Repository Software Evaluation using the Audit Checklist for Certification of Trusted Digital Repositories

Joanne S. Kaczmarek
University Archives, University of
Illinois at Urbana-Champaign
1408 W. Gregory Drive
Urbana, IL 61801
+1 217-333-6834

jkaczmar@uiuc.edu

Thomas G. Habing
Grainger Engineering Library,
University of Illinois at Urbana-
Champaign
1301 W. Springfield Avenue
Urbana, IL 61801
+1 217-244-4425

thabing@uiuc.edu

Janet Eke
Graduate School of Library and
Information Science, University of
Illinois at Urbana-Champaign
501 E. Daniel Street
Champaign, IL 61820-6211
+1 217-333-4701

jeke@uiuc.edu

ABSTRACT

The NDIIPP ECHO DEPository project [1] digital repository evaluation will use an augmented version of the draft Audit Checklist for Certification of Trusted Digital Repositories (Audit Checklist) [2] to provide a framework for examining how well currently popular repository software applications support the notion of a "trusted digital repository." The evaluation will also demonstrate the application of a scoring software evaluation methodology similar to one developed by the Center for Data Insight (CDI) at Northern Arizona University [3], used for evaluation data mining software. This scoring methodology in conjunction with the Audit Checklist can be used as a tool by librarians, archivists, and other data custodians to make informed decisions as they develop digital preservation management services.

Categories and Subject Descriptors

H.3.7 Digital Libraries, Systems Issues

General Terms

Measurement, Documentation

Keywords

Digital Preservation Management, Repositories, Evaluation

1. INTRODUCTION

The ECHO DEPository is a 3-year Library of Congress National Digital Information Infrastructure and Preservation Program project at the University of Illinois at Urbana-Champaign. The project is undertaken in partnership with the Online Computer

Library Center (OCLC) [4] and a consortium of content provider partners. One component of the project is the evaluation of various open source repository software applications. The evaluation will focus on how these applications support activities of an institution or organization interested in providing services associated with a trustworthy digital repository. The framework for the evaluation has been developed from the Audit Checklist undertaken by widespread efforts coordinated by the Research Libraries Group (RLG) [5] and NARA [6]. Repository software application evaluations previously conducted have included initiatives on behalf of the Open Society Institute [7] and as part of other NDIIPP-related activities [8]. While these efforts have focused on some technical attributes of some of the repository applications to be evaluated in this study, they predate the release of the Audit Checklist. The highly successful Digital Preservation Management workshops of Cornell University concisely articulate a three-pronged approach to digital preservation, including technology, resources, and management. [9] The ECHO DEPository project's adaptation of the Audit Checklist is undertaken with an interest in evaluating repository applications in their context as components within the larger organizational commitment toward trustworthy digital preservation. The evaluation also aims to inform decisions in the future development of digital preservation management services.

2. AUDIT CHECKLIST

The Audit Checklist, still under development, provides a means by which an institution can perform a self-evaluation to determine how well it is positioned to provide an expected level of trustworthiness. Project team members reviewed the list to determine which items might apply specifically to repository software applications. For each item on the Checklist the question was asked, "How might a repository software application support an institution to meet this criterion?" Items that did not appear to apply to repository software applications were ignored and items that did seem to apply were expanded.

Expanded items have specific details listed. These details are the benchmark criteria used with each application. Original checklist document language has been modified where appropriate to conform to language used in the Reference Model for an Open Archival Information System (OAIS).[10]

B. Repository Functions, Processes & Procedures
B.1.3. Repository has an identifiable, written definition for each SIP or class of Content Information ingested by the repository. How well does the repository software document its submission requirements?

Figure 1: Sample of Modified Audit Checklist

3. SCORING METHODOLOGY

In addition to a narrative qualitative evaluation presented using the Audit Checklist as a framework, we plan to explore the use of a scoring methodology. based on standard decision matrix concepts. [3] We do not intend to provide quantitative scores of the repositories under consideration within the context of the project, but rather to suggest a methodology that may provide an additional useful tool for those charged with developing digital preservation services for their institutions. As an added aid, we will present an example case study showing the application of the scoring methodology based on a local repository scenario.

The selection criteria of the methodology will be our modified Audit Checklist, available through our project website. In applying the scoring methodology, each of the selection criteria is weighted. These weights must necessarily be assigned according to local needs and intended uses of the software; therefore, the weights will vary across different uses of the methodology. For purposes of our example evaluation, sample weights will be assigned according to our example scenario. The individual repository software packages are then scored based on how well they meet the criteria. The methodology requires assigning one software package an average score for all criteria. This becomes a reference repository used to compare additional repositories rated as to whether they are: much worse (1), worse (2), the same as (3), better (4), or much better (5) than the reference. The scores across all criteria are then totaled to give an overall score for each repository. The criteria may also be categorized to give subtotals for different categories of criteria with each category potentially having its own weight.

Criteria	Wgt.	Repo A	Repo B
B1. Ingest/acquisition of content B1.1 Repository identifies prop…	0.10		
B1.3 Repository has an identi…	0.20		
…			

Figure 2: Sample of Scoring Matrix

4. DATA COLLECTION

Data used for evaluating the repository software applications is gathered by project team members throughout the project timeline. The repository software applications to be evaluated by the ECHO DEPository project include DSpace, Eprints, Fedora, Greenstone, and the OCLC Digital Archive. Other repository systems may be included as project resources permit.

Initial data gathering has been undertaken during the course of installing each repository within the project environments provided by the University of Illinois Grainger Engineering Library and the Graduate School of Library and Information Science. Other data is being collected during the course of ingesting digital content into each repository software application as well as during dissemination of digital content between repositories. Collected data will be used to provide narrative feedback using the Modified Audit Checklist for each repository.

5. ANTICIPATED OUTCOMES

Anticipated outcomes for this portion of the ECHO DEPository work include a simple qualitative methodology and framework that can be used to assist in decision-making when considering digital preservation management services. A spreadsheet or an interactive web application may be developed to assist decision-makers in applying the framework and methodology within their own environments. Separate white papers will also be produced to articulate specific details of our experiences with each repository software application, as well as recurring themes noted during data ingest and data exchange activities.

6. ACKNOWLEDGMENTS

7. REFERENCES

[1] ECHO DEPository Project http://ndiipp.uiuc.edu

[2] An Audit Checklist for Certifying Digital Repositories http://www.rlg.org/en/page.php?Page_ID=20769

[3] Collier, K., Carey, B., Sautter, D., and Marjaniemi, C. A Methodology for Evaluating and Selecting Data Mining Software. In *Proceedings of the 32nd Hawaii International Conference on System Sciences – 1999*. IEEE. http://csdl2.computer.org/comp/proceedings/hicss/1999/0001/06/00016009.PDF

[4] OCLC http://www.oclc.org

[5] RLG http://www.rlg.org/index.php

[6] NARA http://www.nara.gov

[7] The Open Society Institute Guide to Institutional Repository Software http://www.soros.org/openaccess/pdf/OSI_Guide_to_IR_Software_v3.pdf

[8] Archive Ingest and Handling Tests (AIHT) Reports and Appendices http://www.digitalpreservation.gov/index.php?nav=3&subnav=14

[9] Digital Preservation Management Workshop http://www.library.cornell.edu/iris/dpworkshop/

[10] OAIS Reference Model www.ccsds.org/documens/650x0b1.pdf