Building from Trust: Using the RLG/NARA Audit Checklist for Institutional Repository Planning and Deployment

Carolyn Hank, Helen R. Tibbo, and Heather Barnes; School of Information and Library Science, University of North Carolina at Chapel Hill/USA

Abstract

This research reports on application of the RLG/NARA August 2005 draft of An Audit Checklist for the Certification of Trusted Digital Repositories (TDR). *Reasons for use and non-use were explored along with specific application. Overall, participants' perceived the TDR as a valuable framework for IR planning and high-level decision making but few used it. This study also investigated participants' perceptions of other IR planning documentation. Fifty-nine unique resources were identified as "valuable," with the TDR as one of the three most cited resources.*

Introduction

In August 2005, the Research Libraries Group (RLG) and the National Archives and Records Administration (NARA) released a draft version of *An Audit Checklist for the Certification of Trusted Digital Repositories* (TDR) [1]. It was built on a framework of attributes for trusted digital repositories that was presented in a 2002 RLG/OCLC collaboration, [2] based on the Open Archival Information System (OAIS) Reference Model [3]. The TDR was revised and expanded and published in a non-draft version, *Trustworthy Repositories Audit & Certification (TRAC): Criteria and Checklist*, in March 2007 [4]. The present investigation reports solely on application of the TDR in draft version as the research was conducted prior to the formal publication of the new checklist.

While the primary audience for the TDR is those involved in the evaluation, certification, and auditing of trusted digital repositories, the authors of the draft checklist also note the wider appeal of their audit instrument, including to those planning digital repositories. The intent of this study is to demonstrate the use of the TDR outside of the instrument's primary user audience and intended purpose, and report on institutional repository (IR) planners' perceptions and experience in applying the document's attributes and guidelines at the IR planning stage. This investigation also identifies other resources ARL member libraries used at the planning stage.

Operational IRs are still emerging, with only 37 (30%) of ARL members [5] reporting an operational IR as of January 2006. A broader survey of 446 academic library participants [6] reported operational IR deployment at 48 (11%) institutions, with 236 (53%) of survey participants reporting no current IR plans. Overall, the results of this research will provide evidence to support improved IR and other digital repository planning, as well as for those institutions with existing IRs that have plans for expansion and new platforms.

Methodology

To assess use of the TDR as an IR planning resource, the research team selected ARL members as the study population. Currently, there are 123 ARL member institutions [7]. We employed a mixed-method approach, using both qualitative and quantitative data collection techniques. First, we invited 122 ARL member libraries to participate in a web-based survey during a three-week period between January and February 2007. We excluded UNC-CH, the researchers' home university, from the study. Second, the research team conducted follow-up telephone interviews with five survey respondents during a three-week period between February and March 2007. The intent of these semi-structured interviews was to acquire additional information on use and perception of the TDR, and to further investigate application of other IR planning resources.

Findings

The web-based survey responses contained both quantitative and qualitative data. Presented below are results from the web-based survey, followed by an aggregated analysis of the five telephone interviews.

Survey

Of the 122 ARL institutions invited to participated, 47 completed the web-based survey, for a response rate of 39%. Early in the survey, respondents were asked to identify the status of their current IR activities as a deployed operational IR; a pilot IR; or neither. We characterized an operational IR as, "being open to your institution's community members for depositing, storing, disseminating, and preserving your institution's digital assets," excluding instances of pilot or proof-of-concept IRs. A pilot IR was characterized as, "a proof-of-concept IR for testing and/or garnering institutional support before an institution-wide, operational IR is deployed." Forty-four (94%) respondents reported either current IR development plans or a deployed pilot or operational IR. In response to clarifying questions, 22 institutions reported an operational IR and 10 reported a pilot IR. Six respondents reporting that their institution had neither an operational or pilot IR. These respondents were asked to comment on anticipated plans to deploy a pilot or operational IR. Four institutions planned deploy of a pilot IR; three had anticipated launch dates in 2007 and one had no specified launch date. Two reported plans for an IR with no specification of pilot status, one with a launch date within the next year and one with no specified launch date.

Institutions at the Planning Phase

Respondents were asked to comment on the estimated length of time required for their institutions' pilot and/or operational IR planning activities. Reported length of time for pilot IR planning was shorter than operational IR planning. Of 28 respondents reporting on pilot IR planning time, 14 (50%) reported a planning period of one year or less. For 25 respondents reporting on operational IR planning time, only five (20%) reported a planning period of one year or less, with the majority of respondents, 15 (60%), reporting a planning period of one to two years. The complete results appear in Table 1.

Table 1: Length of IR planning time

Estimate of time	Pilot IR (n=28)	Operational IR (n=25)
Under 6 mos.	3 (11%)	1 (4%)
6 mos. – 1 yr.	11 (39%)	4 (16%)
1 to 1.5 yrs.	7 (25%)	10 (40%)
1.5 to 2 yrs.	5 (18%)	5 (20%)
2 to 2.5 yrs.	2 (7%)	3 (12%)
2.5 to 3 yrs.	0 (0%)	0 (0%)
Over 3 yrs.	0 (0%)	2 (8%)

Institutions reporting on pilot IR planning activities were asked to estimate the time between pilot IR deployment and a planned wide-scale operational IR. Of the 16 respondents, five (31%) estimated deploying an operational IR within one year, post-pilot IR phase. Five (31%) respondents were unsure the length of time between pilot IR deployment and operational IR deployment, and one (6%) reported his institution as planning only a pilot IR, with no plans for a post-pilot operational IR.

Trusted Digital Repositories (TDR)

Respondents were asked if their institutions referred to the TDR in their IR planning activities. Of 36 respondents, 12 (33%) reported having referred to the TDR in their IR planning activities, while 24 (67%) of institutions did not. For respondents reporting an initial reference to the TDR, four (33%) did so before 20% of their IR planning was complete, with four (33%) reporting reference to the TDR between 20 and 40% of IR planning complete. Three (23%) reported initial reference to the TDR with 40 to 60% of their IR planning complete, with the remaining respondent (7%) reporting reference between 60 and 80% of IR planning complete.

For those institutions (24) that did not reference the TDR, respondents reported one or more reasons for this exclusion. Eleven (44%) reported IRs that were planned or deployed prior to the TDR's August 2005 release date. Nine (36%) did not know of the existence of the TDR following its release. One (4%) of the respondents did not find the TDR applicable to their planning activities, while one (4%) did not use the TDR because of other IR planning priorities. Three (12%) of respondents reported no use of the TDR because they lacked sufficient time, staff or other resources. Additionally, two (8%) were not aware if the TDR had been referenced by others in their IR planning activities, although neither had used the TDR themselves.

Respondents that made use of the TDR were presented with a series of five statements, and asked to indicate their agreement or disagreement to these statements on a likert-type scale: strongly agree, agree, neutral, disagree, and strongly disagree. Respondents were provided the opportunity to provide open-ended commentary for each statement.

In response to the statement, "The TDR informed your IR planning activities," nine (75%) of the respondents agreed and two (17%) responded as neutral, with one (8%) not responding. Comments regarding this statement were provided by seven (58%) respondents. Respondents noted the TDR's utility for informing IR policies and service development, describing use of the TDR as a planning and reference tool. Respondents commented that the TDR was informative but that, due to its theoretical, conceptual nature, it was difficult to operationalize the TDR's attributes. Several respondents specified no current plans for implementing the TDR as an audit checklist.

In response to the statement, "The TDR can be/has been applied to your planned or operational IR with minimal or moderate effort," five (42%) agreed and three (25%) disagreed. Three (25%) responded as neutral, and one (8%) did not respond. Comments regarding this statement were provided by six (50%) respondents. Respondents described the TDR as an excellent source for informing IR planning and business practices, but anticipated that application of the TDR as an audit tool would be difficult. While the content of the TDR is very informative, some reported more success in understanding and making use of TDR recommendations later in the IR planning process. The TDR was not recommended as an initial IR resource document for those new to the planning process, as it requires a moderate to extensive level of familiarity with digital repository scenarios in order to be effectively used in planning.

In response to the statement, "The TDR's technical outcome attributes, specifically, informed your IR planning activities," seven (58%) strongly agreed or agreed. One (8%) disagreed, three (25%) responded as neutral, and one (8%) did not respond. Comments regarding this statement were provided by four (33%) respondents. Respondents characterized the TDR as valuable in prioritizing technical attributes in IR planning, allowing for ongoing review of IR instances. Some respondents predicted that, as the service model for IRs at local institutions matures, dependence on the TDR would increase.

In response to the statement, "The TDR's repository management outcome attributes, specifically, informed your IR planning activities," four (33%) agreed, seven (58%) responded as neutral, and one (8%) did not respond. Comments regarding this statement were provided by four (33%) respondents. Respondents commented on the continued utility of using the TDR for reference following their initial IR development activities, as IRs and their underlying architectures continue to mature.

Regarding the potential utility of the TDR for certification audits, one respondent commented on the need for IR planners to refer specifically to the OAIS reference model. Others noted the usefulness of applying the TDR to the many IR policies and workflows necessary for implementation, regardless of intent (pilot or operational), and that identification and creation of such evidence is a necessary, but time-demanding process.

Overall, across all questions, a majority of respondents affirmed the TDR's usefulness and application to IR planning. In summary comments on the TDR, respondents commended the TDR as an excellent source for informing IR planning and

deployment activities. Future IR development at local levels, as well as clarification of the TDR at an operational level, would help in moving application of the TDR from a planning guide to an eventual audit tool.

Other Resources

Respondents were asked to identify up to five resources they perceived to be valuable in IR planning, including the TDR if applicable, and to rank these resources in descending order, from most to least valuable. In total, respondents identified 97 resources. Twenty-four respondents identified their most valuable resource, followed by 23 identifying the second most valuable, 21 the third, 15 the fourth, and lastly, 14 the fifth. Of these 97 resources, 59 (60%) were unique.

In regard to commonalities in resource selection regardless of value perception, eight of the 59 unique resources (14%) were selected by four to six respondents. Six (10%) were selected by two respondents, and the remaining 45 (76%) were only selected by one respondent. Table 2 shows the eight resources identified by four to six persons, providing resource name and frequency of selection.

Table 2: Most cited resources, by frequency

Resource	Frequency among total resources (n=97)
SPARC Institutional Repository Checklist and Resource Guide[8]	6 (6%)
Creating and Institutional Repository: LEADIRS Workbook[9]	6 (6%)
TDR[1]	6 (6%)
Reference Model for an Open Archival Information System (OAIS) [3]	5 (5%)
Institutional Repositories: Essential Infrastructure for Scholarship in the Digital Age [10]	5 (5%)
Trusted Digital Repositories: Attributes and Responsibilities: An RLG/OCLC Report [2]	4 (4%)
The Case for Institutional Repositories: A SPARC Position Paper [11]	4 (4%)
A Guide to Institutional Repository Software v 3.0 [12]	4 (4%)

As shown in Table 2, the TDR was one of the three most referenced resources along with the *SPARC Institutional Repository Checklist and Resource Guide* [8] and the *Creating an Institutional Repository: LEADIRS Workbook* [9].

Interviews

In the survey, 32 of the 47 respondents (68%) indicated a willingness to participate in a follow-up phone interview. Subsequently, the research team conducted interviews with 5 (10%) of these eligible respondents. Institutions chosen to participate include three that made use of the TDR in their IR planning activities and two that had deployed their operational IRs prior to the TDR's August 2005 release date. Each interview followed a semi-structured interview script. Respondents were asked questions in four general areas: Application of TDR, Certification Intent, Attribute Integration, and Planning Resources.

Application of TDR

Three participants used the TDR as a framework for informing their pre-deployment IR planning activities. In the words of one respondent, the TDR was used "to make sure we covered our bases and that we were considering all the issues that needed to be considered." Other than reference as a general framework, the TDR was used for informing development of IR preservation and archival capacities, providing a "useful breakdown of how to think about preservation planning." One respondent put out a call for specific operation-oriented recommendations to complement the theoretical model of the TDR.

Two participants with deployments preceding the TDR draft release (August 2005) were asked to comment on perceptions and subsequent use. One reported first exposure to the TDR as preparation for this research study interview, noting that, "following deployment, the IR planning committee was dissolved, resulting in no forum for discussing the TDR." The other participant has referred to the TDR and, while finding it very useful, finds it to be limited in scope of audience and intent. There is no clear consensus on goals of IRs, so different institutions have different objectives. The TDR would not be as useful for those that lack a focus on preservation. Use of the TDR, in general, requires institutions to be at a "deeper level of planning."

Certification Intent

Two participants reported intentions for achieving certification at some future point, in one instance emerging mid-way through their IR planning process, and in the other, resulting from a second planning phase, with initial IR planning and deployment having taken place three years prior.

Four participants did not express an intention to seek certification at this point or in the future. One commented that, "I don't think I personally know enough about benefits for certification, nor does our community of users. Whether IRs are certified or not wouldn't change their behavior." Another participant stressed that their objective is not to achieve a vetted certification, but to acquire a level of trustworthiness. Although these participants report no planned objectives for certification, two acknowledge that it may be necessary in the future as IRs continue to mature. At that time, those participants estimate that a call for certification from their institutions would be met.

Additionally, participants commented on requirements for certification across all IRs. Participants disagreed with such a hypothetical recommendation, noting that the extent and outcomes of IR instances are difficult to generalize, and that certification should not be an assumed goal for all IR deployments. Some institutions are less prepared for the rigor of certification, especially in regard to resource dollars and staff. The decisions to achieve certification should remain a local one.

Attribute Integration

Participants commented on the particular value of the technologies, technical infrastructure, and repository functions sections of the TDR for delineating build requirements for their IRs, especially in reference to preservation policies and practices. On participant commented on using the TDR "for assessing preservation features, not to assess across all repository features." Looking at the attributes of the TDR, several participants commented on the utility of application across a "high-level" view for influencing decision-making.

Participants commented on the desirability of recommendations for evidentiary documentation for denoting IR trustworthiness and success, but as best practices guidance rather than as requirements. Different system structures and components will require different strategies. Again, participants stressed the shared perception that IRs are not equal, and that differences in dollars, staff and support from each IR's respective community will contribute to an IR administrator's ability to incorporate such recommendations. Also, different institutions have different polices, so not all recommendations would be attainable. One participant provided a copyright scenario as an example.

Planning Resources

Participants shared perceptions of the quantity of available IR planning resource documentation for their planning activities, but varied on their perception of quality. For early deployers, participants relied primarily on their community of implementers for documentation. All participants agreed that there was now a sizeable amount of documentation available for informing IR planning activities. One noted that this body of documentation was "all useful in some way." Others, however, stated that, "there was more information out there than we could read and digest," and "there is information around that is just talking about the same thing and if we already have that information, then we just move on." Several participants discussed the benefits of having access to other IR planners' anecdotal reporting, via listservs, conferences, workshops and other methods of interaction. Several called for more IR management-focused resources, specifically models for staffing, collaboration and sustainability planning. To complement software-specific documentation available, there is an ongoing need for best practices and success stories, as well "less-success" stories. One respondent suggested the development of an online IR managers' community for exchanging information. Two weeks after this interview, such a community premiered, consisting of a blog, web forum, and listserv. [13]

Conclusions

One cannot view the data from this study and not be struck by how early it is in the development of institutional repositories and by the gap that exists between academic libraries and national and international efforts to ensure repository trustworthiness. Overall, participants in this study who made use of the TDR in their IR planning activities reported that the TDR was a valuable resource, providing a solid framework for planning, deployment, and subsequent IR evaluations. Yet, upon closer examination we see that few major research libraries in the US with IRs or plans for IRs have used the TDR in any way and fewer still view certification to be important to their efforts. The TDR was one of 59 unique planning resources denoted as valuable by participants, achieving a top rank across all resources in terms of frequency of selection. However, this measure does not imply wide use across the IR planning community since only 6 participants self-reported it as one of five valuable planning resources. Of the 33 respondents with planned or deployed IRs following the TDR's 2005 release date, only 12 (36%) made use of it. Generally, the current state of IR planning and deployment makes formal calls for attribute standardization and certification premature with compliance unlikely for all but perhaps the largest institutions such as national libraries.

At the same time, efforts such as the TDR do not seem to address the current needs and developmental stage of typical institutional repository efforts at colleges and universities. The fact that the Center for Research Libraries is now the home of the TDR's successor document, *Trustworthy Repositories Audit & Certification (TRAC): Criteria and Checklist* [4], may change this situation significantly, along with the release of two other, recent documents: the Digital Curation Center's (DCC) [14] and DigitalPreservationEurope's (DPE) [15] *DRAMBORA: Digital Repository Audit Method Based on Risk Assessment* Toolkit [16] in March 2007, and the Network of Expertise in Long-term STOrage of Digital Resources' (nestor) [17] *Catalogue of Criteria for Trusted Digital Repositories (CCTDR)* [18], released in an English-language version in December 2006. Libraries may see institutional repositories increasingly as a core activity and these more specific tools may help in this process. The general maturation of institutional repository efforts may also bring about more concern for technical development, trustworthiness, and auditable policies and procedures.

This study leads to several recommendations for future research into IR planning, and recommendations for conducting investigations of IRs, in general. First, with the recent publication of TRAC [4], DRAMBORA [16], and the CCTDR [18] it would be interesting to follow up this study with one on participants' perceptions of these publications to determine if some of the concerns arising from the draft TDR, such as lack of operational recommendations and measurements, are satisfied. DRAMBORA takes a risk-based approach for assessment, supporting self-administration and providing tasks for each associated stage of assessment. The latter is an audit feature requested by participants in this investigation, allowing for operationalize of the theoretical model provided in the TDR. DRAMBORA also provides recommendations for evidentiary documentation and templates for coordinating documentation management and acquisition. These ten areas include digital repository objectives, operations, polices, agreements, standards in use, assets, depositors, and risks.

The CCTDR took an open approach in the development of their audit checklist; it is intended for use across all types of digital repositories, and supplements the instrument's conceptual criteria descriptions with concrete examples, albeit not an exhaustive list. Such a follow-up investigation of participants' perceptions may need to be expanded to target actual participant experience with these three recent releases, particularly in regard to the DRAMBORA and CCTDR. Of the 59 planning resources denoted as valuable by participants, few originated outside of the United States. DRAMBORA and the CCTDR are UK and German initiatives, respectively.

Second, efforts to align and identify the checklist's attributes with manifestations of evidentiary documentation in a digital

repository environment are problematic [19]. A future study, building upon this current initiative, would be a thorough investigation of the specificity for academic IR planners that is contained within the generalized context of the recently released *Trustworthy Repositories Audit & Certification (TRAC): Criteria and Checklist*. An outcome for such a study would be recommendations for a common approach at IR planning and building down to more specific levels of detail. Another outcome of such a follow-up investigation would be an aggregate of accounts in IR attribute documentation, leading to recommendations for the collection and documentation of appropriate evidence for subsequent audits and certification. While this current research study attempted to probe for this type of information, only a portion of survey respondents (12%) reported use of the TDR in their planning activities, with 9 (20%) reporting no knowledge of the TDR following its release. It may be too early in IR development for any of these studies to yield significant results, but they could prove valuable in the future.

Since approximately half of ARL institutions report deployed or planned IRs, per this work and the recent ARL deployment study [SPEC], there is still a substantial audience in this community of institutions yet to report IR initiatives that would benefit from additional planning-focused research. A thorough review of the selected IR planning resources reported in this study would work toward creation of a bibliography for IR planners, consisting of vetted, relevant and necessary information for informing successful planning activities.

Several study participants noted the continued lack of consensus across IR activities in general, including scope, objectives, policies, and management. A few invited participants contacted the research team to assess if they were eligible to participate, unsure if their local definitions for an IR matched that of the study team. Lynch and Lippincott [20] reported lack of definitional consensus as an issue in their survey attempting to enumerate the number of IR deployments across American universities. While absolute consensus on a restricted set of terms for identifying the different attributes of digital repositories at universities, nationally and abroad, may not be feasible, developing a clear set of parameters for classification levels, such as extent of collection scope, content contributors, and access, might be attempted through future investigations and a synthesis of past IR-related studies. It would benefit subsequent reports on IR deployment, usage, and other IR-related measures.

Acknowledgment

The authors wish to thank the National Historical Publications and Records Commission (NHPRC) for supporting this study through the NHPRC Electronic Records Fellowships program (2006-09).

References

[1] RLG/NARA, Taskforce on Digital Repository Certification, Audit Checklist for Certifying Digital Repositories: Draft for Public Comment (2005). http://www.rlg.org/en/pdfs/rlgnara-repositorieschecklist.pdf

[2] RLG/OCLC, Working Group on Digital Archive Attributes, Trusted Digital Repositories: Attributes and Responsibilities (2002). http://www.rlg.org/legacy/longterm/repositories.pdf

[3] CCSDS, Reference Model for an Open Archival Information System (OAIS): Blue Book CCSDS 650.0-B-1 (2002). http://public.ccsds.org/publications/archive/650x0b1.pdf

[4] OCLC, CRL, and NARA, Trusted Repositories Audit and Certification: Criteria and Checklist (2007). http://www.crl.edu/PDF/trac.pdf

[5] University of Houston Libraries Institutional Repositories Task Force, SPEC Kit 292: Institutional Repositories (SPARC, Wash. DC, 2006).

[6] Karen Markey, Soo Young Rieh, Beth St. Jean, Jihyun Kim, and Elizabeth Yakel, Census of Institutional Repositories in the United States: MIRACLE Project Research Findings (CLIR, Washington DC, 2007). http://www.clir.org/pubs/reports/pub140/pub140.pdf

[7] Association of Research Libraries. http://www.arl.org/

[8] Raym Crow, SPARC Institutional Repository Checklist & Resource Guide (SPARC, Washington DC, 2002). http://www.arl.org/sparc/IR/IR_Guide.html

[9] Mary R. Barton and Margaret M. Waters, Creating an Institutional Repository: LEADIRS Workbook (2004). http://www.dspace.org/implement/leadirs.pdf 2004.

[10] Clifford Lynch , "Institutional Repositories: Essential Infrastructure for Scholarship in the Digital Age," ARL Bimonthly Report (226) (2003). http://www.arl.org/newsltr/226/ir.html

[11] Raym Crow, The Case for Institutional Repositories: A SPARC Position Paper (SPARC, Washington DC, 2002). http://www.arl.org/sparc/IR/ir.html

[12] Budapest Open Access Initiative. http://www.soros.org/openaccess/

[13] IR-Managers. http://oaresearch.org/

[14] Digital Curation Center. http://www.dcc.ac.uk

[15] DigitalPreservationEurope. http://www.digitalpreservationeurope.eu/

[16] DRAMBORA: Digital Repository Audit Method Based on Risk Assessment. http://www.repositoryaudit.eu/

[17] Network of Expertise in Long-term STOrage of Digital Resources. http://www.langzeitarchivierung.de/index.php

[18] nestor, Working Group Trusted Repositories – Certification, Catalog of Criteria for Trusted Digital Repositories, version 1.0 (2006). http://edoc.hu-berlin.de/series/nestor-materialien/8en/PDF/8en.pdf

[19] Seamus Ross and Andrew McHugh, "The Role of Evidence in Establishing Trust in Repositories," D-Lib Magazine, 12 (7/8). (2006). http://www.dlib.org/dlib/july06/ross/07ross.html

[20] Clifford Lynch and Joan Lippincott, "Institutional Repository Deployment in the United States as of Early 2005," D-Lib Magazine, 11(9). (2005). http://www.dlib.org/dlib/september05/lynch/09lynch.html

Author Biography

Carolyn Hank is a Triangle Research Libraries Network (TRLN) Doctoral Fellow at the University of North Carolina at Chapel Hill. She is the project manager for the Digital Curation/Institutional Repository Committee, charged with developing a pilot institutional repository at UNC-CH. She helped organize the 2006 JCDL workshop, Digital Curation & Trusted Repositories: Seeking Success. She is a co-author on the State Library of North Carolina's recent five-year LSTA evaluation report.

Helen R. Tibbo is a Professor in the School of Information and Library Science at the University of North Carolina at Chapel Hill. She teaches in the areas of archival management, and digital curation. She is the co-primary investigator for the NHPRC Electronic Records Research Fellowships Program, and principle investigator for the "Preserving Access to Our Digital Future: Building an International Digital Curation."

Heather Barnes is a current student in the graduate program at University of North Carolina at Chapel Hill's School of Information and Library Science. She is the Project Assistant for Dr. Helen Tibbo in the NHPRC Electronic Records Research Fellowships program. Her primary areas of research interest include digital libraries and audio/visual collections.