

Open Bibliography and Community Information Services

Jim Pitman

Mathematics and Statistics, U.C. Berkeley

Workshop on Mathematics Journals

February 16, 2011

Open Bibliography

The [Open Bibliographic Data Working Group](#) of the [Open Knowledge Foundation](#) has published a set of principles for open bibliographic data:

<http://openbiblio.net/principles/>

These principles express a philosophy of openness for bibliographic data in support of research and knowledge enhancement:

*For society to reap the full benefits from bibliographic endeavors, it is imperative that bibliographic data be made open
— that is, available for anyone to use and re-use freely for any purpose.*

- Cross-disciplinary movement, comparable to the OA movement in potential impact
- Endorse the principles <http://openbiblio.net/principles/endorse/>
- Practice the principles by publication of significant biblio datasets with a [CC0 License](#)

Two quotes:

- **Books should only be catalogued once.** Currently the public purse pays for having the same book catalogued over and over again. Librarians should act as they preach: data sets created through public funding should be made freely available to anyone interested. Open Access is natural for us, here at CERN we believe in openness and reuse... By getting academic libraries worldwide involved in this movement, it will lead to a natural atmosphere of sharing and reusing bibliographic data in a rich landscape of so-called mash-up services, where most of the actors who will be involved, both among the users and the providers, will not even be library users or librarians
- Jens Vigen, Head of the CERN Library ([CERN Library publishes its book catalog as Open Data](#)) ([YouTube](#))

- It may be helpful to visualize my motivation as from a scientist who until recently had no interaction with mainstream library practice. The motivation springs from the fact that secondary publishers use metadata to control our actions and also charge us money for it. **We live in occupied territory.**
 - [Peter Murray-Rust](#) on the Open Bibliographic Mailing List, December 19, 2010. ([petermr's blog](#))

Community Information Services

Problem posed at this meeting by Bernard Teissier:

- **How to organize all the data?**

Partial solution (developed with NSF sponsored [Bibliographic Knowledge Network Project](#))

- **Empower communities of practice to take control of their bibliographic data**
- **Enable network effects in such communities by open biblio data principles**

Other components: Google, Microsoft Academic Search, ...

Exemplars:

- **RePEc:** <http://repec.org/>
- **Probability Abstract Service** <http://pas.imstat.org>
- **Probability Web** <http://probweb.berkeley.edu/>
- **Departmental Services** (Berkeley, Toronto, Oxford, Sydney,)

Requirements:

- **Adequate software, editorial commitment, and open biblio data.**

Fundamental Problem for AMS:

How to accomodate community information services?

Proposed solution: provide monthly metadata and identifier dumps from MathSciNet at some URL, under CC0

Benefit to the community: Allows others to

- extract and develop whatever data they care about
- link back to MathSciNet for reviews, authoritative identifiers, library quality service
- mash-up the data in community-specific ways (ratings, rankings, ...)
- provide APIs and visualizations over the data
- connect to the web of Linked Data <http://linkeddata.org/>

Cost to AMS: Close to zero provided libraries continue to support MathSciNet

Benefit to AMS: Engagement with community supported interest groups (cf. ASA, SIAM)

What if MathSciNet (or any other data provider) does not cooperate with community services?

- The tide of open biblio data will rise anyway ([ORCID](#), CERN, British Library, ...)
- The value of closed sources will diminish.
- Community hostility to closed information services will rise.
- Libraries should refuse to sign license agreements which restrict scripted access or substantial downloads.
- Communities can work with Thomson WoS and Scopus which already have APIs allowing companies like [Symplectic](#) to provide university-wide bibliographic services.

Basic Strategy for Bibliographic Data Liberation

Given a raw bibliographic item, machine match it into a variety of information services:

- save copies of all records obtained
- publish a composite record back to the web under CC0 with links to all sources
- defy any source to complain

Legal Justification (http://en.wikipedia.org/wiki/Feist_v._Rural)

Under U.S. law, most copyright and licensing restrictions on bibliographic metadata appear to be largely voided by

[*Assessment Technologies v. Wiredata*](#) (2003) in which the [Seventh Circuit Court of Appeals](#) ruled

- a copyright holder in a compilation of public domain data cannot use that copyright to prevent others from using the underlying public domain data, but may only restrict the specific format of the compilation, if that format is itself sufficiently creative.
- It is a [fair use](#) of a copyrighted work to [reverse engineer](#) that work in order to gain access to uncopyrightable facts.
- It is a [copyright misuse](#) and an [abuse of process](#) if one attempts to use a contract or license agreement based on one's copyright to protect uncopyrightable facts.

