



Center *for* Research Libraries
.....
GLOBAL RESOURCES NETWORK

Report on Portico Audit Findings

Executive Summary

The Center for Research Libraries (CRL) conducted a preservation audit of Portico (www.portico.org) between April and October 2009 and hereby certifies Portico as a trustworthy digital repository. CRL has found that Portico’s services and operations basically conform to the requirements for a trusted digital repository. The CRL Certification Advisory Panel has concluded that the practices and services described in Portico’s public communications and published documentation are generally sound and appropriate to the content being archived and the needs of the CRL community. Moreover the CRL Certification Advisory Panel expects that in the future, Portico will continue to be able to deliver content that is understandable and usable by its designated user community.

This certification is based upon a site visit and sampling of archives content, and upon the review of information gathered by CRL and its Certification Advisory Panel and of documents and documentation provided by Portico. CRL’s analysis was guided by the criteria included in the *Trustworthy Repositories Audit and Certification* checklist, and other metrics developed by CRL through its analyses of digital repositories.

CRL conducted its audit with reference to generally accepted best practices in the management of digital systems; and with reference to the interests of its community of research libraries and the practices and needs of scholarly researchers in the humanities, sciences and social sciences in the United States and Canada. The purpose of the audit was to obtain reasonable assurance that Portico provides, and is likely to continue to provide, services adequate to those needs without material flaws or defects and as described in Portico’s public disclosures. The CRL audit provides a reasonable basis for these findings.

CRL has assigned Portico the following levels of certification (the numeric rating is based on a scale of 1 through 5, with 5 being the highest level, and 1 being the minimum certifiable level):¹

CATEGORY	PORTICO SCORE
Organizational Infrastructure	3
Digital Object Management	4
Technologies, Technical Infrastructure, Security	4

In the course of the audit, the Certification Advisory Panel identified a number of issues that Portico will need to address to more fully satisfy the concerns of CRL libraries. Those issues are described in Section B, *Detailed Audit Findings*, below, and pertain to: a) specific criteria in the TRAC checklist and b) the preservation interests and requirements of the CRL libraries, including the scope of the archive, functionality and services, and future costs and risks. Portico has agreed to address the issues identified by the panel and to make certain disclosures to CRL periodically, as a condition of continued certification. Those ongoing requirements are outlined in Section C of this report.

¹ A working version of the schema CRL uses in providing summary ratings of a repository’s compliance with the TRAC criteria is available at: <http://www.crl.edu/archiving-preservation/digital-archives/metrics-assessing-and-certifying/crl-ratings>.

About the Audit Participants

PORTICO

Portico (www.portico.org) is a not-for-profit digital preservation service providing a permanent archive of electronic journals, books, and other scholarly content. Portico is a part of ITHAKA (www.ithaka.org), a not-for-profit organization helping the academic community use digital technologies to preserve the scholarly record and to advance research and teaching in sustainable ways. As of October 2009, the Portico archive preserved over 14 million e-journal articles and 1,900 e-books. Portico worked with CRL in 2006 on a test of the *RLG/NARA Draft Audit Checklist for the Certification of Trustworthy Digital Repositories* and other metrics developed by CRL under a grant from the Andrew W. Mellon Foundation.

CENTER FOR RESEARCH LIBRARIES

The Center for Research Libraries (CRL - www.crl.edu) is an international consortium of university, college, and independent research libraries. CRL supports advanced research and learning in the humanities, sciences, and social sciences by ensuring the survival and accessibility of source materials vital to those disciplines. In order to enable its community to accelerate the shift to electronic-only resources in a careful and responsible manner, CRL has a hybrid strategy of preserving and maintaining shared physical collections of materials and certifying digital repositories of interest to its community.

CRL analysis of Portico documentation and operations was undertaken by Marie Waltz and Bernard Reilly. Additional technical support for the site visit and assessment was provided by James A. Jacobs, Data Services Librarian Emeritus, University of California, San Diego.

To guide its Portico audit, CRL formed a panel of advisors representing the various sectors of its membership. The Certification Advisory Panel is constituted so as to ensure that the certification process addresses the interests of the entire CRL community. The Panel includes leaders in collection development, preservation, library administration, and digital information technology. The members of the CRL Certification Advisory Panel are:

MARTHA BROGAN (CHAIR)

Director of Collection Development & Management
 University of Pennsylvania

ANNE POTTIER

Associate University Librarian
 McMaster University

WINSTON ATKINS

Preservation Officer
 Duke University

OYA Y. RIEGER

Associate University Librarian for Information
 Technologies
 Cornell University

BART HARLOE

Director of Libraries
 St. Lawrence University

PERRY WILLETT

Digital Preservation Services Manager
 California Digital Library

WILLIAM PAROD

Senior Repository Developer
 Northwestern University Libraries

A. Audit and Assessment Methodology and Criteria

This assessment evaluates whether or not Portico meets the commitments that it has made in regard to the long-term preservation of digital scholarly content for the academic community. A CRL audit includes a site visit, a sampling of content from the archives, a review of the information thus gathered, and a review of documents and documentation provided by digital repository.

CRL conducted its audit with reference to:

- generally accepted best practices in the management of digital systems
- the interests of the CRL community of research libraries
- the practices and needs of scholarly researchers in the humanities, sciences, and social sciences in the United States and Canada.
- the criteria included in *Trustworthy Repositories Audit & Certification: Criteria and Checklist*²
- the Open Archive Information System reference model³ (OAIS)
- other metrics developed by CRL through its analyses of digital repositories

CRL assigned to Portico a level of certification in each of three categories. The numeric rating is based on a scale of 1 through 5, with 5 being the highest level, and 1 being the minimum certifiable level.

The general metrics used by CRL in assessments are based on the *Trustworthy Repositories Audit and Certification* checklist, and on other metrics developed by CRL through its analyses of digital repositories. TRAC was developed by a joint task force created by the Research Libraries Group (RLG) and the National Archives and Records Administration between 2003 and 2005, to provide criteria to be used in identifying digital repositories capable of reliably storing, migrating, and providing access to digital collections. TRAC represents best current practice and thinking about the organizational and technical infrastructure required to be considered trustworthy and worthy of certification.

2 TRAC - http://www.crl.edu/sites/default/files/attachments/pages/trac_0.pdf

3 OAIS - <http://public.ccsds.org/publications/archive/650x0b1.pdf>

B. Detailed Audit Findings

As a result of the audit, CRL has identified areas in which Portico should improve. This includes specific line items within TRAC and also areas that have been identified as important to the CRL membership. The latter areas of interest include scope of the archive, functionality and services, and future costs and risks.

FINDINGS RELATED TO TRAC

There are three primary areas to be assessed within TRAC. CRL has assessed Portico in each of these areas and assigned a level of certification. (The numeric rating is based on a scale of 1 through 5, with 5 being the highest level, and 1 being the minimum certifiable level):

TRAC CATEGORY	PORTICO SCORE
Organizational Infrastructure	3
Digital Object Management	4
Technologies, Technical Infrastructure, Security	4

Within TRAC, there are 84 individual criteria. CRL has concerns about Portico's status on 12 of the 84 criteria. Below we describe each of those criteria and CRL concerns.

Criteria - A1.2 Repository has an appropriate, formal succession plan, contingency plans, and/or escrow arrangements in place in case the repository ceases to operate or the governing or funding institution substantially changes its scope.

At present there is no designated Portico successor organization. Portico should identify such an organization. Then Portico should put in place and disclose a plan for the disposition of Portico archival content, technology, and other assets in the event of discontinuation of the program by the parent organization. This is particularly important because the ongoing business viability of Portico as a service is not yet assured, judging from financial information disclosed to date.

Criteria - A2.2 Repository has the appropriate number of staff to support all functions and services.

At this time, Portico's documentation is not sufficient to allow CRL to verify compliance with this metric. Portico should implement a process to keep job descriptions up to date and better document how the roles and responsibilities of people and positions change over time.

In addition, the archive's procedures are almost entirely designed around, and end with, ingest and should be modified to include specific responsibilities for ongoing testing and maintenance of archived content. The archive should, for example, add responsibilities for testing and responding to problems with already ingested content to the Portico Roles and Responsibilities document, the Portico Automated Workflow, the E-Journal Workflow 1.9 documentation, and related policies and procedures.

Criteria - A3.2 Repository has procedures and policies in place, and mechanisms for their review, update, and development as the repository grows and as technology and community practice evolve.

Portico policy infrastructure has improved considerably since the test audit in 2006, but some of these policies still suffer from internal contradictions and inconsistencies, specifically in the area of roles & responsibilities and job descriptions.

As rate of growth and size of the archive increase to as much as a terabyte per day, the archive may have to revise its current policies to accommodate the increased time needed to refresh and migrate content.

Criteria - A3.6 Repository has a documented history of the changes to its operations, procedures, software, and hardware that, where appropriate, is linked to relevant preservation strategies and describes potential effects on preserving digital content.

Portico has much technical information about the systems and environment encoded in the metadata of the individual content items preserved in the Archive. However, there is no separate, complete documentation tracking the changes to the hardware and software of the ingest and archive systems over time. This documentation is an important tool in assessing any repository and should be created.

Criteria – A4.1 Repository has short- and long-term business-planning processes in place to sustain the repository over time.

While it is apparent that Portico has business-planning processes in place, it was not possible for CRL to assess those processes because of the unavailability of documentation of same.

Criteria - B1.6 Repository provides producer/depositor with appropriate responses at predefined points during the ingest processes.

Portico should put in place procedures and mechanisms to routinely notify and provide other appropriate responses to licensors/publishers as materials are ingested into the archive and develop workflow documentation for this process.

Criteria - B2.10 Repository has a documented process for testing understandability of the information content and bringing the information content up to the agreed level of understandability.

Portico needs to continue to identify what its community believes is necessary for “understandability” or usability of the preserved content. Portico should develop a process to support ongoing research into the needs of its community and determine what the Portico stakeholders think is an understandable e-journal, e-journal article, e-book, etc. As those needs evolve, Portico should develop test scenarios to evaluate how well the archive meets those needs.

This will be particularly important as Portico archives genres of content other than e-journals. Portico is actively exploring the requirements and potential funding and business models to support archiving of genres such as e-books, digitized newspapers, and other types of databases. (At present Portico does not plan to target the preservation of entire genres of video or other audiovisual formats, though files in these formats are preserved within the archive). As these genres may be less compatible with existing Portico workflows and technologies and may impact workflow and data management techniques, meeting “understandability” requirements for these could affect Portico’s costs and pricing. This is an area to which the CRL audit team will pay particular attention over the coming years.

Criteria - B2.12 Repository provides an independent mechanism for audit of the integrity of the repository collection/content.

The Portico audit interface contains a subset of the content of the entire Archive and at this time, it is not possible to independently “look” into the Portico archive and determine if the requested digital object is complete. Portico is working on a new audit interface and system that may address these concerns.

Criteria - C1.10 Repository has a process to react to the availability of new software security updates based on a risk-benefit assessment.

Portico needs to establish and provide to CRL a risk register that identifies what software and hardware patches for the Portico systems are available and what the risk assessment and plan is for each.

Criteria - C2.2 Repository has software technologies appropriate to the services it provides to its designated community and has procedures in place to receive and monitor notifications, and evaluate when software technology changes are needed.

Portico’s ability to disseminate content to the users in the event of a major “trigger event” (for example, where all content from a large publisher with a large user base must be made available) is limited. This relates to Portico’s status as a dark archive. Aside from the “audit” interface provided to enable subscribers to verify the presence of content in the archive, there is a delivery interface that is rudimentary at present. Portico states that it would rely upon the existing JSTOR infrastructure for support of the Portico Web site, which would deliver Portico content after such a major “trigger event.” However, it is not clear how quickly this delivery infrastructure could scale to meet user needs in the event of a major trigger event.

Criteria - C3.1 Repository maintains a systematic analysis of such factors as data, systems, personnel, physical plant, and security needs.

The security of Portico systems was not tested, although Portico has already conducted a penetration test. While we have no reason to believe that the Portico systems are at risk, the Certification Advisory Panel believes that Portico should undergo a security audit.

Criteria - C3.3 Repository staff have delineated roles, responsibilities, and authorizations related to implementing changes within the system.

Portico needs to maintain more accurate and up-to-date documentation of the roles and responsibilities of key repository personnel, particularly the roles and responsibilities of those involved in technology watch activities.

FINDINGS RELATED TO ADDITIONAL CRL CONCERNS

If Portico is to continue to be recognized as one of the CRL community’s permanent archives of scholarly content, then Portico should address the following concerns of the CRL Audit Advisory Panel.

Scope of the archive

If Portico is to provide CRL libraries a comprehensive, long-term preservation archive of e-journals, then it is still short of archiving a “critical mass” of journal content. In 2006 Portico had 13 publishers, representing 3,557 electronic journal titles. As of October 2009, Portico had 83 committed publishers representing 10,461 titles (although as of the same date content from only 7,682 e-Journal titles were actually preserved within Portico). Even if all electronic issues of those titles were included in Portico, this is still, however, only 50% of the ~20,900 journal titles in CrossRef. CRL will work with Portico to determine the percentage of the journal titles in CrossRef that would constitute a critical mass of journal content.

It should be noted here that there is no way to independently verify and monitor the presence and integrity of content in a repository like Portico comprehensively or on a meaningful scale. Such verification and monitoring is a challenge inherent in “dark” archives, which are unable to be accessed for such purposes. Portico provides an audit archive interface, designed to enable users to “view” the content archived. However, the interface accesses not the actual archived information, but rather information that is a replica of the archive. Portico’s process for generating the replica information appears to be sound, based on a demonstration of that process performed during the site visit. Yet given the amount of content in the repository, such demonstrations are not a practical means of verifying and monitoring the presence of content on a comprehensive basis. Therefore, the auditing and assessment community will need to devise a satisfactory means of independently monitoring the archive’s content. This will require Portico cooperation in further exposing its content, or its metadata, to scrutiny.

Functionality and services provided by the repository

The holdings comparison tool has limitations and should be improved. It has been reported that it is difficult to compare Portico holdings with those of a given participating library. This difficulty renders the scope of content preserved by the repository unclear and undermines the ability of a library to fully determine the value of the Portico service. The value and usability of the comparison reports would be enhanced if Portico provided a glossary of definitions for the different fields in the spreadsheet and a summary of the overall findings (i.e. the extent of gap/overlap).

Moreover, it is not clear that the minimal delivery standards Portico has set for itself fully conform to the expectations of all of its designated user communities. One area of concern is the lag time between a “trigger event” and delivery of content by Portico. The lag time of up to 60 days, specified in Portico agreements with publishers and libraries, is less likely to be acceptable in some fields, like medicine, where a hiatus of this duration would have a greater impact on users than a comparable loss of access to a journal in the humanities. As reasonable over time, the archive should tailor its agreements with publishers to better accommodate use cases in all fields.

Future costs and risks

Portico is a part of ITHAKA (www.ithaka.org), an independent not-for-profit organization. Portico and JSTOR are both not-for-profit services that are part of ITHAKA. Portico is fiscally dependent upon ITHAKA, and thus its relationship with JSTOR presents both a risk and an opportunity. This affiliation with JSTOR might be a deterrent to some publishers’ willingness to deposit journal content in the Portico Archive. On the other hand, JSTOR delivery capabilities could offer Portico a robust network and server environment through which post-trigger access to the archived journals content might be provided on a large scale.

C. Ongoing Requirements

CRL and Portico have agreed that ongoing certification is contingent upon Portico making certain disclosures every two years. These disclosures should entail, at minimum:

- Inventory of new content added to the repository since prior certification;
- Description of significant changes in repository system architecture or configuration, critical software, or software platforms;
- Software and hardware patch risk register;
- New agreements and contracts with key depositors of content, content users, major funders or sources of revenue, and providers of critical repository services;
- New key policies regarding acquisition, management, and disposition of archived content and related files and metadata;
- Records of significant events and changes in the nature and condition of digital content since the most recent audit, such as server logs;
- Records of significant events and changes in the operations of the repository;
- Most recent financial statements for the repository organization or service unit. The financial statements should indicate the categories and, where appropriate, sources of revenue and the level of same; the functional allocation of expenses; and changes in the financial position of the organization supporting the service unit; and
- Revenue and expense projections by function, for the repository organization or service unit, for the next three years.

In addition, Portico must allow a periodic, systematic sampling and inspection of the repository's archived content by CRL, or by a third party designated by CRL, using either a manual or automated process as determined by mutual agreement between CRL and the repository. This will allow CRL to independently verify and monitor the presence and integrity of content in the repository comprehensively or at least on a meaningful scale. We expect the ongoing disclosures provided by Portico to CRL to show their progress in acting on the CRL issues of concern identified through this audit.

D. A Recommendation for Further Work

Portico is compliant with TRAC criterion B1.2 ("Repository clearly specifies the information that needs to be associated with digital material at the time of its deposit, i.e., SIP"). However, per the Portico E-Journal Content Type Action Plan, Portico takes on the burden of transforming publisher-supplied content into Submission Information Packages (SIPs): "e-journal content as received from the publisher is batched and processed into Submission Information Packages (SIP) by the Portico ConPrep System." In addition, Portico has minimal requirements of publishers for metadata (an ISSN and a publication date). CRL believes it would be more effective for Portico to work closely with individual publishers and the publishing and standards communities to develop a standard, consistent SIP for journal articles and a more complete set of metadata that will facilitate uniquely identifying individual journal articles. These steps would be more consistent with OAIS 2.3.2, 3.2.1, and 4.3.1.

Portico may be doing this already to some extent, particularly with regard to the development of the NLM XML standard. We encourage Portico to continue that work, but also suggest that Portico articulate this as a long-term goal and work specifically with individual publishers to provide more consistent, standardized delivery of usable SIPs. This would benefit Portico by making it easier (and therefore less expensive) to ingest content and it would benefit the larger preservation community by furthering standards for content creation that would facilitate preservation.