



# Free Library Data?

**Raymond Bérard**

Director, Agence bibliographique de l'enseignement supérieur (ABES),  
227 avenue Professeur Jean-Louis-Viala, 34193 Montpellier cedex 5, France,  
[berard@abes.fr](mailto:berard@abes.fr)

## Abstract

As library materials are catalogued by public organisations and librarians are active promoters of the principles of open access, one would expect library data to be freely available to all. Yet this is not the case. Why then do so few libraries make their data available free of charge? This article reviews the diverging, often restrictive policies and the interests (commercial and strategic) at stake. It presents a panorama of the current situation, the actors and interests involved. It addresses the legal aspects and the obstacles and it shows how data produced by libraries can be made freely available to other knowledge organisations while retaining and developing the collective organisations and services built by library networks over the years.

The aim of the 'free the data movement' is to share and reuse bibliographic data in a new ecosystem where all the actors are involved, both users and providers, not just librarians.

**Key Words:** bibliographic records; library data; open access; WorldCat; OCLC

## 1. A Hot Topic

Topics involving bibliographic records have long been restricted to the circle of librarians and to an even smaller circle: cataloguers. Cataloguing is a feature that is no longer fashionable nowadays with library managers looking for ways to cut costs to deal with major other challenges: digitisation,

institutional repositories, electronic resources etc. Yet library data are back on the stage: At the Berlin7 Conference in Paris (December 2009), metadata were placed on the same level as academic literature in a leap of the open access movement to library catalogues. And although involuntarily, OCLC has made metadata a subject of controversy with their abortive attempt at introducing a new policy for WorldCat records in October 2008. The new policy prompted an outcry from the library community around the world. Even the venerable *Guardian* dedicated its headline to metadata!<sup>1</sup>

Why this new craze for data produced by libraries? What are the academic and economic issues? Who are the actors involved? What are the claims and the expected changes?

## **2. Issues**

The issues raised by library data are commercial, ideological and political in nature:

- Commercial: there is a market for selling records to libraries and booksellers. The peculiarity of this market is that the records are produced mainly by public actors.
- Ideological: a growing number of actors believe that records should escape the business logic and be free and freely accessible. This move corresponds to the rise of web technologies to facilitate innovative uses of records. The ideological issue is an extension of the open access movement, as supported by Jens Vigen, Head of the CERN Library, who announced on 20 January 2010 that the records of the CERN library are now made available under the Public Domain Data License: '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.'<sup>2</sup>
- Political: library data are public data. Several governments are developing a policy to make public data freely available to promote innovation through the use and re-use of government data sets. The purpose is to increase public access to high-value, machine-readable data sets. *Data.gov* in the USA, *Data.gov.uk* in

the United Kingdom, and *Mashup Australia* are good examples while other countries are planning similar services for their public data.

### 3. The Context of Linked Data

Linked data must be placed in the context of a powerful movement that started with commercial products and tourism and opens the way to a new public service of raw data. The nature of linked data requires that you abandon control of your data: you expose them; you accept to lose control over who will use them, for what purpose; you allow new, innovative uses; you allow mash-ups. It is no coincidence that the first catalogues that applied the principle of linked data (Libris, Hungarian National Library) come from organisations well known for their commitment to open data.

The World-Wide Web Consortium (W3C) announced on 21 May 2010 the launch of a Library Linked Data Incubator Group ‘whose mission is to help increase global interoperability of library data on the Web, by bringing together people involved in semantic web activities — focusing on linked data — in the library community and beyond.’<sup>3</sup> The W3C Members who sponsored the charter for this group are well known for their innovations: Helsinki University of Technology, DERI Galway, the Competence Centre for Interoperable Metadata (KIM), the Library of Congress, Los Alamos National Laboratory, MIMOS, OCLC, Talis, the University of Applied Sciences Potsdam, and the Vrije Universiteit Amsterdam.

### 4. Actors

National libraries are the major suppliers of library records (see section 6). However, new stakeholders have emerged who actively promote open access to and reuse of library records:

- Open Library helps individuals build their own catalogues. It is a project of the non-profit Internet Archive built on open software and data, funded in part by a grant from the California State Library and the Kahle/Austin Foundation. To date, Open Library has gathered

over 20 million records from a wide variety of catalogues as well as single contributions.

- Biblios.net 'is a free cataloging service with a data store containing over thirty million records. Records are licensed under the Open Data Commons Public Domain Dedication and License, making the service the world's largest repository of freely-licensed library records'.<sup>4</sup> The CERN library announced that it will provide its data via Biblios.net. The service was created and is maintained by LibLime. A French company that has partnered with LibLime states that it does not actually sell records, because the fee they charge to libraries covers the *online service* (access to the cataloguing tool), not the records downloaded by libraries.
- LibraryThing is aimed at individuals rather than libraries. It 'is a social cataloging web application for storing and sharing personal library catalogs and book lists.'<sup>5</sup> LibraryThing was developed by Tim Spalding and it now comprises 920,000 users and nearly 45 million books catalogued. Data are imported through Z39.50 connections from booksellers and libraries including the Library of Congress, the National Library of Australia, the Canadian National Catalogue, the British Library, and Yale University. LibraryThing no longer belongs exclusively to Tim Spalding. Commercial companies have taken an interest in it with online bookseller AbeBooks (now owned by Amazon) buying a 40% share in LibraryThing in May 2006. In January 2009, Cambridge Information Group acquired a minority stake in the company and their subsidiary Bowker became the official distributor to libraries.<sup>6</sup> This development may have an impact on the use of records imported from external sources.

The private sector is also active on the market of library records: private companies seek to collect records for resale to their customers. OCLC, a not-for-profit-organisation, dominates the market, with metadata still representing 36% of its revenue in 2008/2009 (2003/4: 44%).<sup>7</sup> The metadata are produced by libraries and keyed into OCLC library systems; OCLC resells them. Other actors include booksellers or companies close to publishers: Casalini in Italy, Electre in France (a company owned by the French book trade association). New players have recently appeared on the market to threaten OCLC's dominant position: Skyriver is the most prominent

of them. It was established early 2010 and it promises to cut library expenditure for bibliographic services by as much as 40%. It claims it is 'a new bibliographic utility that offers a low-cost alternative for cooperative cataloging.'<sup>8</sup> Several US libraries, hit by cuts in public funding, have switched from OCLC to Skyriver, which holds 20 million records from the Library of Congress and the British Library. Skyriver was founded by Jerry Kline, owner of Innovative Interfaces, which provides administrative and infrastructure support to Skyriver. Innovative Interfaces recently filed an anti-trust suit against OCLC.<sup>9</sup>

## 5. Are Bibliographic Records Protected by Copyright?<sup>10</sup>

Complex legal issues surround the exchange of bibliographic records: records are produced by libraries, but libraries do not produce all the records in their catalogues themselves: they download a significant portion of them from external sources: national libraries, vendors, union catalogues, bookstores etc.

Three sets of legal rules may apply to library records:

### Copyright

An intellectual creation is protected by copyright when it materialises in an original form created by its author (e.g. in the choice of presentation, forms, colours, words used). Conversely, data are not protected by copyright when they are the result of technical constraints, either legal or contractual. Thus, the protection by copyright does not apply to raw information which only gives the facts without any interpretation or organization, e.g., lists of names, cities, figures, stock information, statistics.

An individual bibliographic record is not an 'original work', as the cataloguer should certainly not be creative: he is asked to enter strictly objective information in each field, in a fixed, standardised way. Copyright therefore does not apply to individual records.

Now what about sets of records? A whole set of records can only be protected by copyright if the data it contains are selected or arranged in a unique way.

As data in a bibliographic database are chosen and organised according to specific standards and are supposed to be exhaustive, a database of bibliographic records is not protected by copyright.

### **Protection of the Database Producer**

The content of a data base is protected by copyright when its producer can prove that he has made substantial investments to create and maintain the database (financial, technical and human resources). In this case, copyright benefits the investor, not the author. Copyright on databases prohibits any extraction or reuse of qualitatively or quantitatively substantial content from the database. The producer may claim his right to sell the data. A bibliographic database like WorldCat is protected by the producer's right.

### **Specific Case of Reuse of Public Data**

Records produced by libraries are public data. Public information is freely reusable for any purpose, whether private or public, commercial or not, free of charge or not. Any economic operator can reuse and redistribute public data in order to create a commercial value-added product. Public organisations can charge a fee for the reuse of public information by a private company. Under this rule, if a library is the sole producer of its records, it may transfer and make them available to anyone on a commercial basis or a non-commercial basis. This rule does not apply, however, to the records the library may have derived from external sources (national libraries, WorldCat, etc.): in this case it must respect the rights of the producer. Some national libraries are planning to outsource the production of some of their records or to reuse records produced by publishers. This will make it an even more complex issue as public actors will not be the sole producers of their records.

## **6. Overview of Some Suppliers' Conditions<sup>11</sup>**

Table 1 gives an overview of suppliers' conditions. Presently, quite a few national libraries are changing their business model: both the BL and the DNB have indicated that they are moving away from seeing records as a revenue source although they still restrict use at the moment. There is a general trend to a more open environment, publicly funded, along the lines reflected by Sweden (Libris).

Table 1: Suppliers' conditions

|  | Availability of metadata for reuse   | Cost  |
|--|--|---|
| British Library (BL)   | Records supplied exclusively under license   | Cost recovery in the UK, for-profit overseas for priced service options. Free for online access |
| Deutsche National Bibliothek (DNB)   | The business model is being changed right now. Until now the metadata may not be relicensed or redistributed for money | Cost recovery for special services that involve further manual labour                           |
| Swedish National Bibliography (Libris)   | No restrictions  | Free of charge  |
| Danish National Bibliography   | No restrictions  | Metadata are not priced, but handling costs related to delivery of records in files are         |
| Japan (National Diet Library) ISSN   | Records are supplied exclusively under license<br>License<br>Transfer to other libraries not allowed                   | Cost recovery. Free of charge online access<br>€€€€€  |
| WorldCat (Guidelines for the Use and Transfer of OCLC-Derived Records, 1987)                               | License  | €€€   |
| WorldCat (WorldCat Rights and Responsibilities for the OCLC Cooperative, Draft for community review. 2010) | Code of good practice for members  | €€€   |

## 7. The New OCLC Policy

Until July 2010, the policy for the use and transfer by libraries of OCLC-derived records was subject to 'Guidelines' dating from 1987. The text required revision to update it, reflect technological developments and take into account the new information landscape. A new draft policy for records was presented in 2008 to the OCLC Global Council. It sparked massive protests as the text was seen by the library community as a unilateral attempt to establish a monopoly and to restrict members' freedom to exchange data. The reactions prompted OCLC to consult its members once more and more widely. The Association of Research Libraries

(ARL) issued a well-argued report on the proposed new policy. Building on the ARL's recommendations, OCLC decided in September 2009 to withdraw the proposed new policy and establish a council of thirteen librarians, the so-called Record Use Policy Council (RUPC). Its charge was to propose new guidelines for the use and transfer of records. The RUPC produced a draft for community review and the final document was approved by the OCLC Board of Trustees in June 2010. It became effective 1st August 2010.

What are the main features of the new policy?<sup>12</sup>

- Its approach is different:
  - It is not a legal document but a code of good practice for members of a cooperative based on shared values, trust and reciprocity in understanding rights and responsibilities;
  - It focuses on member rights and responsibilities instead of detailed provisions or restrictions, with the general aim to foster innovation in our ever-changing information landscape;
  - Members can transfer their data to other libraries, cultural and academic institutions including OCLC members and OCLC non-members. Members can transfer their data to agents acting on their behalf;
  - It focuses on the value of the WorldCat database as a whole and its value to members in visibility of holdings, in support of resource sharing and other services without distinction between original cataloguing and WorldCat-derived records, or the ownership of individual records as the focus;
  - It includes a process for collective, regular review of the policy;
  - It details steps OCLC can take to address inappropriate use by members, the Global Council being the advisory body on how to proceed if no earlier resolution is available.
- The policy intends to encourage the widespread use of WorldCat bibliographic data while also supporting the ongoing and long-term viability and utility of WorldCat and WorldCat-based services; to enable and facilitate innovation; to maintain a balance between openness and boundaries.
- It considers WorldCat as a club (or membership) good, not a public good. A club good is shared by a community of stakeholders;

it defines conditions for access to benefits; it manages the ongoing supply of the good through mechanisms that distribute the cost of providing the good. A public good is freely available to all without restrictions; once available, there is no feasible way to exclude anyone from the good's benefits.

This policy marks a significant step forward. But in making WorldCat a 'good club', the policy will not satisfy the militants of open data. It is all a matter of balancing the interests of free sharing of records, enhanced by innovative uses that are emerging in many libraries, and the limits set to this freedom to preserve the economic viability of WorldCat.

In this context, one must distinguish between the database itself (support for multiple services) and the records, which are created by members. WorldCat is not just a reservoir; for libraries worldwide it represents a guarantee of international visibility and a range of services across the web (resource discovery in tens of thousands of libraries, harvesting by Google and Yahoo, APIs, tools for collection analysis etc.).

Opinions about WorldCat vary according to the uses made of it. The shift from WorldCat as a record supply service to a global network of data and services is a new way of thinking which is understood better in Europe than it is in the US. Many European networks have uploaded their catalogues to WorldCat but at the same time they have their own cataloguing platforms and browser interfaces (Sudoc in France, GBV in Germany etc.). The issue of control over records is more sensitive in Europe than in the US where libraries catalogue directly into WorldCat — the de facto North American union catalogue. The Europeans will not relinquish control over their records once they are in WorldCat. The RUPC has sought to strike a balance.

## **8. A Pragmatic Approach to Sudoc**

The French Agence bibliographique de l'enseignement supérieur (ABES) has taken a pragmatic approach with regard to Sudoc records. As suppliers' contracts can be very different, allowing for different uses, Sudoc members were sometimes confused because they failed to read the small print in the contracts and sometimes infringed upon their clauses. To make things easier, ABES asked Sudoc members to define their minimum requirements

## *Free Library Data?*

for the use of data. Sudoc members came up with five requirements. ABES then wrote to all its suppliers asking them to grant permission for the uses required by members.

Below is the list of uses ABES submitted to its suppliers (OCLC, DNB, ISSN, Helka, BnF and INSERM):

1. refer to all bibliographic records in the Sudoc catalogue;
2. copy and modify all bibliographic records describing documents from the library's collection in the Sudoc catalogue;
3. download all bibliographic records describing documents from the library's collection in its integrated library system;
4. download all bibliographic records describing documents from the library's collection in a union catalogue in which one or several libraries take part;
5. put online on the library's website the bibliographic records describing documents from its collection. In this case, bibliographic records have to be in a non-professional format and the library has to mention on its website the origin of the records.

All suppliers agreed to the five uses, except for ISSN, which did not agree to use no. 4.

## **9. Conclusion**

It is difficult to predict the future, but the movement for free access seems driven to win the game for library data, mainly because national libraries, which are the largest producers of data, are gradually moving to this new model.

Will the free access model challenge community achievements such as OCLC? I estimate that this will not happen in the near future, because the commitment of libraries to OCLC is strong. However, competition is developing in a climate of declining public budgets that may force libraries to explore the possibilities of competition between OCLC and vendors. OCLC urgently needs to invent a new economic model that allows it to rely less on the provision of records and more on services to libraries.

## Notes

---

<sup>1</sup> The Guardian, 22 October 2009.

<sup>2</sup> <http://library.web.cern.ch/library/Library/announcement.html>.

<sup>3</sup> <http://www.w3.org/2005/Incubator/lld/>

<sup>4</sup> Wikipedia.

<sup>5</sup> Wikipedia.

<sup>6</sup> Wikipedia.

<sup>7</sup> OCLC annual report 2008/2009.

<sup>8</sup> <http://theskyriver.com/>

<sup>9</sup> <http://www.librarytechnology.org/ltg-displaytext.pl?RC=14916>.

<sup>10</sup> The information on copyright is based on a study commissioned by ABES to Cabinet Alain Bensoussan: *Rapport d'audit sur la propriété de la base de données SUDOC*, 15 décembre 2008 (internal document); see also: Michèle Battisti, Anne-Laurence Stérin, 'A qui appartiennent les notices bibliographiques?', *Arabesques* 58, avril-juin 2010, <http://www.abes.fr/abes/DocumentsWebAbes/abes/arabesques/Arabesques58.pdf>

<sup>11</sup> <http://www.ifla.org/en/node/2216>.

<sup>12</sup> WorldCat Rights and Responsibilities for the OCLC Cooperative. OCLC, 2010. <http://www.oclc.org/us/en/worldcat/recorduse/policy/default.htm>