

Trust, but verify

Barbara Sierman

This afternoon we will have a discussion about audit & certification for which we invited several experts. In my presentation I'll attempt to pave the path for this discussion.

For several years now, different initiatives have started to deal with the concept of a trusted, or better, a trustworthy repository, the TDR. A first initiative to provide guidelines and to raise the issue of certification, was the joint publication in 2002 of RLG and OCLC of *Trusted Digital Repositories: Attributes and Responsibilities*. followed up in 2007 with the *Trustworthy Repositories Audit & Certification: Criteria and Checklist*. Based on the latter version, an initiative started to propose it for ISO standardization. To involve as many interested digital preservation people as possible, everyone could participate in this working group, and the discussions held there are still publicly available.

The proposal was offered July this year for review to CCSDS (the same body that prepared the OAIS model) before it will be send to the ISO standard committee. In the mean time the group is preparing the Requirements Handbook for auditors and initiatives have been started to do some pilot audits to see whether the draft was workable.

The German “nestor” group, the German competence centre for digital preservation, published in 2007 their Kriterienkatalog. First it was published in German for their German speaking audience but an English translation can be found on their website as Catalogue of Criteria for Trusted Digital Repositories. This document is also submitted for standardization, in this case to the German DIN standardization group.

A third initiative came from the EU founded project Digital Preservation Europe. In collaboration with the Digital Curation Center in the UK, the DRAMBORA method was created: Digital Repository Audit Method based on Risk Assessment. This method and the accompanying tool focuses on organizations willing to perform a self-assessment to get an overview of the risks in their organization. Here also a working group is preparing this method to become an ISO standard for risk assessment of record management.

The three main initiatives should be seen as complementary and they are closely cooperating together. As a result of this cooperation the *ten core principles of trust* were formulated in 2007, as a leading principle for trustworthy repositories.

Around these three initiatives some other related products were created. Just to name some examples. IBM developed their Long-Term Digital Preservation Assessment (LTDPA) tool, aimed at assessing the maturity level of Long-Term Digital Preservation archives, based on TRAC and to assist the management of the organization to draft a roadmap for action.

Related to DRAMBORA the PLATTER tool (Planning Tool for Trusted Electronic repositories) was developed to help organizations that were building up digital preservation environments, to be able to meet audit and certification requirements.

In the Netherlands the TRAC document was translated into Dutch by the Provincial Archive Inspection “Landelijk overleg Provinciale Archief Inspectie LOPAI” in an adapted version for their intended community.

DANS, the Dutch Data Archiving and Networked Services, published their Data Seal of Approval in which guidelines for data suppliers are formulated.

To summarize: the audit and certification landscape: there are at least 3 tools for TDR audit and certification, all applying to become a recognized standard.

Audit and certification are not goals in themselves. Audit and certification exercises service a purpose. Key is the word “trust”. Trust is not something an organization will acquire easily; it needs to be proved that an organization can be trusted, again and again. As a Dutch saying goes: trust comes step by step, but goes on horseback.

Who need to be convinced of the trustworthiness of a repository? The introductions of the aforementioned methods all speak about establishing trust for the stakeholders. Stakeholders could be, among others, producers, users and funding bodies.

The producers are the ones sending material to the repository (the digital archive). They require trust in the organization that will take care of their digital data. Although many memory organizations have shown their competencies and right of existence well in the analogue world, they need to re-establish this trust in the digital world, as the digital world faces them with new challenges.

The users, using the digital material, now and in the future, are another group of stakeholders. Among them, researchers who want to re-use datasets. Or the future public of a digital library preserving snapshots of websites, related to important national events. Researchers and interested public, who want to explore governmental behavior in digital archives. They need to be convinced that the data have been taken care of in a responsible manner that the data are complete and will be available for long term.

The funding bodies of the digital repository need to be convinced that their money is well spent and that the repository stands by the agreements and fulfils its related tasks in a responsible manner.

From a different perspective there is yet another major stake holder: the management of the organization itself, responsible for the digital repository. The IBM tool, for example, is, according to the website, specifically designed to measure the maturity level of the LTP archive and could give the management an idea about the trustworthiness of the repository.

Both TRAC, DRAMBORA and the Kriterienkatalog explicit state that they can be used for self assessment: Drambora is designed to help organizations to identify risks and to manage these risks. The Kriterienkatalog advertises itself as a “guidance for planning, set-up, self-evaluation etc.”

But are organizations really doing self-assessment exercises? And if so, what are the results of these exercises? Did it result in specific changes of their procedures and approaches?

There is hardly any literature to find about this. This may be obvious, as no one is willing to admit that misjudgments have been made, that not everything is according to expectations and, worst of all, that data have been lost or damaged.

The Drambora website tells us that the manual was downloaded around 2000 times and the toolkit around 1200 times, and states that around 80 audits took place. On the website, organizations are mentioned who implemented Drambora. I could not find similar figures for TRAC or the Kriterienkatalog. So may be these documents are used for self-assessment, but evidence and outcomes are not fed back visibly into the preservation community.

Do the guidelines serve other purposes than that of self-assessment or certification? Sometimes. One example. The Center for Research Libraries initiated an assessment based on TRAC of both Portico and HATHITrust in order *“to promote understanding of and, where justified, confidence in, digital repositories. In today’s economic climate libraries must realize the greatest possible return on their investment in electronic scholarly resources and digital preservation services, and need to move aggressively to reduce the costs of redundant print holdings.”*

Here the confidence in a digital repository as in contrast with paper storage is at stake. But is certification already the right approach to convince organizations that digital preservation is feasible?

Another example: the MetaARchive West said on the IPRES last September: *“The project has identified specific components of TRAC as a measurement of*

*success, including ones in the area of professional development, repository scalability, documentation, assessment, and software technology.”* TRAC as a measurement of success.

A lot of organizations use TRAC, DRAMBORA and the KRITERIENKATALOG as a reference point and a guidance. For them this seems at this stage more feasible than starting a certification process: they just want to be guided in setting up their repository. Often they don't have a large staff, and are looking for practical, ready to implement instructions. They seem to look for a reference point to decide whether they are on the right track. And sometimes they need to raise awareness in the organization to get things done by referring to what is expected by future certification.

Twenty years of digital preservation has brought us a wealth of information, many papers can be found on internet, lots of interesting discussions were held, projects have been started and published their outcomes. But did it bring us ready to use solutions for repositories great and small? Is it clear for everyone how to approach digital preservation? Is it possible for organizations to write a clear policy in which their goals are described, accompanied by a description of how they intend to reach these goals? Only a few organizations published their policies on the internet. Some of them formulate in their policy that they will use “best practices”. But do these “best practices” already exist and where can one find them?

As Reinhard Altenhöner said on the IPRES in September: *Progress was and is being made especially in the EU-funded projects like CASPAR, PLANETS and SHAMAN, but there is a significant lack of progress in establishing a common approach to solving the problems of preservation across the spectrum of memory institutions.”*

Yes, there are standards and de facto standards in the community. To name just the most popular I'll take OAIS and Premis.

OAIS is seen in many organizations with a digital repository as a starting point for the design of their system and as a reference point in the terminology and concepts they use in communication with their digital preservation colleagues. But I think that we also need to admit that, especially in smaller organizations, with a limited staff, OAIS is sometimes too abstract to translate to the day to day work. And for these organizations there is no place where one can go to with an OAIS related question.

Another example is Premis, for preservation metadata, seen by many repositories especially in the library community as a de facto standard. The Premis website offers practical products to translate the Premis concept in real life products and there is a lively discussion forum on which every organization, starting with preservation metadata, can raise basic questions and will often receive answers from experienced community members.

These are just two examples of standards related to digital preservation.

But organizations that will receive digital objects in obscure file formats to preserve for long term, need to find the optimal solutions themselves. They have to do this by combining several sources on the internet with information from their personal network of preservation colleagues and then draw their own conclusions, unable to check whether this is the optimal approach. There is no database with “best practices” yet.

This might gradually change however. The approach in the Planets project, where there will be a Testbed in which one can find the results of earlier performed preservation actions, together with the parameters and tools used, might be a first attempt to make life a little bit easier for repository managers.

But if an organization wants to follow the 10 rules of TDR's, is reading DRAMBORA, TRAC or the Kriterienkatalog enough? Do these documents offer enough background information to support the repository management to take the right decisions?

Take rule 6 for example. Many discussions are currently related to authenticity. If the “experts” in digital preservation still need lots of debate to define authenticity and to discuss an approach to warrant this, how can one expect new or small organizations to find the right approach to meet the requirement in rule 6?

So if there is not a wealth of information for repositories to rely upon for their day to day activities, how can they prepare themselves for certification? And based on which “evidence based” information could they be judged by auditors?

The current audit initiatives distinguish basically three areas for certification: the organization itself, the (digital) object management and the technical infrastructure and security.

They all three see their audit and certification document as related to other ISO standards, for example standards for record management or ISO 9000, placing digital preservation in a context of “responsible management”.

The “significant lack of progress” in Reinhardts citation, may raise the question whether all three areas of digital preservation mentioned in the certification documents (organization, object management and infrastructure) can be certified.

Currently there is no certification body at all. But the attempts to get the three initiatives certified, implies that within a foreseeable time there will be certification bodies. Stakeholders will require their repositories to become certified repositories.

The current documents certainly are helpful to assist organizations to make up their mind and set up an organization that is prepared to perform “digital preservation”. It induces them for example to plan budgets for longer term. To think about escrow, in case the repository might cease to exist under the management of the organization. Or to create a preservation policy for their repository in which goals, collections, and future users are described. But these

aspects create the environment for digital preservation, but still won't prove that digital collections will be saved over time.

To tackle this last and main challenge, it is important to bridge the gap between what is expected from organizations from an audit and certification point of view and what is offered to organizations as practical solutions. How to bridge this gap, could be subject for debate.

Would it not help if the audit and certification documents were accompanied by a knowledge base to support the decision process for preservation actions with practical information, based on evidence of current state of knowledge about it? Take for example bit level preservation. Many digital preservation repositories claim to do this. But a starting organization will have great difficulties in finding do's and don'ts for bit level preservation. A knowledge base might support them in taking the right decision.

Part of the content of the knowledge base could be a compilation of anonymised information of repositories that met the (self) audit and certification requirements, although they might have dealt differently with some challenges. Such a central place of practical information could bring digital preservation really forward. It would level the gap between the audit and certification documents on the one hand and the practicalities of digital preservation on the other hand.

It would support repositories great and small in their goals to preserve digital objects for the long term in a responsible and evidence based manner. Thank you