ELAG 2006 Paper: Long-term electronic archiving as part of a digital library solution – an overview of products

Johan van Halm, Library Consultant, Amersfoort (NL)[1].

1. Introduction

Long-term electronic archiving or preservation, hereinafter referred to as LTAP, has become a more or less routine operation within digital library projects, whereby now – after a period of national libraries dominance of this scene, often then with experimental projects – a variety of open-source and commercial products have emerged.

Within the context of ELAG it is useful to know, that most ILS[2] vendors have a Digital Library Module, such as DigiTool of Ex Libris, which covers a component like LTAP. Often is LTAP on the basis of a 3rd party product, such as ArchivalWare of PTFS, which is the underlying product of the SIRSIDYNIX digital library solution.

Digital libraries, notably in a university or research institution, are not dealing only with published electronic resources and institutional electronic collections, but are also becoming institutional repositories. The data produced by university's researchers are a valuable resource and academic and research institutions are coming to the realization that their digital collections should also include the web pages and software pertaining to a researcher's work and not just the refereed journal articles that resulted from the study.

Unfortunately, while information technology has made it possible to generate and manipulate large quantities of scientific data, changes in the management of this data have not kept pace. Researchers have neither the time nor the inclination to ensure that the data supporting past projects is compatible in the future. By becoming the repository for an institution's documentation, research libraries are stepping into the gap to preserve scientific data and other "grey" materials before technological change renders the data irretrievable.

Institutional repositories are but one example of the recent efforts of digital libraries to offer access to an ever-broader array of materials. Digital libraries have brought data together by making seamless connections/links between formats such that information is no longer restricted by institutional or technical barriers.

Along with expanding the breadth of its collections, the library's mission is to make the collection serve the needs of its users. Since the mainframe era, libraries have adopted new information technologies to facilitate the location and linking of records. In 2002 Clifford Lynch drew a distinction between *digital collections* which offer only limited interpretive tools, and *digital libraries*, which are "systems that make digital collections come alive, [that] make them usefully accessible, [and] that make them useful for accomplishing work.

In this presentation we will deal with a selection of software products in the library and corporate domain, which don't differ that much, other than by the idioms used:

> Digital library software
> Digital assets management software
> Digital rights management software
> Electronic document & records management systems (EDRMS)

---

[1] johanvanhalm@cs.com
[2] Integrated Library System

2.   ILS Software products

With our colleague Anna Maria Tammaro's article (see ref. 2) as a starting point, we have searched for similar products and we have found also products of ILS vendors, as well as products of EDRM companies. We have not evaluated the products and we are merely using listings based on vendors' information (brochures, web, talks)[3].
In this paragraph we list products from the library and information domain:

Name of software	Name of provider	Website/URL
Agora	SRZ Berlin	www.agora.de
ArchivalWare	PTFS	www.ptfs.com
CDSWare	CERN	http://cdsware.cern.ch
Content DM (OLIVE)	OCLC	www.oclc.org
Convera	Convera, Inc.	www.convera.com
Curator	Endeavor Info Systems	www.endinfosys.com
DigiTool	Ex Libris	www.exlibrisgroup.com
docWORKS	CCS	www.ccs-gmbh.de
DSpace	DSpace Federation	http://dspace.org/federation/index.html
EPrints		www.eprints.org
FEDORA	University of Virginia Cornell University	www.fedora.info
HIP (Horizon Information Portal	SirsiDynix	www.sirsidynix.com
LOCKSS	Harvard University Library	http://lockss.stanford.edu/
OPUS	-	http://elib.uni-stuttgart.de/opus/
Scholnet	ERCIM	www.ercim.org/scholnet

3.   Requirements

In picking a software, a library has to consider a number of questions:

- What material should be stored in the repository?
- Is long-term preservation an issue?
- Which software should be chosen?
- What is the cost of setting the system up? and
- How much know-how is required?
- Can digital works using a custom workflow process be captured and described?
- Can institution's digital works be distributed over the Web, so users can search and retrieve items in the collection

and this can be translated into the need for the following technical specifications of a LTAP software
- Authentication capabilities

---

[3] When I did indicate my willingness in 2005 to present a paper in 2006 I took for granted, that I would present evaluated information, but that project has been postponed for the time being

- Configurable Interface
- Digital content of all types (text, images, audio, video)
- Dublin Core Support
- Duplicate management
- Export in non-proprietary formats (METS/ALTO XML)[4]
- Format independence[5]
- Full-text search
- Image browser viewing
- Input (extended data entry capabilities)
  - format independent
  - high volume enterprise scanning
  - low volume departmental scanning (MS Windows desktop
- MARC XML/XML editors
- Metadata capture & harvesting
- Multilingual capabilities
- Multi-site/community management
- Navigation toolbars
- OAI compliant, incl. linking
- Online admin and capture
- OpenURL compliant
- PDF indexing
- Search engine
- Secure Access
- SGML support
- User personalization
  - Document basket
  - e-mail
- web-archiving tool
- Web-based
  - export
  - import
  - metadata editor
- Workflow management

Web2.0 developments have to be closely watched as serious impacts may be expected of easier and better interoperability components.

An article by Prudlo (reference 1) discusses LOCKSS, EPrints and DSpace, so does Tammaro, which are some of the most widely known repository management tools, in terms of who uses them, their cost, underlying technology, the required know-how, and functionalities.

---

[4] METS = Metadata Encoding & Transmission Standard
   ALTO = Analysed Layout & Text Object
[5] **GDFR**
The Harvard University Library (HUL) will develop a registry of authoritative information about digital formats, because detailed information about the format of digital resources is fundamental to their preservation. For current information and updates on GDFR, visit the project web site at http://hul.harvard.edu/gdfr/

4.   How do vendors call their product

Name of software	
Agora	Content Management System
ArchivalWare	Digital archiving software for information management professionals
CDSWare	Integrated Digital Library to run your document repositories
Content DM (OLIVE OCLC)	Digital content & document management system
DigiTool	Digital Asset Management – develop your institutional repository
docWORKS	Digitisation and conversion management system
DSpace	Repository management tool
EPrints	Repository management tool
Fedora	Open-source digital object repository system
HIP (Horizon Information Portal	Digital library and information portal
LOCKSS	Repository management tool
Meridian/Curator	Digital management solution
OPUS	Online Publikationsverbund der Region Stuttgart
Scholnet	Building a digital library infrastructure to support scholarship

5.   Related software products used by corporations

Most so-called Electronic Document and Records Management Systems (EDRM systems) have more corporate applications, but could – most likely – be used in a library environment as well. We have taken this listing from the March 2006 Sector Update "Electronic Document & Records Management Systems (EDRMS) in Information World Review.

Examples are

Anacomp	docHarbor	www.docHarbor.com
Diagonal solutions	Diagonal Solutions	www.diagonal-solutions.com
EMC	Documentum	www.emc.com
FileNet	FileNet	www.filenet.com
Hummingbird	Hummingbird	www.hummingbird.com
Interwoven	Interwoven	www.interwoven.com
Iron Mountain	Iron Mountain	www.ironmountain.co.uk
Mobius	Mobius Manag. Systems	www.mobius.com
Open Text	Open Text	www.opentext.com
Stellent	Stellent Records Managem.	www.stellent.co.uk
Tower Software	Tower Solutions	www.towersoft.com/emea
Version One	Version One	www.versionone.co.uk
Vignette	Vignette	www.vignette.co.uk
Zantaz	Zantaz Solutions	www.zantaz.com

6. My message

Librarians have to look at developments outside the topical library IT world and many features (and much more) of the ERDM systems market are applicable in the library world. Eventually proven open-standard components – as 3rd party solution – should be acquired.

7. References

1 Prudlo, M., E-Archiving: An Overview of Some Repository Management Software Tools. Ariadne, Issue 43, 30-April-2005.
(Originating URL: http://www.ariadne.ac.uk/issue43/prudlo/intro.html)
2 Tammaro, A.M. & T. De gregori: Ruolo e funzionalita dei depositi instituzionale. Biblioteche Oggi. Dicembre 2004, pp. 7 - 19
3 Information World Review, March 2006