**Discovery-Systems and e-Books.**

**Gerard Bennett**

Information Systems and Library Services

# igelu *Newsletter*

*International Group of Ex Libris Users*

**Welcome to Madrid 2008 and**

- I-NUG Meeting 7th September 2008

- 3rd IGeLU Meeting 8th to 10th September 2008

## In this issue:

## Discovery-Systems and e-Books

*By Gerard Bennett*

There were two topical plenaries at the conference this year. The first explored 'Discovery Tools' and brought together 3 speakers, 2 of whom reported on Primo implementations (Marshall Breeding, Vanderbilt University, and Jørgen Madsen, Danish Primo Project), and the third of whom addressed a major Endeca implementation (Michele Newberry, Florida Center for Library Automation).

I have had the pleasure of hearing Marshall Breeding speak a few times at conferences and was once again impressed by his command of his subject, and his articulate, direct and engaging delivery. He gave us a very enthusiastic account of Primo and the implementation at Vanderbilt. Primo does away with the shortcomings of traditional systems that provide too many places to search, and of federated search which is too slow, complex and gives shallow results. The breadth of scope of Primo was praised – giving equal scope to print and non-print, breaking down the silos that each is usually confined to, and furthering the abolition of the divide between local and remote. Primo search was defended against the charge of dumbing-down the catalogue: a well-designed keyword search enhanced by the inclusion of words from authority headings, plus faceting and relevance ranking, more than made up for the loss of browse searching (which is not possible with this architecture). The success of Primo in being able to deal with both structured and non-structured data was shown by the content in Vanderbilt's Primo implementation: their Unicorn catalogue (Acorn) plus a database of TV news abstracts. But what about the addition of subscribed content such as the data in A&I databases? This is planned, and will be reliant on the attitude of database vendors to the harvesting of their content. Marshall anticipated that their attitude would be positive – eventually.

Their attitude may be revealed sooner by the Danish Primo project that Jørgen Madsen presented. In some ways a more ambitious project involving a consortium of institutions led by the Royal Library, Copenhagen, it aims, among other things, to test the scalability of Primo, searching a database of 200 million records, and to integrate article searching seamlessly into Primo, including through the use of 3rd node technology. This was a new and exciting term to many of us and very simple once explained. Instead of each Primo institution harvesting databases that are also used by others, why not a central node or nodes hosted by Ex Libris where commonly used databases are harvested and made searchable by many Primo installations? The Danish project also had national ambitions to include a data bank of all licensed material for all Danish research libraries searchable by Primo. Jørgen also spoke of 'atomising' Primo where this made sense that is to present part of its functionality in varying contexts, using Primo web services. We had a glimpse of interesting possibilities here such as the ability to call up a Primo search box from a citation in, for example, a PDF document.

Michele Newberry's thoughtful account of the Endeca implementation at FCLA (Florida Center for Library Automation) threw further light on the possibilities of search technology. Their adoption of Endeca was in part motivated by exploring alternatives for upgrading 11 university library OPACs from Aleph 15 to 18, and has provided them with a union catalogue with rich functionality including guided navigation, faceted search, stemming and spell check, along

# Discovery-Systems and e-Books (cont.)

with OPAC features such as display of copy availability, ability to renew and place reserves (features that the Primo user at present must drop into the OPAC to use). Michele also addressed such aspects as building the Endeca file, record de-duplication, data integrity, and a range of change management issues.

After coffee the 3 speakers joined a panel which included Dale Flecker and Oren Beit-Arie to reflect on the presentations and take questions from the audience. We learned interesting things about trends in discovery tools, for example, there is a project in the US to produce a protocol to standardise the ways in which discovery tools talk to the library system back-end such that reservations and renewals can be done through the discovery tool interface; and about different ways that Primo can handle copy availability – to deliver it on the fly, or provide it via indexing/harvesting mechanisms, the latter allowing the information to enrich the faceting thus permitting the user to view results based on availability. There were also honest evaluations of these discovery tools as they do not offer a 100% solution, and are limited by the quality of data they index. But all on the panel were very positive about their advantages and left the audience with much to ponder.

The e-books plenary on Wednesday afternoon had 2 speakers, Hazel Woodward from Cranfield University in the UK addressing the topic 'Impact of E-books on Library Services', and Dale Flecker reporting on the work of the IGeLU/ELUNA/Ex Libris E-book Focus Group.

Hazel gave us an overview of the challenges faced by librarians developing e-book collections. Her basic case was that of the prosecution: the conditions needed to develop e-book collections are not being met. Not enough core materials are being provided; too many different deals are on offer; pricing models are not flexible and not often affordable; discovery and access are hindered by the variety of sources, and interfaces and by the poor quality of metadata, hindering integration and linking; and digitization projects though welcome are burdened by

copyright restrictions. All of this leads libraries adjusting to a difficult situation rather than influencing developments. One attempt to lead developments is the JISC Observatory Project in the UK, which will spend £600,000 on core readings in 4 subject areas. The books will be available to all in UK academic institutions for 2 years, and their usage will be logged to assess impact and usage. The pricing issue was underlined when Hazel revealed that less that 100 titles were included.

Dale's report listed some initial findings of the Focus Group and it was clear that it had carried out some trenchant analysis. Few would be surprised by the finding that the complexities of e-books require separate treatment by library management systems. Among the reasons identified for this were the differences in scale. For example the quantity of e-books compared to e-journals; differences in business models; problems with identifiers in that too few e-books have DOIs, and large numbers of books being digitized were published before the era of identifiers. The incorporation of e-books in the OPAC is a major issue and there are many aspects to this including the different strands of digitization projects – full text of out-of-copyright materials, component parts for in-copyright titles – as well as the role of TOC and abstracts for catalogue enrichment. Dale then listed the tasks before of the Focus Group. Following the submission of its recommendations the Group would direct a series of implementations. This was interesting news. Dale concluded his report with a call to us for our input to the focus group. "Tell us what hurts most", was the apposite phase I remember!



Brno. Photo: Ragnar Helin



Panel discussion with from the left: Dale Flecker, Oren Beit-Arie, Michele Newberry, Marshall Breeding and Jørgen Madsen.
Photo: Ragnar Helin