2008 Annual Conference of CIDOC
Athens, September 15 – 18, 2008
Andrew Mchugh, Perla Innocenti, Seamus Ross

ASSESSING RISKS TO DIGITAL CULTURAL HERITAGE WITH DRAMBORA

Andrew McHugh, Perla Innocenti, Seamus Ross
HATII at the University of Glasgow
11 University Gardens, Glasgow G12 8QQ
Scotland
UK
{ a.mchugh, p.innocenti, s.ross}@hatii.arts.gla.ac.uk

Abstract

With the goal to provide a practical, evidence-based toolkit for assessing repositories and digital libraries, the UK Digital Curation Centre (DCC) and DigitalPreservationEurope (DPE) jointly developed the Digital Repository Audit Method Based on Risk Assessment (DRAMBORA). The toolkit adopts a bottom-up approach that takes risk and risk management as its principle means for determining repository success and charting improvement. This paper introduces the methodology as well as its associated online tool, DRAMBORA Interactive, and describes their broad, but flexible applicability, describing the pressing need for tangible assessment methodologies. Its coverage includes a description of experiences accumulated and lessons learned from the series of pilot assessment programmes that have made possible the development, validation and evolution of the methodology. DRAMBORA draws on experiences accumulated throughout 18 evaluative pilot assessments undertaken in an internationally diverse selection of data repositories (including the National Archives of both the Netherlands and Scotland, Gallica at the National Library of France, CERN's Document Server, the Netarkivet, and the Google MBooks project at the University of Michigan).

Introduction to the Digital Curation Challenge

The curation of digitally encoded content has become a fundamental priority within the cultural heritage community, as researchers, curators and consumers increasingly rely on technological means for content delivery and reproduction. The most natural responses to the unerring transition towards an increasingly digital world view focus on the implicit opportunities. Broader and more interactive access to all kinds of information assets, additional possibilities for information analysis, filtering and discovery, and widespread empowerment of information consumers are natural benefits of the digital

platform. However, there are a number of equally notable challenges posed by the wholesale adoption of digital technologies, and on our increasing reliance on data availability.

Put simply, our digital data is at risk, from the very moment of its creation. Numerous influences threaten the integrity of digital data in the face of an uncertain future. Physical deterioration over time is a familiar concept from traditional analogue archiving environments. The heightened complexity implicit within most contemporary digital storage is accompanied by a commensurate increase in the likelihood of hardware and media failure. The problem is further compounded as even isolated errors occurring on digital platforms or within encoded materials can restrict access to disproportionate quantities of information or at worst render content completely inaccessible. The fundamental vulnerability of data extends beyond comparatively trivial questions of carrier integrity. Unlike in the analogue world, data require appropriate software, hardware and semantic mediators to enable their comprehension, digestion or manipulation. Data obsolescence describes situations where cultural or technological trends have limited their availability. Numerous circumstantial triggers can provoke fears of information obsolescence, restricting the ease with which it can be accessed and limiting the likelihood of its ongoing survival. In the absence of appropriate contingencies, technological, social, organisational and political changes pose considerable threats to data availability, integrity, authenticity and usability. Vendors may discontinue support for legacy formats or media within their latest software or hardware releases; cultural trends may attract users away from traditional platforms, lessening their ubiquity and limiting the likelihood of their ongoing accessibility; intellectual property or licensing restrictions may limit the circumstances within which data can be accessed; knowledge upon which information interpretation is dependent may be lost; and diminishing organisational commitment to data management can threaten its availability.

CULTURAL HERITAGE CURATION CHALLENGES

Fundamentally, one must acknowledge that the challenges facing our digital cultural heritage are seldom solely technical. Bit-level preservation describes those activities aimed at maintaining the integrity of the binary streams that represent the physical reality of our digital files. Our collective understanding of information security issues has reached a considerable level of maturity, and the challenges associated with bit-level preservation are now regarded as rather trivial. *Information*

preservation is much more difficult; can we maintain the *understandability* of our digital assets, whatever that may mean? How, in a highly dynamic digital context, can we continue to ensure that our digital data can be rendered, processed, searched, or computed? How, when information exchange possibilities might result in data sharing by highly diverse and disassociated communities (including those separated by a temporal dimension), can we be confident of maintaining consistent semantic interpretability and structural authenticity? How, when the commercial realities of information technology favour proprietary solutions and digital rights management provisions, can we ensure the continued usability of digital materials long after software vendors, service providers or facilitators have ceased trading or discontinued support for legacy products. Given the realities of data value, where even the most well established data creators and curators are likely to be survived by the value of their data, who can we charge with the responsibility of providing stewardship for our often invaluable information? Will Flickr.com still exist in one hundred years' time? Many of us have access to family photographs dated from over a century ago. In many cases their relatively untarnished form suggests that they will easily last as long again.

The experience of the Arts and Humanities Data Service (AHDS), until recently funded by the Arts and Humanities Research Council (AHRC) to provide archival stewardship for data generated from Arts and Humanities research in the UK provides a stark, and compelling insight into the organisational frailty that can fatally undermine information sustainability. Stripped of its funding in late 2007, the AHDS was forced to acknowledge its own finite status. Needless to say, a considerable proportion of its data retains considerable value, and questions surrounding the identity of those to whom archival responsibilities will be conferred continue to persist. Some suggest that accountability will be federated to individual institutional repository environments. If this expectation reflects the intentions of the AHRC, then it is essential that infrastructures exist to determine whether or not specific repositories are adequately positioned to meet the many challenges implicit in digital preservation.

BENEFITS OF A SEMI-CENTRALISED APPROACH TO INFORMATION MANAGEMENT

Most contemporary mainstream curation strategies demand a highly active approach to the associated

challenges, and a keen awareness of the emergence of threatening contextual factors. The environment within which preservation is undertaken, and the policies and processes to which encoded digital information is exposed are assumed to be of critical importance. Digital repositories have emerged with explicit preservation mandates, environments within which data can enjoy safe stewardship, designed to minimise their exposure to the problems that are likely to compromise their value. The mooted transfer of preservation responsibility from the nationally focused AHDS to regional universities and research centres can of course be considered an act of decentralisation. But nevertheless, the concept of digital repositories is predicated upon the benefits of an at least semi-centralised approach to preservation. Irrespective of whether individual discrete repositories are structured and perimeterised according to institution, discipline or data type, these are environments intended to confront challenges that would otherwise be met by information creators and consumers. Information management generally, and preservation in particular, demands considerable resource availability; preservation's most fundamental dependency is money. Organisational infrastructures designed to exhibit sustainability are theoretically better equipped to command ongoing financial commitment. Furthermore, given the perceptible global shortfall of digital preservation competencies, centralisation of skills makes sense. Implicitly, repositories must be capable of demonstrating competencies that reflect their own purpose; where preservation is a priority then repository administrators, staff and systems must be equipped to manage each of the contextual and object specific factors that influence data availability.

The value of digital information, within the cultural heritage environment and beyond, is potentially limitless, and realised in a broad variety of ways. The digital environment facilitates complex information analysis, offering means for legitimising research conclusions and empowering scholars to challenge, validate or relate results. It can offer new opportunities for distribution and interactivity, where multiple sources can influence the physicality of information content and context. Digital enables the deployment of novel content with no feasible analogue within more traditional environments. Digital data boasts an implicit flexibility, and means are required to enable varied, perhaps unanticipated reuse of digital materials, at an undetermined point in the future. Given the uncertainty that characterises that which is yet to come, the curation problem may be one of near limitless proportion. The manifestations of data value are almost limitless whether scholarly, evidential, creative, economic, or otherwise, and having acknowledged their potential value, there is a natural will

to seek means ensure persistent data availability.

A COMPETITIVE CURATION ENVIRONMENT

Curation is a challenging business, as indicated above, but is further complicated by the recent emergence of numerous diverse prospective solutions, service providers and best practice resources. Stewardship responsibilities are increasingly being embraced by a range of organisations, either in response to escalating rates of data production (the 'data deluge' [11]), perceived commercial opportunities in providing preservation services, or shortfalls caused by the dissolution of more established, highly centralised preservation environments (such as the AHDS). But the effort expended procuring, installing and running repository software and claiming stewardship over a collection of digital assets is dwarfed by the fundamental challenges implicit in providing long term, sustainable preservation services.

The choices for those seeking to ensure that their content remains available (and seeking to exploit the value of preserved materials) are more numerous than ever. Amid such a diverse range of apparently viable preservation options, it is vital that choices are well informed, and based on the results of formal assessment. Numerous stakeholders have an interest in identifying the extent to which particular preservation services are successful. The most obvious group consists of those seeking to actually provide such services. The preservation discipline continues to lack maturity; the uncertainty that characterises preservation over time is reflected within the efforts of many repository practitioners. Notions of best practice for information archiving equate to little more than best guess, such is the unpredictable technological, social and cultural landscape that awaits us in even the relatively short-term future. Doubts are a natural consequence and many preservation practitioners have felt compelled to seek assurances of the suitability, comprehensiveness and sustainability of their efforts. Of critical importance is the repository's ability to demonstrably improve over time, and exhibit sufficient transparency and responsiveness to minimise the threats posed to collection sustainability. Beyond the preservation environment, many other stakeholders seek similar assurances; information creators, owners and consumers each legitimately demand insights into the extent to which preservation repositories are capable of meeting their stated missions. Their agreement to deposit valued content or trust the integrity and authenticity of accessed materials will be contingent on demonstrably adequate

performance levels. Assessment results will be of similar interest to repository financiers, both with respect to those repositories already directly funded, and to alternative infrastructures that might merit cash injection or are likely to inherit responsibilities in the event of existing repository services being decommissioned.

OUTCOMES OF THE ASSESSMENT PROCESS

Numerous outcomes of repository assessment have been mooted, with both local and more globally resounding impact, accompanied by a sliding scale of complexity in terms of infrastructural dependencies. *Audit* and *Certification* are two terms that have been frequently related, as if inseparable and synonymous. In fact, the two terms describe individual discrete activities, both with highly specific demands. The issue of certificates to corroborate organisational success or compliance is commonplace within many disciplines, from information security to environmental health and safety. The intention is primarily to engender trust; if organisations are certified by a trustworthy accredited body, according to agreed and objective criteria, then stakeholders are empowered to compare alternatives or favour individual examples. Depending on the stakeholder, within the repository context that may mean choosing to deposit or subscribe to content or to contribute to their economic foundations. There are three implicit requirements for certification to be successful, three pivotal mechanisms that must be available as a precursor to its viability. The first is an appropriate set of criteria; these represent the definition of success, and must be evident within certified repositories. Such criteria already exist in the shape of several repository check-list documents, such as the *Trustworthy Repository Audit and Certification Criteria and Checklist*, and the nestor Catalogue of Criteria for Trusted Digital Repositories, both described in more detail below. The second dependency is an accredited individual or organisation that enjoys sufficient widespread esteem to elicit community respect and provoke minimal dissent with its awards of certification. It must be permanent, or at least expected to outlive the organisations subject to its judgements. Given the scant regard shown to political national boundaries by digital materials on the Internet, it should ideally be politically neutral, and international. Finally, it must be sufficiently well resourced to absorb liabilities or challenges to its decisions to award or withhold certification. The organisation that might occupy this role is currently far from clear; few if any bodies within the preservation community appear to qualify, and it has been suggested by some commentators that only by engaging with public, intercontinental organisations, such as the European

Union, can certification really be viable. Although the identities of potential certifiers remain clouded in uncertainty, steps are being taken to legitimise more broadly the metrics that they might base their conclusions upon. A CCSDS Working Group is currently being assembled to take forward work to standardise existing objective metrics within ISO. Other disciplinary areas have seen the emergence of profitable commercial certification agencies following the publication of similar standards, and this may be a valuable consequence within the repository context.

A third dependency for successful certification services is a recognised formal means for undertaking repository assessment, in order to determine where conformity does and does not exist. This defines the main flesh of the 'audit', and although not intrinsic within existing certification metrics, is a necessary precursor, in order to ensure the validity and reproducibility of their application. What must be noted is that audit can be of tremendous value even without subsequent certification. The concluding part of the process need not be the award or otherwise of endorsements or plaudits. The ability to identify shortcomings and monitor ongoing improvement is in and of itself of potentially tremendous benefit. A focused assessment, reflecting the specific realities of individual environments need not be structured solely to satisfy the weight of comparison with other discrete repositories. At this time, given the practical difficulties associated with the award of globally meaningful certification, and notwithstanding the fact that primary beneficiaries are the repositories themselves, this is arguably a much more worthwhile aim to pursue. The *Digital Repository Audit Method Based on Risk Assessment* (*DRAMBORA*), released by the UK *Digital Curation Centre* and *DigitalPreservationEurope* provides a formal method for conducting repository self assessment, with this express aim [3].

DEFINING A COMMON CONSENSUS FOR REPOSITORY ASSESSMENT

As hinted at above, a number of mainstream reference materials are now available to support the assessment of digital repository environments. Considerable work has been undertaken to develop audit check-lists that will eventually provide an intellectual basis for awarding certification to sufficiently capable repository service providers. There are two principle examples currently available. Released in 2007, the *Trustworthy Repositories Audit and Certification* (TRAC) *Criteria and Check-list* [1] was developed by a consortium jointly overseen by the US *National Archives and Records Administration* and the *Research Libraries Group* (prior to its absorption within OCLC), and is now

maintained by the *Center for Research Libraries*. *TRAC* describes approximately ninety characteristics that must be demonstrable by repositories that aspire to a certifiable, trustworthy status. The second example, also released last year, adopts a more regionally specific focus. The *nestor Catalogue of Criteria for Trusted Digital Repositories* [2] was developed in Germany by the *Network of expertise in Digital long-term preservation* (*nestor*). Structured similarly to the *TRAC* document, this provides examples and perspectives that are more representative of a German operational, legal and economic context.

More focused on the mechanics of audit than in the determination of objective certification criteria, the *Digital Repository Audit Method Based on Risk Assessment* (*DRAMBORA*) was released by the *Digital Curation Centre* and *DigitalPreservationEurope* in a document form in 2007 and as an interactive online tool in early 2008. It adopts a bottom up approach, enabling repositories to relate explicitly their benchmarks for success to their own aims and contextual environment. Capable of being used both independently and in association with more objective guidelines, *DRAMBORA* describes a formalized process that encourages repositories to consider and document their mission, objectives, constraints and activities, before deriving, expressing and planning to address the fundamental challenges that threaten overall success.

The developers of each of these resources met in early 2007 with a view to formalizing the repository problem space, in order to maintain compatibility, and facilitate the comparison of their respective results. Despite the difficulties associated with determining an objective and universally reflective perspective of ‘digital repositories’ the benefits in undertaking this exercise were clear. An accepted understanding of what digital repositories actually are is a necessary precursor to any work that seeks to determine their effectiveness. Adopting a broad view that echoes the work undertaken by RLG/OCLC in their seminal 2002 “*Trusted Digital Repositories – Attributes and Responsibilities*” [4], ten general principles of repositories [5] were conceived, capable of encapsulating all the organizations and organizational components that might be subject to assessment using the three tools. In isolation, the list of principles is insufficient to support assessment but provides a structure that informs the processes and outcomes of *TRAC*, *nestor* and *DRAMBORA*, and contributes to their compatibility.

Irrespective of environment, discipline or specific priorities, any repository seeking to use these tools

ought to have provisions to demonstrate the following broad characteristics:

1. Mandate & Commitment to Digital Object Maintenance;
2. Organizational Fitness;
3. Legal & Regulatory Legitimacy;
4. Efficient & Effective Policies;
5. Adequate Technical Infrastructure;
6. Acquisition & Ingest;
7. Preservation of Digital Object Integrity, Authenticity & Usability;
8. Metadata Management & Audit Trails;
9. Data Dissemination;
10. Preservation Planning & Action;

Clearly the coverage of these extends beyond just technology: issues of organizational competence, legal legitimacy and adequacy of policies are all similarly prioritized. From an object management perspective, mappings can be identified between the principles' explicit requirements with the functional model described in the Reference Model for an Open Archival Information System [6].

OBJECTIVE MECHANISMS FOR REPOSITORY ASSESSMENT

Both *TRAC* and *nestor* reflect a top-down assessment philosophy. Both seek to define an objective consensus of the priorities and responsibilities that should exist within any repository environment (albeit, in *nestor*'s case, mainly limited to Germany). To adopt *only* this perspective is to some extent dangerous, since it implicitly disregards the great variety visible across contemporary digital repository platforms. Diversity in terms of funding, scale, legislative responsibilities and restrictions, content types, technology, and policy are identifiable in even a localized sample. The preservation of cultural heritage materials presents a number of varied challenges. A considerable proportion of these will be generic, and prioritised similarly in alternative disciplines. However, there are other, often critical issues that are uniquely prevalent for cultural heritage preservation that must be taken into consideration when determining success. Given this practical reality, generically defined criteria are difficult to conceive; if expressed too vaguely they tend to lack meaning, but if too specific will be

rendered irrelevant for a significant proportion of potential users. Feedback from the repository community has demonstrated concerns. Although both of these check-lists were developed by diversely assembled individuals committed to *reflecting* existing good practice (and not to mandate novel or theoretical approaches to preservation), complaints of non-applicability have been evident. In several cases this reflects short-sightedness on the behalf of those working within the repositories; criteria have been painstakingly phrased to maximise their flexibility, and facilitate optimal general applicability. But nevertheless, it is evident that within the community there is the need for a more tailored assessment solution that takes into account atypical repository qualities, as either a companion piece, or alternative, to the existing guidelines.

The most fundamental problem with an objective approach is the implicit assumption that all repositories share a singularity of purpose, and that their priorities are uniform, irrespective of where or why they exist. But the diversity evident within repositories is also identifiable in the ways that success can be demonstrably realized. Listing blue-sky criteria for digital repositories is a valuable process; *TRAC* and *nestor* are both compelling reference materials, selection boxes for organizations seeking to develop new repository features, or to subject their existing infrastructures to gap analyses. However, both of these criteria check-lists are expressed in necessarily vague terms, and it is therefore quite challenging from the perspective of repository practitioners to understand how conformity might be adequately measured. The process of certification is well served by documents such as *TRAC* and *nestor*. The conferment of a universally acknowledged recognition of success presupposes the availability of an objective benchmarking mechanism. One cannot compare apples to oranges, and similarly a certification process that is based variably upon the specific issues associated with individual repositories would immediately sacrifice its weight of legitimacy. The discussion of whether or not certification is indeed a high priority within the preservation community is not the primary focus of this discussion – although there are significant questions to resolve. Nevertheless, the most compelling benefits of certification demand comparability of results to enable an objective view of individual repositories' successes in a wider context.

Best practice guidelines and check-lists provide an undoubtedly useful intellectual foundation upon which to construct an audit, but in their current form, neither *TRAC* nor *nestor*'s documents provide, in explicit or implicit terms, a sufficiently tangible structure for determining where conformity and

success actually exist. Neither offers sufficiently detailed insights into the mechanics of the audit. Which individuals should be involved? What questions should be posed? How should experimental evaluation of systems be conducted? What are the quantitative or qualitative evidence expectations that will adequately demonstrate sufficient check-list compliance?

BOTTOM-UP REPOSITORY ASSESSMENT METHODOLOGIES – DRAMBORA

DRAMBORA adopts a contrasting, but ultimately compatible approach, intended to overcome many of the problems associated more objective, and not completely reflective criteria. It adopts a bottom up approach, enabling repositories to relate their benchmarks for success more explicitly to their own aims and contextual environment. Capable of being used both independently and in association with more objective guidelines, *DRAMBORA* describes a formalized process that encourages repositories to consider and document their mission, objectives, constraints and activities, before deriving, expressing and planning to address the fundamental challenges that threaten overall success.

Its development was prompted in 2006 by fears that there had been insufficient commitment to define formally the practical process of analysis that would determine where success fundamentally lay, irrespective of the specific purpose of the preservation environment being exposed to scrutiny.

The starting point for the work was a series of pilot repository assessments undertaken by the Digital Curation Centre in 2006 and 2007. The selected participants demonstrated considerable diversity, with repositories exhibiting a range of varied characteristics. A signficant initial priority was to explore concerns surrounding the adequacy of a one-size-fits-all solution, deployed in isolation. 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 cultural heritage focused *Beazley Archive* (BA) at the *University of Oxford* were among those that kindly agreed to take part. As well as providing the participating organizations with an objective and expert insight into the effectiveness of their operations, and determining the robustness and global applicability of existing draft criteria, it was

acknowledged quickly that the pilots provided an excellent opportunity to explore in detail the optimal means for *conducting* assessment. The focus evolved, until the main research priority became securing a comprehensive understanding of how evidence is practically accumulated, assessed, used and discarded throughout the audit process. Research yielded numerous conclusions about the ways in which practical, objective sense can be made of the potentially limitless evidence types that might be submitted in support of certification, and of ways to classify evidence examples according to origins, form and weight of legitimacy [7]. Regularizing disparate evidence equips the auditor to effectively cross-compare, corroborate and priorities the full range of proof and testimony that is provided throughout a repository's bid for certification.

A methodology for performing repository audit was quickly established and subject to considerable subsequent refinement. In March 2007 the process was formalized as *DRAMBORA*. The methodology itself is flexible, and responsive to the structural and contextual peculiarities of individual repositories, its metric for success directly linked with repositories' own aims. More objective guidance materials such as *TRAC* and *nestor* (as well as other domain-neutral criteria) can, and should be used in combination, in order to inform the process, and prompt analysis of particular issues, but no criteria are considered mandatory. Consisting of two discrete primary phases, the *DRAMBORA* process places considerable emphasis on demonstrable, and not just inferred, success. The initial phase is a process of information accumulation, aggregation and documentation. Numerous responses must be provided to describe in detailed terms the repository's strategic purpose, its action planning, and any contextual factors that influence or limit its ability to meet its objectives. This is a detailed and highly structured scene-setting exercise. A hierarchical analysis is undertaken, beginning with consideration of the repository's mandate. This is its fundamental mission, expressed in a document, legislative instrument or policy that describes and justifies its existence and legitimizes its purpose. From this starting point, the organization is subject to increasingly focused scrutiny, requiring detailed and documented descriptions of fundamental repository objectives as well as the activities aimed at their completion and any assets, the availability of which they depend. Finally, each of the repository's contextual constraints must be made explicit. These may include legislative requirements, technological limitations, or policies resulting from strategic planning - anything that significantly influences or limits any aspect of the repository's business activities should be documented. The ten principles described above provide an intellectual infrastructure that facilitates efforts to describe, document and

relate the various responses. For example, objectives must be identified to correspond with maintaining organizational fitness, legal legitimacy, and technological adequacy as well as every aspect of digital object management workflow. The outcome of this phase is a comprehensive organizational overview, which immediately leads into the latter phase, concerned with the identification of risk.

The risk identification, assessment and management part of the *DRAMBORA* process is where conclusions are derived from the organizational picture detailed within the previous phase. Risk is utilized as a convenient means for visualizing repository success – those repositories most capable of demonstrating the adequacy of their risk management (as opposed to those facing the least number of risks) are those that can more reasonably claim a trustworthy status. Preservation is fundamentally a risk management process. Numerous uncertainties or threats relating to any number of social, semantic and technological factors are capable of inhibiting long term access to digital materials. Successful repositories are those that plan for these uncertainties, and convert them to risks that can be managed to mitigate the likelihood of problems occurring and limit their potential impact. Risks are implicit in every aspect of an organization's goals and activities, and can be borne or influenced by any number of internal or contextual factors. Perhaps most importantly, repository risk is assessed as an all-encompassing issue. In common with the ten principles, consideration must be made of not just the service-oriented procedures and policies, but also of organizational, legal, resource-related and technological risks.

Of course, one might assume that the results of such assessment will be of little value in a global sense, and will limit opportunities for repository comparison. Following the *DRAMBORA* assessment process, how, for example, can one compare two repositories with dramatically different preservation goals? In fact, to pose such a question is to misunderstand the complex realities of the digital repository landscape. ‘Digital repository’ is a convenient, broadly applicable term, that unfortunately, when subject to even gentle analysis, means very little, as evidenced by the necessarily broad ten principles described above. Repositories are now so widespread within such diverse disciplines that increasingly granular classification has become necessary. Websites, databases, CRM systems, banking software, eLearning or eResearch environments, digital libraries, blogs, wikis and even personal desktops can be feasibly described as repository environments, with identifiable mappings to the ten principles, OAIS functional model or any other defining instrument that one elects to reference. Even notwithstanding

the smaller subset of repositories that exist within the 'preservation community', or even the cultural heritage environment, there is sufficient diversity to make questions of comparability between disparate or unrelated repositories virtually moot. A compelling argument against the importance of establishing a single tier ranking system is that, given the current state of repositories, the primary value of evaluation is probably not to *sell* the repository. There is little evidence currently to suggest that certification is a particularly high priority for repository administrators, depositors, or customers. Conversely, the results are best suited to internal use, a means to facilitate the planning efforts of repository administrators and practitioners, and support sustained, structured and responsive improvement. For this reason, *DRAMBORA* is mainly deployed as a self assessment tool. In many respects, its implicit processes are indistinguishable from good repository management procedures. Repositories *should* maintain an organizational self-awareness, and continuously monitor their status, and exposure to potentially disruptive forces. Maturity modelling is at *DRAMBORA's* very heart - its cyclical nature facilitates structured evolution and ongoing improvement. Each iteration through the *DRAMBORA* process references that which has gone before. Over time, a diminishing level of risk severity illustrates repository improvement, without doubt the most fundamental prerequisite to the establishment of trustworthiness. The completion of the *DRAMBORA* audit does not result in the generation or conferment of a certificate. Repositories seeking an endorsement to place proudly on their website or a flag to wave in order to woo potential customers or funders will not find these as *explicit* outcomes of the *DRAMBORA* process. It undoubtedly equips repositories extremely well to subsequently obtain such expressions of success, if and when they become available, but the most important reward is in the streamlining and optimization of repository infrastructures.

BUILDING CONSENSUAL COMMUNITIES THROUGH AUDIT

Fundamental to *DRAMBORA's* effectiveness is its bottom-up approach; within its defined self audit process the parameters for success are associated directly with the objectives and activities of the audited repository. Similarly, specific contextual factors and constraints are considered only where they are relevant. This ensures that the results of the process are, from the participating repository's perspective, wholly applicable and immediately useful. However, this approach is not immune to criticism; as discussed above, without objective consensus on the definition of success, the comparability and reproducibility of results is lessened. This is of course tolerable; *DRAMBORA's*

primary purpose is to provoke better repository management through the results of its process. Of more immediate concern with respect to a wholly subjective approach is that the potential for repositories to improve may be limited by their own horizons. Self assessment alone can only indicate problems within the bounds of what repositories believe that they should be doing. Problems arise when organizations are oblivious to their shortcomings, or unaware of the available possibilities that they might usefully seize. How indeed can repositories comment on the likelihood or potential impact of unanticipated risks that they are yet to fall foul of? These issues have all been identified within a series of facilitated repository assessments conducted since *DRAMBORA*'s launch by *DCC* and *DPE*, and by the *DELOS Digital Preservation Cluster*. Feedback from these activities has indicated that the process of self assessment has been universally valuable for participating organizations. However, a consistent concern that has been mooted by participating repositories is that if required to conduct the process without the assistance of experienced audit facilitators, the results would have been less comprehensive. This was a problem identified prior to the first release of the *DRAMBORA* methodology, in its initial document form, and some efforts were made to alleviate its effects by incorporating a list of around eighty example risks that might be modified by repositories for inclusion in their own risk responses. This is perhaps insufficient however – the list of risks is a top-down concession within an otherwise bottom-up focused approach, and suffers from the same criticisms leveled at objective metrics in a diverse realm that are described above. Recent developments within DRAMBORA are expected to largely overcome this issue however. In early April 2008 a second version of the methodology was released as *DRAMBORA Interactive*, an online tool that offers an intuitive form based interface, peer-comparison features, sophisticated and extensible reporting mechanisms and maturity tracking. By requiring users to describe the characteristics of their own repositories the tool presents 'comparable organizations' with insights into the priorities and challenges of their peers, in order to help ensure a more comprehensive coverage. This information will form the basis for a series of repository profiles capable of encapsulating core roles, responsibilities, functions and risks for a variety of repository types. The availability of these profiles is expected to facilitate and further legitimize both repository assessment and development. Currently, repository profiling measures correspond with a number of descriptive fields already utilized within the *DigitalPreservationEurope* project's Registry of Repositories. These include:

- Institution Type;

- Country;
- Description;
- Domains and Disciplines Covered;
- Scope;
- Material Types;
- Languages;
- Technical Properties (including software);
- OAI-PMH Properties;
- Legal Properties;
- Ingest and Preservation Strategy;

By requiring repositories to define their own characteristics, the *DRAMBORA* software is able to make appropriate recommendations, based on the responses of their peers. If cultural heritage repositories in France, Germany and Belgium have each described similar European legislative requirements, and another UK based example has not done so, then the system will be capable of drawing this to their attention, in case they have omitted a significant detail from their own self assessment. The list of characteristics suggested above is unlikely to be exhaustive, and it is hoped that it can be extended in the future to enable increasingly granular and optimally meaningful repository classification. The ultimate outcome will be the evolution of an ontology of repository attributes. Some theoretical work has already indicated the feasibility of these efforts. Philosophically, the approach is an amalgam of top-down and bottom-up; to some extent suggestions that can follow based on intrinsic conclusions are prescriptive, but there is a careful acknowledgement of the specificity of individual types of repositories. The intention is always to reflect the current state of repositories, and not to mandate a classification scheme with its genesis in research theory. Community profiles that emerge will by their very nature be defined with the consent of practitioners themselves. Information creators, depositors or consumers will not select repositories based on the results of certification alone. Their first consideration will be to determine which of the available repositories appear committed to providing a service that meets their requirements and expectations. As individual classes of repository are increasingly identified and described, their common services and characteristics can be understood and ultimately subjected to comparison. *DRAMBORA* enables such classification to take place prior to and during an organizational assessment. In order for its legitimacy to be accepted, any such classification

must be representative of practice, and not prescriptive, evolving *from* the repositories themselves. *DRAMBORA* empowers repositories to define their own position within a repository landscape of potentially limitless diversity, spacing themselves in a context of comparable repositories that are, in terms of organisation, function or policy, similar. By doing so they can influence, inform and benefit from the tailored, evolved perspective of 'best practice' that exists within their particular sector of the 'repo-sphere'. No two repositories are likely to be identical, but if a repository shares insights from one repository with a comparable funding model, another preserving similar file formats, and a further example that operates within the same legislative context, the potential benefits are obvious. Cultural heritage archives are just one possible repository type that might be identifiable from the shared audit results. Other relationships based on content-neutral factors may be equally or even more compelling.

Within the context of the *DELOS Digital Preservation Cluster* four audits of Digital Library environments were undertaken, using *DRAMBORA*, with a view to determining common characteristics of Digital Library repositories, in order to facilitate both knowledge transfer and comparison. The report, due to be published imminently at the time of writing describes a range of common objectives, constraints, roles, responsibilities, activities and risks within the *University of Michigan Library's MBooks*, *CERN's Document Server*, *Gallica at the Bibliothèque Nationale de France* and the *Swedish National Library's Digital Library* [8].

CONCLUSIONS

The purpose of DRAMBORA, and more specifically DRAMBORA Interactive, is to facilitate the self audit process; to do so it must be injected with sufficient scope and functionality to direct respondents and as far as possible ensure the comprehensiveness of their responses. That it can do so by referring users to the responses provided by peer organizations is of considerable value, which will only increase as the number of respondents documenting their own repository experiences continues to grow. Either in association with objective guidelines or in isolation, *DRAMBORA* offers benefits to repositories both individually and collectively. As a means of opening lines of communication between discrete, but related repositories, *DRAMBORA* is capable of determining and disseminating expressions of both general and more specialist best practice. Categories of repositories can be constructed to reflect and inform practical realities. The cultural heritage community is just one of potentially limitless

classifications, and since none must be mutually exclusive, repositories can reflect on the relevant aspects of any peer repositories to develop a consensual definition of best practice in every aspect of repository management. In what remains an immature discipline, where the naivety and uncertainty of core practitioners remain considerable barriers to progress, the circulation of emerging insight tailored to specific priorities, context and constraints has the potential to be of tremendous benefit.

REFERENCES

[1] Center for Research Libraries and RLG OCLC Programs, (February 2007), *Trustworthy Repositories Audit & Certification (TRAC): Criteria and Checklist*, Version 1.0, http://www.crl.edu/content.asp?l1=13&l2=58&l3=162&l4=91;

[2] *nestor* Working Group, (December 2006), *Catalogue of Criteria for Trusted Digital Repositories*, Version 1 (draft for public comment), http://www.nbn-resolving.de/?urn:nbn:de:0008-2006060703;

[3] Digital Curation Centre & DigitalPreservationEurope, (March 2007), *Digital Repository Audit Method Based on Risk Assessment (DRAMBORA)*, Version 1.0, http://www.repositoryaudit.eu/ (accessed May 31 2008);

[4] RLG/OCLC Working Group on Digital Archive Attributes, (May 2002), *Trusted Digital Repositories: Attributes and Responsibilities*, http://www.oclc.org/programs/ourwork/past/trustedrep/repositories.pdf (accessed May 31 2008);

[5] CRL/OCLC/NESTOR/DCC/DPE, (January 2007), *Core Requirements for Digital Archives*, http://www.crl.edu/content.asp?l1=13&l2=58&l3=162&l4=92 (accessed May 31 2008);

[6] Consultative Committee on Space Data Systems, (2002), *Reference Model for an Open Archival Information System (OAIS) – ISO 14721*, http://public.ccsds.org/publications/archive/650x0b1.pdf (accessed May 31 2008);

[7] S. Ross & A. McHugh, (2006), *The role of evidence in establishing trust in repositories*, D-Lib, Vol 12 No 7/8, ISSN: 1082-9873. doi:10.1045.

[8] DELOS Digital Preservation Cluster, (Forthcoming 2008), *Investigation of the Potential*

[9] S. Ross & A. McHugh, (2006), *Audit and certification of digital repositories: Creating a mandate for the Digital Curation Centre*, RLG DigiNews, Vol 9 No 5, ISSN: 1093-5371.

[10] Kenneth Thibodeau (2007), *If You Build it, Will it Fly? Criteria for Success in a Digital Repository*, Journal of Digital Information, Vol. 8, No. 2, Texas Digital Library.

[11] JISC Briefing Paper, (2004), *The Data Deluge*, http://www.jisc.ac.uk/publications/publications/pub_datadeluge.aspx (accessed May 31 2008).