

79th IFLA General Conference and Assembly - "Future Libraries: Infinite Possibilities":

Inspiring solutions emerging from open source

From OPAC to Archive: integrated discovery and digital libraries with open source

Edmund Balnaves

ealnaves@prosentient.com.au

Prosentient Systems



Copyright © 2013 by **Edmund Balnaves**. This work is made available under the terms of the Creative Commons Attribution 3.0 Unported License:

<http://creativecommons.org/licenses/by/3.0/>

Abstract:

Institutions have traditionally neglected the intellectual potential of their research and publishing output. The traditional model for academic publishing neglects the importance of the institutions contribution to the creation and innovation of intellectual output. The open source movement represents an alternative to a monolithic vision of the web information resource offered by Google, Amazon and other content behemoths. Since the turn of the century innovation in open source systems has enabled libraries to extend their information management to digital libraries and to integrate these resources with other resource management and discovery tools. This presents an opportunity for institutions to make better use of the inherent knowledge in publications by consolidating these resources in an institutional repository managed by an effective metadata framework and suitable for long term discovery of content.

This paper presents a model for development of unified institutional electronic resources delivered through open access and open source models through a combination of discovery tools using the OPAC and the repository capabilities of digital libraries.

Examples of the cross-integration of open source systems will be provided as implemented in both developing and developed countries.

This approach provides a sustainable and low cost model for implementation of institutional repositories. Libraries, with strong understanding of the importance of metadata and services standards, should be well positioned to managed these resources as long as they realign their skills to be actively engaged in the open source arena. This is a distributed vision – where quite small and very large collections can be bought together in a semantically meaningful way. It is also one where the library and librarian mediate internal datasets, digital and physical collections and external

resources to the benefit of their organisation in effective knowledge management, education and entertainment.

Keywords: Open Source, Open Access, Integrating systems

1 INTRODUCTION

Libraries are changing: from physical to digital management, from “library” to “information centre”. Library as managers of physical collections risk being marginalised in the organisation. However as effective managers of complex resource hybrid print/electronic resources, libraries can assist institutions to make more effective use of their neglected the intellectual potential of their research and publishing output. The traditional model for academic publishing neglects the importance of the institutions contribution to the creation and innovation of intellectual output. The access source movement represent an alternative to a monolithic vision of the web information resource offered by Google, Amazon and other content behemoths. In conjunction with Open Access, the library open source movement has blossomed since the turn of the century. Innovation in open source systems has been an enabler for libraries to extend their information management to digital libraries and to integrate these resources with other resource management and discovery tools. This presents an opportunity for institutions to make better use of the inherent knowledge in their reports and publishing activity by consolidating these resources in an institutional repository managed by an effective metadata framework.

Many institutions have a continuous now wish to capitalise on their publishing output and make better use of their research and information, and it is appropriate for libraries to step up to the mark and engage in this area. For some libraries, this involves becoming a “publisher”. This is an important transition from the librarian as information “curator”. The tools to support this transition are available now, and accessible to both small and large organisations. Organisational websites tend to be short-term and subject to major makeover, whereas library systems have a strong record for long term content management and are effective as repositories for metadata and documents.

Such is the evolution of these systems, with 15+ years of development under their belts, that these systems are not only robust and stable, but also increasingly capable of integration This integration capability introduces possibilities for a new generation of integration between open source systems, and between these systems and external web-based resources.

2 AN OPEN SOURCE MODEL

One of the great duo’s that underlie modern libraries has been the rise of open source systems (OSS) and open access (OA) (Keast and Balnaves 2009). .Open source systems for libraries have flourished in the last 15+ years, and especially in the area of Digital Library and Library Management System software. While even as late 2008 as the study by JISC found that the “LMS is not workable for most institutions” (Adamson et al. 2008), this finding has been belied by the extraordinary take-up of open source, not only in the LMS space but also for Digital Library and more recently discovery tools. Open source can be an enabler for the adoption of open access in an institution. Equally a successful OA project can justify the ongoing improvement of the OSS implementation. Second, OSS can provide a

level of certainty for an institution in their operation costs when combined with open source. The larger the community of adopters of open source the stronger the overall support. Third, the OSS can provide a level of security in that there is no proprietary lock-in and the code is visible (and therefore can be corrected). The functional depth of this security will be improved by the work of those adopting the open source model.

A common confusion is that open source means “free”. While it may be lower cost, no information technology system operation is free. The ongoing nurturing of a system, software upgrades over time, support for customisations and enhancements, server administration, network costs are just a few of the base-line elements of managing an information system. Nevertheless, the amortisation of the software support across a wide installed base makes for an effective cost model for smaller institutions. IT Departments are typically the chief barrier in acceptance of an open source model, and are reluctant to implement open source systems. This barrier has been overcome in many cases with the emergence of many hosting provider who provide a packaged open source solution. This has itself stimulated considerable innovation in the open source area as it has enhanced the commercial elements of the community .

Open source projects such as Koha have had hundreds of contributors around the world who enhance the project on an on-going basis.

The increasing access to digital resources entails growing complexity in the resources management by the library, including digital libraries/collections, digital news feeds, digitisation of organisational resources. This additional complexity can lead to requirements for a federated search capability (integrating into a single portal the major information resources available to library clients) and workflow management systems (to manage the complex processes in electronic collection development) and single sign-on (to hide the complexity of access to multiple underlying database resources). The rapidity of technological development brings long-term difficulties in the management of intellectual and creative output in digital form. Libraries and museums have a key role in the preservation of analytical and creative endeavours over the long term. However, most libraries are ill equipped to undertake research into the preservation of new media artefacts and creations. Digital Library collection building has associated with it inherent risks of technological obsolescence. In addition to the systematic risks associated to critical information technology architecture, are the problems of software and hardware obsolescence. Issues of obsolescence are not inherent obstacles to the move to management of electronic resources – but they are issues that need to be addressed by the institution in the management of the disparate resources that constitute an electronic collection. Information systems inevitably go through a continuous series of transformations over time, as do digital objects stored in an information system, and an open source community provides a diversity in innovation of the system and its capabilities to manage resources over the long term.

This model can be illustrated through the combination of three open source systems. There is not a single “open source” solution to everything that the library needs. However there are ways in which combinations of open source can be deployed to achieve highly functional outcomes. For instance, we have combined the following systems effectively to provide rich library solutions.

- VuFind – a discovery tool that can harvest across multiple information sources. Using Apache SOLR, it would scale well to large search sets. VuFind is relatively

new to the scene, but provides a nice search front-end and can harvest and import data from a range of sources.

- Koha, – a Library Management System. It can scale to several search across several million objects, can be effective as a medium-size LMS and Federated search.
- DSpace – a Digital Library System. A scalable digital library system; that can support very large collections but is very single-purpose in design.

Of these Koha has the most diverse community supporting it, with hundreds of developers and contributors around the world. DSpace has strong institutional development backing from Universities and a slower release process.

The following is a role analysis of the ways in which these open source systems can be deployed:

	Digital Library	Library Management	Federated Search	Single Sign-on
Koha		Koha	Koha can be used to aggregate different sources to provide a single search point across resources.	Koha can be both the target and the source for single –sign-on
DSpace	Scalable digital library system that supports a role both for digital library management and records management.			DSpace can be the target of single sign on
VUFind			VUFind, unlike Koha, is designed from ground up as a tool to search aggregation tool.	

3 INTEROPERABILITY AND METADATA

Metadata is the information describing objects in the Digital Library. For instance, the item title, author, dimensions and format are all examples of metadata. Metadata serves three purposes in the Digital Library:

- Descriptive metadata - as with traditional cataloguing, digital objects need to be described and identified so that they can be discovered within the Digital Library. Digital Library metadata standards for describing objects serve the same role as AACR2 and MARC standards do for traditional catalogues. Examples of

Descriptive metadata standards commonly used in Digital Libraries are Dublin Core Metadata Initiative (DCMI), Metadata Object Description Schema (MODS), and Metadata Encoding & Transmission Standard The (METS). While DCMI is probably more widely used by Digital Libraries, MODS and METS provide a fuller descriptive framework as a successor to MARC. DSpace and Greenstone use DCMI as their descriptive metadata framework.

- Semantic metadata - the semantic metadata provides the subject classification and relationship information for objects in the Digital Library. While this may be based on a traditional name/value pair of identifiers (subject = 'Parliamentary History'), the current trend is to move to Resource Description Framework (RDF). RDF underpins many projects that are realising the possibilities of the Semantic Web for purposes of stronger metadata description of documents on the web (and in archives). A semantic metadata description goes beyond the name/value descriptive pair to describe metadata in a series of “statements” in a subject, object and predicate statement (the title of the book is 'The history of Parliaments'). Central to the concept of RDF is the ability to unify concepts across many resources in a meaningful way. Fedora Commons implements RDF as its underlying schema.
- Harvesting metadata - There are many Digital Library systems - commercial, open source and bespoke (home grown). Irrespective of the internal metadata approach for description and subject classification of the objects in the library, support for a harvesting metadata standard provides a means for inter-operability between Digital Library systems. The most widely implemented harvesting system is Open Archives Initiative Protocol for Metadata Harvesting (OAI/PMH). This scheme supports metadata “harvesting” between digital libraries to allow discovery of digital resources between systems. Kete uses OAI/PMH for its internal schema. DSpace, Greenstone, Fedora Commons and Kete support an OAI/PMH harvesting interface.
- The long term inter-operability of your resources with other digital resources being developed in-country and regionally will be enhanced or impeded by the level and quality of the metadata collected and linked using digital resources. The selection of a metadata framework should be undertaken with reference to existing projects nationally and regionally.

4 REGIONAL CASE STUDY

Libraries are making a progressive transition from physical resources to electronic resources.

The digital library system acts as an enabler for the library to:

- Provide local and public access to the intellectual output of the organisation
- Provide local access to born digital resources and digitised that are locally cached in the digital library

The National Parliament of Solomon Islands illustrates just this methodology. In March 2012 they implemented a combined open source solution of open source digital library software (DSpace) and library management software (Koha) to create an electronic library service providing:

- Access to legislation, committee reports, gazettes and Hansard records of debate
- Access to media clippings and media releases by parliamentarians.

The Parliamentary Library will need to manage an increasingly diverse range of electronic resources, including:

- online electronic databases services - external provides with search services and possibly full text
- digital libraries - providing a facility for managing digital documents created by the library (either born-digital or converted to digital form)
- syndication feeds - information feeds providing current information resources based in information preferences (for instance news feeds)
- electronic subscriptions - including e-books and e-journals.

A framework built around open source provides a cost-viable framework for access to complex information resources.

It can be implemented in several models depending on the Information Technology resources available to the organisation:

- through outsourced hosting through an open source hosting provider – there are many of these now emerging
- through co-operative IT co-hosting with other similar departments or organisations to share the IT maintenance cost

* through direct hosting and management in the organisation. If the library has a supportive IT area, this can be an effective way to build internal knowledge and take full ownership of the systems. It is sometimes still possible to draw on the expertise of hosting providers to provide technical support for the internally hosted systems.

5 CONCLUSION

Open source systems present a new opportunity for libraries to directly engage in the technology that they provide and to deliver functionally rich solutions. This new generation of open source systems provides open services layers to support an increasingly digital, mobile, client community for libraries. Libraries such as the National Parliament of Solomon Islands can use this technology to jump ahead a generation of systems to provide new access services through to digital resources in a practical a cost effective manner.

6 REFERENCES

Adamson, Veronica, Paul Bacsich, Ken Chad, and Jane Plenderleith. 2008. "JISC & SCONUL Library Management Systems Study."
[Http://www.jisc.ac.uk/publications/generalpublications/2008/librarymanagementflyer.aspx](http://www.jisc.ac.uk/publications/generalpublications/2008/librarymanagementflyer.aspx)
(April).

Keast, D., and E. Balnaves. 2009. "Open Source Systems Bring Web 2.0 to Special Libraries." In International Conference of Medical Libraries. Brisbane, 2009.
<http://espace.library.uq.edu.au/view/UQ:179870>.

Edmund Balnaves is the Information Officer for the Information Technology Section of IFLA, is a Director of ALIA and unashamedly dabbles with library information technology in his company Prosentient Systems