



PARSE.Insight

Deliverable D6.1

Workshop Report

Project Number	223758
Project Title	PARSE.Insight. INSIGHT into issues of Permanent Access to the Records of Science in Europe
Title of Deliverable	Workshop report
Deliverable Number	D6.1
Contributing Work package	WP6: Sustainability
Deliverable Dissemination Level	Public
Deliverable Nature	Report
Contractual Delivery Date	28 Feb 2009 (M12)
Actual Delivery Date	18 May 2009
Author(s)	STFC

The PARSE.Insight project is partly funded by the European Commission under the 7th Framework Programme, Research Infrastructures.

Abstract

This document describes the work undertaken within PARSE.Insight to further the development of a pair of standards on which an audit and certification process for digital repositories can be based.

Keyword list

Audit of digital repositories; certification of digital repositories.

Contributors

Person	Role	Partner	Contribution
Simon Lambert		STFC	Document owner and author

Document Approval

Person	Role	Partner
David Giaretta	Coordinator	STFC

Distribution

Person	Role	Date	Partner
Simon Lambert	Author	13 May 2009	Distributed amongst all PARSE partners

Revision History

Issue	Author	Date	Description
0.1	Simon Lambert	20 April 2009	Initial draft version
1.0	Simon Lambert	17 May 2009	Final version

Table of Contents

Page

<u>1 INTRODUCTION AND BACKGROUND.....</u>	<u>4</u>
<u>2 THE ORGANISATION OF THE WORKSHOP.....</u>	<u>6</u>
<u>3 THE OUTCOME OF THE WORKSHOP.....</u>	<u>7</u>
<u>4 NEXT STEPS.....</u>	<u>10</u>

1 Introduction and background

The objectives of Work Package 6 are to:

- address issues of sustainability of data resources
- bring together international approaches for the evaluation and certification of long term preservation repositories in order to promote practical experiences with the evaluation and certification
- to work with the global community on developing a standard on which a data repository certification process can be based.

Tasks 6.1 and 6.3 are focussed on building on the ongoing certification work and this report is about the important workshop which was sponsored by PARSE.Insight to take this work forward at a critical time. The outcome of this workshop is expected to be a full ISO standard which will form the basis of the international audit and certification process, thereby being a major influence in the digital preservation marketplace.

Task 6.2 will follow on, based on the results of the survey and case studies, with specific focus on the funding horizon of targeted institutions and the technical, sociological and legal constraints on the pooling of resources to share the burden of sustaining these resources and associated services.

This Work Package is therefore initially concerned with the trustworthiness of digital repositories for long-term storage of digital information. It is evident that the judgement of trustworthiness is a difficult problem; only with hindsight may it be truly assessed. When making investment decisions or choosing where to deposit digital material, however, there is a need to make an assessment in the present. This cannot be a definitive assessment, but it is possible to evaluate repositories on the basis of evidence of their constitution, practices and technology, supporting a judgement of how likely they are to be sustainable into the future. Such an evaluation could be formalized into a procedure for audit and certification of the repositories, leading to a system of accreditation.

The ISO standard Open Archival Information System (OAIS) Reference Model has become a *de facto* standard for building digital archives. Section 1.5 of OAIS (“Road map for development of related standards”) includes an item for accreditation of archives, reflecting

the long-standing demand for a standard against which repositories of digital information may be audited and on which an international accreditation and certification process may be based. It was agreed that RLG and the US National Archives and Records Administration (NARA) take a lead on this follow-on standard. This they did, forming a closed panel which produced *Trustworthy Repositories Audit & Certification: Criteria and Checklist* in 2007¹.

This report was based on two the OAIS Reference Model and the *Report on Trusted Digital Repositories: Attributes and Responsibilities* (RLG-OCLC, 2002). The latter complemented OAIS by focussing on the administrative, financial and organisational requirements for the body undertaking the preservation activities.

In order to progress this work, a Working Group within the Consultative Committee on Space Data Systems (CCSDS) was set up, with its charter agreed in January 2007. The goal was to obtain ISO approval of a standard that establishes the criteria that a repository must meet to be designated an ISO Trusted Digital Repository. This was following the route of OAIS itself; CCSDS is the “working arm” of TC20/SC13 of ISO. The group’s work is based on the TRAC document, with the aim of reviewing and refining the criteria therein.

The group is following an open process with a publicly accessible wiki at <http://www.digitalrepositoryauditandcertification.org>. Weekly online discussions take place, covering the evolving criteria and other documentation required. The notes are recorded on the wiki with all working documents. This forum was also agreed as the “clearing house” for the private discussions of the other groups in this area (e.g. the nestor initiative in Germany).

Two documents are being edited:

- *Metrics for Digital Repository Audit and Certification*. This is the basic document that lists the criteria against which a repository will be judged.
- *Requirements for Bodies Providing Audit and Certification of Digital Repositories*. An additional document setting out requirements on those organizations that provide the audit and certification of repositories. Clearly such bodies themselves must be properly constituted and effectively managed.

¹ Available at <http://www.crl.edu/PDF/trac.pdf>

In addition to the weekly online discussions, occasional face-to-face meetings of the Working Group take place. Task 6.1 in PARSE.Insight called for an international expert workshop to progress this work, involving key players from the EU and USA plus others. The present deliverable is the record of that workshop. It is deliberately brief but reflects the progress made.

2 The organisation of the workshop

In October 2008 members of the Working Group had met face-to-face in Berlin during the CCSDS meeting taking place there. The notes of that Working Group meeting can be seen on the wiki at <http://www.digitalrepositoryauditandcertification.org> under “Meetings” and then “Face to face meeting in CCSDS 14 October 2008”. On that occasion work started on the *Requirements for Bodies Providing Audit and Certification of Digital Repositories*, and there was collective marking up of parts of the metrics document for follow-up in subsequent online discussions.

In order to maintain the momentum achieved at this meeting, the PARSE.Insight workshop was set for February 2009, before the next CCSDS meeting in April 2009. The objective of this workshop was to make substantial progress working through the sections of the metrics document, with a view to finishing with a version close to being suitable for submission to the ISO process.

The workshop took place on 11–13 February 2009 at the National Archives at College Park in Maryland, USA. The following persons took part in the workshop.

Name	Affiliation	Country
Bruce Ambacher	University of Maryland	USA
Mark Conrad	National Archives and Records Administration	USA
Robert Downs	CIESIN, Columbia University	USA
Riccardo Ferrante	Smithsonian Institution	USA
John Garrett	Goddard Space Flight Center, NASA	USA
David Giaretta	STFC	UK
Simon Lambert	STFC	UK
Reagan Moore	San Diego Supercomputer Center	USA
Bernie Reilly	Center for Research Libraries	USA
Don Sawyer	NASA	USA
Barbara Sierman	Koninklijke Bibliotheek	The Netherlands
Katia Thomaz	Instituto Nacional de Pesquisas Espaciais	Brazil
Helen Tibbo	University of North Carolina	USA
Marie Waltz	Center for Research Libraries	USA

Others, especially individuals from other European efforts in this area, such as nestor, or those involved in the earlier TRAC work, were invited but were unable to attend because of other commitments.

The initial agenda is shown below.

Wednesday 11 Feb	
<i>Morning</i>	Overview of agenda for documents Discussion of division of material between the documents Initial discussion of test audits Discussion of support organisation
<i>Afternoon</i>	Work on metrics
Thursday 12 Feb	
<i>Morning and afternoon</i>	Work on metrics
Friday 13 Feb	
<i>Morning</i>	Metrics Guidelines for Auditors Test Audits
<i>Afternoon</i>	Finalisation of document drafts Review of actions and deadlines

3 The outcome of the workshop

The metrics document that the Working Group had been examining comprises three major sections:

- Section A: Organisational Infrastructure
- Section B: Digital Object Management
- Section C: Infrastructure and Security Risk Management

Each section comprises a number of metrics setting out the criteria that a repository must fulfil. It had earlier been agreed that each metric would have a uniform structure:

- Statement of requirement
- Supporting text
- Examples of ways the repository can demonstrate it is meeting this requirement
- Discussion

The following screen shot of the wiki gives a flavour of the metrics document at the start of the workshop. It shows the record of previous discussions, edits and remarks. The metric itself is the text following the section number "B1.3".

B1.3 Repository has mechanisms to authenticate validate the source of all materials.

Supporting Text

The repository must ensure that the sources of the materials it intends to preserve are who/what they claim to be.

This is necessary in order to avoid providing erroneous provenance to the information which is preserved.

Examples of Ways the Repository can Demonstrate it is Meeting this Requirement

Submission agreements/deposit agreements/deeds of gift; workflow documents; evidence of appropriate technological measures; logs from procedures and authentications, legally binding submission agreements/deposit agreements/deeds of gift

-- [BarbaraSierman](#) - 17 Mar 2008: How can a submission agreement demonstrate the source of the material? The logs are more convincing. There is a difference between the managerial decision to archive material (- submission agreement) and the actual receipt of that material. I would assume that it is the last situation that is meant in B1.3.

Discussion

The repository's written standard operating procedures and actual practices must ensure the digital objects are obtained from the expected source, that the appropriate provenance has been maintained prior to submission, ~~and that the objects are the expected objects~~. Confirmation can use various means including, but not limited to, digital processing and data verification and validation, and through exchange of appropriate instrument of ownership (e.g., legally binding submission agreements/deposit agreement/deed of gift). Different repositories will adopt different levels of proof needed; the Designated Community should have the opportunity to review the evidence.

-- [KatiaThomaz](#) - 20 Mar 2008 - What does "source of all materials" really mean? A physical or corporate person responsible for issuing the materials?

During the course of the workshop the metrics document was thoroughly reviewed. Important decisions or areas of progress include the following.

- It was confirmed that a hierarchical substructure would be used for metrics, that is, each metric may optionally have sub-metrics which are sufficiently important to merit being described separately. These sub-metrics are necessary but not sufficient for the fulfillment of the overall metric. An example is shown below.

B2.4 Repository has and uses a convention that generates persistent, unique identifiers for all AIPs.

In particular the following aspects must be checked.

B2.4.1 The repository must be able to show how any AIP can be uniquely identified within the repository.

B2.4.1.1 The repository must be able to demonstrate that the identifiers are unique.

B2.4.1.2 Documentation must show how the persistent identifiers of the AIP and its components are assigned and maintained so as to be unique within the context of the repository.

B2.4.1.3 Documentation must also describe any processes used for changes to such identifiers.

B2.4.1.4 The repository must be able to provide a complete list of all such identifiers and do spot checks for duplications.

B2.4.1.5 The system of identifiers must be seen to fit the repository's current and foreseeable future requirements for things like numbers of objects.

Supporting Text

These requirements are necessary in order to ensure that each AIP can be unambiguously found in the future. They also are necessary to ensure that each AIP can be distinguished from all other AIPs in the repository.

- It was agreed that the terminology describing policies and plans would be the hierarchy: Mission statement – Preservation strategic plan – Preservation policy – Preservation implementation plan.
- Review of the three sections was completed
- Remaining checking and tidying was assigned to individuals

There was also progress with the test audits that are necessary for validating and refining the metrics, through putting them into practice against a number of real repositories. It was agreed that these test audits would result in the following outcomes:

- points of confusion/difficulty in the text
- proposed clarifications
- what is missing and what may be eliminated
- any additional notes to auditors
- estimates of time needed
- proposed phased implementation plan for improvements of the archives, i.e. what does the audit suggest could be improved

-
- names people use for the evidence, i.e. a consistent set of terms so that in all the Evidence sections the same words are used to mean essentially the same document
 - suggestions for Glossary definitions

4 Next steps

Following this workshop, a CCSDS meeting took place in Colorado Springs in April 2009 at which the metrics document was revisited and further progress made. The output document was intended to be of the right form to submit to CCSDS, though some corrections to the wording were still needed. Additionally an online form was prepared for use in collecting information during the test audits.

The plan now is to conduct a number of test audits in May and early June. It is expected that the metrics document will be submitted for ISO review in June, and the review process will probably take around eight months. Plans for the international accreditation and certification process will be completed during the ISO review.

We recognise that it is possible that this standard will be ignored in the wider world. However the widespread demand for such certification leads us to believe that we have a very real chance of playing a major role in determining the way in which digital preservation is evaluated and the development of the digital preservation market. Production of the standards is a first, important, step. The development of internationally recognised accreditation and certification institutions and committees are perhaps even more important.