# TECHNICAL ALIGNMENT

Michael Seadle (Humboldt-Universität zu Berlin)
Andreas Rauber (Vienna University of Technology)
Adam Rusbridge (University of Edinburgh)
Sabine Schrimpf (Deutsche Nationalbibliothek)
Matt Schultz (MetaArchive Cooperative)

**Abstract**

*This essay discusses the importance of the areas of* infrastructure *and* testing *to help digital preservation services demonstrate reliability, transparency, and accountability. It encourages practitioners to build a strong culture in which transparency and collaborations between technical frameworks are valued highly. It also argues for devising and applying agreed-upon metrics that will enable the systematic analysis of preservation infrastructure. The essay begins by defining technical infrastructure and testing in the digital preservation context, provides case studies that exemplify both progress and challenges for technical alignment in both areas, and concludes with suggestions for achieving greater degrees of technical alignment going forward.*

## Introduction

This essay considers two critical areas in which the maturing digital preservation field should seek to advance technical alignment both within and across national boundaries: infrastructure and testing.[1] Aligning work in these areas will help practitioners more effectively meet stakeholders' demands for high-levels of reliability, transparency, and accountability. The infrastructure for digital preservation has reached a stage of development that enables interoperability and benchmarking. To accomplish the former, we must continue to encourage transparency and collaboration between technical frameworks, and

[1] Infrastructure in the context of this essay refers to the technological components of an organization's infrastructure that are required for digital preservation. Other essays in this volume address additional components of infrastructure for digital preservation, e.g., organizational, economic, and education.

it is important to demonstrate and document the ways that the field benefits from digital archiving framework interoperability efforts.

To enable benchmarking and to establish a culture of infrastructure testing, we must first convince the community of the need for quantitative analysis, arrive at agreed upon metrics, and then gather and publish empirical results. Coordinated action across the community (particularly if it is combined with future requirements from funding agencies to incorporate testing into government funded projects) could lead to an evolving public test-bed in which we can fairly and accurately evaluate various archiving systems and preservation solutions. This essay discusses the importance of such developments: 1) by defining technical infrastructure and testing in the digital preservation context, 2) by providing case studies that exemplify both progress and challenges for technical alignment in both areas, and 3) by concluding with suggestions for achieving greater degrees of technical alignment going forward.

## Infrastructure

For technical alignment, the term *infrastructure* can encompass far more than the hardware and software necessary for managing digital archiving systems and the communication protocols for sharing resources across a network or system. It can also extend to the ways in which digital information is structured: both separate data objects and the linkages within applications and environments that make them function as a visible and usable whole. In that sense, infrastructure also relates to the metadata used to describe digital information or the systems used to generate descriptive information on an as-needed basis. Using this broad definition, *infrastructure* may also include the software used for migration and emulation processes (although these depend heavily on assumptions about how archived information will be used in the future and thus require a strong user-behavior assessment component). Standards, organizational elements, and economic factors also play a role in infrastructure as well, since they influence the design process for infrastructure development. Each of these elements is addressed in regards to their own alignment issues in separate essays within this volume. The following discussion seeks to account for facets of these broader influences on the digital preservation field's technical infrastructure alignment activities.

### *Alignment of Infrastructure*

This discussion of the alignment of infrastructure begins with a concrete consideration of existing examples of technical implementation, focusing on four specific digital archiving systems and support networks as case studies:

- UK LOCKSS (Lots of Copies Keep Stuff Safe) Alliance;
- kopal (Kooperativer Aufbau eines Langzeitarchivs digitaler Informationen) / koLibRI (kopal Library for Retrieval and Ingest) & DP4Lib in Germany;
- nestor in Germany; and
- LuKII (LOCKSS und KOPAL: Infrastruktur und Interoperabilität) in Germany.

These system infrastructures are highlighted here as one set of exemplars and case studies in the digital archiving field. They are not intended to serve as an exhaustive overview of the field, but rather as a useful subset that can help us to consider some of the principles and criteria that might foster and advance technical alignment.

As we consider these case studies below, we focus on the following questions:

- What infrastructure components comprise these digital archiving systems?
- Are their code bases open source and thus reusable for other archiving systems?
- To what degree do these infrastructures enable and/or foster interoperation?
- To what degree are these systems "complete" or "incomplete" for digital archiving purposes?

Taken together, these case studies exemplify the advantages we may gain through aligning infrastructures across multiple borders and barriers. Though there is some overlap on a software level between these initiatives, the projects and programs themselves have very different national priorities, organizational contexts, and archiving priorities. They are especially useful for the purposes of this discussion of infrastructure for achieving technical alignment because of their developers' insistence upon pushing the limits of the underlying technology's interoperability,

and each of the system's corresponding degree of openness and potentials for doing so.

**Case Study 1: UK LOCKSS Alliance**

The UK LOCKSS Alliance (UKLA)[2] is a cooperative membership organization whose goal is to ensure continuing access to scholarly work in ways that are sustainable over the long term. It represents the collaborative activity of UK libraries that are interested in building national "network-level" infrastructure and coordinating the preservation of electronic material of local and UK interest.

The UKLA seeks to ensure libraries remain central to the process of scholarly information management by enabling its members to take custody of the assets for which they have paid in order to build—not simply lease—local collections of published scholarly material. The UKLA uses the LOCKSS (Lots of Copies Keep Stuff Safe)[3] software to enable UK Higher Education libraries to develop journal preservation infrastructure and collections and to engage with journal preservation issues at a tangible, local level.

The LOCKSS technology is an open source, peer-to-peer, decentralized digital preservation infrastructure. LOCKSS preserves all formats and genres of Web-published content. It works by collecting a direct copy of digitally published scholarly content such that the intellectual content, including the historical context (the look and feel), is preserved. This content is collected by a network of geographically distributed servers that actively monitor the content through iterative cycles of voting and polling (using SHA-1 hashes) to establish the continued authenticity and veracity of the collected content over time.

The strategic goals of the UK LOCKSS Alliance for the period 2010-2013 are to:

1. Identify, negotiate and make available for preservation a collection of journal titles relevant to need;
2. Increase usefulness and relevance of the UK LOCKSS Alliance community activity; and to

---

[2] See UK LOCKSS Alliance: http://www.lockssalliance.ac.uk/ (last accessed 03-14-2012).

[3] See LOCKSS: http://www.LOCKSS.org (last accessed 03-14-2012).

3. Sustain and develop a well-founded UK national cooperative library organization to assist with ensuring continuing access to scholarly material.

EDINA, JISC's National Data Centre at the University of Edinburgh, is leading the provision of support for the UK LOCKSS Alliance. A dedicated team at Stanford University Library develops the LOCKSS software and leads and supports its US and international development.

Libraries are required to supply their own hardware upon which the LOCKSS software is installed. Staff responsibilities tend to be split between librarians responsible for collection development and IT staff responsible for system maintenance. UKLA found that these roles are not always under the same administration structures, and so responsibilities for maintenance are not always clear and well understood. This can lead to the marginalization and neglect of infrastructure. To overcome this, ongoing education and training helps motivate staff and some libraries have found that introduction of an explicit e-journal preservation policy has helped secure the engagement of both library and IT staff and secure commitment of resources, embedding local preservation activity into staff workflows and job descriptions.

For some members, the value of participation in the UK LOCKSS Alliance is best demonstrated through access to content. In early 2012, integration of LOCKSS with link resolver systems was released and the components are now undergoing community test and deployment. Demonstrating access will help secure future funding and resources to add additional functionality and undertake further testing.

A number of e-journal preservation initiatives have emerged over the last decade, and monitoring statements regarding "who is preserving what" is becoming increasingly important. EDINA and the ISSN International Centre have partnered to develop the Keepers Registry, which provides easily accessible information about inclusion of journals in preservation services and will help to identify gaps in coverage. This service aggregates information from archiving initiatives, currently using the information made publicly available (often in spreadsheet formats, with some adhering to the KBART guidelines). As the service develops, it is

proposed that journal metadata will be collected using the recent ONIX for Preservation standard.[4]

Testing of LOCKSS in the UK environment has focused on aspects needed to improve service-level qualities of the approach: how to improve coverage and access to content, and how to demonstrate value from participation. All content goes through a quality assurance test process before being preserved in the LOCKSS network. LOCKSS collects content from a wide variety of publishing platforms, and content must be collected according to licensing boundaries (i.e., delimited by volume). A "plugin" defines the URLs to be collected, fetching the relevant full text, PDFs, images, etc. A test process then confirms that everything that should be collected has been collected. We are now at a stage where further testing of the UKLA network is needed, for example to assess the quality and completeness of the content held by UK machines, and of the effectiveness of the software to provide access to content as and when it is needed. Practical tests of this nature will provide libraries with more assurances that a switch to e-only is reliable, and allow the LOCKSS approach to further develop economies of scale to work with a greater range and quantity of material.

**Case Study 2: kopal/KoLibRI & DP4Lib**

Parallel to these technical alignment developments in the UK, discussions about a digital preservation infrastructure for Germany have from the beginning emphasized a distributed model. The system of memory institutions in Germany is traditionally decentralized with well-established state and regional libraries and archives. Technical alignment is thus critical to cooperation in this environment in order for several disparate organizations to be enabled and empowered to contribute to a larger national directive and initiative for accomplishing digital preservation.

Schwens and Liegmann stated this most eloquently in 2004:

> A cooperative structure for digital preservation, corresponding to the structure of the analogue realm, ought to be developed, which ensures preservation and availability of all digital resources published in Germany (in German language or about Germany) [,

[4] The ONIX for Preservation Holdings draft standard is available online at http://www.editeur.org/127/ONIX-PH/ (last accessed 07-05-2012).

> which] ensures preservation and availability of the most important objects in all scientific fields, no matter if it is text, facts, images, or multimedia, [and which] ensures the preservation and availability of digital archival records.[5]

The *kopal* project ("Co-operative Development of a Long-Term Digital Information Archive") and its successor DP4Lib (see below) represent important building blocks for achieving this alignment.

The aim and purpose of the *kopal* project was to develop and test a long-term preservation system for co-operative use. The system is based on DIAS, at that time a standards-oriented implementation of the OAIS reference model using established IBM software (more on standards and infrastructure implementations below). The DIAS system was designed as an in-house long-term archive for the Koninklijke Bibliotheek (KB) and was extended in the *kopal* project to support co-operative use and remote access. The open source "*kopal* Library for Retrieval and Ingest" (koLibRI) connects individual users with the archival system and it can be configured to meet the needs of those users. As such it allows users with various different selection profiles and with different types of digital objects to share a single archival system, while retaining control of their data.

koLibRI validates the objects' file formats, and packages the objects together with their technical metadata as Submission Information Packages (SIPs) using the Universal Object Format (UOF). The UOF SIP files are imported, and, in OAIS terminology, stored as Archival Information Packages (AIPs) in the DIAS archival storage unit. Each *kopal* user can, via koLibRI, address and retrieve only its own data. Migration was tested as a preservation action within the *kopal* project. Other preservation actions are still to be developed.

After the end of the project, the *kopal* archival system had two active users: The German National Library (DNB) and the Göttingen State- and University Library (SUB). The DNB and SUB have subsequently allied with six different additional partners with varying use scenarios. One partner, the German Institute for International Pedagogical Research, is a research institute with

---

[5] U. Schwens, H. Liegmann: Langzeitarchivierung digitaler Ressourcen, (2004). The paragraph quoted is originally in German.

large specialized holdings, including digitized and born digital journals as well as databases. Another partner, the Library Service Centre of Baden Württemberg, offers long-term preservation as a service to its customers and seeks a safe harbor for the data for which it has assumed responsibility.

The purpose of the DP4Lib project ("Digital Preservation for libraries") is to open up the *kopal* system to these additional users mentioned above and to extend its functionality. The overall goal is to establish and run a ready-to-operate service for long-term preservation. While co-operative use of the *kopal* system is generally technically feasible, various organizational issues had to be clarified and are addressed in the project. The DP4Lib partners are, for example, conjointly compiling a catalogue of requirements for long-term preservation as a service, and are developing business and cost models, as well as process models for co-operative long-term preservation operations. Further work is also being done to enhance functionality, namely evaluating tools for generating technical metadata, and tools for converting and normalizing digital objects. These additional evaluation activities, particularly those focusing on re-use, interoperability, and collaboration factors are made possible and given promising potential thanks to *kopal's and* DP4Lib's intentional emphases on developing a co-operative infrastructure from the outset.

**Case Study 3: nestor**

Closely associated with *kopal* and DP4Lib, and worth mentioning briefly, is *nestor*, the national competence network for digital preservation in Germany. *Nestor* was originally established in 2003, in the same year that the *kopal* project kicked off. While *kopal* intended to establish the technical preconditions for a co-operative and shared preservation infrastructure in Germany, the *nestor* network aimed at setting the organizational framework and infrastructural foundations. *nestor* brings together experts and institutions active in digital preservation. The *kopal* users and several of the DP4Lib partners take part in *nestor*, as well as the Bavarian State Library, which has implemented a digital long-term archive based on Ex Libris Rosetta. Last not least, the three national subject libraries, which intend to set up a shared digital preservation solution for their purposes together, joined *nestor*. *nestor* contributes to ensuring the conditions through which developers of archiving systems can collaborate to ensure their infrastructures and systems are complete for accomplishing their stated purposes. When considering the value and importance of

coalescing trends toward common infrastructures of broad applicability, such groups and models should not be overlooked or undervalued.

*nestor* hosts several working groups on relevant preservation related questions and standards and it fosters knowledge exchange and advancement. It offers a platform for memory institutions to discuss and align roles and responsibilities in the digital realm. *nestor* also runs a cooperation with the German Institute for Standardisation (DIN), to help crystallize standards in the relatively new field of digital preservation.

Together with several higher education partners, *nestor* develops initial and further training courses in the field of digital preservation in Germany, so that qualified staff are available to deal with the digital preservation challenge.

*nestor* has also been actively involved in developing an audit and certification system for trusted digital archives. Trust is an important prerequisite for co-operation (more on trust below). Especially in a shared and networked preservation system, partners want to be sure that their information is safe with the respective partners' institution. Because it is impossible to predict in which state a piece of digital information will be in, for example, 50 years, it is important to evaluate the set-up of existing archives.

**Case Study 4: LuKii**

The LuKII (LOCKSS und KOPAL: Infrastruktur und Interoperabilität) initiative bridges the LOCKSS and KOPAL systems, providing an interoperability model for digital archiving. LuKII is an infrastructure and research project with staffing at Humboldt-Universität zu Berlin and at the German National Library in Frankfurt. The project began in 2009 with funding from the German Research Foundation (Deutsche Forschungsgemeinschaft). The project lists the following goals in its original proposal:

- To establish a cost-effective LOCKSS network in Germany including infrastructure to provide ongoing technical support and management for LOCKSS and its variants (e.g. CLOCKSS);
- To conceptualize and implement interoperability between LOCKSS and KOPAL in order to combine cost-effective bitstream preservation with well-developed usability preservation tools; and

- To test the interoperability prototype by archiving data from German institutional repositories.

An important element of the first goal was to get a minimum of seven partner libraries to be able to implement a Private LOCKSS Network (PLN) within Germany.[6]

A competence center at Humboldt-Universität zu Berlin offers German speaking technical assistance about LOCKSS to the German partners and to others in the German-speaking community. The competence center runs out of the university's computer center (called Computer and Media Service) and is in regular contact with the Stanford LOCKSS team. LOCKSS refers all problems in the German-speaking regions to Berlin.

Programmers are also working at both the DNB and at Humboldt-Universität on modifications to koLibRI and LOCKSS respectively to enable interoperability. One modification is to enable LOCKSS to make use of METS metadata. LOCKSS can, of course, store METS (it can store any form of digital information) but has not previously also used it as actionable metadata. Another modification is to shift the storage containers to the new WARC format. KoLibRI staff have collaborated with the Berlin LOCKSS team to make progress on the WARC conversion, as well as on enabling koLibRI's migration manager to work with LOCKSS. The goal is to introduce prophylactic migration to LOCKSS and to let *kopal* data be able to use on-the-fly migration through LOCKSS. Developing local expertise with the core LOCKSS code also helps to decentralize LOCKSS maintenance and expansion. LuKII is a successful effort to test and validate the value and importance of open source re-use of existing technologies, pursuing interoperability where advantageous, and selecting infrastructure options that are flexible for promoting multi-institutional collaborations on behalf of digital preservation.

The harvesting of works in German open access repositories is about to begin. The first wave of harvesting will be using unmodified LOCKSS software and the second wave will harvest

---

[6] As of mid 2011, LuKII has ten official partners: Bayrische Staatsbibliothek, Deutsche Nationalbibliothek, hbz - Hochschulbibliothekszentrum NRW, Humboldt-Universität zu Berlin, Karlsruher Institut für Technologie, Sächsische Landesbibliothek - Staats- und Universitätsbibliothek Dresden, Universität Konstanz, Universitätsbibliothek Stuttgart, Universitäts- und Landesbibliothek Münster, Niedersächsische Staats- und Universitätsbibliothek Göttingen.

the same sources using LOCKSS in order to be able to test the new programs. This testing will foster a better understanding of the modifications the project team has made to the LOCKSS framework, both within our team and throughout the broader community of digital archiving practice. The empirical data we collect and publish regarding these tests will mark an important development in establishing technical benchmarking for digital archiving systems.

Each of the above case studies demonstrates the advantages gained through aligning technical infrastructures across multiple borders and barriers. In the case of the *UKLA*, use of the open source LOCKSS software has enabled UK Higher Education libraries to build a national "network-level" infrastructure and coordinate the preservation of electronic material of local and UK interest. The focus of *kopal/KoLibRI & DP4Lib* on developing a co-operative infrastructure at the outset models the value of establishing a firm foundation for benefitting later from factors such as re-use, interoperability, and collaboration. *Nestor* demonstrates the organizational dimensions of technical alignment through facilitating interactions across groups to ensure that developers can mutually collaborate to the benefit of their archiving systems. And *LuKII* has demonstrated how to combine open source technologies to enrich preservation activities while bridging multi-institutional environments. In the course of each of these on-going technical alignment developments, iterative testing was recognized as being of critical importance to their maturation and adoption, and remains so. The next sections explore the importance of testing to improve technical alignment.

## Towards Testing: Standards and Infrastructure Implementations

The importance of standards to alignment more broadly is discussed in a separate essay within this volume. Here, we focus our discussion specifically on the need for standard approaches to establishing interoperability between digital archiving infrastructures. Such standard approaches ultimately will improve the chances of bridging systems. They make can make ingest and retrieval simpler by reducing the number of choices and special adaptations needed. Standards should also, in an important sense, reduce risk because they represent choices that have in theory undergone extensive design considerations and testing. This is ideal.

There are instances, however, where technical standards for digital archiving have failed to achieve these goals for a variety of reasons. At the ANADP conference a member of the panel on standards admitted that the problem with standards is that there are too many of them. If there are too many "standards" for interoperability and/or for testing of technical components, the result may be no common standards at all. In the technical landscape, some official standards fall into virtual disuse soon after they receive approval, because a new standard supersedes them or because the technical environment changes. This is less the fault of standards-setting organizations like the W3C or ISO than it is the fault of commercial market factors, which determines in fact whether a standard will be used or ignored. Libraries, archives, and other memory institutions have in general too small a market share, even collectively, to influence commercial vendors to accept the standards that the community favors. The exception is firms that market only to memory institutions.

Technical standards tend also to be somewhat misunderstood in the digital preservation community. OAIS (Open Archival Information System) is a classic example. The Consultative Committee for Space Data Systems (CCSDS) documentation about OAIS clearly discusses it as a reference model.[7] That means that it labels the key elements of an archiving system to enable common discourse about the services that that element provides and the role it serves. Many in the digital preservation community continue to conflate this reference function with a system design. A system could be designed specifically with components that use the OAIS model, but more typically it is a matter of changing names on established designs. Commercial vendors use the OAIS label more for marketing than for engineering. This does not make their systems worse, but nor does the label make them better. OAIS compliance has minimal design meaning in most cases, and these claims sometimes obscure as much as they reveal.[8]

Closing the gap between the over-abundance of technical standards that exist today and more widely adopted standards that

---

[7] The OAIS Reference Model document includes a definition of the term Reference Model (page 1-14) and throughout Section 1 refers to the role and significance of reference models (CCSDS, 2009).

[8] Developers are, however, beginning to build and test open source digital archiving systems that aim to be OAIS compliant—DAITSS and DAITSS2, as well as Archivematica, as just a couple of examples. The adoptability and use of these systems is in need of further implementations and tests.

would enhance interoperability and reduce risk involves testing on a large scale. Merely testing to find out whether a proposed standard functions as it should, and whether it has the potential for addressing technical needs, is only a starting point. A more important test is whether multiple system-vendors are willing to adopt a standard, implement it in their software, and then determine whether it meets their needs. This form of testing could also gather actual empirical information about the functioning of a standard. Standards that did not get a minimum number of adopters would fail the test automatically.

The technical standards that matter most for digital preservation can in fact be determined on these empirical levels. For example, formats that are used today to publish contents on the World Wide Web (that is, contents accessed via HTTP services over the Internet) represent de facto format standards after a certain level of adoption, which includes incorporation into established browsers such as Firefox, Internet Explorer, Chrome, and Safari. These browsers have a strong record of enabling backward compatibility. The number of file formats published online and readable by browsers in the 1990s that cannot be read today is negligible. It does not matter whether these formats represent official standards or not—they are the way in which content was and is shared. It is important to distinguish between the longevity of these publication formats and the formats used by text editing systems such as MS-Word. Word was never meant to be a publication format or anything more than an intermediate editor for content. Few MS-Word documents play a publication role except (ironically) in institutional repositories, which are generally run by universities and are meant (at least in part) for digital preservation (Rosenthal 2010).

The long-term use and testing of metadata standards can also contribute to advancing technical alignment on an infrastructure level. However, applying a similar empirical test to metadata is somewhat harder, because metadata tends to be less visible. Clearly, Dublin Core plays a significant role in information exchange on the Internet. METS, and some elements of PREMIS, are increasing in popularity within the digital preservation community, perhaps in part because both schemas are extensible in the capabilities and features that they offer. Whether METS or PREMIS have achieved a similar status more broadly is less likely. In the broader commercial world relatively few METS (and virtually no PREMIS) implementations exist, except among

vendors like Rosetta that market directly to the digital preservation community and arguably use METS because of its appeal to customers.

Publication formats and metadata are only two examples of areas where the existence of de facto standards impacts the implementation of digital preservation systems. What is important here is the need to distinguish between those standards established by standards setting agencies that, despite all good intentions, fail to play any functional role as standards, and those that, sometimes without official approval, are in fact so commonly used that digital preservation implementations need to recognize and accept them. In all cases, the role of sound testing is critical for closing gaps, enhancing interoperability, and reducing risk. Testing is needed on a routine basis throughout various implementation phases.

## Testing

***Testing*** involves reproducible experiments using, if possible, real data to show whether software and hardware perform under conditions that reflect a reasonable hypothesis about the future. Testing can take several forms and depend on design goals and targeted outcomes (functional vs. non-functional; static vs. dynamic; unit vs. systems, etc.). The first and most basic test is whether a system functions at all—that is, whether the code compiles and runs without errors. A second level test might establish whether the system scales appropriately—the testing should involve not merely storage capacity, but also ingest and access processes. One example would be a stress test, in which large numbers of access requests (including permissions decisions and search/retrieval) are made of a system in a short time. A third and more complex type of test would involve conditions that can be anticipated for future digital environments. One example might be bit rot, which can be predicted mathematically and emulated to age storage systems virtually. Future storage may propose to eliminate bit-rot, but no current evidence suggests such a development. Other examples could be user-tests involving format migration to adapt to evolving e-reader devices.

Testing is one of the key ingredients to making progress in technical alignment in digital preservation. To date, a great deal of the research in this domain lacks the solid ground provided by thorough and consistent testing. Solutions are being developed and presented, yet little is done to ensure that the underlying systems

actually address the right problems and address them in ways that have a high probability of long term success.

When it comes to aligning, sharing, and collaboratively furthering tools and infrastructure (both technical as well as knowledge bases) it is essential to be able to rely upon the individual building blocks. This requires reproducible testing of tools and know-how, as well as thorough documentation of the circumstances under which the software was tested. Currently, most tools and most techniques are simply "evaluated" by people without the necessary technical skills or background to judge to what degree it fits the intended purpose.

The problem with this type of evaluation is that it is not replicable, not scalable, not reusable and provides limited (if any) basis for technical alignment and continuous development. The library and archiving community needs to move from ad-hoc evaluation to solid testing and benchmarking. A similar focus on solid and thorough testing has brought huge boosts in other disciplines, specifically information retrieval and machine learning. Testing provides a scientific basis, well-understood measures and limitations, and a sense of the fitness-for-use via its various benchmarks and measurements.

### *The Role of Trust and the Importance of Distrust*

There is a useful tension between trust and distrust in the technical aspects of digital preservation. The *nestor* efforts to certify trusted repositories offer a valuable basis for any form of digital preservation, because certification ensures that basic procedures are followed and that process descriptions exist. A repository whose update or backup procedures are sloppy or one that fails to document key features in system management is not a repository that is likely to provide data with reliable integrity or authenticity over long periods.

Certifying that a repository currently carries out appropriate procedures (opening to review or inspection and expressing conformance to recognized standard practices) does not, however, mean that it should be trusted to reliably preserve digital information over prolonged periods of time. Certification gives a snapshot in time. Typically, organizations make special efforts to clean up their procedures before a certification visit takes place and may let them slide again afterwards. Good practice between certification visits may remain in place, but certification cannot guarantee that. Certification is a form of audit, but one that does

not typically include auditing the data for integrity or evidence of authenticity—in part because these are technically complex and difficult issues that the audit teams may not be prepared to handle. The cost in time and effort would increase significantly. Only a few systems, notably LOCKSS, have a built-in integrity-checking process that functions as an ongoing internal audit (described in more detail below and in Rosenthal, 2010).

Distrust presents itself as a much safer basis than trust for designing systems and for planning long term digital preservation, as long as that distrust means building in sufficient redundancy to make reasonable allowance for error, accident, external attack or deliberate internal damage—all of which are known problems. Precisely how much redundancy is needed can currently only be guessed at, since few companies or even non-profit organizations want to admit or publicize their internal problems. The most-cited study in this area (Power, 2002) is now outdated and those with computer center experience believe that the results probably understate the actual magnitude. There is no reason to think that the dangers have changed substantially, though the balance of risks may have changed because of increasing external attacks.

Redundancy also has a geographic component. Recent natural disasters such as the earthquake and tsunami in Japan in March 2011 and even Hurricane Irene in the US in August 2011 show the danger of trusting any one particular location. While no data was known to be lost in either case, electricity was interrupted, services broke down, and the nuclear power plants failed despite extensive and well-tested protections. A repository with all of its data in a single location or even a single geographic area subject to adverse weather, seismic, economic, or political conditions should be considered to be at risk.

The limits of distrust are equally important to recognize. Librarians understand from their experiences with print and microfilm that every additional copy in a different and secure location and on a different physical medium increases the chances of long-term survival. The assurances inherent to static physical mediums that are missing due to the vulnerabilities of electronic content often privilege trust in the physical over the digital. The problem is that information no longer comes exclusively in static text and image formats with clear beginnings, endings, and sequences from start to the finish. They forget also the vulnerability of paper and film to damage by users, to say nothing of a vulnerability to environmental conditions such as humidity or

insects. A form of distrust that goes to the extreme of discounting digital archiving errs in its trust of physical media, just as a form of trust in a particular "trusted" archive errs in misjudging the long term vulnerabilities of any one organization. Balance is key.

## *Requirements for Testing*

To achieve effective testing for digital preservation, the digital preservation community needs to begin with a range of scenarios that have:

- CLEAR GOALS: this includes a description of a specific purpose or purposes for the testing.
- BENCHMARK DATA: benchmark data should have the range and complexity of real data and be checked whether they fit the purpose and goals;
- MEASUREMENT SCALES: these scales and measurements need to remain stable over time, even with improvements, so that comparisons are possible;
- KNOWLEDGE BASE: the knowledge base provides a location to collect and make available the test results.

Each of these points will be discussed further below.

### Goals for Testing

Testing needs to be specific in terms of what is being tested and what the outcomes mean. Effective testing may have multiple well-focused goals but should not become a catch-all that attempts to cover everything. Defining common goals that are meaningful across multiple software platforms could pose a major challenge to the highly heterogeneous digital preservation community. It may be necessary to focus on some subsets, rather than trying to address too many goals at once.

The goals for testing can exist on multiple levels. At the highest level they should perhaps focus on broad concepts such as establishing how well archiving systems can perform on issues such as:

- maintaining the integrity of the digital content;
- retaining evidence of the authenticity of that content; and
- demonstrating that the content can be used (read) under potential future circumstances.

None of these goals are easy to test, in part because no consensus exists even about how to define terms like *integrity* or *authenticity* in a digital environment. *Use* is particularly problematic because many librarians define use simply as reading the way they read today, without considering how reading has changed over time and without taking other kinds of use (interactive games, for example) into account. *Integrity* comes closest to having an established technical definition based on the comparison of check-sum calculations, though *integrity* is also used in a broader sense by managers of digital content in ways that may confuse this specific technical use of the term.

At a lower level, testing may need to have goals that can vary with particular types of systems, while still enabling broader comparisons among results. A good example of this is the SIP stress test for the Rosetta software, where they tried to find out how many documents they could add in a specific amount of time (Ex Libris, 2010). This was an excellent example of public testing, but to make comparisons with other systems possible, the goals for such tests need to specify the conditions under which they take place. A load test using fiber channels on closely linked systems is, for example, very different than a test loading data via standard Internet services.

**Benchmark Data**

Standard benchmark data are one of the most important elements in a systematic testing program and are among the hardest to establish. The temptation is to manufacture data that fits a particular system, but artificially manufactured data tend to fail to represent the variety and complexity of real data. This means that systems may work flawlessly with manufactured data and less well with actual cases. Even real data can be flawed if the set does not include the full range of types and formats. In fact, a key first step is defining the range and type of complexity that the benchmark data should have. In some cases this is best done empirically with sampling to avoid overly simplistic assumptions, while in others it may be better to design artificial data sets with well-defined and known characteristics.

Typically library-based digital preservation systems have focused on archiving those text-oriented formats that are

successors to print publications.[9] A print-image PDF may seem like a reasonable representation for this form of data, but this may already be an outdated assumption. Publishers typically offer HTML-based versions as well as PDFs. The number of researchers in the UK who get their information from online sources is now up to 85% according to a recent study, and about 45% of them read online rather than print (Tenopir, 2011). Online reading may be PDF, but the screen-friendly online formats using HTML, CSS, Javascript, JPEG, etc. may be more attractive for reading and PDF for printing. The data and the interactions in these HTML-based formats are more complex than content in single file and multimedia data or data from interactive systems are more complex still, especially since the "data" may include executable code.

Knowing what types, varieties, and formats of data to collect still does not mean that it will be easy to gather appropriately representative data. Legal issues may create permissions problems, especially for making the data available as benchmark data to multiple systems. Quantity can also be a problem. A stress test or a scaling test needs relatively large quantities of data.

## Measurement Scales

Measuring the success of a test is complex because the scales need to be meaningful in terms of both the goals and the data. There is a strong tendency to approach measurement with a binary mentality: success or failure. This oversimplifies most real situations and is more of a marketing tool than a scholarly assessment. A stress test for an ingest system could have a measurement scale in items per hour, if the items are comparatively homogenous. It could also have MB per hour, if size varies or is a significant factor—though separating performance between large and small items could be necessary too. But if size is relatively stable and the complexity of the digital content varies, then the scale may need to take complexity into account. An overly simplistic scale can show misleading results.

Measurement scales need to be stated in a way that meaningful comparisons are possible when multiple systems run the same test. Anonymous participation in benchmark evaluations

[9] With the rapid expansion into research data, this is beginning to change to some degree.

has been shown to be successful in other domains, with only voluntary disclosure of a participant's identity after the evaluation. Commercial vendors may be reluctant to engage in this kind of controlled comparison of systems fearing adverse results.

### *Knowledge Base*

If one of the key reasons for testing is comparison, then the results, the data, the measurement scales, and the goals need to be publicly and openly available. This does not mean in this era of distributed computing that a single server needs to host this information, but it does mean that some form of linkage and easy discovery is needed. While it is tempting to say that there should be established standards for testing and that some institution needs to maintain them, it is also important that testing standards not encounter the same problem as other technical standards where there are so many that actual comparison (the testing equivalent to interoperation) becomes meaningless. It may be better to perform widespread testing first and to build on that experience when establishing standards specifically for digital preservation testing.

In practical terms a subset of the digital preservation community needs to take the lead in creating data, in developing testing scenarios and measures to address specific goals, and in sharing openly all the elements that went into the testing. One incentive for doing this is that the subset that takes the lead could get an advantage of setting the terms by which archiving is tested. It will also be doing the community a service. The task is not trivial, however, and results may take years before the mass is sufficient to be useful.

### *Learning from Other Domains*

In testing, the digital preservation community can also learn from other domains, such as for example the medical domain, where strong compliance requirements exist and are frequently tested beyond mere conformance checks. DICOM standard compliance testing, for example, includes the Connectathon (http://www.ihe.net/Connectathon), which is a week-long interoperability-testing event where system developers must demonstrate their ability to exchange data and to interoperate via common communication protocols using ad-hoc task settings. Similar lessons can be learned from the Machine Learning and Information Retrieval communities, both of which have strong traditions in automated, objective benchmark evaluation, in test

data and ground truth compilation, and in scientific competitions, all of which form the basis for scientific progress (Kalgren, 2011).

## *Examples of Testing for Digital Preservation*

So far, several important steps have been made in this direction of establishing a culture of testing for the digital preservation community. Below are a series of case studies that demonstrate progress in this direction and offer approaches that can be built upon and re-applied.

### Case Study 1: LOCKSS

LOCKSS has a long history of public testing. Two tests in particular stand out. One looked at measures to resist attacks on LOCKSS as a peer-to-peer preservation system. The issue is especially important for LOCKSS because the LOCKSS servers work in the Internet environment and can routinely be subject to attack. For this reason it was worthwhile to test their robustness and to demonstrate publicly their ability to withstand intrusion attempts (Manaitis, 2004).

The second LOCKSS test of special importance was the test of on-the-fly migration. Migration is a matter of special concern within the library community because of bad experiences with word processing formats. The LOCKSS approach to migration did not rely on converting contents to new formats and storing the resulting version, but built in the ability to convert a format in real time, as the demand arises. LOCKSS demonstrated that the process worked seamlessly and efficiently and published the results (Rosenthal 2005). Storing the code to convert a format is also more space-efficient and makes it easy to implement quality improvements in the migration.[10] That said, format obsolescence remains an area of constant research and particularly for more obscure formats and use cases may require more sophisticated monitoring and migration measures.

### Case Study 2: Rosetta

Rosetta (from Ex Libris) did a "scaling proof of concept" for the Church of the Latter Day Saints, and the results of this test are available online. The test used up to 50 million synthetic records of

[10] For more on format migration, see David S.H. Rosenthal. "Format Obsolescence: Assessing the Threat and the Defenses," *Library High Tech*, Special Issue, vol. 28, no.2, 2010, pp. 195-210. doi:10.1108/07378831011047613 (last accessed 06-11-2012).

varying sizes. The goal was to demonstrate that they could "meet organizational objectives of loading two petabytes of data within one year" (Ex Libris, 2010). The test was (as Ex Libris explains) a compromise between a full-scale demonstration and one that was economically feasible.

**Case Study 3: PLANETS**

PLANETS (Preservation and Long-Term Access through Networked Services) offers a test-bed for experiments. The test-bed runs on a Dell PE 2950 III server running Ubuntu with 900 GB of storage. This clearly limits the kind of experiments that are possible and excludes tests involving production-level systems like LOCKSS, Rosetta, or Portico. Its strength is that it offers a standard location and formal methodologies for testing and makes it easy for others to comment. The Planets Preservation Planning Tool PLATO (http://www.ifs.tuwien.ac.at/dp/plato/intro.html) allows testers to share evaluations of the performance of specific preservation actions such as migration and emulation tools, some of which may be called from within a controlled environment.

**Case Study 4: CASPAR**

CASPAR (Cultural, Artistic and Scientific Knowledge for Preservation, Access and Retrieval) also has a test-bed implementation plan that focuses on "evidence that the CASPAR approach is doing something useful for digital preservation in several different domains in several different organizations." (CASPAR, 2009) CASPAR's goals are, among others:

- Enhance the techniques for capturing Representation Information and other preservation related information for content objects.
- Design virtualization services supporting long-term digital resource preservation, despite changes in the underlying computing (hardware and software) and storage systems, and the Designated Communities.
- Integrate digital rights management, authentication, and accreditation as standard features of CASPAR.
- Research more sophisticated access to and use of preserved digital resources including intuitive query and browsing mechanisms (CASPAR, 2011).

**Case Study 5: TRAC**

TRAC is the short name for the "Trustworthy Repositories Audit & Certification: Criteria and Checklist" that was produced by a task force convened by the Research Libraries Group (RLG) and the US National Archives and Records Administration Task Force on Digital Repository Certification in 2007 and since maintained by the Center for Research Libraries (CRL). The goal was clearly stated:

> The goal of the RLG-NARA Task Force on Digital Repository Certification has been to develop criteria to identify digital repositories capable of reliably storing, migrating, and providing access to digital collections. The challenge has been to produce certification criteria and delineate a process for certification applicable to a range of digital repositories and archives, from academic institutional preservation repositories to large data archives and from national libraries to third-party digital archiving services.

The TRAC checklist has been used by CRL in performing audits of digital preservation systems. TRAC provided the basis for "ISO standard 16363: Audit and certification of trustworthy digital repositories" (ISO, 2012).

## Testing: Opportunities for Technical Alignment

To align, share, and further tools and infrastructure collaboratively, the digital archiving community must mature past ad-hoc evaluations and establish a culture of testing, so that the community can trust the technological solutions being offered. This requires solid evaluation of tools and know-how, as well as thorough documentation of the circumstances under which the software was evaluated. These evaluation strategies need to be replicable, scalable, and re-usable. The purpose of this essay is not to provide a detailed roadmap, but to demonstrate the need for testing and to stimulate thinking about practical solutions. The testing scenarios described and depicted above in the various case studies are a step in the right direction. Building upon their efforts, a couple of further approaches are suggested below.

One approach might be for the cultural memory community to work towards establishing sustainable environments and neutral platform to initiate benchmarking strategies. This could have the added side benefit of creating a market of sorts for emerging

solutions. This environment would also serve to drive technical alignment goals such as interoperability. Progress in this direction would require:

- knowing and defining what to test and what is fit for testing;
- thinking about how to test these components and principles;
- defining such tests: including goal specification, measures, data, etc.; and
- running an initial set of pilot tests.

Another approach would be for libraries and other memory institutions, with the help of funding agencies, to progressively and collectively insist on tests and comparisons before they make decisions about choosing long-term preservation solutions. This customer-driven approach might be less systematic and likely many of the tests would turn out to be suspect, but merely insisting on public tests would begin to create a culture of testing and of decision-making based on empirical data that would make systematic benchmarking such as described in the first approach more realistic.

## Conclusions

As detailed above, the key technological accomplishments in digital preservation thus far mostly involve the coalescing and maturation of a variety of digital archiving systems, services, and solutions that have demonstrated qualities for achieving technical alignment on national levels across multiple organizational borders and boundaries. This variety should help to protect against the failings of any one system. Two emerging themes demonstrate the power of aligned, heterogeneous approaches: first, initiatives in which data exchanges have been tested between digital archiving frameworks and programs in order to ensure that if a system fails, its data may be safely transitioned into another system option (e.g. MetaArchive and Chronopolis completed a technical bridge between their LOCKSS- and iRODS-based infrastructures for this purpose in 2011, see http://www.metaarchive.org/projects/nhprc). And second, service providers are building technical and organizational partnerships that enable participants to preserve their content in multiple, heterogeneous digital archiving systems (e.g., DuraCloud and Chronopolis are collaborating to offer a combined service). Complimenting this variety of technical approaches, many systems share design features and infrastructure.

This has the advantage of enabling reusability, interoperation, and collaboration.

As we work to align our technical approaches to digital archiving, we also need to design and implement common infrastructure testing practices. This testing needs explicitly to address the technical components and approaches of digital archiving systems. To date, technical testing has largely occurred at the program level. LOCKSS especially has put an emphasis on public testing and peer-reviewed publication of the results. Ex Libris (Rosetta) has also conducted public tests. These are small but significant steps toward establishing an evaluative process for digital preservation that relies on empirical data and reproducible results. This would compliment such audit frameworks as the TRAC standard, and it would provide evidence that libraries and publishers could use as they make decisions to choose one or another archiving system or framework for particular types of content. Significant progress in this area is needed.

Establishing a culture of testing and benchmarking represents a key technical alignment challenge. There are a number of reasons for this. One is that our community currently lacks a culture of testing or using empirical data for decision-making. One reason may be that existing testing scenarios have been poorly developed and that few well-established metrics exist for evaluating success. Another might be that institutions have not yet understood the need and value of such empirical testing, and instead are relying heavily on more qualitative analytic tools such as the TRAC standard or the DRAMBORA approach.

The culture of testing is weak in part because testing is both difficult to do and even more difficult to get funding to implement. Particularly in the early stages of field development, funding agencies are happier to support building a new resource than they are to spend money to test how well the resources they are funding perform. Yet without systematic testing, no archiving system should be considered reliable. Commercial archiving systems have shown little interest in engaging in public testing on their own initiative. They put the emphasis instead on marketing that addresses librarians' concerns and fears. If that trend continues, the risk to digital content will not diminish over time, and our field will not reach appropriate levels of success in our preservation of digital content.

Success is an endlessly moving target, best measured by the continued access to content. Long-term digital preservation ultimately can never be considered complete, because there will (presumably) always be a future with new circumstances and new problems to address. A reasonable five-year goal would be to establish a culture of testing and of basing decisions about digital preservation on empirical data as well as qualitative/organizational data. A major step in that direction would be for funding agencies to encourage, fund, and implement systematic public testing of archiving systems and preservation solutions.

## References

Caplan, Priscilla, (2010) "The Florida Digital Archive and DAITSS: a model for digital preservation," *Library Hi Tech*, Vol. 28 Iss: 2, pp. 224 – 234. Available: http://www.emeraldinsight.com/journals.htm?articleid=1864750&show=pdf (last accessed 07-05-2012).

CASPAR (2009), "CASPAR Draft Testbed implementation Plan." Available: http://www.casparpreserves.eu/Members/metaware/Deliverables/caspar-draft-testbed-implementation-plan/at_download/file.pdf (last accessed 07-05-2012).

CASPAR (2011), "The CASPAR Project." Available (August 2011): http://www.casparpreserves.eu/caspar-project.html (last accessed 07-05-2012).

DAITSS [Dark Archive in the Sunshine State] (2011) Website. Available: http://daitss.fcla.edu/ (last accessed 07-05-2012).

Ex Libris (2010) "The Ability to Preserve a Large Volume of Digital Assets: A Scaling Proof of Concept." Available: http://www.exlibrisgroup.com/files/Products/Preservation/RosettaScalingProofofConcept.pdf (last accessed 07-05-2012).

Hockx-Yu, Helen (2006), "Establishing a UK LOCKSS Pilot Programme," *Serials: The Journal for the Serials Community, Issue*: Volume 19, Number 1 / March 2006, Pages: 47 – 51. Available: http://serials.uksg.org/content/c431kl9ya6qcpl80/fulltext.

pdf (last accessed 07-05-2012).

International Organization for Standardization, "ISO standard 16363: Audit and certification of trustworthy digital repositories," Edition 1, 2012. Available: http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=56510 (last accessed 07-05-2012).

Karlgren, J. et al., 2011. Use cases as a component of information access evaluation. In Proceedings of the 2011 workshop on Data infrastructurEs for supporting information retrieval evaluation. pp. 19–24.

Knight, Steve, (Preservation Research and Consultancy, National Library of New Zealand, Wellington, New Zealand) 200, Early learnings from the National Library of New Zealand's National Digital Heritage Archive project, *Program*

Koçer, Dipl.-Inf. Kadir Karaca and Dr. Thomas Wollschläger, "Evaluierung von Strategien für lokales Entpacken und Übertragen komprimierter Objekte eines digitalen Archivs," *Frankfurt am Main*, 2005. Available: http://kopal.langzeitarchivierung.de/downloads/kopal_Evaluierung_Entpacken.pdf (last accessed 07-05-2012).

Library of Congress, "Metadata Encoding and Transmission Standard (METS)." Available: http://www.loc.gov/standards/mets/ (last accessed 07-05-2012).

Library of Congress, "PREservation Metadata: Implementation Strategies (PREMIS) Maintenance Activity." Available: http://www.loc.gov/standards/premis/ (last accessed 07-05-2012).

Maniatis, P. et al., 2004. Impeding attrition attacks in P2P systems. *Proceedings of the 11th workshop on ACM SIGOPS European workshop: beyond the PC*. Available: http://portal.acm.org/citation.cfm?id=1133572.1133601 (last accessed 07-05-2012).

Portico (2011) "Digital Preservation Defined." Available: http://www.portico.org/digital-preservation/services/preservation-approach/ (last accessed 07-05-2012).

Power, R., 2002. CSI/FBI computer crime and security survey, Computer Security Institute.

RLG/NARA Task Force on Digital Archive Certification (2007), "Trustworthy Repositories Audit & Certification: Criteria and Checklist." Chicago: CRL. Available: http://www.crl.edu/sites/default/files/attachments/pages/trac_0.pdf (last accessed 07-05-2012).

Rosenthal , D. S.H et al., (2003), "Economic Measures to Resist Attacks on a Peer-to-Peer Network," *Workshop on Economics of Peer-to-Peer Systems*. Available: http://berkeley.intel-research.net/maniatis/publications/P2P-Econ.pdf (last accessed 07-05-2012).

Rosenthal , David S. H. et al (2005), "Transparent Format Migration of Preserved Web Content," *D-Lib Magazine* 11, no. 1. Available: http://www.dlib.org/dlib/january05/rosenthal/01rosenthal.html (last accessed 07-05-2012).

Tenopir, Carol et al., (2011), "Data Sharing by Scientists: Practices and Perceptions," *PLoS One*, Available: http://www.plosone.org/article/info:doi/10.1371/journal.pone.0021101 (last accessed 07-05-2012).

Schwens, Ute and Hans Liegmann, 2004. "Langzeitarchivierung digitaler ressourcen," In: *Grundlagen der praktischen Information und Dokumentation*. München: K.G. Saur, pp. 567–570, Available: http://nbn-resolving.de/urn:nbn:de:0008-2005110800 (last accessed 07-05-2012).

Walters, Tyler and Katherine Skinner, (2011), "New Roles for New Times: Digital Curation for Preservation," *Association for Research Libraries Report*, March 2011. Available: http://www.metaarchive.org/reading-room (last accessed 07-05-2012).