68th IFLA Council and General Conference
August 18-24, 2002

Cataloguing and organizing library workflow - New ways

Stina Degerstedt
Section of Bibliographic Development and Coordination
The Royal Library - National Library of Sweden
Stockholm, Sweden

Abstract:

The National Library of Sweden has come to agreements with Swedish book publishers about delivering full text versions of new e-books together with bibliographic information. As a service in return catalogue records are quickly made available in LIBRIS – the National Union Catalogue. In order to handle this flow of text files and metadata, new technical solutions has been developed – a first step toward new interesting possibilities.

Introduction

The Royal Library (KB) - Sweden's National Library - has at present an agreement with a small number of publishers to deliver e-books. In return, the Royal Library will promptly supply catalogue records to the national union catalogue LIBRIS. The terms state that the publishers are obliged to also deliver specific bibliographic data about the e-books. In this paper I will attempt to describe a technical solution and the workflow from publisher to a completed catalogue record. A workflow, which we have established in

collaboration with Uppsala University Library. I will also comment developmental options as we distinguish them today, but will begin by presenting some facts and background.

Background

The National Bibliography of Sweden and LIBRIS

The National Bibliography of Sweden is a part of LIBRIS, which stands for the holdings of Swedish research and special libraries as well as a number of public libraries. Access to the service via WebSearch [1] is free. The National Bibliography consists of catalogue records based on printed matter issued by publishers and delivered to the Royal Library by decree of the Legal Deposit Act.

The printers customarily deliver legal deposits. Nevertheless, as these deliveries usually appear long after the book has been published, the commercial publishers (via their distributors) will themselves deliver an extra copy to the Royal Library to assure that the bibliographical data is included in LIBRIS simultaneously to the book reaching the retailers. Swedish publishers need to see their products registered in LIBRIS - or else, suffer anonymity.

The Legal Deposit Act

The Legal Deposit Act came into existence 1661 and encompasses all printed matter intended for nationwide distribution. This requires us not only saving publishing house products, but also "ephemera" or "grey matter" such as flyers, timetables, postcards, annual reports etc. Latter day revisions to this Act have come to encompass image, sound recordings and electronic publications. However, the Act does not yet cover electronic publications published on the web, but a new Act should be in the pipeline in the near future.

Kulturarw[3] – collecting Swedish web pages

Kulturarw[3] [2] is a project aimed at long-term preservation of electronic information. With the aid of search robots, the Royal Library validates and files Swedish web pages in an archive. Since its implication in 1996, the robots have performed two sweeps each year. This comparatively slow pace of collecting unfortunately results in us not finding everything on the Internet, but sufficient enough to supply us with "snap shots" of how the Swedish Internet appeared at that particular time. The archive will be accessible for searches at the Royal Library.

In addition to these sweeps the robots are programmed to regularly perform selected measures at certain addresses, such as those of newspapers and the government sector.

Parallel to the initiation of Kulturarw[3] the Royal Library launched yet another project, Svesök (SweSearch) [3]. Svesök is a search service for Swedish web resources on the Internet whereby all web pages with live links collected by Kulturarw[3] are indexed. Searches can be carried out using metadata (HTML-metadata or Dublin Core) most likely found in the web page code.

Initially, Kulturarw[3] and Svesök were meant to function as a national bibliography for the Swedish part of the Internet. It did not quite turn out as we had expected, and there were a number of reasons as to why this was. To begin with the search robots do not collect everything they come across, such as password-protected sites or contents of databases. There have also been a number of crucial questions which until now, have been unanswerable, such as - Should the publications be catalogued in the existing catalogue or in a separate database?

From publisher to national catalogue - describing a work flow

The voluntary deliveries of e-publications from publishers

There are certain categories within electronic publishing which we, whilst awaiting better solutions, would want to see catalogued in LIBRIS. Such as online periodicals and newspapers, dissertations, reports from public authorities, e-books etc.

As the Legal deposit Act does not cover the material published on the web, the Royal Library has decided to attain separate agreements with each independent publisher. As a matter of fact, it was the publishing community who initially approached the Royal Library, anxious to have their recent publications made visible in LIBRIS.

Around this time experimental work was taking place at the Electronic Publishing Centre at Uppsala University Library [4], whereby new approaches were tested for the electronic publication of dissertations. They have instigated a workflow pattern whereby the different products are given the opportunity to re-cycle the information previously supplied by the authors of the dissertations themselves. Collaboration between the Royal Library and Uppsala University Library has extended the range of this model to include other kinds of material. At present the target group is the established commercial publisher but the aim also includes the public authorities.

E-books - a case in point

When both parties have reached an agreement, the publisher sends a computer file, in the format of their choice, to the Royal Library. The file should include such metadata as is required for the bibliographical description. In return, the Royal Library ensures expedient delivery to LIBRIS. The delivered file is the property of the Royal Library and is filed for future reference, as are the deliveries of printed books. On the other hand, the e-book cannot be borrowed, printed or copied in its entirety. It can only be read at especially assigned data terminals located on library premises.

When the time comes to place a newly issued e-book on a publisher's web site, the publisher can fill in the registration form, designed by the Royal Library, on-screen. Each publisher has its own registration form requiring a unique password enabling individual adaptation. To exemplify this we have added name of publishers and place of publication as primary default values to each registration form.

The process consists of the following stages (see also figure 1):

1. The publisher enters the required bibliographical data on the registration form: title and statement of responsibility, language, edition, ISBN, file size, format, price, etc. The publisher is also required to inform as to the name of the text file being sent. When the information has been dispatched the following will take place:

2. The bibliographic data found in the registration form is forwarded in XML to a database for storage. The stored information can then be converted to any suitable metadata set. In this particular case the data is converted to a preliminary record in MARC 21 and sent on to LIBRIS. The record will be available the day after and can be found via LIBRIS WebSearch.

3. The text file is further despatched via FTP to the Royal Library's digital archive for future preservation.

4. E-mail will be posted to inform staff at the Royal Library of the arrival of a new book file and a new catalogue record.

The process described in points 2-4 are fully automated and have at this stage not been edited by a cataloguer.

5. When the message of delivery has reached the national bibliographical department at the Royal library, the preliminary record will be produced online from LIBRIS. Control of the text file will then commence updating the record with classification codes and subject headings. There is also a name authority control. The record is then upgraded to national bibliographical level.
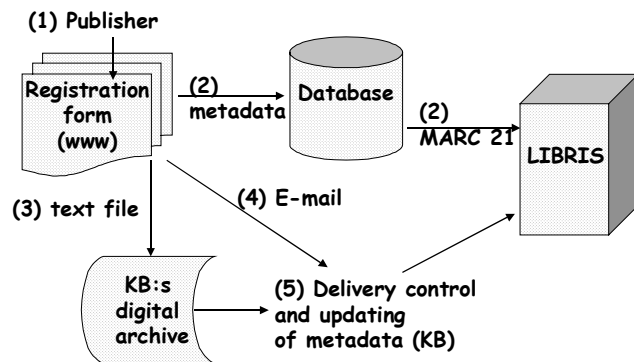

Figure 1

The entire process should take no more than a few days. The aim is to deliver the catalogued record to LIBRIS in unison with the publisher's update to their web site informing the public of the new title.

Development

Many publishers have been in touch and are in line to join up and it is my firm belief that we will see a dramatic development in this enterprise. In the long run there are several available options to extend the present range of services. For instance, we should be able to receive metadata directly from the distributors without taking the long way round of filling out a web registration form. The data we require from the publishers is already there in one shape or another, and the act of filling in the form once more is superfluous. The Royal Library should not only be able to receive metadata, but also to export it back to the publishers if they so wish as well as to other databases and search services. We should also be able to offer this in differing formats: MARC, ONIX, Dublin Core, etc.

Bibliographic control or chaos?

The situation involving national bibliographic work and the Internet can at times seem a little chaotic. Today, we can only speculate about format and structure of future national bibliographies, but we will most likely need to find new solutions for various kinds of material. I have described a possible model for the cataloguing of electronic publications in the national database LIBRIS. The model suits a number of publications. Most sites found on the Internet are freely available and might just as well be collected by Kulturarw[3] search robots. But, who will assume the responsibility of making these documents searchable?

In the midst of this apparent chaos, it remains of the utmost importance for national libraries to retain bibliographic control. There are several ways this can be achieved. The extensive use of metadata and automated routines should be applied to ensure manageability of the vast amounts of documents residing

on the Internet. National libraries must assume an increasingly active and participatory role to influence publishers to apply more metadata, such as Dublin Core. The companies creating programmes for document archives should be encouraged to improve their working tools. National libraries often implement their own devices, which could be promoted on the Internet and made freely accessible. The devices I am referring to are: classification systems, subject heading lists and authority files. We must persevere in international standardisation work and increase propagation for the use of standards. URN Resolution Service and the availability of NBN are other examples related to services facilitating searches and identification of documents.

Above all, we need to expand our operational relationships with publishers.

<div align="center">*****</div>

English translation: Jonathan Pearman

[1] LIBRIS Websök, http://www.libris.kb.se/
[2] Kulturarw^3 , http://www.kb.se/kw3/
[3] Svesök, http://www.svesok.kb.se/
[4] Electronic Publishing Centre at Uppsala University Library, http://publications.uu.se/epcentre/