DIGITAL REPOSITORY SOFTWARE

SOME NOTES AND CITATIONS

H.M. Gladney

HMG Consulting
Saratoga, CA 95070

7 August 2025

Abstract: **[enter abstract here]**

This report is an unpublished draft provided for critical discussion and no other purpose. It has been produced at the private expense of the author.

1 REQUIREMENTS ANALYSIS

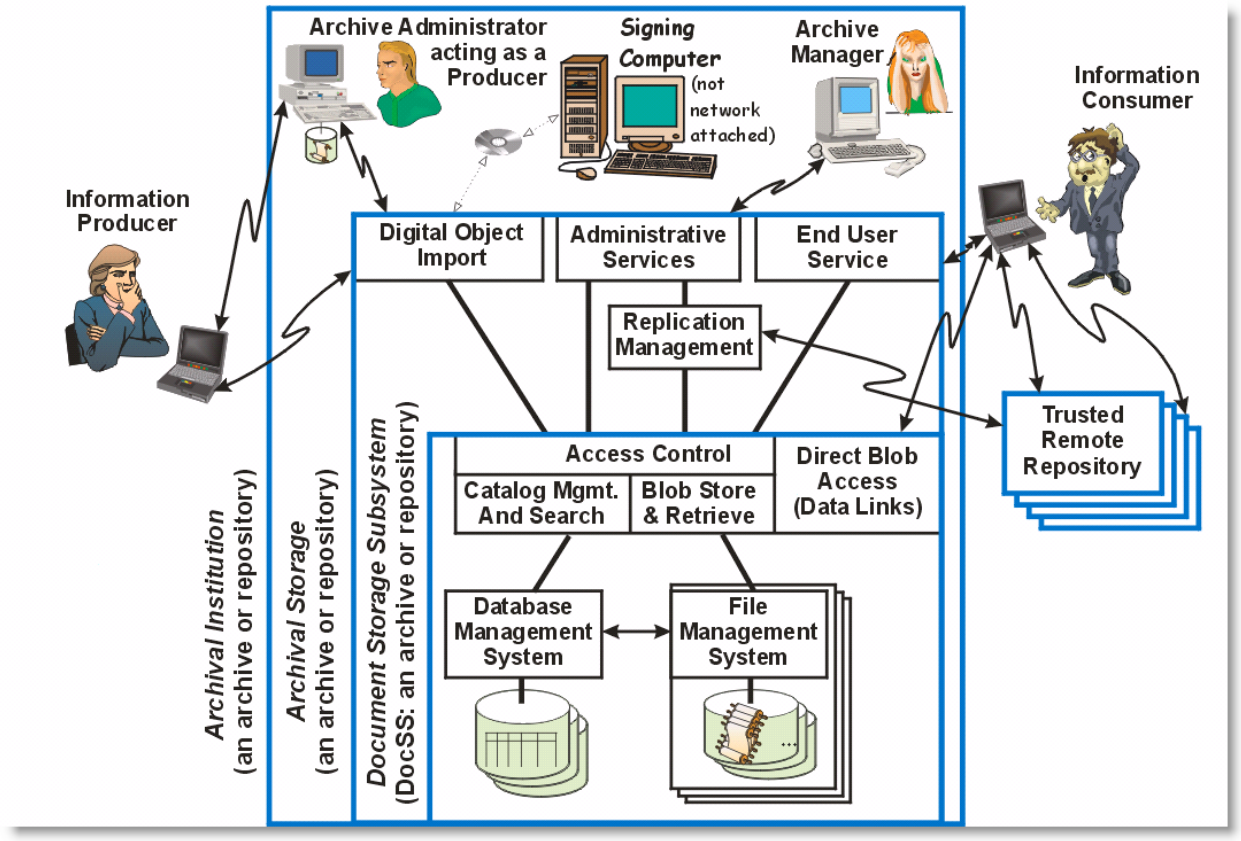

Figure 1: Repository nesting architecture
illustrating generic nesting structure in which local customization should occur in outer layers.

1.1 The [U.K.] National Archives

Home page at http://www.nationalarchives.gov.uk/archives/

The National Archives, *Standard for Record Repositories*, 2004.
http://www.nationalarchives.gov.uk/electronicrecords/generic.htm

Generic Requirements for Sustaining Electronic Information over Time, 2003. accessible via
http://www.nationalarchives.gov.uk/electronicrecords/reqs2002/

1.2 DSpace

WWW home page http://www.dspace.org/

Presentation by MacKenzie Smith:
http://www.dpconline.org/graphics/events/presentations/pdf/DSpaceatDPCOct2002.pdf

1.3 Fedora

WWW home page http://www.fedora.info/

1.4 Greenstone

WWW home page

See also *The New Zealand Digital Library* at http://www.sadl.uleth.ca/nz/cgi-bin/library

2 RELATED SW

2.1 Open Archives Initiative

WWW home page: http://www.openarchives.org/

3 COMPARISONS

3.1.1 Search for a Content Management System (U. Arizona, Feb. 2004)

http://dlist.sir.arizona.edu/archive/00000623/01/digital_content_management.pdf [Han]

4 MISCELLANEOUS

4.1 Other Analyses

4.1.1 International Oceanographic Data and Information Exchange (IODE)

IOC/IODE-MIM-VIII/07 - Hardware and Software for Library and Information Centres at http://ioc3.unesco.org/iode/files.php?action=dlfile&fid=396 This is a UNESCO initiative.

A professional organization not especially well informed about software has produced a good list (with some useful evaluative comments) of the software that a small organization should consider. (This might also be useful for the CHM SCC taxonomy topic.)

4.2 E-mail

4.2.1 Storage for preservation (from a DSpace author, 21-Jan-05)

To: dspace-preservation mailing list dspace-preservation@mit.edu

http://mailman.mit.edu/mailman/listinfo/dspace-preservation

> 1) performed no compression

> 2) did not change the re-encode of the contents (to make human reading of the contents as easy as possible).

> 3) had a table of offsets to enable efficient extraction of files

> 4) had a widely adopted and implemented open standard of minimum complexity. (most likelihood that implementations will survive, or that replacements will be easy to build)

> It's quite possible that there isn't a format available atm that comes close enough to these, but I think the preservation benefits of using archive files warrants us considering them.

Been doing some research to help the discussion: I've had a look at many of the archive formats available, and it doesn't seem any of them fulfill all these requirements. Also, my assumption about single files being more likely to be stored intact than directories turns out to be untrue for many filesystems. I'm not going to look into archive formats further, unless anyone knows something I missed.

For preservation beyond the lifetime of a running DSpace the asset store software is obviously important. Perhaps all we need to do about this is make administrators of DSpace aware of the potential problems in using experimental / proprietary filesystem software?

Jim Downing <ojd20@cam.ac.uk>
DSpace@Cambridge

4.2.2 From Technical Univ. of Denmark (July 2004)

ajh@cvt.dk I do not yet know the Greenstone archiving system, but do work with Dspace.

Do I understand right - you ask for a program/solution that can get the two systems to talk together, possibly exchange data (meta-data and digital objects - collections ...)?

If I understand right, you can export METS-package from DSpace. Unfortunatelly there is no METS importer, but we, from the Technical University of Denmark, are interested in this import feachure too. The question is now, if Greenstone supports METS.

To the discussion with Mr. Dauphin: If there should be made any comparison between archiving systems by or for UNESCO, one should certainly look at the Fedora work from Cornell, as another candidate. (sorry to the MIT people...)

In my opinion, using Dspace for learning object repository does have a weak point - yet - the fact that it is rather difficult to change: a) the metadata field set b) the workflow for collecting the metadata.

The first is rather difficult to solve but is demonstrated in the EUL extension by Thesis Alive. The second you can solve by batch importing, hopefully soon METS-importing and such like. …

(Al)Fred
Center for Knowledge Technology
Technical University of Denmark

4.3 Q and A

4.3.1 Larry Masinter (21-Jan-05)

Q: "Why should CHM actually be the service provider for hosting the collection?"

A: I do not know that it should be, except that it is not traditional for a museum or library to contract out the holding of its collection. What other cultural archives have contracted out their repositories. (This is a consideration that did not arise in the discussion at either the December or January SCC meetings.)

Even if the core of the repository were held in one or more (replication for the usual reasons, etc.) external repositories, would not some repository front end still be required for CHM special aspects? (See Figure 1, which reminds us that the word 'repository' is ambiguous, denoting at least three different levels.) I.e., whatever the hosting decision might be, some SW choice decisions (including choices of the “buy or build” type) seem to be necessary.

Q: "How big do expect the entire collection to be, within the next 3-4 years?"

A: I have no idea of the collection size, perhaps because I am a SCC 'newbie'. However, I doubt that it is among the high priority things that will influence decisions for this topic, given the low price of storage space and supporting hardware.

5 BIBLIOGRAPHY

See also an Open Source Software and Libraries Bibliography at
http://www.vuw.ac.nz/staff/brenda_chawner/biblio.html

See also Lorcan Demsey’s Weblog at
http://orweblog.oclc.org/archives/cat_digital_asset_management.html

[Bass 02] Michael J. Bass, David Stuve, Robert Tansley, Margret Branschofsky, Peter Breton, Peter Carmichael, Bill Cattey, Dan Chudnov, and Joyce Ng, *DSpace – A Sustainable Solution for Institutional Digital Asset Services – Spanning the Information Asset Value Chain: Ingest, Manage, Preserve, Disseminate*, October 2002.

[Buchanan 04] Buchanan, George. David Bainbridge. Katherine Don. Ian H. Witten. A New Framework for Building Digital Library Collections, 2004.

This paper introduces a new framework for building digital library collections and contrasts it with existing systems. It describes a radical new step in the development of a widely-used open-source digital library system, Greenstone, which has evolved over many years. It is supported by a fresh implementation, which forced us to rethink the entire design rather than making incremental improvements. The redesign capitalizes on the best ideas from the existing system, which have been refined and developed to open new avenues through which users can tailor their collections. We demonstrate its extensibility by showing how digital library collections can be extended and altered to satisfy new requirements.

[Han 04] Han, Yan. *Digital Content Management: the Search for a Content Management System*, 2004.

Digital Content Management System is a software system that provides preservation, organizat ion and dissemination services for digital collect ions. By adapting the systems analysis process, the Universit y of Arizona Library analyzed its needs and developed Content Management System requirements for finding a suitable information system that addresses the increasing needs of digital content management. Dozens commercial and open source candidates were examined to match against the requirements. This article provides detailed analysis of three major players (Greenstone, Fedora, and DSpace) in key areas of digital content management: preservation, metadata, access, and system features based on the needs of the Universit y of Arizona Library. This paper describes the process we used to analyze and evaluate potential candidates. We have included results of our analysis to illuminate our process.

[Nestor] Nestor, the 'Network of Expertise in Long-Term Storage of Digital Resources'[1], is a national German project, founded by the German Ministry of Education and Research. It is the initial phase of building up a permanent distributed infrastructure for long-term preservation and long-term accessibility of digital resources in Germany comparable to the Digital Preservation Coalition (DPC[2]) in the UK.

Nestor will not archive anything – it will provide support in questions of long time preservation of digital objects. Aims of the nestor project are:

- to heighten awareness of the problem
- to produce and distribute expertise and information
- to facilitate cooperation
- to establish a durable organisational form

The nestor Information Platform consists of the nestor Subject Gateway[3], a Calendar[4] of events, a News-Site[5], a Review-Site[6] and a Newsletter[7].

The Information Platform is complemented by the nestor Communication Platform, which offers Mailing Lists[8], Forums and Workspaces[9].

Further Information can be at the homepage homepage[10], which offers project information's and a glossary[11].

The nestor partners are:
Die Deutsche Bibliothek[12] (German National Library) as the leading institution for the project
Niedersächsische Staats- und Universitätsbibliothek Göttingen[13] (Goettingen State and University Library)
Computer and Media Service of Humboldt-University[14], Berlin
Bayerische Staatsbibliothek[15] (Bavarian State Library)
Institut für Museumskunde[16] (Institute for Museum Information)
Generaldirektion der Staatlichen Archive Bayerns[17] (Bavarian State Archive – Head Office)

[1] http://www.langzeitarchivierung.de
[2] http://www.dpconline.org
[3] http://nestor.sub.uni-goettingen.de
[4] http://nestor.sub.uni-goettingen.de/calendar/index.php
[5] http://nestor.sub.uni-goettingen.de/aktuell/index.php?lang=en
[6] http://nestor.sub.uni-goettingen.de/review/index.php
[7] http://nestor.sub.uni-goettingen.de/newsletter/index.php
[8] http://www2.hu-berlin.de/nestor/mail/index.php
[9] http://nestor.cms.hu-berlin.de/tiki/tiki-index.php
[10] http://www.langzeitarchivierung.de
[11] http://www.langzeitarchivierung.de/index.php?module=Encyclopedia&func=lettersearch&letterget=All&vid=1
[12] http://www.ddb.de
[13] http://www.sub.uni-goettingen.de/
[14] http://www.cms.hu-berlin.de
[15] http://www.bsb-muenchen.de
[16] http://www.smb.spk-berlin.de/ifm
[17] http://www.gda.bayern.de

[Payette 02] Sandra Payette and Thornton Staples, *The Mellon Fedora Project: Digital Library Architecture Meets XML and Web Services,* ECDL'02 submission, July 2002. See **FEDORA page**.

The complete technical specifications for the Fedora software are available at: http://www.fedora.info/techdoc.shtml.

Fedora's sub-systems are described using the Web Services Description Language (**WSDL**), as are all auxiliary services included in the architecture. The system communicates over HTTP and supports the Simple Object Access Protocol (**SOAP**). Additionally, the project has adopted the Metadata Encoding and Transmission Standard (**METS**) as the means to encode and store digital objects as XML entities.

[Staples 03] Thornton Staples, Ross Wayland, and Sandra Payette, *The Fedora Project: An Open-source Digital Object Repository Management System*, D-Lib Magazine 9(4), April 2003.

[Witten 00] Witten, I.H.; McNab, R.J.; Jones, S.; Apperley, M.; Bainbridge, D.; Cunningham, S.J., Managing complexity in a distributed digital library, Computer - IEEE Computer Magazine, 32(2), Feb, 1999, pp 74-79

[Witten 00] Ian H. Witten, Stefan J. Boddie, David Bainbridge and Rodger J. McNab, Greenstone: a comprehensive open-source digital library software system, in Digital Libraries 2000, pp175-184, 113-121, June, 2000. Also in Comm. ACM 44(5), xxx-xxx, May 2000.

[Witten 01] Witten, Ian H. (University of Waikato), How to Build a Digital Library Using Open-Source Software, JCDL 2001. See also The New Zealand Digital Library and pubs & downloads.

This tutorial describes how to build a digital library using the Greenstone digital library software, a comprehensive, open-source system for constructing, presenting, and maintaining information collections. Collections built automatically include effective full-text searching and metadata-based browsing facilities that are attractive and easy to use. They are easily maintainable and can be rebuilt entirely automatically. Searching is full-text, and different indexes can be constructed (including metadata indexes). Browsing utilizes hierarchical structures that are created automatically from metadata associated with the source documents. Collections can include text, pictures, audio, and video, formed using an easy to use tool called the Collector. Documents can be in any language: Chinese and Arabic interfaces exist. Although primarily designed for Web access, collections can be made available, in precisely the same form, on CD-ROM or DVD. The system is extensible: software "plugins" accommodate different document and metadata types. The Greenstone software runs under both Unix and Windows, and is issued as source code under the GNU public license. Attendees will receive an extensive user manual and should learn enough to download the software and set up a digital library system. Those with programming skills should be able to extend and tailor the system extensively.

[Witten 03] Witten, Ian H. David Bainbridge, *How to Build a Digital Library,* Morgan Kaufmann, 2003. ISBN 1-55860-790-0

How to Build a Digital Library is the only book that offers all the knowledge and tools needed to construct and maintain a digital library-no matter how large or small. Two internationally recognized experts provide a fully developed, step by step method, as well as the software that makes it all possible. *How to Build a Digital Library* is the perfectly self-contained resource for individuals, agencies, and institutions wishing to put this powerful tool to work in their burgeoning information treasuries.

Sketches the history of libraries-both traditional and digital-and their impact on present practices and future directions

Offers in-depth coverage of today's practical standards used to represent and store information digitally

Uses freely accessible Greenstone open-source software-available with interfaces in the world's major languages (including Spanish, Chinese, and Arabic)

Written for both technical and nontechnical audiences

Web-enhanced with software documentation, color illustrations, full-text index, source code, and more at *www.mkp.com/DL*

