Evaluation of a Digital Repository

Hugo Quisbert; Luleå University of Technology; Luleå, Sweden

Abstract

This paper is the result of an evaluation of an organisation as a Trusted Repository. The evaluation was carried out in November 2007, using parts of the existing guidelines concerning the certification of Trusted Repositories. The contributions of are 1) findings on organisation concerning its commitment towards its stakeholders and archival holdings and 2) findings about repository evaluation. The findings show that a highly committed and competent organisation is reliable as a trusted repository and there is a need of continuity evaluating digital repositories since an evaluation does not show the reliability of an organisation as a trusted repository in the long-term.

Introduction

Long-term Digital Preservation is prerequisite for long-term accessibility to digitally recorded information. Long-term Digital Preservation has become prioritised as a strategic and research issue for different institutions the latest years [5], [8]. There are several reasons for approaching digital reservation; one of them is software and hardware obsolescence [6], [12]. Different requirements have developed different "solutions" to Long-term Digital Preservation. Garret & Waters [7] stated, "A critical component of the digital archiving infrastructure is the existence of a sufficient number of trusted organizations capable of storing, migrating and providing access to digital collections." The Research Library Group & the Online Computer Library Center [14] define a Trusted Repository, as "… one whose mission is to provide reliable, long-term access to managed digital resources to its designated community, now and in the future". This component has been identified as a Trusted Repository. This statement implies two sides of the same coin; what is a trusted repository. A trusted repository cans be a technical as well as an organisational infrastructure. The technical side is managed by a trusted (organisational) repository. Kaczmarek et al. [10] refers to the technical infrastructure as the "little r" whereas the organisational infrastructure is referred as the "Big R". R refers, of course to "Repository". Such repositories are implemented in institutional level offering services such as access to archival collections. The "little r" consists of the actual software executing processes of preservation, migration or retrieval and so forth. The title of this paper refers to a "Digital Repository" while the section above is describing a "Trusted Repository". A Trusted Repository (in this paper) is indeed a Digital Repository, but a Digital Repository is not necessarily a trusted Repository. Nevertheless, the definition of Trusted Repository also highly resembles the definition of Digital Archives provided by the OAIS Reference Model [2]. However, these concepts are treated equally until the Conclusion section of this paper.

The description of Trusted Repositories corresponds to definitions of Information Systems proposed by Checkland [3], Buckingham [1], or Langefors [11] amongst others. Consequently, such Information Systems, when implemented, must also be object of evaluation.

On Information Systems Evaluation

What is evaluation? I think this question is connected to "why" the Information System exists and for "whom", but also "what" is to be evaluated and "when". Evaluation can be carried out in many ways. Cronholm & Goldkuhl [4] distinguishes three types of evaluation 1) Goal-based evaluation, focusing on the intended services and outcomes of an Information System 2) Goal-free evaluation creating an understanding of the nature of what is evaluated and, 3) Criteria Based evaluation, that uses checklists, heuristics or quality ideals . Smithson & Hirscheim [17] identify five levels of evaluation. The macro level focuses on a national or international perspective. The sector level focuses on the acting domain or market sector. The firm level approaches the performance and impact of an Information System within a firm or organisation. The application level evaluates the impact of a particular application and finally the stakeholder level that matches stakeholders concerns and values of the system towards the Information System itself. Serafeimidis [16] offers the understanding of evaluation from a technical stream looking at efficiency or technical performance, an economic stream looking at effectiveness or quality and utilisation of the outputs of Information Systems and interpretive alternatives looking at the understanding of what is to be evaluated, meaning social action and organisational learning. Evaluation can also be understood as the measurement of Information Systems success. There is, therefore, a range of success factors (measurable or observable in some way) for any information system. Wateridge [22] identifies six of them: meets user requirements, achieves purpose, meets timescale, meets budget, (produces) happy users and meets quality. Some of them are easy to observe and measure such as "meets budget" and "meets timescale". Factors such as "achieves purpose" or "produces happy" users (i.e. stakeholders) are far more difficult to measure or observe due to the nature of long-term digital preservation. Evaluation is therefore a snapshot over the status of an Information System under certain circumstances and at a specific moment.

Information System according to Verjin-Stuart [21] can be seen in a broader and a narrow sense; in the broader sense, it is seen the totality of data processes within an organization, including the communication surrounding the environment, alluding a social phenomenon in a social context; in the narrower sense a computer-based systems. This corresponds to the terms of "Big R" and "little r" respectively.

Some evaluation methods that are focused to evaluate the software used by an organisation (i.e. a curator organisation) "… does not take into account factors such as hardware, time, manpower, money and other resources, as these may vary depending on the implementing organization or individual" (Goh, 2006). This means that the "little r" can be evaluated separately

from the "Big R" and perhaps vice versa. The definition of Trusted Repository implies that it is an organisation. Consequently and in my opinion, the evaluation of any Trusted Repository has to be carried on by evaluating it as an organisation (i.e. an information system in the broader sense of its definition). However, evaluating Trusted Repositories cannot be done in the same way as traditional information systems. It has to be preceded by framing the characteristics of a Trusted Repository (what) and the aim of the evaluation (why and for whom). Nevertheless, a method has to be developed or adapted (how) and a level of evaluation has to be set up (how much). These constraints differ slightly to the dimensions proposed by Saracevic [15] in which the *Context* dimension explains to be the richest dimension of all. The writer explains it to answer to why and I explain as why and for whom.

Two conclusions can be made from the above statements. The first is that evaluation of Trusted Repositories has to be made following specific guidelines that embrace the concept of Trusted Repository and a focus (perspective or intention) on what, how, how much and why and for whom. Approaching the first statement is within the state of the art, easy. There are three internationally well-established guidelines for the evaluation of Trusted Repositories (responding to how). These guidelines also have high degree of acceptance within the Lon-term Digital Preservation Community. The guidelines are described in the next section.

Approaching the second statement seems to be tricky. Nevertheless, my approach to evaluation is guided by several ideas described above. The first is connected to why. I think that a valuable result of an evaluation is that the evaluated organisation should learn something from it (a long-term perspective). The second is guided by for whom. Paraphrasing Wateridge [22], I should say that one of the success factors of a Trusted Repository is to *produce happy stakeholders* (eventually, happy in a long-term). An evaluation cannot really be goal-free, but the goals for an evaluation not necessarily have to show some direct economical benefits. Within the Long-term Digital Preservation community, evaluation in a macro level is highly desirable and a stakeholder level evaluation as well. Furthermore, organisational learning is necessary in the long-term.

Guidelines for the Evaluation of Trusted Repositories

The Digital Repository Audit Method Based on Risk Assessment toolkit (DRAMBORA) was developed as a collaboration joint task force of the Digital Curation Centre (DCC) and Digital Preservation Europe (DPE) [18]. A central aim for the development of the DRAMBORA toolkit was to conceive criteria, means and methodologies for audit and certification of Trusted Repositories. The focus on risks is very central for the DRAMBORA and developed as a toolkit for self-audit. The toolkit is based on the ISO 27001 standard, "Information Security Management - Specification with Guidance for Use", and comprise six stages 1) Identifying organisational context which is, identify repository's role, goals and objectives. 2) Document policy and regulatory framework which is: collecting and compiling a set of documents regulating the work of the repository. 3) Identify activities, assets and their owners, which is: developing a conceptual model over the work and activities of the repository identifying key assets, technology and staff. 4) Identify risks in which a set or list of potential risks are identified, 5) Assess risks in which the risks and their mutual relationship are characterised, and 6) Manage risks, in which business decision are made in order to approach risks.

In 2003, the Research Library Group (RLG) and the National Archives and Records Administration (NARA) in the United States of America, created a joint task force to work specifically on creating guidelines for the certification of digital repositories. The goal of the RLG-NARA Task Force on Digital Repository Certification has been to develop criteria to identify digital repositories capable of reliably storing, migrating, and providing access to digital collections. The resulting audit tool is the work of many experts representing an international range of communities in research, governments, data archives, and cultural heritage organizations. The Trustworthy Repositories Audit & Certification (TRAC) guidelines [19] are based on five basic principles for application of criteria 1) Documentation or evidence showing objectives, basic concepts, specifications, implementations, and quality and security standards. 2) Transparency meaning publishing appropriate parts of the documentation, 3) Adequacy related to the context of each implemented individual archiving task according to its feasibility and 4) Measurability showing measurable characteristics if possible or in some cases instead indicators that demonstrate the degree of trustworthiness. The checklist is divided into three sections: 1) Organizational infrastructure 2) Digital object management and 3) Technologies, technical infrastructure, and security.

The Working Group on Trusted Repositories Certification of the Network of Expertise in long-term STORage (nestor) developed also the Catalogue of Criteria for Trusted Digital Repositories [19] (henceforth, referred as the nestor guidelines). This catalogue identifies criteria, which simplify the evaluation of digital repository trustworthiness, both at organisational and technical levels. The criteria are defined in close collaboration with a wide range of different memory organisations, information producers, experts and other interested parties. This open approach is the basis for achieving a high degree of universal validity and practical applicability and facilitates broad-based acceptance of the results of any evaluations conducted based on these criteria. The present criteria catalogue for public comment represents an important milestone on the road towards achieving the working group's goals. The memory organisations should be given a well-constructed, coordinated and practical tool for achieving and demonstrating their trustworthiness. However, the intention is also to present the opportunity for repository certification within a standardised national or international process as a formal endorsement of an organisation's trustworthiness. The document's current draft also supports active participation in existing international standardisation efforts. The nestor guidelines are based on five basic principles for application of criteria. Both the principles and the guidelines catalogue are identical to the TRAC guidelines.

I could state that the TRAC and nestor guidelines conceptually are very similar; they differ in structure, the name of the checklist items and the sequence of these.

The Case

The division of Digitization and Media at the Mikkeli University of Applied Sciences (MiUAS) in Finland during 2005-2007 has developed an application for digital archiving named

Digilab Archive System (Digilab). The Digilab consists of a set of the services to several designated communities. The main objective of these services is to provide Digilab customers with a reliable environment for long-term digital archiving, fulfilling archival conventions and requirements. Separately from these services, the Digilab also provides search, migration and digitizing services for archived material. An aim of the division is to be considered by its stakeholders as a Trusted Repository. Since the MiUAS have committed to offer these services to a designated community, then the "whole picture" fits into the definition of a Trusted Repository. The evaluation was carried out in November 2007.

Method

The evaluation is to be regarded as a case study because it focuses on contemporary events. Focusing on contemporary events distinguishes the case study from the history research and makes it suitable to for a case study according to Yin [23]. Naturally, other methods, such as surveys or historical analysis, can be brought in, to complement the findings from the case study (ibid). The aim of the evaluation was to show the status of the MiUAS as a Trusted Repository. This is equivalent to the description of contemporary events and therefore an approach to descriptive case study was suitable. Data were collected by asking the personnel filling forms, interviews, informal conversations and history documents. Further, the collection of data took place in at the premises of the MiUAS avoiding in this manner strange and uncomfortable environments. Some interviews where preceded by slight chats about every day problems "at the office". This approach to generate empirical data is qualitative by nature because of the intellectual, analytical and interpretive aspects of it [13].

The "questionnaire" (the fill in forms) elaborated are entirely the templates provided by the DRAMBORA community. The analysis form is a slightly re-formatted form from the templates provided by TRAC guidelines.

During the evaluation, I used some parts of the DRAMBORA guidelines, namely stage 1, 2 and 3. These three stages are in concrete 1) identify organisational context including goals, objectives, mandates and mission 2) identify and collect policy documents and regulatory framework and 3) identify activities, assets and their owners. The reason for using just these stages was that they address principally the collection of data about the repository in written form (i.e. policy documents, written contracts, systems descriptions, etc.). I did not use the stages 4-6 of the DRAMBORA guidelines because they try to give a metric or measurement to qualitative aspects of an organisation. According to Wateridge [22], some (these) aspects are difficult to observe and measure. In addition, I used the TRAC and nestor guidelines for analysis and documentation of results of the evaluation. The reason for doing in this (novel) manner is that I needed a comprehensive way of gathering documentation, confirming the knowledge of the existence of this documentation and inquiring the level of awareness about their activities. The DRAMBORA guidelines were extremely helpful in this quest because the offer the possibility of fill and explain, extensively and exhaustively, the matters I searched for. Data were collected using the proposed worksheets for the stages investigated. Simultaneously, I collected documents (in paper and electronically) concerning the organisation, policies, systems descriptions and so on. The involved personnel at the division filled in the worksheets, according to their position and knowledge.

Analysis

The real evaluation was performed by analysing, the collected data and answering "yes" or "no" and when needed writing comments or observation about the item. I did the analysis sequentially from section A to C from the structure of the TRAC guidelines. To every item in the checklist, I looked up within the answers of the personnel, wherever there was a relevant answer. To some items were several answers. The answers are presented in a later section in tables. I used a simplified version of the TRAC guidelines. The nestor guidelines were used to control whether the text of some items were unclear for me or not. However, most of the items, especially in sections A and B of the TRAC guidelines were used and answered. Below, I present the results of the analysis of the evaluation.

Organisation Infrastructure

Aspects investigated were 1) governance and organisational viability, 2) organisational structure and staffing, 3) procedural accountability and policy framework, 4) financial sustainability, and 5) contracts, licenses and liabilities. The commitment of the MiUAS towards its stakeholders is made in awareness of a very long-term perspective, specified in the contracts policies and practices that show the extent of the commitment. The staffing is adequate in relation to the size and the solvency of the repository. However, some of the personnel are contracted in project basis and the length of the contract is determined by the length of the project they are working with. New contracts are dependant of new projects that are financed by external economic support. This might be a major risk due to people might be looking for a permanent position in order to secure it own economical situation in a long term. Short-term contracts do not offer this "economical security".

Most of archival and curation institution are dependent on external financing. Is not really the case on this audited repository, However, the awareness over the financial situation is not only clear but also crucial and the initiative to the development of new ideas and projects in the field show that the repository is aware of financial sustainability to "survive" as repository institution. As far as I could see, the economical sustainability of the repository in a short term is good. The contracts are clear and specify responsibilities of both contractors. There are no occurrences of unlicensed software.

Digital Object Management

Aspects evaluated concerning digital object management are 1) ingest: acquisitions of content 2) ingest: creation of archival package, 3) preservation planning, 4) storage, preservation and maintenance, 5) information management, and 6) access management. The answers of the personnel of the repository show a very deep knowledge in these matters. Beside the topics covered above, is also necessary to point out the occurrence of other technical activities such as "Digitization of documents, motion pictures, video", "Automatic metadata capture" etc. The granularity of the answers in the topics of acquisition of content shows a very good understanding and very good skill in managing these activities. Something that needs to be emphasised is that the

OAIS Reference Model has been used as a tool for description of the material to be preserved. (From this section on, reference to the model will be done implicitly).

In this section of ingest: creation of archival packages the answers of the personnel of the repository show a very deep knowledge in these matters. The granularity of the answers in these topics shows a very good understanding and very good skill in managing these activities. Preservation Planning is present at the repository and followed in a very good manner.

In the section of storage, preservation and maintenance, once again, the answers of the personnel of the repository show a very deep knowledge in these matters. Even in this section, granularity of the answers in these matters shows a very good understanding and very good skill in managing these activities. These activities must be considered not to be "normal" backup activities. The plans and activities in this matter are far more developed, in accordance to long-term digital preservation practices.

The section of information management shows that the answers of the personnel of the repository show a very deep knowledge in that matter. The granularity of the answers in these topics shows a very good understanding and very good skill in managing these activities.

Finally in the section of access management, the access policies and arrangements are described clearly in the system description and in the answers of the personnel. In this case access policies are strongly connected to system security policies, managed in a very god manner with very good skills.

Technologies, Technical Infrastructure & Security

The aspects evaluated are: 1) system infrastructure, 2) Appropriate technologies and 3) security.

The technological infrastructure is also necessary to be commented. It is almost impossible to state that "this or this" infrastructure is the "very best". Choices are made in accordance to financial solvency and prevailing state of the art technologies. Anyway, the chosen and implemented technology, in my opinion is reliable, as long as there are policies that guide the personnel using in optimal manner the soft- and hardware. The skills of the personnel and the commitment and responsibilities of the people at MiUAS show that the stakeholders and the designated communities should be confident that the MiUAS would take care of the delivered digital assets in the best of ways. Although there are no documented plans for handling some security aspects of the repository, those aspects are considered parts of the core activities of a trusted repository, are followed strictly by the personnel at the division

It is shown that to document is not an activity that is carried out automatically and consequently by technicians, especially IT personnel. It is also confirmed that documentation is crucial and considered important by the people at a trusted repository. The documentation shows a clear insight about the activities of the trusted repository, as the policies and regulations to follow as well. In the case of the MiUAS, the documentation about policies and preservation exists. However, the documentation could be more centralised. Some documents are to be found in personal areas of storage (example: C:\Users\Mårten\Documents\Pms_Avain\Aton\DigitalArchiveSystem_2611.doc). The question is if these documents are available to other personnel whenever needed. A more regulated way of naming documents might be also desirable. These "problem" can be solved by a clear naming policy and the arrangement of storage of documents in order to make important document available to all personnel.

Conclusions

The evaluation shows important considerations about the MiUAS and the instrument of evaluation.

Concerning the status of MiUAS it can be said that: a) The MiUAS need to adopt a more stringent policy about handling documents necessary for the organisation; this concern a more centralised and controlled access (and availability) to documents even concerning naming and used language. b) It seem to be that new project are needed it order to generate bigger financial sustainability. There is risk that this "money haunting" might compromise the reliability of MiUAS as a trusted repository. However, a real market for archival services has not appeared and most of the Long-term Digital Preservation community is dependant on external financing. c) There is a high level of competence and commitment at the MiUAS towards its role as "Trusted Repository". The commitment concerns its stakeholders and archival holdings and the competence concerns handling technologies. The overall judgment of this part is that the repository shows a strong awareness in the field of long-term digital preservation and the direction of the repository is heading for deeper commitment in the field. The organisation, in this aim, acts in accordance to open cultural and transparent accountability.

Concerning the evaluation instrument, the method used for the evaluation was created from parts of the three guidelines. This combination functioned well due to these guidelines are based in the same idea, certification of Trusted Repositories, and have influenced one another. It is difficult to adapt an evaluation according to the initial requirements of the repository. Changes might have occurred over time and some conflicts towards the initial requirements might exist. It is also going to be very difficult to validate if an organisation will be a trusted repository organisation in a long-term perspective. This demands continuity not only doing new evaluations of the repository, but also reviewing the guidelines of evaluation. Using the guidelines for the study confirms that the guidelines need to be adapted specifically to an actual case. There is other reason to have used these guidelines, namely that there is no existing model or method for the evaluation of digital archives. I don not say that the abandoned stages of the DRAMBORA guidelines are useless. I could not use as instrument that renders quantitative measures when I searched for qualitative aspects of the MiUAS. I see the need of methods or guidelines for the evaluation of Trusted Repositories that addresses qualitative aspects of it according to qualitative approaches and quantitative aspects of it according to quantitative approaches.

Some reflections follow: A) It is confirmed that documentation is crucial and considered important by the people at a trusted repository. The documentation shows a clear insight about the activities of the trusted repository, as the policies and regulations to follow as well. The access to documents has to be easy and transparent. B) To document seem not to be a cognitive automatic process for technicians, especially IT personnel. Organisations teaching IT should be more demanding in this subject. C) The MiUAS shown that even though there are no

documented plans for handling some preservation or security aspects of the repository, those aspects are considered highly important and followed strictly by the personnel at the division. This implies that i) some activities lives "in the walls" and are part of the culture of the organisation and its role as Trusted Repository making it a reliable organisation and ii) related activities are part of the profession are practiced naturally.

References

[1] Buckingham R., Hirschheim R., Land L. and Tully C. eds. Information Systems Education: Recommendations and Implementation, (Cambridge University Press, 1987).

[2] CCSDS: Reference Model for an Open Archival Information System (OAIS) – Blue book. (National Aeronautics and Space Administration., 2002)

[3] Checkland, P and Holwell, S. Information, Systems and Information Systems – making sense of the field. (Wiley & Sons Ltd, Chichester, England, 1998)

[4] Cronholm S. & Goldkuhl G. "Strategies for Information Systems Evaluation". Electronic Journal of Information Systems Evaluation. 6, 2. (2003). pg 65-74

[5] Deegan M. & Tanner S. Digital Preservation. (Facet Publishing, 2006).

[6] Dollar C. M. Authentic Electronic Records: Strategies for long-term access. (Cohasset Associates Inc. USA, 2000).

[7] Garrett, J. and Waters, D., ed. Preserving Digital Information: Report of the Task Force on Archiving of Digital Information. (Washington DC: Commission on Preservation and Access and Research Libraries Group,1996)

[8] Gladney, H. M.: Preserving Digital Information. (Springer, 2006)

[9] D. H.-L. Goh, A. Chua, D.A. Khoo, E. B.-H. Khoo, E. B-T. Mak & M. W-M. Ng (2006). "A checklist for evaluating open source digital library software". Online Information Review, 30(4), pg 360-379.

[10] Kaczmarek J.,Hswe P., Eke J., Habing T. G. "Using the Audit Checklist for the Certification of a Trusted Digital Repository as a Framework for Evaluating Repository Software Applications". D-Lib Magazine. December 2006 Volume 12 Number 12

[11] Langefors, B. Essays on Infology. (Dept. of Information Systems. Göteborg University, Göteborg, 1993).

[12] Lazinger S., Tibbo, H. Digital Preservation and Metadata: History, Theory, Practice. (Libraries Unlimited Inc, 2001).

[13] Mason J. Qualitative Researching. (SAGE Publications, 1996).

[14] Trusted Digital Repositories: Attributes and Responsibilities. An RLG-OCLC Report. Reseach Library Group. (2002)

[15] Saracevic, T. "Evaluation of digital libraries: an overview". In Fuhr N., Tsakonas G., Aalberg T., Agosti M., Hansen P., Kapidakis S., Klas C-P., Kovács L., Landoni M., Micsik A., Papatheodorou C., Peters C., Sølvberg I. "Evaluation of Digital Libraries". International Journal of Digital Libraries. Springer. (2007). 8. Pg 21-38

[16] Serafeimidis V. A Review of Reseach Issues in Evaluation of Information Systems. In Van Grembergen (Ed.). Information systems Evaluation Management. (Idea Group Publishing, 2002) pg. 167-194

[17] Smithson S. and Hirschheim R., "Analysing information systems evaluation: another look at an old problem". European Journal of Information Systems, 7 pg. 158-174. (1998)

[18] The DRAMBORA Guidelines, (2007) http://www.repositoryaudit.eu/ Accessed: 2008-01-25

[19] The nestor Guidelines (2007). http://journals.tdl.org/jodi/article/view/199/180 Accessed: 2008-01-25

[20] The TRAC Guidelines, (2007). http://www.crl.edu/PDF/trac.pdf Accessed 2008-01-25

[21] Verrijn-Stuart, A. Information Systems Concepts: in E.D. Falkenberg, P. Lindgreen (eds.): Information System Concepts - An In-depth Analysis; Proc. IFIP WG 8.1 Conf. ISCO-1, Namur, Belgium. North Holland (1989).

[22] Wateridge, J (1998). "How can IS/IT projects be measured for success?" International Journal of Project Management 16(1): 59-63.

[23] Yin, R. K. Case Study Research: Design and Methods (3rd Edition Sage Publications, Inc, 2003).

Author Biography

Hugo Quisbert is a Ph D Student at the Luleå University of Technology in Sweden. His area of research concerns the development of Digital Information Preservation Systems. He is connected to the Swedish Centre of Long-term Digital Preservation.