
Copyright © 2009 Association for Computing Machinery

A copy can be downloaded for personal non-commercial research or study, without prior permission or charge

Content must not be changed in any way or reproduced in any format or medium without the formal permission of the copyright holder(s)

http://eprints.gla.ac.uk/45793/

Deposited on: 18 November 2010
Assessing Digital Preservation Frameworks: the approach of the SHAMAN project

Perla Innocenti, Seamus Ross
HATII at the University of Glasgow
11 University Gardens
G12 8QQ UK
+443304453
P.Innocenti@hatii.arts.gla.ac.uk

Elena Macecuvite, Tom Wilson
Swedish School of Library and Information Science
University of Borás, 50190 Sweden
+46334354000
Elena.Maceviciute@hb.se

Jens Ludwig, Wolfgang Pempe
Goettingen State and University Library
Papendiek 14
37073 Goettingen, Germany
+ludwig@sub.uni-goettingen.de

ABSTRACT
How can we deliver infrastructure capable of supporting the preservation of digital objects, as well as the services that can be applied to those digital objects, in ways that future unknown systems will understand? A critical problem in developing systems is the process of validating whether the delivered solution effectively reflects the validated requirements. This is a challenge also for the EU-funded SHAMAN project, which aims to develop an integrated preservation framework using grid-technologies for distributed networks of digital preservation systems, for managing the storage, access, presentation, and manipulation of digital objects over time. Recognising this, the project team ensured that alongside the user requirements an assessment framework was developed. This paper presents the assessment of the SHAMAN demonstrators for the memory institution, industrial design and engineering and eScience domains, from the point of view of user’s needs and fitness for purpose. An innovative synergistic use of TRAC criteria, DRAMBORA risk registry and mitigation strategies, iRODS rules and information system models requirements has been designed, with the underlying goal to define associated policies, rules and state information, and make them wherever possible machine-encodable and enforceable. The described assessment framework can be valuable not only for the implementers of this project preservation framework, but for the wider digital preservation community, because it provides a holistic approach to assessing and validating the preservation of digital libraries, digital repositories and data centres.

Categories and Subject Descriptors
H.4 [INFORMATION SYSTEMS APPLICATIONS]

General Terms

Keywords
Digital preservation, Digital preservation frameworks, Policy frameworks, Web based Digital Ecosystems, SHAMAN project, Assessment criteria, Data grids, Digital libraries, Digital repositories, Data centres, Persistent archiving.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

Conference ACM MEDES’09, October 27–30, 2009, Lyon, France.
Copyright 2009 ACM MEDES

1.2 The SHAMAN Project

The SHAMAN (Sustaining Heritage Access through Multivalent ArchiviNg) is a Integrated Project [3], part of the European Union's 7th Framework Programme. The aim of this project is to investigate the long-term preservation of large volumes of digital objects in a distributed environment, by developing a preservation framework that is verifiable, open and extensible. Our research addresses digital preservation from ingestion to dissemination in an environment where the collections, producers, consumers and curators are geographically distributed and the content of the collections is of a dynamic nature.

SHAMAN is developing associated preservation tools for analyzing, ingesting, managing, accessing and re-using digital objects across libraries and archives. Three prototypical applications will support evaluation and validation of the results in memory institutions, in industrial design and engineering settings and in the domain of e-science.

To achieve this aim, the project is investigating data grid, digital library, persistent archive and information knowledge and content representation technologies, to create preservation system prototypes that characterize the preservation process in ways that make it feasible to replace preservation services without impact upon the digital objects, or access to it and re-use of it (Figure 1).

SHAMAN’s Digital Preservation Framework is based upon a combination of technological, organizational, and R&D methods. This paper introduces the main principles applied for evaluation of the integrated elements of the SHAMAN framework from the point of view of user’s needs within all three domains of focus.

2. ASSESSMENT OF A PRESERVATION FRAMEWORK

2.1 Project goals and outputs to be assessed

The SHAMAN project seeks to:

- provide a vision and rationale to support a comprehensive Theory of Preservation that may be used to develop systems for the storage of and access to any type of digital objects, based on the integration of digital library, persistent archive, and data management technologies;
- supply an infrastructure that provides expertise and support for users requiring the preservation and re-use of digital objects; and,
- develop and implement a grid-based production system, which will support the virtualisation of digital objects and services across archival, scientific and engineering domains.

This hierarchy of tasks determines the levels of evaluation of the project and outputs.

2.2 Assessing criteria derived from TRAC and DRAMBORA

SHAMAN will deliver a preservation framework, which will supply the infrastructure for users requiring long-term preservation services, and which will develop and implement a grid-based production system to support the virtualisation of digital objects and services in a variety of user domains. The process of evaluation of its success might therefore be considered in terms of the benchmarking and risk assessment tools that have been proven in other projects and initiatives.

As a crucial part of this process the two noted TRAC (Trustworthy Repositories Audit & Certification: Criteria and Checklist) [4] and DRAMBORA (Digital Repository Audit Method Based on Risk Assessment) [5] were mapped to iRODS (i Rule Oriented Data Systems) [6] and the objectives of the SHAMAN work packages.

The Trustworthy Repositories Audit and Certification (TRAC) Criteria and Checklist is configured as a checklist, meant to help institutions objectively to evaluate responsibilities against capabilities and identify potential risks to digital content. TRAC takes OAIS (Open Archival Information System) [7] as its foundation, and the benchmark for measuring success in terms of trustworthiness.

The main goal of DRAMBORA, which is an interactive online support management tool at repository level, is enabling evidence-based risk management for digital repositories. The DRAMBORA assessment process focuses on risks, and their
classification and evaluation according to individual repositories’ activities, assets and contextual constraints (Figure 3), to determine a particular repository’s ability to contain and avoid the risks which might threat its ability to receive, curate and provide access to authentic and contextually, syntactically and semantically understandable digital information.

As a result of the evaluation conducted as part of this work, we concluded that it would be appropriate to set as a target evaluation criteria the question: Will we be able to prove that a system designed and deployed according to the SHAMAN framework properly supports the TRAC/DRAMBORA rules and criteria? This evaluation will be possible when the ‘integrative sub-projects’ of SHAMAN have been delivered and are available to test against the TRAC/DRAMBORA criteria.

2.3 Assessing criteria derived from iRODS rules

SHAMAN will use Integrated Rule-Oriented Data System (iRODS™) data grid technology as a storage substrate for digital preservation.

iRODS implements rules for grid-based data management. The whole process of digital objects ingestion, manipulation, access and use can be managed over a grid-based system through the application of the “iRODS Rule Engine”. Consequently, the nature of the rules and the ability of digital preservation systems to operate under those rules will guide the development of the SHAMAN preservation framework. Evaluating whether or not the individual elements of the framework satisfy this requirement will be a feature of the assessment process.

Finally, the individual work packages of SHAMAN will produce a variety of outputs, including software and conceptual schemes, designed for grid-based operation. Evaluation of the outputs will be according to the general criteria for successful information systems development and should be tested against the iRODS rules.

2.4 Assessing criteria derived from information system models

The outputs of the individual work packages will be assessed using information systems success criteria developed by DeLone and McLean [8,9]. The framework proposed by DeLone and McLean [8,9] is widely accepted as appropriate for the evaluation of information systems. We intend to use their original McLean [8,9] is widely accepted as appropriate for the evaluation of information systems. We intend to use their original McLean [8,9]. The framework proposed by DeLone and McLean [8,9], is widely accepted as appropriate for the evaluation of information systems. We intend to use their original McLean [8,9]. The framework proposed by DeLone and McLean [8,9], is widely accepted as appropriate for the evaluation of information systems. We intend to use their original McLean [8,9]. The framework proposed by DeLone and McLean [8,9], is widely accepted as appropriate for the evaluation of information systems. We intend to use their original McLean [8,9]. The framework proposed by DeLone and McLean [8,9], is widely accepted as appropriate for the evaluation of information systems. We intend to use their original McLean [8,9]. The framework proposed by DeLone and McLean [8,9], is widely accepted as appropriate for the evaluation of information systems. We intend to use their original McLean [8,9]. The framework proposed by DeLone and McLean [8,9], is widely accepted as appropriate for the evaluation of information systems. We intend to use their original McLean [8,9]. The framework proposed by DeLone and McLean [8,9], is widely accepted as appropriate for the evaluation of information systems. We intend to use their original McLean [8,9]. The framework proposed by DeLone and McLean [8,9], is widely accepted as appropriate for the evaluation of information systems. We intend to use their original McLean [8,9]. The framework proposed by DeLone and McLean [8,9], is widely accepted as appropriate for the evaluation of information systems. We intend to use their original McLean [8,9]. The framework proposed by DeLone and McLean [8,9], is widely accepted as appropriate for the evaluation of information systems. We intend to use their original McLean [8,9]. The framework proposed by DeLone and McLean [8,9], is widely accepted as appropriate for the evaluation of information systems. We intend to use their original McLean [8,9]. The framework proposed by DeLone and McLean [8,9], is widely accepted as appropriate for the evaluation of information systems. We intend to use their original McLean [8,9]. The framework proposed by DeLone and McLean [8,9], is widely accepted as appropriate for the evaluation of information systems. We intend to use their original McLean [8,9]. The framework proposed by DeLone and McLean [8,9], is widely accepted as appropriate for the evaluation of information systems. We intend to use their original McLean [8,9]. The framework proposed by DeLone and McLean [8,9], is widely accepted as appropriate for the evaluation of information systems. We intend to use their original McLean [8,9]. The framework proposed by DeLone and McLean [8,9], is widely accepted as appropriate for the evaluation of information systems. We intend to use their original McLean [8,9]. The framework proposed by DeLone and McLean [8,9], is widely accepted as appropriate for the evaluation of information systems. We intend to use their original McLean [8,9]. The framework proposed by DeLone and McLean [8,9], is widely accepted as appropriate for the evaluation of information systems. We intend to use their original McLean [8,9]. The framework proposed by DeLone and McLean [8,9], is widely accepted as appropriate for the evaluation of information systems. We intend to use their original McLean [8,9]. The framework proposed by DeLone and McLean [8,9], is widely accepted as appropriate for the evaluation of information systems. We intend to use their original McLean [8,9]. The framework proposed by DeLone and McLean [8,9], is widely accepted as appropriate for the evaluation of information systems. We intend to use their original McLean [8,9]. The framework proposed by DeLone and McLean [8,9], is widely accepted as appropriate for the evaluation of information systems. We intend to use their original McLean [8,9]. The framework proposed by DeLone and McLean [8,9], is widely accepted as appropriate for the evaluation of information systems. We intend to use their original McLean [8,9]. The framework proposed by DeLone and McLean [8,9], is widely accepted as appropriate for the evaluation of information systems. We intend to use their original McLean [8,9]. The framework proposed by DeLone and McLean [8,9], is widely accepted as appropriate for the evaluation of information systems. We intend to use their original McLean [8,9]. The framework proposed by DeLone and McLean [8,9], is widely accepted as appropriate for the evaluation of information systems. We intend to use their original McLean [8,9]. The framework proposed by DeLone and McLean [8,9], is widely accepted as appropriate for the evaluation of information systems. We intend to use their original McLean [8,9].

3. ASSESSMENT PLAN

3.1 Key Performance Indicators

All SHAMAN Work Package leaders were invited to indicate and describe relevant assessment criteria and Key Performance Indicators (KPIs) for their WP, according to the following fields:

- Title of the KPI
- Definition of the KPI
- Measurement criteria for the KPI
- Target to be achieved.

Information systems may be interpreted as having three dimensions: a technical dimension, a management dimension and a user dimension and assessment criteria are needed for all three, in order to provide a full evaluation of the systems produced within the framework. The feedback from the project’s partners [3] was analysed in the light of the combined approaches and groups of criteria (Technical, Management, and User-related KPIs) were defined for the evaluation process.

4. MAPPING TRAC, DRAMBORA AND iRODS

Since the SHAMAN data grid implementation will build on iRODS, a set of iRODS rules can be used to ensure the most comprehensive compliance with the various criteria/requirements defined by DRAMBORA and TRAC. These rules are small units of software, which can execute server-side operations, so called micro-services, in several ways (e.g. triggered by certain events, under certain conditions, manually, or according to a user-defined time schedule). Policies especially for digital objects and data management can therefore be expressed by iRODS rules. Reagan Moore and Adil Hasan have composed a set of rules to enable/support the compliance of an iRODS-based repository with the TRAC criteria [11].

In a second step, these rules have been prototypically assigned to DRAMBORA risk mitigation strategies and transferring the iRODS rules of a TRAC criterion to its corresponding DRAMBORA risk. The SHAMAN assessment framework has to verify the implementation and proper functioning of the complete set of rules.

The assessment workflow as derived from Moore [12] can be described in six steps:

1. Definition of assessment criteria (in our case: TRAC, DRAMBORA)
2. Definition of policies enforcing the assessment criteria
3. Definition of rules that apply the policies (iRODS rules)
4. Definition of capabilities that implement the required (preservation) functions (microservices)
5. Definition of (preservation) metadata that capture information about the application of the preservation functions (persistent state information, e.g. audit trails)
6. Query the (preservation) metadata to assess whether the assessment criteria have been satisfied.

Step 1 has already been performed by choosing the TRAC, DRAMBORA and information systems’ success criteria as assessment criteria for SHAMAN, steps 2 and 3 by developing the TRAC-iRODS and DRAMBORA-iRODS mappings [3].

Steps 4 and 5 will be performed by the responsible workpackages and can be seen as project outputs that are the elements of the SHAMAN technologies.

Step 6 is the actual process of performing the assessment. It can only be performed in the context of an implementation of the SHAMAN preservation framework, i.e. a running system like the demonstrators to be developed by the ISPs.

Since iRODS is a data management system, the implementation of iRODS rules focuses on management and technical aspects. Organisational and financial aspects as also addressed by the TRAC and DRAMBORA checklist can be supported, but only on the level of digital objects management. This restriction may be acceptable because SHAMAN is developing a preservation framework and not a ready-to-use preservation system. It is not within the scope of the project to develop business plans, mission statements etc. However, enough information can be gathered to make an assessment of the extent to which the fully implementable and the partially implementable criteria and risk mitigation strategies are implemented.

5. CONCLUSION

SHAMAN is a project involving eighteen partners in Europe and North America from both private and public sectors. The research conducted in the first workpackage innovates in the underlying approach to defining and validating the SHAMAN preservation mechanisms as well as in terms of how it enriches our knowledge about the characteristics that preservation systems must have. An assessment framework has been produced within this workpackage, providing assessment criteria for the whole project, for individual work packages and for outputs. While the work on the user requirements and on the Assessment Framework was defined with the work in the subsequent phases of analysis, design and implementation in mind, they will be invaluable to others who may be developing preservation systems as they provide an example of well-founded and validated requirements.

Through the public release of this deliverable we hope to assist other preservation initiatives.

6. ACKNOWLEDGMENTS

SHAMAN (Sustaining Heritage Access through Multivalent ArchiviNg) Integrated Project is co-funded by the European Union (Grant Agreement No. ICT-216736).

7. REFERENCES


[11] The TRAC-iRODS mapping tables were conceived, designed and created by Reagan Moore (DICE, University of North Carolina at Chapel Hill) with contributions from Adil Hasan (University of Liverpool). Moore and Hasan provided SHAMAN WP1 partners with this TRAC-iRODS mapping for a further analysis as part of the investigations of this deliverable.