholarly journals, in both more and less developed countries registered in DOAJ);
- **Open repositories for an institution's scholarly works** (+ > 4000 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);
- **E-government initiatives** (Data.gov in US, many others worldwide); and
- **Emerging integrated discipline or applications commons, peer production of information, and integrated thematic open knowledge environments** (e.g., virtual observatories, wiki encyclopedias, sub-discipline OKEs).

CC BY

Open Knowledge Environments

Integrated Model: Open Knowledge Environments (OKEs) at Universities

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

- **Mostly publicly funded** through grants and contracts and operated **as full commons**.
- **Organized around OA resources**—journals, gray literature, data, 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.
- **Common-use licensing** of content and tools or waivers of rights (e.g., Creative Commons, GNU), **and technically optimized 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 primarily professors and students**, and possibly external consultants and services (e.g., STM publishers, but that do not capture the content).

CC BY

Open Knowledge Environments

One example of an OKE:

1) Genomic Standards Consortium (GSC) See: gensc.org

- Development of community standards for (meta)genomics
- Worldwide members/contributors
- Many research projects
- SIGS—open access journal
- Wiki
- Publications/news
- Platform for data exchange/integration
- Convening international meetings
- Funding is mostly voluntary; some grants from government, foundations, and industry
- All open

CC BY

Open Knowledge Environments

Limitations on creating OKEs at universities and other public institutions:

- Implementation and acceptance of new policy and institutional frameworks, frequently with **conservative management practices**.
- 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 and high positive externalities to attract users and funding).
- **Overcome pressures to commercialize** the OKE (e.g., by University Presses).
- In all cases, must **balance with legitimate countervailing values** and related legal restrictions (protection of national security, privacy, confidentiality, and IPRs in bona fide commercial opportunities). Some portion of an OKE may be subject to proprietary periods of protection, either prior to publication or as a trade secret (if commercializing), or it may be handled as a closed semi-commons among members of a consortium.

CC BY

Open Knowledge Environments

Additional works by the author on this general 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).
- Bailey-Mathae and Uhlir , eds., *The Case for International Sharing of Scientific Data: A Focus on Developing Countries* (NAS, 2011).
- Reichman, J.H., Paul F. Uhlir &Tom Dedeurwaerdere, *Governing Digitally Integrated Materials, Data, and Literature: Global IP Strategies for a Redesigned Microbial Research Commons:* (Cambridge University Press, 2016).

CC BY