



InterPARES 3 Project

International Research on Permanent Authentic Records in Electronic Systems

TEAM Canada

Title: General Study 08 – Open-Source Records Management Software: Final Report

Status: Final (public)

Version: 1.4

Date Submitted: November 2010

Last Revised: October 2013

Author: The InterPARES 3 Project, TEAM Canada

Writer(s): Corinne Rogers
School of Library, Archival and Information Studies, The University of British Columbia

Elizabeth Shaffer
School of Library, Archival and Information Studies, The University of British Columbia

Project Unit: Research

URL: http://www.interpares.org/ip3/display_file.cfm?doc=ip3_canada_gs08_final_report.pdf

Document Control

Version history			
<u>Version</u>	<u>Date</u>	<u>By</u>	<u>Version notes</u>
1.2	2010-11-01	C. Rogers, E. Shaffer	Draft of final report.
1.3	2013-05-02	R. Preston	Copy and minor content edits.
1.4	2013-10-31	R. Preston	Added page numbers and other minor copy and content edits for public version.

Table of Contents

Introduction.....	1
Electronic records management and digital preservation	1
What is open-source software?	2
Open-source software and digital preservation.....	3
Open-source software and records management	4
A brief survey of open-source electronic records management tools.....	6
<i>DISSCO</i>	6
<i>KnowledgeTree</i>	7
<i>Alfresco Enterprise Content Management</i>	7
Conclusion	10
References.....	11
Appendix 1: The Open Source Definition	13

Introduction

In many organizations, the lack of electronic records management capability can be one of the most serious impediments to creating, maintaining and preserving authentic electronic records. Numerous proprietary electronic document and records management systems (EDRMS) are currently on the market, but their high cost place them out of reach of small and medium-sized organizations. In recent years, however, several open-source EDRMS have emerged, introducing the possibility of implementing electronic records management without paying heavy software licensing costs. This case study originally proposed to evaluate these products and “[map]...their functionality to the InterPARES Creator and Preserver Guidelines, and records management and archival standards, MoReq2, ISAD(G) and ISO 15489.”¹ After some review, the case study team decided to focus on Alfresco Records Management, the most widely used product and the only one certified to comply with the *Design Criteria Standard for Electronic Records Management Software Applications* (DoD 5015.2). However, doubts soon arose about the open-source nature of Alfresco’s records management product and about its usability, quality of user documentation and availability of support. Because of this, the case study ultimately focused not on mapping the product’s functionality to international standards but on determining whether an organization would be likely to deploy it successfully. For the reasons outlined in this report, the case study team ultimately concluded that Alfresco does not provide an open-source electronic records management tool that is feasible for use in small to medium-sized organizations.

Electronic records management and digital preservation

InterPARES has long been aware that good electronic records management supports the creation and maintenance of authentic records. The original UBC Project, which ran from 1994 to 1997, established rules for records classification, registration and consignment to a central record keeping system.² These rules were developed into DoD 5015.2, a well-accepted international standard for the design of electronic records management software tools. InterPARES 1 followed this work with a set of benchmark requirements for supporting the presumption of authenticity of electronic records, requirements which relate to the ability to identify records (including the authoritative version of the record if there are multiple copies) and to determine the controls on the creation, access to, modification and relocation of the records. Other key requirements include evidence of procedures to “prevent, discover, and correct loss or corruption of records.”³

Electronic records management systems are designed to allow an organization to implement and enforce these requirements. An EDRMS provides an environment in which a document can be declared a record, metadata can be added to establish its identity, and the record

1 General Study Research Proposal: Open Source Records Management Software, Version 1.1, 22 November 2009, available at http://www.interpares.org/ip3/display_file.cfm?doc=ip3_canada_gs08_research_proposal_v1-1.pdf.

2 See Luciana Duranti, Terry Eastwood and Heather MacNeil (1997), “The Preservation of the Integrity of Electronic Records,” <http://www.interpares.org/UBCProject/index.htm>.

3 Requirements for Assessing and Maintaining the Authenticity of Electronic Records, Authenticity Task Force, InterPARES 1 Project, March 2002, p. 6, available at http://www.interpares.org/display_file.cfm?doc=ip1_authenticity_requirements.pdf.

can be filed into a functional classification scheme in order to provide information about its provenance, nature and purpose. A DoD 5015.2-compliant EDRMS provides a complete audit trail of actions taken against the record, from registration to removal from the system, in order to provide evidence that the record has not been inappropriately modified or otherwise corrupted during the active and semi-active stages of its lifecycle. Verifying authenticity of records generated outside such systems is sometimes possible but it is difficult and time-consuming, relying on such measures as “a comparison of the records in question with copies that have been preserved elsewhere, or with back-up tapes; comparison of the records in question with entries in a register of incoming and outgoing records; textual analysis of the record’s content; forensic analysis of the medium, script, and so on; a study of audit trails; and the testimony of a trusted third party.”⁴

For small to medium-sized organizations, the focus of the InterPARES 3 case studies, the ability to verify the authenticity of records acquired from poorly managed record-keeping systems is a daunting challenge. The shared network drives of even a small organization may contain tens or hundreds of thousands of poorly identified, disorganized, fragmented and redundant documents and records. Unfortunately, commercially available EDRMS products tend to be prohibitively expensive, both in licensing fees and in integration and training costs. The recent emergence of open-source EDRMS tools, however, appears to offer the possibility that resource-poor organizations will be able to implement good electronic recordkeeping while avoiding heavy software licensing and integration costs, relying on publicly available developer support and a community of users for assistance.

What is open-source software?

In general terms, open-source software is software that can be freely used, modified and redistributed through access to its source code. Open-source software is defined by the license that makes it available to the public; although the code writers may retain copyright, the license waives most rights typically associated with copyrighted work. At a minimum, in order for the software to be considered open-source, the license must allow access to the source code, not just compiled versions of the code, and users must be able to modify the source code and redistribute the modified versions (derivatives). There are several variations on this theme: for example, some licenses allow selling derivatives or combining the software with other software which is then sold together as a package. However, a key requirement of all open-source licenses is that a user must not redistribute the code under terms that are more restrictive than those under which it was originally released, ensuring that software that started out as open-source does not eventually change into proprietary or otherwise restricted software.⁵

⁴ Ibid., p. 3.

⁵ There are many articles that explain the intricacies of various licenses. A good summary can be found at <http://www.linux.com/news/biz-os/legal/28138-licensing-101-for-open-source-projects-pick-a-license>. For a list of open-source licenses approved by the Open Source Initiative, see <http://www.opensource.org/licenses/category>.

In practice, compiled versions of the code are also typically made freely available on the Internet; users are able to download the software and use it without charge. User support may be provided by freely available on-line documentation, listservs, user and developer discussion lists and similar means. A mature discussion list or listserv often provides an arena for the software developers to help users and for users to help other users, and lengthy exchanges on these lists can often result in improvements to future releases of the software. Some open-source software may attract the interest of commercial companies that provide support and training contracts; the software is free but services surrounding the use of the software need not be. Other service models can be built around a large organization developing expertise in the use of the software and sharing that expertise with smaller organizations, or a number of organizations collaborating to implement the software and share expertise and even resources with one another.

Open-source software and digital preservation

The library and archival communities have embraced open-source software for digital preservation. In the U.S., California Digital Library, Harvard University, University of Florida, Stanford University, Cornell University, the Massachusetts Institute of Technology, the University of California at Berkeley and San Diego and other leading institutions have developed and distributed open-source tools for digital preservation, including repository software and tools for format identification and validation.⁶ The National Archives of the United Kingdom has developed a file format registry and an open-source tool for format identification using the registry,⁷ and a large-scale collaborative digital preservation project using open-source repository software was undertaken at a number of British universities in 2007.⁸ The National Archives of Australia has produced tools for digital preservation workflow management and format normalization,⁹ and has collaborated with the National Library of New Zealand and the UK Web Archiving Consortium to produce an open-source web archiving tool.¹⁰ In Canada, a collaborative project is underway to design an OAIS-based preservation system that integrates a suite of open-source tools and makes them available via a single user interface.¹¹

Individual projects for developing open-source software tools for digital preservation are beginning to coalesce into stable, long-term national and multi-national undertakings. In Europe, a recently concluded research project called Planets (Preservation and Access through Long-

6 See JHOVE-2, The Next Generation Architecture for format-aware Characterization, <https://bitbucket.org/jhove2/main/wiki/Home>; DSpace, <http://www.dspace.org/>; Florida Digital Archive, http://fclaweb.fcla.edu/FDA_landing_page; FedoraCommons, <http://fedora-commons.org/>; and Lots of Copies Keep Stuff Save (LOCKSS) homepage, <http://lockss.stanford.edu/lockss/Home>.

7 The Technical Registry, PRONOM, <http://www.nationalarchives.gov.uk/PRONOM/Default.aspx> and DROID (Digital Record Object Identification) at <http://droid.sourceforge.net/>.

8 See the OpenLOCKSS Project based at Glasgow University, <http://www.lib.gla.ac.uk/Research/openlockss/index.shtml>.

9 National Archives of Australia Tools for Digital Preservation, <http://www.naa.gov.au/records-management/preserve/e-preservation/at-NAA/software.aspx>.

10 The Web Curator Tool. See <http://www.natlib.govt.nz/services/get-advice/digital-libraries/web-curator-tool>.

11 The Archivematica project, <http://www.archivematica.org>. OAIS is the Reference Model for an Open Archival Information System, an ISO standard for digital preservation systems. See http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=24683.

Term Networked Services), which analyzed and developed open-source tools for digital preservation planning and file format conversion, has transformed itself into a non-profit organization hosted by the British National Library. Dr. Adam Farquhar, Planets Project Coordinator, writes that he expects the new organization “to encourage take-up of Planets technology, provide stable hosted access to Planets Services [and] coordinate further open-source development....”¹² In the U.S., the Library of Congress, which released its first open-source tool in 2008, recently announced the establishment of new internal procedures for streamlining the process of creating open-source software, in order to “allow the Library and its partners to more fully participate in the open source development community.”¹³

Open-source software and records management

In contrast, open-source software has made very few inroads into the world of records management. Until recently, in fact, electronic records management systems have been exclusively proprietary. There are a several reasons for this, which serve to underscore some of the key differences between the records management and archival professions and the different communities they serve:

1. Archives are collaborative, records management is institution-based

Redundancy is becoming one of the cornerstones of digital preservation. Redundancy means that one institution can preserve the digital objects of another institution, either at the same time for the sake of backup and security or as a successor organization in the event the original repository ceases to exist. One type of popular repository software, called Lots of Copies Keeps Stuff Safe (LOCKSS), is built around the premise that a consortium of institutions, preferably no fewer than seven, works together using the same systems to preserve, back up and provide public access to each others’ content. Many other preservation projects are collaborative, including Toward Interoperable Preservation Repositories (TIPR), a joint effort of New York and Columbia Universities and the Florida Digital Archive designed to develop a set of standards for the interoperability of heterogeneous digital preservation systems.¹⁴ In some cases, a single organization acts as a centralized repository for a network of linked institutions; an example is the Florida Digital Archive, which is a repository for the digital objects of all university libraries in the state.¹⁵ This type of collaborative environment heavily favours the development and use of open-source software because institutions that are working together to accomplish the same tasks fare better when they are using the same software tools. In some collaborative projects, one

¹² Planetarium, the New Bulletin of the Planets Programme, December 2009, www.planets-project.eu/publications.

¹³ Library of Congress Explores Ways to Release Open Source Software, January 14, 2010, http://www.digitalpreservation.gov/news/2010/20100114news_article_open_source.html.

¹⁴ See Priscilla Caplan, “Repository to Repository Transfer of Enriched Archival Information Packages,” *D-Lib Magazine* November/December 2008, <http://www.dlib.org/dlib/november08/caplan/11caplan.html>. The published draft specification is now available at <http://wiki.fda.edu:8000/TIPR/21>.

¹⁵ Florida Digital Archive, <http://www.fcla.edu/digitalArchive/index.htm>. See also California Digital Library’s UC3Merritt, <http://www.cdlib.org/services/uc3/merritt/>.

institution develops tools to share with the others; in others, particularly now that there are so many tools available, a group of like-minded organizations need only agree on the tools to use and download them from the Internet in order to get started.

In contrast, records management programs typically work in isolation. Inter-institutional sharing of records for the purpose of providing redundant storage and care of active records is rare because of security and privacy concerns, and, since planning for permanent preservation is not the focus of a records management program, successor planning (i.e., designating another institution to take custody and control of the records in the event the originating institution ceases to exist) is rarely considered. In this environment, sharing of software tools is largely irrelevant because the records do not have to be shared between systems.

2. There is big money to be made developing EDRMS software

Large, resource-rich organizations which would never consider putting money into an archival program are required by practical and legal considerations to manage their electronic records. This means that there is a much larger pool of potential clients for private companies developing proprietary software, and in fact proprietary software for electronic records management has reached a high level of maturity. Only the largest of archives, however, has the resources to pay potentially millions of dollars for software licenses. Moreover, EDRMS implementations require anyone within an organization who creates and uses records to have a desktop license to interact with the records repository; an archives may need only a few specialized staff to interact with a digital preservation system. Thus a municipal government may require 5,000 EDRMS software licenses and only six digital preservation repository licenses. The differences in expenditure for software licenses mean that commercially licensing EDRMS software is highly lucrative while commercially licensing digital repository software may be hardly worth the effort.

3. Active records are not considered cultural assets

Archives and libraries hold cultural assets. In practical terms this means that they often receive government funding, typically in the form of short-term grants, requiring outputs which will provide a general benefit to society—such as the production of new software tools that can be shared freely with others. However, organizations that hold records only for the purposes of conducting their daily business and meeting their legal obligations are not perceived as providing an immediate, tangible cultural benefit to society at large through their records, and must therefore come up with their own money, removing any incentive to develop tools that can be shared for free with other organizations.

4. EDRMS tools are integrated with other software products

In digital preservation, the digital objects are removed from their originating systems and handed over to the custody and control of the preserver, which may use software systems that the producers have never heard of and with which they need not concern themselves. The preserving archives may have a well established culture of using non-standard, niche and open-source tools to accomplish its tasks, because its activities are so highly specialized. EDRMS software, on the

other hand, is by necessity tightly integrated with the operating systems, office products and other software tools used by the parent institution. Managers may feel that an open-source product is not a good fit with the existing software environment, and IT departments may be unwilling to support software that does not have the backing of large, well-established and familiar software vendors.

Despite these considerations and concerns, a few open-source tools for electronic records (and document) management have emerged in the last two or three years. The rest of this report discusses these tools and their suitability for use by small to medium-sized institutions.

A brief survey of open-source electronic records management tools

The case study team investigated three open source records management systems: Document Management Integrated System for Scientific Organizations (DISSCO); KnowledgeTree; and Alfresco Records Management.

DISSCO

According to documentation on the project website, DISSCO is an electronic records management system designed to support “basic administrative processes within scientific institutions.”¹⁶ The development of DISSCO was a project of the Multiannual Information Society Support Programme (2001-2008) financed by the Belgian Federal Science Policy Office. A four-year project that ended on 31 May 2007, DISSCO was intended to meet the needs of public scientific institutions requiring “open source software and open formats.”¹⁷ and was designed to comply with ISO 15489 and *Model Requirements for Electronic Records Management Systems* (MoReq) records management standards and the *General International Standard Archival Description* ISAD(G). DISSCO is modular and consists of four main functionalities: metadata, information, workflow and security (user rights management).¹⁸

According to their site, the four project partners—Centre for Historical Research and Documentation on War and Contemporary Society; Royal Meteorological Institute (IRM-KMI); University Libre de Bruxelles; and Vrije Universiteit Brussel—participated in the development and will test implementation of DISSCO. However, while basic information about the DISSCO project and software is available on their website, there is no data about the project beyond 2006.¹⁹ The case study team could not locate any data on implementation or availability of DISSCO, nor any implementations of the DISSCO system. This lack of information and the specificity of DISSCO to a niche community (public scientific institutions) led the team to rule out DISSCO as a potential open-source EDRMS solution for small- and medium-sized organizations.

¹⁶ DISSCO, “Objectives,” available at <http://www.meteo.be/DISSCO/objectives.html>.

¹⁷ DISSCO Extended Report, available at <http://www.meteo.be/DISSCO/publications.html>.

¹⁸ DISSCO, “Main Functionalities,” available at <http://www.meteo.be/DISSCO/functionalities.html>.

¹⁹ The case study team attempted to reach the contact on the DISSCO site to obtain further information on the current status of the project and determine whether the implementation took place, however, an email to the contact went unanswered.

KnowledgeTree

KnowledgeTree is a cloud-based document management system using the Amazon EC2 platform. It offers both a paid enterprise and free, open-source community edition. The community edition is licensed under a GNU-GPL (see Appendix 1 for an explanation of free software and open source software licenses). KnowledgeTree markets itself as a secure and affordable online document sharing and control system for small- and medium-sized businesses.²⁰ The community edition is written in PHP and uses the Apache web server and MYSQL database management system.

A survey of the community edition's capabilities²¹ shows that it does not support the electronic records management requirements identified by the UBC Project and InterPARES 1 as being necessary to establish a presumption of authenticity. KnowledgeTree is in fact a document management system: there is no defined records management module and this tool lacks the ability to integrate with existing desktop applications and to provide compliance capabilities and security sufficient for it to act as an EDRMS. Additionally, there appears to be a lack of support for the open source community edition that would be necessary for small and medium-sized organizations to effectively deploy an open-source records management solution.²²

Alfresco Enterprise Content Management

Alfresco offers an enterprise content management (ECM) solution, stated on its website to be “the leading open source alternative... [that] couples the innovation of open source with the stability of a true enterprise-class platform.”²³ The records management module is DoD 5012.02 certified and “has been implemented on top of a generalized records management metadata model, allowing other standards (such as MoReq2, NOARK, etc.) to be supported.”²⁴

Alfresco records management comes in two forms - a Community Edition and an Enterprise Edition. The Community Edition is available as a free download from the website, and is supported by a downloadable manual. The Enterprise Edition is available only to customers who purchase an Enterprise subscription, which provides access to Alfresco Enterprise and a number of subscription-only services, including Alfresco technical support, access to online resources and services through “The Alfresco Network,” maintenance releases, patches and hot fixes, a quality assurance program, platform support, and warranty and indemnification. The Enterprise Edition is licensed under a commercial license, the terms of which are not available on Alfresco's website.

Alfresco hosts a wiki for developers who want to view source code and contribute “platform fixes and enhancements” to the product line.²⁵ Contributors are encouraged to upload their work to be considered for inclusion by Alfresco, and although they are welcome to make

²⁰ KnowledgeTree, “About KnowledgeTree,” available at <http://www.knowledgetree.com/company>.

²¹ KnowledgeTree, “Compare Products,” available at http://www.knowledgetree.org/Compare_Products.

²² The KnowledgeTree product sheet indicates that there is no training, set-up, phone or web-based support available for the community edition. KnowledgeTree, “Compare Products,” available at http://www.knowledgetree.org/Compare_Products.

²³ Alfresco, “About Alfresco,” available at <http://www.alfresco.com/about/>.

²⁴ <http://www.alfresco.com/industries/government/>

²⁵ See http://wiki.alfresco.com/wiki/Source_Code.

their work available under any license they choose, they must sign a contribution agreement with Alfresco for their work to be included in Alfresco's source tree. This agreement "define[s] the intellectual property license granted by persons or entities that contribute code to [Alfresco] for the Project," and grants to Alfresco

- a perpetual, irrevocable, non-exclusive, worldwide, fully paid-up, royalty-free, unrestricted license to exercise all rights (including sublicensing) under all worldwide copyrights, copyright applications and registrations in the Contribution; and
- a perpetual, irrevocable, non-exclusive, worldwide, full paid-up, royalty-free patent license to make, have made, use, offer to sell, sell, import, and otherwise transfer your Contribution and derivative works thereof, where such license applies only to those patent claims licensable by you that are necessarily infringed by your Contribution alone or by combination of your Contribution with the Project to which you submitted the Contribution.²⁶

In January, 2010, the case study team attended a "Lunch and Learn" session in Richmond, BC, hosted by members of the Alfresco sales team. At this session, the Alfresco representatives stated that although the source code was freely available for Alfresco's Community Edition, it was controlled for the Enterprise Edition through their subscription service. This was confirmed in an e-mail the following week: in response to a question about whether organizations purchasing an Enterprise subscription license were free to modify and redistribute the source code, the Alfresco representative replied "The Alfresco Enterprise product is available through a subscription which includes support, maintenance, and an open source commercial license. The commercial license protects the product from [being] redistributed [by] customers and vendors."²⁷

This information led the case study team to conclude that although the Enterprise Edition of Alfresco is marketed as open-source software, only the Community version is actually open-source. Accordingly, in September, 2010, the case study team downloaded Alfresco Community Edition software and the user's manual. Of immediate concern was the copyright notice that covered the manual, expressly forbidding anyone to reproduce or transmit any portion of the document by any means, binding users to strict US trade and patent laws, and forbidding use of the document "for the purposes of evaluating [the software's] functionality or for any other competitive purposes."²⁸ Another concern was the length of the manual: at only 32 pages, it seemed unlikely to contain all the information needed to use Alfresco to successfully implement electronic records management in even a small organization.

²⁶ Alfresco wiki, "Contribution Agreement," available at <http://wiki.alfresco.com/wiki/File:Contribution-Agreement-22-Jan-2010.pdf>.

²⁷ E-mail from Joe Morrison, January 18, 2010.

²⁸ Alfresco Community Edition, "Getting Started with Alfresco Records Management," 2010, http://wiki.alfresco.com/wiki/File:Getting_Started_with_Alfresco_Records_Management_for_Community_Edition_3_3.pdf.

In addition to the user manual, the Community Edition is supported through the Alfresco forum, to which users can post their questions.²⁹ Members of the Alfresco development team do monitor this forum, but do not respond to all queries. As the case study team navigated the program, following the installation and implementation steps outlined in the manual, several questions arose about the usability of the software. As a result, the team posted two questions to the records management wiki on September 21, as follows:

Using Community Edition in an organization

by **corinne.rogers** » 21 Sep 2010, 22:50

Hello,

We have gone over the Community Edition (3.3) RM manual and were wondering how users could integrate records management into their regular workflow. Does the user have to open up Alfresco and upload documents into the file plan? Or, for example, can they save a document directly from something like MS Word into the file plan?

Thanks for your help.
corinne.rogers

automatic metadata input?

by **corinne.rogers** » 21 Sep 2010, 22:38

Hello,

We are trying out the Community version of Al Fresco (3.3) and are wondering about manual vs. automated metadata input. In order to declare a record there seem to be quite a few fields for the user to fill out. Can this be automated or inherited in some way?

Thank you.
corinne.rogers

As of August 5, 2011, the questions had been viewed 289 and 228 times respectively, but no answers had been posted.

Part of the success of open source software implementations depends on the availability of good user documentation and timely and knowledgeable response of the user community to questions as they arise. The case study team felt that the absence of detail in the user manual coupled with the lack of responses to its questions on the user forum indicated that user support for Alfresco Community Edition was too limited to continue with further implementation of the software, even as part of a test-bed case for InterPARES 3. In fairness to Alfresco, the company

²⁹ See <http://forums.alfresco.com/en/>.

does admit that the Community Edition is not in fact a product designed to be implemented easily by organizations lacking resources and a high degree of technical expertise: according to its website, “Alfresco Community is recommended for developers and highly technical enthusiasts in non-mission critical environments. As this version is unsupported it is intended to be used by those happy to spend time and resources solving issues independently.”³⁰ However, this statement may come as a surprise to those who only hear that Alfresco is the world’s first DoD-compliant open-source electronic records management tool.

Conclusion

The case study team’s initial exploration of available open source records management software led it to conclude that there was only one option that appeared to be open-source and in compliance with records management standards: Alfresco Enterprise Content Management. However, for the reasons outlined in this report, the team does not believe that even Alfresco offers a viable open-source electronic records management system for most small and medium-sized organizations. Alfresco’s interpretation and application of the open-source licensing model for its subscription-supported Enterprise Edition does not appear to comply with the intent or spirit of free and open-source software licensing. In theory, the Community Edition allows organizations to download and use electronic records management software for free. In practice, however, the lack of support for the Community Edition means that it is unlikely to be useful to organizations lacking the resources to purchase a commercial EDRMS.

An examination of the available products led the case study team to conclude that the use of an open-source tool may not currently be a viable model for electronic records management. The development and distribution model for electronic records management software is capital-intensive and profit-driven, unlike software for digital preservation which is collaborative and community-driven. The proprietary nature of the model provides little incentive for the development of a truly open-source solution, at least at the present time. This can only change when the same types of institutions developing open-source preservation software recognize that preserving authentic electronic records begins at creation, and begin to lead the way in developing the necessary community support for electronic records management software. Even Alfresco Records Management Community Edition could become a viable open-source electronic records management tool if a large institution, or group of institutions, chose to fork the existing source code (i.e., copy it and further develop it under a different name, which is permitted by its open-source license) and support it in the same way that digital repository software such as LOCKSS is supported. If and when this happens, the archives community could look forward to using software tools that support the full life-cycle of electronic records, rather than just the stage at which the records have been removed to the custody of a trusted preserver.

³⁰ Alfresco Enterprise and Alfresco Community Edition Comparison, <http://www.alfresco.com/products/networks/compare/>.

References

[Note: All URLs last accessed August 9, 2011.]

Alfresco, "About Alfresco." <http://www.alfresco.com/about/>.

Alfresco, "Contribution Agreement." http://wiki.alfresco.com/wiki/Source_Code.

Alfresco, "Getting Started with Alfresco Records Management" (2010).
http://wiki.alfresco.com/wiki/File:Getting_Started_with_Alfresco_Records_Management_for_Community_Edition_3_3.pdf.

The Archivemata Project. <http://www.archivemata.org>.

California Digital Library, UC3Merritt. <http://www.cdlib.org/services/uc3/merritt/>.

Caplan, Priscilla, "Repository to Repository Transfer of Enriched Archival Information Packages," *D- Lib Magazine* 14 (November/December 2008).
<http://www.dlib.org/dlib/november08/caplan/11caplan.html>.

DISSCO, "Extended Report (March 2004)." <http://www.meteo.be/DISSCO/publications.html>.

DISSCO, "Main Functionalities" <http://www.meteo.be/DISSCO/functionalities.html>.

DISSCO, "Objectives." <http://www.meteo.be/DISSCO/objectives.html>.

DROID (Digital Record Object Identification). <http://droid.sourceforge.net/>.

DSpace. <http://www.dspace.org/>.

Duranti, Luciana, Terry Eastwood and Heather MacNeil (1997), "The Preservation of the Integrity of Electronic Records." <http://www.interpares.org/UBCProject/index.htm>.

Florida Digital Archive. http://fclaweb.fcla.edu/FDA_landing_page.

Free Software Foundation, "The GNU Project – Free Software Definition,"
<http://www.gnu.org/philosophy/free-sw.html>.

GNU Operating System, "GNU General Public License." <http://www.gnu.org/licenses/gpl.html>.

InterPARES 1, Authenticity Task Force, "Requirements for Assessing and Maintaining the Authenticity of Electronic Records." (March 2002).
http://www.interpares.org/display_file.cfm?doc=ip1_authenticity_requirements.pdf.

InterPARES 3, "General Study Research Proposal: Open Source Records Management Software, Version 1.1" (22 November, 2009).
http://www.interpares.org/ip3/display_file.cfm?doc=ip3_canada_gs08_research_proposal_v1-1.pdf.

JHOVE-2, "The Next Generation Architecture for format-aware Characterization."
<https://bitbucket.org/jhove2/main/wiki/Home>.

KnowledgeTree, “About KnowledgeTree.” <http://knowledgetree.com/company>.

KnowledgeTree “Compare Products.” http://www.knowledgetree.org/Compare_Products.

Library of Congress, “Library of Congress Explores Ways to Release Open Source Software” (January 2010).
http://www.digitalpreservation.gov/news/2010/20100114news_article_open_source.html.

National Archives of Australia, “Tools for Digital Preservation.”
<http://www.naa.gov.au/records-management/preserve/e-preservation/at-NAA/software.aspx>.

National Library of New Zealand, Web Curator Tool, <http://www.natlib.govt.nz/services/get-advice/digital-libraries/web-curator-tool>.

OpenLOCKSS Project, Glasgow University Library.
<http://www.lib.gla.ac.uk/Research/openlockss/index.shtml>.

Open Source Initiative, “Mission.” <http://www.opensource.org/>.

Open Source Initiative, “Licenses.” <http://www.opensource.org/licenses/category>.

“Planetarium, the New Bulletin of the Planets Programme” (December 2009).
<http://www.planets-project.eu/publications>.

The Linux Foundation, “Licensing 101 for Open source Projects - Pick a License.”
<http://www.linux.com/news/biz-os/legal/28138-licensing-101-for-open-source-projects-pick-a-license>.

The National Archives, “PRONOM: The Technical Registry.”
<http://www.nationalarchives.gov.uk/PRONOM/Default.aspx>.

Appendix 1: The Open Source Definition³¹

Introduction

Open source doesn't just mean access to the source code. The distribution terms of open-source software must comply with the following criteria:

1. Free Redistribution

The license shall not restrict any party from selling or giving away the software as a component of an aggregate software distribution containing programs from several different sources. The license shall not require a royalty or other fee for such sale.

2. Source Code

The program must include source code, and must allow distribution in source code as well as compiled form. Where some form of a product is not distributed with source code, there must be a well-publicized means of obtaining the source code for no more than a reasonable reproduction cost preferably, downloading via the Internet without charge. The source code must be the preferred form in which a programmer would modify the program. Deliberately obfuscated source code is not allowed. Intermediate forms such as the output of a preprocessor or translator are not allowed.

3. Derived Works

The license must allow modifications and derived works, and must allow them to be distributed under the same terms as the license of the original software.

4. Integrity of The Author's Source Code

The license may restrict source-code from being distributed in modified form *only* if the license allows the distribution of "patch files" with the source code for the purpose of modifying the program at build time. The license must explicitly permit distribution of software built from modified source code. The license may require derived works to carry a different name or version number from the original software.

5. No Discrimination Against Persons or Groups

The license must not discriminate against any person or group of persons.

6. No Discrimination Against Fields of Endeavor

The license must not restrict anyone from making use of the program in a specific field of

³¹ Open Source Initiative, "Open Source Definition," available at <http://www.opensource.org/docs/osd>.

endeavor. For example, it may not restrict the program from being used in a business, or from being used for genetic research.

7. Distribution of License

The rights attached to the program must apply to all to whom the program is redistributed without the need for execution of an additional license by those parties.

8. License Must Not Be Specific to a Product

The rights attached to the program must not depend on the program's being part of a particular software distribution. If the program is extracted from that distribution and used or distributed within the terms of the program's license, all parties to whom the program is redistributed should have the same rights as those that are granted in conjunction with the original software distribution.

9. License Must Not Restrict Other Software

The license must not place restrictions on other software that is distributed along with the licensed software. For example, the license must not insist that all other programs distributed on the same medium must be open-source software.

10. License Must Be Technology-Neutral

No provision of the license may be predicated on any individual technology or style of interface.