Preservation Pressure Points: Evaluating Diverse Evidence for Risk Management

Seamus Ross, Director HATII University of Glasgow and Associate Director of the UK's Digital Curation Centre & Andrew McHugh, Audit and Advisory Manager of the Digital Curation Centre

Establishing a comprehensive understanding of the effectiveness and trustworthiness of a digital repository requires a broad range of evidence. Preservation can be considered as a complex spatial and chronological network of challenges, and associated risks. For example the organisational, financial, technological and operational contexts within which a repository resides and the extent to which it is capable of managing related threats must be ascertained if an audit is to determine the likelihood of the institution's success. Significant effort must be directed towards the definition of methodologies for identifying appropriate classes of evidence, their evaluation, and their weighting. Formal means are required to facilitate the analysis and comparison of disparate evidence types in order to enable auditors to accommodate a diverse range of physical, testimonial and experiment-based proof. In addition to binary systems of inquiry (e.g., does the organisation have a mission statement?) auditors must display an ability to distinguish the most persuasive examples from those that provide less substantive evidence of organisational competence. Similarly, if, for instance, a significant proportion of staff reveals that they have no idea of the content of their organisation's mission statement then this must be reflected in the overall organisational assessment. Comprehensive insights and informed evaluation can only be achieved after fully exploring the evidential basis upon which compliance is to be founded. This discussion of evidential appraisal techniques for repository audit reflects a series of pilot audits undertaken by the Digital Curation Centre within a selection of UK and worldwide data centres and archives, including the Beazley Archive, the British Atmospheric Data Centre, the National Digital Archive of Datasets, the National Library of New Zealand's National Digital Heritage Archive, the Florida Digital Archive, and, in collaboration with the Center for Research Libraries' Certification of Digital Archives project, the Dutch Koninklijke Bibliotheek.

As noted in previous work on this subject[1], any mechanisms to facilitate the assessment of trustworthiness in digital information preservation repositories must be supported by sound and transparent evidence demands. Efforts to date[2] have concentrated on the establishment of check-lists to document the key criteria that ought to be identifiable within successful and ultimately trustworthy repositories. However, questions have arisen over the practical applicability of such instruments; without associated metrics for determining the extent and effectiveness of organisational compliance it remains impossible to conceive reliable means for comparing and assessing repositories that are heterogeneous in terms of scale, scope or mission. The identification of appropriate evidence requirements to demonstrate check-list compliance is a necessary starting point, and work has already been undertaken to classify and describe the kinds of documentary, testimonial and observation evidence that may be both

1 See Ross, S & A McHugh, '*The Role of Evidence in Establishing Trust in Repositories*', D-Lib Magazine, Volume 12 Number 7/8, ISSN 1082-9873, available from http://www.dlib.org/dlib/july06/ross/07ross.html [Accessed 19 September 2006]

2 Led most notably by the RLG/NARA Certification Task Force, and the German *nestor* group

admissible and insightful in determining the extent to which an organisation is trustworthy. Nevertheless, identifying the appropriate kinds of evidence is only the first step in a process that must ultimately establish a platform for comparing disparate evidence types, making explicit the weight of legitimacy and persuasiveness associated with individual evidence examples, and the extent to which organisations may be compelled to submit specific evidence or combinations of evidence in order to satisfy particular check-list criteria. Of similarly pressing concern is the current absence of any formal means by which single examples or collections of evidence can be either evaluated or accepted as having maximum impact by virtue of their mere existence and availability. Finally, in the event of contradictory evidence an impact hierarchy must be conceived to determine which kinds of evidence should ultimately supersede.

A Context for this Work

In order to approach these issues, the Digital Curation Centre has committed to a series of pilot audits within a range of information repositories, which, as well as providing the participating organisations with an objective and expert insight into the effectiveness of their operations, is aimed at exploring the means by which assessment will take place and determining the robustness and global applicability of those metrics, criteria and methodologies that have already been conceived[3]. This paper is derived from experiences accumulated throughout six audits. The first was undertaken as part of the Center for Research Libraries' Certification of Digital Archives project[4], which involved three similar comparable audits at Portico and the Inter-university Consortium for Political and Social Research in the United States and at the Koninklijke Bibliotheek's e-Depot in the Netherlands. The DCC was present during the latter, mainly in an observational capacity. In addition, the DCC has completed five of its own pilot audits, deliberately approaching a heterogeneous selection of host organisations, based on a will to determine the extent to which generic assessment criteria and methodologies might be viable. This paper describes organisational assessments undertaken at the British Atmospheric Data Centre (BADC); the National Digital Archive of Datasets (NDAD); the National Library of New Zealand's National Digital Heritage Archive (NDHA); the Florida Digital Archive (FDA) at the Florida Centre for Library Automation; and the Beazley Archive (BA) at the University of Oxford. This represents a diverse subset of curatorial contexts, with each contrasting in scale, scope, funding basis, means of deposit, user community and in terms of the nature and origins of their digital holdings.

The Participating Repositories

The Koninklijke Bibliotheek's relationship with two internationally established Dutch publishers has been integral to the establishment of its e-journal preservation storage resource, the e-Depot. Successful negotiations with Elsevier and Kluwer concluded in 1996, and subsequent collaborating publishers include Oxford University Press, Taylor & Francis, Sage, and Springer. At the end of 2005 the e-Depot accommodated around 3,500 e-journal titles, comprising some 5 million articles and totalling around 6.3 Tb.

The BADC provides electronic archiving facilities for a range of data producers, perhaps most notably the research projects funded by the Natural Environment Research Council (NERC) but also significant international meteorological organisations such as the UK Met Office and the European Centre for Medium-range Weather Forecasts. Based at STFC's Rutherford Appleton Laboratories the Centre was established in its original incarnation as the Geophysical Data Facility in 1985. Within the subsequent two decades it

3 Most notably, at this stage, the RLG/NARA Draft Audit Check-list for Certification of Digital Repositories

4 http://www.crl.edu/content.asp?l1=13&l2=58&l3=142 [Accessed 20 September 2006]

has grown to represent the NERC's primary and sole data centre for data originating from atmospheric research, and consists of over sixty terabytes of data in a variety of formats.

The National Library of New Zealand (in the native Maori, translated as Te Puna Mātauranga o Aotearoa) was formally established by the 1965 National Library Act, when the General Assembly Library, the Alexander Turnbull Library and the National Library Service were combined to form a single National Library. It has a legislative mandate to collect, preserve, and make accessible New Zealand's documentary heritage and, courtesy of the 2003 National Library of New Zealand Act, this mandate has been extended to include electronic materials. The Library exhibits a number of novel characteristics. Unlike most National Libraries it operates as an autonomous government department. The legal deposit legislation that the library is subject to also differs from most other national libraries, placing the onus for collection on the library, and not publishers. The National Library is therefore compelled to accept any content that relates to or originates within New Zealand. The National Digital Heritage Archive is an ambitious and heavily funded project to build a repository fit for preserving New Zealand's digital memory, undertaken in partnership with Ex Libris and Sun Microsystems. It represents a sizeable contribution to the New Zealand Digital Strategy, a five year initiative aimed at ensuring that New Zealand fully grasps the opportunities presented by Information Communications Technology.

The National Digital Archive of Datasets (NDAD) managed by the University of London Computer Centre (ULCC) preserves and provides online access to archived digital datasets and documents from UK central government departments. Data stored by NDAD remain in the legal custody of the National Archives, but are managed by ULCC, who provide preservation and dissemination services. The datasets accessioned by NDAD can vary tremendously. Many of the datasets transferred to the repository are decommissioned databases that have been superseded. In other circumstances they represent snapshots of running servers still in production use. Furthermore, the types of data, their size and their subject matter exhibit considerable diversity.

Based in Gainesville at the University of Florida, the Florida Center for Library Automation (FCLA) provides a range of services for each of the eleven State university libraries within the US Sunshine State. One of their principle services is the Florida Digital Archive, which aims to provide long term preservation archival services for digital materials originating from any of Florida's state university libraries. In 2002 a three year grant from the Institute of Museum and Library Services (IMLS) kick-started the archive's development within FCLA. Preservation functionality is prioritised ahead of access features, and consequently the FDA operates principally as a dark archive. Its commitment is that all files deposited by agreement with its affiliates remain available, unaltered and readable from media, with preservation achieved using the best format migration tools available. FDA's technological foundation is a set of scripts and programs, conceived and developed internally, that are collectively known as DAITSS (the Dark Archive in the Sunshine State), which was recently released under the GPL open source license.

Finally, the Beazley Archive database represents an electronic annex to an older physical archive, which itself consists of several hundred thousand examples of notes, photographs, negatives, drawings, books, catalogues and gem impressions. Three databases represent the bulk of the electronic content; the first, documenting Athenian figure-decorated pottery c. 625-325 BC was initially brought online in 1979 and has evolved to now encompass more than 98,000 records and 120,000 images, and boasts 15,000 registered users. Two additional (albeit smaller) databases document gemstones and casts respectively. Much of the Beazley's electronic content acquisition is more proactive than responsive, with staff encouraged to actively pursue newly available catalogue information and photographs for accession within the digital archive.

Perhaps the single common factor in each of these (other than the consistent participation of individual auditors) was the central intellectual foundation upon which assessment was based. A draft version of the RLG/NARA check-list (subsequently formalised as the Trustworthy Repository Audit and Certification Criteria and Check-list) provided the best practice benchmark that informed each institutional assessment. In five of the six organisational assessments repository administrators were already familiar with the document, and with the work that helped shape it, such as international standard ISO 14721:2003, the Reference Model for an Open Archival Information System[5]. The final repository was less well versed in the surrounding intellectual framework, and this had undoubted implications, explored in more detail below. A key goal of the pilot audits was to determine the legitimacy of the check-list's metrics, their applicability in a range of circumstances, and their usability both as self-assessment criteria and as a tool to structure and support third part repository audit.

Although strict confidentiality agreements limit the extent to which specific details of these audits can be documented, these do not generally impact on the extent to which details of the adopted audit methodology can be disclosed. Nevertheless, in order to protect the participating organisations, references to individual repositories are henceforth anonymised.

Approaching an Objective Compliance

At the very heart of recent work in this area is a notion that has been generally and popularly characterised as repository trustworthiness. This still somewhat ethereal concept is one that might[6] be demonstrated by exposing the targeted repository's organisational, procedural and technological infrastructures to a criteria set that represents an objective benchmark for good, sustainable preservation practice. Conformity ought to indicate the mitigation of risk, and provide reassurance to numerous stakeholder groups, including content depositors, information consumers and repository funders. However, concerns highlighted in previous work continue to persist; irrespective of the breadth or scale of community endorsement of existing check-list criteria (with comprehensive agreement unlikely), the process will ultimately rely upon the availability of established, proven and consistent means for formally determining whether or not conformity exists. The development and enforcement of appropriate evidence requirements is essential; however, simply listing relevant documents alongside each check-list metric is almost certainly insufficient. In addition, the community will rely upon the emergence of much more sophisticated means to identify appropriate classes of evidence in particular circumstances, and of non-discipline-specific methodologies to support the weighting, analysis, combination and comparison of evidence. Admissible proof is likely to exhibit significant diversity in terms of its origins, the form it assumes and the extent to which it is persuasive. Furthermore, alternative kinds of evidence must be hierarchically prioritised to adequately deal with contradictions that might arise, for instance between an employee's personal testimony and documented work-flow policy. Finally, evidence might vary in terms of the level of granularity to which it may be assessed. In some cases the simple existence of a particular document or class of document will be sufficient to determine a degree of organisational compliance. Needless to say, in such cases it is usually vital that staff and stakeholder communities are aware of the existence of this documentation. However, in the majority of cases it is likely that auditors will be expected to scrutinise content much more closely to evaluate the *extent* to which repositories meet the certifiable expectations.

With this in mind we set out to develop an increased understanding of how evidence is

5

6 Assuming that the provision of Audit and Certification services is just one of several ways in which we may satisfactorily evaluate and determine repository competence

practically accumulated, assessed, used and discarded throughout the audit process. We sought the means to make practical, objective sense of the potentially limitless kinds of evidence that might be submitted in support of certification, and to classify evidence examples according to their origins, form and weight of legitimacy. Regularising disparate evidence equips the auditor to effectively cross-compare, corroborate and prioritise the full range of proof and testimony that is provided throughout a repository's bid for certification.

An Evidence Based Approach?

The adopted process has varied significantly during each of the audits that the DCC has participated on or led during this spell of pilots. Broadly speaking, based on previous work, evidence is classified as one of documentary, testimonial (interview-based), and observation-based. In every instance, testimonial evidence played a key role, but in often different ways.

Archive X

At the audit of *Archive X*, despite the large volume of documentation that was made available to the team during the on-site process, the most influential evidence was solicited via themed interviews, structured to broadly correspond with the sections of the RLG/NARA audit check-list, and involving a selection of relevant staff. The majority of the conclusions from this audit were drawn from a combination of written self-assessment (mainly corresponding to the RLG/NARA check-list and completed and submitted to auditors prior to the on-site activities), and the series of staff interviews. Only the self-assessment documents and a further short document describing some perceived problems with the RLG/NARA check-list were available to auditors prior to their arrival on-site. Interview questions were primarily designed to address points of uncertainty within the check-list responses, and the specific criticisms of the metric that had been presented. The general process throughout each session was to question only those responses that appeared to demonstrate non-compliance with the check-list's prescribed metric or that questioned the value of those metrics. Significantly less time was spent questioning those responses that suggested best practice had already been implemented. Finally, notwithstanding the insights afforded during a short tour of the archive, the audit offered no opportunities for staff to demonstrate the operation of the system or for auditors to see or question the specifics of actual physical processes. In that respect then, no primary evidence about the hands-on digital object management undertaken within the archive was available.

Documentation was made available on arrival at *Archive X*, and appeared both extensive and persuasive. However, as it was subject to little formal analysis, and was rarely responsible for the provocation of questions during interview sessions one must be cautious of overstating the role it played during the audit. Within the concluding presentation auditors were congratulatory about the level of documentation that the archive had accumulated and made available. However, one might assert that, from a formal, analytical perspective, its legitimacy was granted based on little more than its quantity. A trustworthy repository will have extensive documentation, but it is not necessarily true to say that any repository with extensive documentation is trustworthy. Only when that paperwork has been proven to document appropriate organisational, financial and technological infrastructures can such assertions be made.

One might reasonably assert that the check-list's potential value was not fully exploited. A more effective strategy would combine the different kinds of evidence available (interviews, documentation and observation of practice) to evaluate the satisfaction of the requirements of individual sections (although not necessarily *every* section), with individual metrics providing opportunities for more specific analysis. The audit of *Archive X* dwelt on the pursuit of evidence to explore points of ambiguity or organisational failings perceived

only from the initial self-assessment. In this sense, the approach was reactive, and only to points of negativity – compliance was frequently determined on the basis of little more self-affirmation, whereas this ought to have been admissible as just one of several kinds of evidence that confirmed the pedigree of the repository. The self-assessment documents undoubtedly represent a valuable starting point for the audit process. However, rather than treat them definitively, the on-site process may benefit more from being an exercise in determining their legitimacy, employing interview, document analysis, observation and experimental techniques to establish a sense of the extent to which each metric's requirements are satisfied.

Archive Y

The exercise undertaken at *Archive Y* differed substantially, with significantly less documentation available throughout the process. A small selection was gathered prior to the on-site activities, mainly encompassing descriptions of system procedures and functionality and some documentation describing the archive's practical commitment to preservation. What distinguished this audit most significantly from the others described in this paper was the manner in which the RLG/NARA check-list was employed. In each of the other examples archive staff had familiarised themselves with the document's metrics and provided, in advance, a series of self-penned responses. In the case of *Archive Y*, the archive was only comprehensively exposed to the check-list during the on-site activities. In the absence of sufficient alternative documentation, the discussion with staff closely reflected the check-list's structure; in that respect the on-site activities resembled a measured, facilitated self-assessment exercise. A useful consequence was that auditors were afforded an insight into the check-list's applicability, relevance and usefulness within an archive that had evolved with little prior knowledge of the specifics of its implicit best-practice metrics. Further deviating from the audit of *Archive X*, were efforts to obtain more *practical* insights into the repository operations – each interview was conducted with a computer workstation nearby, ensuring that the physical processes of ingest, archival storage, data management and access could be demonstrated. Typically, the information gathering process began with the check-list requirements. Posed as questions, one or more of the individual metrics would encourage discussion from appropriate individuals, which included management, technical support individuals and object management specialists. In the absence of comprehensive documentation little recourse was available to printed matter, and instead auditors would frequently request that further illustration be provided by way of practical example within the Archive's digital object management system.

The absence of extensive written documentation, particularly prior to the on-site activities, hampered the ability of the auditors to offer a comprehensive and definitive assessment of *Archive Y*. One might argue that it was the lack of opportunity for repository staff to familiarise themselves with the check-list's terminology – to become immersed in the culture of the check-list – that proved most problematic. But what appears clear is that the check-list does not seek to *define* or *legislate* a new form of best practice; rather its intention is to reflect and encapsulate that which is already broadly accepted within the preservation repository community. An optimal chronology for evidence gathering begins to emerge - the check-list provides an initial and primary focus for repositories, but the primary starting point for auditors has to be documentation. The absence of documentation from *Archive Y* was symptomatic of the same organisational shortcomings that meant a large proportion of check-list metrics appeared unfamiliar and onerous. A self assessment based on the check-list provides a useful mechanism to complement the documentary analysis, which can be backed up by staff or stakeholder testimony and further galvanised through demonstration.

Archive Z

The audit of *Archive Z* was facilitated with the availability of substantial and varied documentation, interview subjects that were both responsive and forthcoming and an organisationally enthusiastic attitude to the demonstration of practical processes undertaken during the archive's normal operation. Near comprehensive documentation was supplied to auditors prior to the visit (including a self assessment based on the RLG/NARA check-list), offering an opportunity to establish considerable foundational understanding of the organisation, its contextual spacing, the nature of its business and its digital holdings, its technological infrastructure and the services and functionality it is committed to providing. In isolation, this falls far short of representing conclusive proof of the trustworthiness of the repository (although its very existence provides a persuasive indicator of managerial effectiveness). As a starting point however, the documentation, which included extensive details about the archive's systems and procedures, technical architecture, staffing, funding, depositor relationships (including legal relationships) and risks represented an essential 'base-camp' from which to undertake further exploration. Equipped with an initial world view, the auditors could spend their limited time on-site seeking confirmation; staff interviews would provide compelling insights into whether the documentation was representative of real day-to-day practice and observation of the completion of tasks, interactions with the system and management process would prove even more conclusive. The on-site interviews roughly corresponded with those undertaken during the audit of *Archive X*, although perhaps only in terms of their chronology. Once more, the check-list provided a helpful structure within which to organise interview sessions. However, with the extensive evidence already submitted and considered it became feasible to conduct the interview process along more adversarial, almost judicial terms. The exercise became akin to cross examination, where truths within the documentation were corroborated, gaps were gradually filled in and hazily alluded-to concerns were embodied, their extent realised. Every interview room had facilities to access the archive's computer system, which facilitated both the demonstration of any concepts or processes that arose as well as the recovery of any additional electronic documentary evidence, which could be checked whenever referenced in conversation.

As the audit continued, the evidence-based focus gradually narrowed; the initial goal to accumulate a broad understanding of the archive evolved into increasingly granular level of inquiry, culminating in the determination of whether selected individual criteria from within the RLG/NARA check-list had been satisfied. In that respect the check-list provided a pivotal structural support – its broad scope determined the parameters of both initial general investigation and its individual metrics the focus of more specific subsequent assessment.

Archive A

The preservation service provided at *Archive A* is undertaken as a contractual obligation under agreement with a national archive, and as a consequence the documentation available within the organisation was perhaps second to none, at least within the context of this pilot process. Perhaps the greatest lesson learned in this exercise focused less on the specifics of the available audit tools or methodology than on the preparatory work that repositories might undertake to facilitate a successful assessment. As a prerequisite and consequence of its contractual relationship, the outcome of a competitive tender process, *Archive A* maintains a tremendous body of documentation relating to almost every part of repository operations. Within an impressive catalogue of policy and procedure manuals, issues such as digital object acquisition, digital preservation, information security, staff training, legal responsibilities, policy review and access were each explained in considerable detail, enabling an auditor to quickly gain a comprehensive picture of the repository, which could be immediately compared with the criteria implicit within the audit check-list. Similarly, the repository had willingly undertaken a challenging process of certification under the ISO 9000 series of

standards, relating to quality assurance across every aspect of the organisation. It was also already subject to detailed inspection by its primary client to determine the suitability of its physical infrastructures. The check-list self assessment document returned by *Archive A* was of particular value since unlike every other received during this programme, it was completed by a broad range of staff representing every level and repository function. This provided an opportunity to confirm that repository policies had absorbed into the consciousness of all staff, and not just a single overseer. Discussions with staff indicated a broad and in-depth awareness of policy and procedure in every area and an organisational cohesion consistent with documentation, enabling auditors to more easily take demonstrations of repository functionality at face value with less need to adopt a more adversarial investigative approach. By embedding a culture of assessment, improvement and transparency firmly within the repository, the demands of inviting external auditors to perform further assessment were minimised. In *Archive A*, these characteristics were implicit as a management objective, and to the fore to facilitate the effective running of the repository. An increase in the organisation's '*auditability'* appears to be a resultant side effect. The goal of auditors is to identify good management practice; the goal of repository staff is to manage their repository effectively. Both are consistent with a requirement for a formally documented and internally expressed self awareness.

Archive B

The repository audit that took place at *Archive B* was unique because it was the only example where the subject of the assessment was currently in development. As a consequence, this exercise provided several valuable insights. Archival assessment is fundamentally about seeking an understanding of where shortcomings or, expressed slightly differently, opportunities for improvement, are evident; in the first instance this can be done by comparing the repository with the best practice benchmarks found within tools like the RLG/NARA check-list. However, within subsequent assessments, the focus will naturally shift somewhat into an analysis of the extent to which those initially identified opportunities for improvement have been realised. No longer is an objective check-list the only reference point, with the previous audit informing to a great extent the specific lines of enquiry that are subsequently pursued. Although *Archive B* was very much in a state of development, an interim system had been in place for some time. Considerable resources had subsequently been made available, and repository staff had taken the opportunity to work towards an improved solution. The initial stages of this development process had a great deal in common with the other archival audits, with implicit shortcomings of the existing system identified, reversed and expressed as functional requirements that would be the basis for a new system. In a sense, this pilot audit had much in common with a 'follow-up' assessment. In reality, it exhibited hybrid characteristics, and two-pronged approach was maintained throughout the three days spent on site. Partially, the process was concerned with validating the conclusions reached internally about the state of the interim system. Exposure of existing repository practice to the check-list criteria was therefore an obvious starting point, but the latter part of the assessment concentrated on the proposed new system. The majority of available documentation that was specific to the new implementation consisted of requirements and planning materials. By validating the original gap analyses (exposing the interim system to the check-list criteria), auditors were then able to determine to what extent the new system addressed the problems implicit within the first. This worked quite effectively, although of course the plans had not been fully implemented (the technological implementation was yet to begin) so the pilot was not completely representative of a full retrospective assessment. An assessment of a system that is in development or planning can take into account only documentary evidence; where questions persist even after inspecting the available documentation one has to take verbal assurances at face value. Therefore, while self assessment can be usefully undertaken within developing environments, and assessment

criteria can be useful deployed as a repository blueprint, the notion of an objective and demonstrable success is less feasible.

A further interesting aspect of this assessment was the suggestion from administrators that the planned repository would be incapable of meeting the requirements of the RLG/NARA check-list. It appears that at least one motivation for welcoming auditors appeared to be to build a formal expression of repository shortcomings in order to command further investment. From that point of view it was interesting to determine the extent to which the RLG/NARA check-list was capable of highlighting failures, as well as successes.

Archive C

Archive C was the final repository to be subject to assessment in the initial DCC pilot programme; the adopted methodology was therefore quite mature by this stage. Like many of the audited organisations, *Archive C* submitted a self assessment document based on the RLG/NARA check-list in anticipation of the audit visit, and this proved to once more be a useful source of insights. By this stage of the program it was clear that self completed check-lists worked best when read after more neutral documentation, providing as they do a quasi personal response, in many ways akin to a dialogue between auditor and repository representative. Accompanying the self assessment document were a variety of additional documents, which included organisational information, policy information, software specifications, and example deposit agreements. The audit began not with the arrival on site, but upon receipt of this documentation, with a thorough analysis providing numerous insights into the repository infrastructure that would be subsequently explored. Two days of on-site activities provided an opportunity for discussions and demonstrations of system functionality and work-flow, and these highlighted a number of implicit concerns. The auditors and audit methodology were by this point sufficiently well established that although interviews were still structured according to the broad categories of criteria within the check-list it was much less necessary to labour over every specific criterion. A more fluid process evolved; although checks for completeness were made by reference to the checklist at the end of each discussion, interviews were mainly structured by the evidence provided by the repository.

The team based at *Archive C* appeared in a number of cases to be broadly conscious of their organisational shortcomings but the audit exercise enabled their more formal realisation and expression, and permitted them to be more effectively addressed. Notwithstanding this, the strongest conclusion reached by the auditors during this assessment was the value of experience accrued during the previous pilots. In isolation, the RLG/NARA check-list criteria offer a useful structure around which to base assessment, and a number of clues about the shape that repository activities might best assume. However, the specific details of how repositories should conform to these criteria are not really expressed within the check-list. Certification is ultimately about comparison, using objective metrics, and with peer organisations. For this to work there is an implicit requirement that tools and methods must support comparability. By exposing his or herself to a range of environments that purport to satisfy the check-list's criteria, the auditor is equipped to determine optimal means of check-list conformity. An additional level of granularity can be expressed, whereby metric conformity or non-conformity is no longer an atomic consideration. Instead, auditors can determine the extent to which specific practical approaches are capable of satisfying individual criteria, and introduce a notional understanding of what this means in terms of a more universal understanding of conformity. For example, exposed to just a single repository, an auditor may see evidence of training provisions where staff may request practical training during an annual skills review session, that appear to satisfy metric A2.3 of TRAC (“Repository has an active professional development program in place that provides staff with skills and expertise development opportunities.”). This may however appear less than satisfactory when the same auditor visits a second repository that offers, in addition to an

annual skills review session, a system that requires each staff member's line manager to monitor performance levels to suggest appropriate training. The latter approach ensures that any training opportunities that staff members may themselves be unaware of remain available, and is therefore preferable. But without the exposure to a range of implementations that aspire to conformity, it remains difficult for auditors to determine where improvement might plausibly be sought. It might be said that the role of a consultant is to distil broad and varied knowledge, accumulated with considerable experience over a significant period, into advice or services for a client that lacks the resources to themselves gather that experience. An auditor's role is similar; in order to understand the practical realities of check-list compliance, one must be exposed to a wide variety of implementations. An aspiration to conform is just half of a picture that must also include a practical capacity to conform.

This suggests that the success of the audit is completely dependant upon the availability of sufficiently expert auditors. It is they who must bear the weight of interpretability of audit criteria, and who must determine the practical meaning of conformity. Any opportunities to objectify the process, and convey this knowledge to those within the repository profession should of course be explored. Accomplished auditors are equipped through their experiences to ask telling questions of repositories, which might be understood as *Key Lines of Enquiry*. There is a danger that unless expressed as at least a semi-formal framework within which evidence can be gathered and assessed, the audit process may appear overly based on 'feel', and dependent on the perceptions of specific auditors, which limits opportunities for comparison.

Key Lines of Enquiry

Over the period of pilot assessments it became increasingly possible to identify key lines of enquiry to correlate with the RLG/NARA criteria. These are structured frameworks intended to facilitate the audit process, a means of relating objectively expressed criteria to the realities of the information infrastructure that is under scrutiny. Taking a check-list criterion as a starting point, one might conceive of example practical expressions of conformity, questions that determine both will and capacity to conform, associated risks, and example risk manifestations. The intention is to make it more straightforward for both auditors and repositories to identify where check-list criteria have been satisfied, and to build an increased sense of the obstacles and problems that might be implicit, although difficult to see within both common and atypical approaches.

TRAC Criterion:	• If unique identifiers are associated with SIPs before ingest, the repository preserves the identifiers in a way that maintains a persistent association with the resultant archived object (e.g., AIP).
Examples of Conformity:	• Objects are renamed to correspond with identifiers • Objects are stored in a directory named to correspond to identifier • Objects are packaged using alternative mechanism with identifier information (e.g., in a zip file with associated text file) • Database table maintains identifier with corresponding field describing full path where object resides, or a sub-path from the root of the archive that remains consistent even if the archive information is transplanted elsewhere, paired with a current path prefix. For example, record the archival path as `/2006/london/record.pdf`, with a current prefix of `/usr/archive` which can be subsequently moved to `C:\Documents and Settings\Archive\` with minimal effort)
Questions:	• Does repository apply its own identifiers or maintain existing ones for

	information packages? ● Under what circumstances could ID collision occur? ● Is a bespoke or off-the-shelf (e.g. Handle, DOI, PURL) identifier scheme employed? ● Are third party resolver services required? ● What overt costs are associated with applying or resolving identifiers? ● How might the identifier become divorced from the related object? ● What redundancy is employed to maintain referential integrity?
Risk:	● Identifier to information referential integrity is compromised. ● It becomes impossible to associate identifiers and information.
Problem Scenarios:	● Repository maintains the use of the file path from the digital object's original environment as the identifier for the archived object, meaning that two distinct objects that originated from different locations share an identifier `/usr/archive/2006/report.pdf`. ● Identifier consists of the time stamp at the point of ingest, but two ingest systems operate simultaneously and duplicate identifiers are consequently applied. ● Archive is migrated to an alternative file system and paths listed within the database are no longer current, resulting in loss of referential integrity; ○ For example, a database record indicates that an object with the unique ID `#123` corresponds to location `/home/archive/report.pdf` on UNIX but it is subsequently moved to `c:\archive\report.pdf` on a Microsoft Windows server, invalidating the stored reference.

Shortcomings in the research methodology

Each of the identifiable shortcomings within these six audits to some extent corresponds to the fact that these were pilot studies, with dual (and perhaps not completely compatible) aims. The evaluation of the audit process was as high a priority as the assessment of the repositories themselves; one might legitimately claim that at times insufficient resource was allocated to fully exploring the evidential basis upon which compliance was determined. It could also be argued in some cases that a disproportionate level of consideration was given to ambiguous responses and concerns relating to the check-list because they indicated shortcomings in the audit methodology. Another criticism reflects the rather non-representative 'balance of power' that existed between auditors and hosts. It would be fair to say that poker for match-sticks and poker for dollars are two completely different games. Similarly, within an established certification environment the incentives for success would be considerable, and auditors would command considerable power to compel repository staff to participate in interviews and provide comprehensive documentation on request. Within the pilot context though auditors were understandably grateful for simply being accommodated and were consequently less inclined to be too disruptive in their activities and demands. Every effort was made to mitigate this latter issue by briefing repositories on their expected responsibilities and arranging a convenient time to minimise disruption, but inevitably, the process fell short of being completely representative of how it would really take place within a formally established certification infrastructure.

From Investigation to Analysis to Assessment

The specific value of individual evidence examples will vary according to the particular aspect of the archive that is being assessed, and this is one of the reasons why evidence impact judgements must be determined in a range of alternative contexts. Given the parameters within which the RLG/NARA check-list operates, these contexts are finite, and one can, with a reasonable degree of success, determine the value of alternative kinds of evidence in each. Before proceeding to define these contexts, and exploring in more depth the types of evidence that are available, it is worth considering why auditors bother to gather such information. There are essentially three questions that evidence must be accumulated to satisfactorily answer, in order to support audit and certification. From one to three these are increasingly enlightening from the perspective of auditor, in his quest to determine truth, but might be said to be inversely useful when attempting to justify an award of certification to a wider audience:

1.What is documented? (high certification value, low truth evidence)

2.What do staff members or other stakeholders believe, think or know happens within the archive?

3.What actually happens within the archive? (high truth value, low certification value evidence)

Without doubt, the initial role of evidence is to set the scene, to enable auditors to build a sufficiently complete picture of the archive to facilitate further, more targeted enquiry. The evidence base that fulfils this role can vary in its form and still maintain its usefulness. Needless to say, the extent to which initial evidence (i.e. that received and analysed prior to the on-site audit activities) is comprehensive will significantly influence the ease with which subsequent on-site analysis, comparison and corroboration can take place. In most cases it is likely that such evidence will be presented as documentation, and generally this will be most useful, providing auditors with the opportunity for detailed study and subsequent reference. In addition, since it is likely to offer a more global (as opposed to individualistic) view of the archive, in contrast with personal testimonial for instance, it is likely to provide a more objective bedrock upon which further investigation can be constructed.

The three questions above broadly correspond to the information gathering work-flow that takes place within the audit process. The question of what is documented is *ex facie* the most straightforward of the three to determine, assuming that adequate document retrieval powers are conferred upon auditors and/or participating repositories are sufficiently forthcoming. Given that the audit process is primarily concerned with identifying trustworthy repositories, transparency is of key significance. Documentation is therefore essential, and ideally will be available not just to auditors, but to every interested stakeholder. Its absence, while not directly condemning the archive to organisational failure, is likely to preclude the award of certification. With respect to several of the RLG/NARA check-list's metrics, the existence of appropriate documentation will provide independent indication of conformity, with little need for recourse to alternative evidential means. In these admittedly limited cases, the content of a relevant document will be of less significance than its very existence. The communication of staff or stakeholder perceptions, beliefs and understanding logically follows on from the accumulation and absorption of documentary evidence. Primary testimony is a useful means of corroborating that written policies, procedures and practices are well understood and representative, determining that staff and stakeholders are aware of the extent of functionality and services offered by the repository, illustrating the level of intrinsic transparency. The identification of a critical mass of stakeholders or staff whose views, beliefs or understanding differs markedly from that which is documented will take precedence and carry greater evidential impact. At least half of all interviewees will have to provide accounts that contradict documentary evidence in order to compel auditors to

immediately label the corresponding documentation as non-definitive. Significant, but lesser degrees of dissent ought to motivate further interviews, or, where it is feasible, provoke an auditor to seek a conclusive practical demonstration. Numerous check-list criteria can be satisfied with little more than first hand (primary) testimony, supported by adequate documentation. Testimony that is unsupported by documentation will be persuasive only to the extent to which it is corroborated by other means, whether by interview with alternative staff or stakeholders or through first hand auditor observation. This observation of practice is likely to carry the greatest evidential impact, although needless to say, is suited to only those check-list metrics that allude to a demonstrable procedure.

Simply witnessing the repository function in accordance with check-list metrics is sufficient to prove its capacity to conform, but not that such behaviour is firmly instantiated within the very fabric of the repository's infrastructure. Both interview and observation evidence offer a more or less transitory view of the world; certification demands persistence, sustainability and the presence of policies, skills, techniques and functionality that are sustainable, assured and persistent. Only the generation and management of comprehensive documentation provide this, documentation's value distanced from contemporaneous illustrations of conformity such as the views of the current staff members or the way things work today. To this extent then, the attainment of trustworthy status seems viable only for those repositories whose evidence is sufficiently intrinsic that it won't be altered by the loss of one or more members of staff. It is argued that certification could be awarded based on any combination of documentary evidence plus one or both of interview evidence and observation. However, in the absence of documentation, although organisational and functional competence might be determinable, it is unlikely that certification could be awarded.

Relieving Preservation Pressure

Preservation pressure points exist throughout repositories, and this is broadly reflected in the RLG/NARA check-list criteria. Preservation can be considered as the careful navigation of a range of risks, challenges, and potentially damaging contextual factors, which must be managed in a considered and expert fashion. These numerous pitfalls threaten to disrupt sustainability and even more destructively, they threaten the trustworthiness attributed by stakeholders to the repository. The availability of appropriate evidence will relieve this pressure, contributing towards the repository's success in recognised archival assessments and providing a transparent degree of reassurance of the repository's commitment, aptitude and sustainability. As noted above, previous work has sought to define some of the individual types of evidence that might play a crucial role within these audits. Formal categories of evidence can be broadly distinguished as one of documentary, testimonial or observation of practice evidence, and these will be usefully applicable in a range of circumstances. Less formal types can provide contextual framing, but are less likely to contribute to the certification process. These include anecdotal evidence, hearsay, or historical experiences that can be expressed only in vague terms (e.g "we've been in business for twenty years, we must be doing something correctly...").

Admissible types of evidence are documented below, with the check-list compliance pressures that each are ideally suited to alleviating.

- **Formal Primary Documentation:** This certainly represents the most useful kind of evidence, although on its own might not be completely conclusive. There are very few examples of check-list criteria compliance that can't be demonstrated at least to some extent with the provision of primary documentation. The scope of understanding of primary documentation is broad, and need not include just paper records, but also on-line content (such as web pages or wikis) and even object or repository metadata.

Irrespective of its form it is vital that auditors identify an organisational commitment to the ongoing management and retention of all documentation. Several RLG/NARA criteria relate specifically to the availability of documentation, and therefore these are satisfied simply by providing acceptable examples that correspond to the appropriate metric. Examples include the requirements to have a mission statement with an appropriate commitment to the long term, to retain appropriate metadata and to document formally the scope of the archive's designated community, their expectations, knowledge base and requirements. Similarly, the provision of comprehensive accounting documentation provides a fairly definitive insight into the repository's financial transparency and their ability to make ends meet. These can be classified as *documentation-based* pressures.

Notwithstanding the example presented above, other aspects of the check-list's financial demands relate less to the repository's means, and more to their commitment to fulfil responsibilities such as the review of business plans and providing analyses and reports on risk, benefit, investment and expenditure. These are just some of the numerous *commitment-based* pressures prevalent within the check-list's metric. Documentation alone is often insufficient to alleviate such pressures, although if auditors can identify the allocation of significant resources towards meeting these commitments this can be persuasive.

The third type of preservation pressures alleviated by the availability of primary documentation relate to the repository's ability to fulfil particular roles or provide particular services. These *capacity-based* pressures are prevalent within the check-list, especially within its second section, which deals with the digital object management activities undertaken within the repository. For example, a repository's ability to record representation information, monitor archival objects' integrity, log access management failures or authenticate the source of materials can all to some extent be illustrated by the provision of appropriate documentation that details the corresponding work flow. It is likely that further elucidation will be required, and interview or observation evidence can provide a vehicle for that.

Further pressures can be categorised as *resource-based*, and these are identified throughout the metrics. Issues such as appropriate staff numbers, appropriate hardware and software and sustainable financial support are key examples. Once more, documentation such as inventories, organisational charts and business plans will provide a considerable insight, but in such cases auditors will generally wish to witness first hand the availability of such resources.

The final set of pressures might be categorised as *planning-based*. These relate to the future; for example, are appropriate disaster recovery or contingency mechanisms in place? Documentation again offers a near complete insight into such factors, but in many cases demonstration (for instance the recovery of 'lost' content from backup) will be necessary to convince auditors of the effectiveness of these plans; their existence alone is likely to be insufficient.

While observation of practice presents the auditor with an unambiguous and authoritative view of the state of the world, primary documentation carries a greater evidential weight. This corresponds to a strong distinction that exists between establishing truth of conformity (the initial role of the auditor) and expressing the fact that there is conformity, or legitimising such an expression (the role of certifier). For certification to be feasible, the documentation must tell a conclusive story. If for example technical work flow manuals describe a process that in practice is undertaken quite differently, then the documentation would have to be changed to reflect reality. The auditor can determine truth but it is inconceivable that they can communicate this

to a wider world without the availability of compatible documentation.

Needless to say, several check-list metrics do not fit comfortably into a single category, and some overlap is expected. Each identified pressure is to some extent contingent on other contextual pressures, and therefore one might legitimately suggest that capacity, resource and planning are necessarily related. Nonetheless, they provide a useful initial insight into the kinds of evidence that we might reach for when assessing particular aspects of a repository infrastructure.

- **Unsubstantiated Stakeholder Testimony:** Unlikely to alleviate pressure in many of the categories detailed above, unsupported testimony is that which lacks corroboration from fellow interviewees, documentation or visible procedure. In this sense then it is quite devoid of formal value, although is valuable as a basis for further investigation. Subsequent investigation will either elevate this testimony to corroborated status or result in its rejection as non-representative. The exception is in terms of resolving commitment-based pressures. These commitments, while generally identified at repository level, can also be determined within the will of individuals, particularly repository management. Unlikely to be conclusive, and necessarily transitory, such testimony is nevertheless to some degree compelling. In the event of contradiction by any other evidence type unsubstantiated stakeholder testimony will in most cases be immediately discarded.
- **Formal Secondary Documentation:** Perhaps the most persuasive second-hand documentation is in the form of reports from prior audits or certification that has already been awarded (or withheld). In such circumstances the range of usefulness is largely dependent upon the aspect of the repository that has undergone assessment, and weight of legitimacy with which the process was undertaken. The latter will be influenced by the renown or level of community acceptance of the auditing organisation and the status of any intellectual foundation upon which the audit was based. Audits conducted by neutral well regarded organisations based on international standards are likely to be particularly persuasive. Conversely, those carried out by local companies and based on internally conceived and undisclosed criteria will be less useful. The check-list does require that repositories commit to formal, periodic review, and therefore with respect to that specific metric, audit reports become primary documentation. Comparable rules will exist for any other secondary evidence supplied in support of certification; individual assessments based on each document's origins and originator ought to be made.

  Secondary documentation has some value in dealing with resource-based pressures; the RLG/NARA check-list demands an appropriate number of staff to support all functions and services. Documentation from comparable archives can offer insights into what ought to be in place. In terms of its evidential impact, secondary documentation will always be superseded by that identified internally during the course of the current assessment.
- **Corroborated Stakeholder Testimony:** The extent to which testimony is corroborated is an influential factor in determining its evidential impact. If several interviewees describe a common world view it immediately implies a degree of credibility. Commitment-based, resource-based, capacity-based and planning-based pressures can all be alleviated to a lesser or greater extent with this valuable form of evidence. In reality, it is likely to represent the most substantial quantity of admissible evidence throughout the process; in a successful audit it will build upon a solid base of documentary evidence, with any remaining doubts satisfied by subsequent demonstration. Its weight is significant and it is compelling; any deviation from documented evidence will not benefit the certification process and may necessitate the

rewriting of contradictory documentation. From the perspective of auditor as truth gatherer, it is difficult to contest the fact that significant corroboration is likely to provide a realistic view of the world as it actually exists.

- **Supervised Practical Demonstration:** Where audited subjects themselves present and direct the demonstration of systems or processes the value of the process is to some extent limited. Known problems may be avoided, and unique and non representative strengths (for example, extraordinarily extensive metadata records for particular objects) might be presented as typical. Nevertheless, in order to alleviate both capacity-based and resource-based pressures such demonstrations have an intrinsic value. It is difficult to mask even reasonably significant system shortcomings or promote that which is atypical during a reasonably in-depth demonstration, and for this reason the results are quite compelling. In terms of their evidential weight, these usurp both testimony and documentation as representations of what really happens, if only in terms of the specific practical issue being presented.
- **Auditor's Personal Interaction or Auditor-Directed Demonstration:** Sharing the strengths of a supervised demo, this has none of its intrinsic weakness. An ideal methodology for assessing practical tools or determining the comprehensive completion of documented work flow this must be regarded as conclusive, superseding both testimony and documentation. Documentation will only carry greater weight in the event of temporary system problems – error status documentation will limit the inferences that auditors should make in these circumstances.

Technical barriers to evidence retrieval

As a footnote to this work, it is essential to bear in mind that evidence retrieval is fraught with numerous difficulties. The most obvious is the reluctance of repository staff to fully cooperate with the process, and their consequent resistance to the release of documentation, disclosure of information in interviews or provision of auditor access to systems, whether due to reasons of sensitivity or otherwise. This is in fact rather simple to alleviate, by conferring upon auditors appropriate and sufficient document retrieval powers, and providing sufficient incentives to repositories to seek certification that they will be completely cooperative. Other barriers are less easily overcome however, particularly when the context for such work is an international one. Documentation is likely to be in native languages, which can represent a significant barrier. Similarly the role of auditor demands considerably diverse skills and knowledge: in most instances, given the breadth of coverage these audits demand, from the point of view of any single auditor, some aspects of documentation, whether technical, financial, legal or archival might just as well be in another language. These are issues that are likely to be increasingly prevalent as this work continues, and to dwell on just the value of evidence is to omit a potentially wider problem.

Conclusion

Evidence is increasingly being acknowledged as being at the heart of the repository audit process, but still little formal understanding exists to determine the means by which it may be combined, collected and deployed. This paper represents a starting point to overcoming that problem. To date, the DCC and CRL audits have been structured to correspond closely to the sections of the RLG/NARA check-list, and this has been a successful means of assessment. However, in addition to this, it is suggested that auditors should bear in mind the kinds of pressures that their efforts are seeking to alleviate; these are directly traceable to the kinds of evidence that will demonstrate check-list compliance in those particular areas. Repositories that are confident of being certifiable will be capable of providing a diverse range of evidence examples in order to support the process. It is vital that

some additional lines are drawn on the audit and certification playing fields to determine how this might be used, interpreted and combined in an optimal, and truly meaningful fashion.