Assessing Digital Preservation Infrastructures: Implementing a Framework for Library, Engineering and eScience Organisations

Perla Innocenti, Leo Konstantelos; Humanities Advanced Technology and Information Institute (HATII); University of Glasgow; United Kingdom

Seamus Ross; Faculty of Information; University of Toronto; Canada

Elena Maceviciute, Tom Wilson; Swedish School of Library and Information Science; University of Borås; Sweden

Jens Ludwig, Wolfgang Pempe; Goettingen State and University Library; Goettingen; Germany

Abstract

Sustaining Heritage Access through Multivalent ArchiviNg (SHAMAN) is an EU-funded project focusing on the development of an integrated preservation framework. Through grid technologies, the SHAMAN framework promotes a distributed approach in preservation systems, whereby ingest, persistent storage, access, presentation and manipulation of digital information is managed for long-term consumption. In order to understand the ever-evolving requirements for functionality in information systems, the SHAMAN team, led by HATII at the University of Glasgow, conducted an in-depth investigation of user needs for preservation solutions. The results were used to inform the development of a corresponding Assessment Framework. The purpose of the Assessment Framework is to evaluate the degree that the SHAMAN outputs are consistent with the identified user requirements and to measure the overall success of the project. The SHAMAN outputs are instantiated as functional prototypes that reflect preservation requirements in three distinct domains: memory institutions, industrial design & engineering and e-Science. Following the specifications of the assessment framework, the software artefacts produced by SHAMAN for each prototype must be assessed to validate their conformance with user and system requirements. To this end, a software validation methodology has been devised, which builds on the SHAMAN Assessment Framework to verify that the SHAMAN software satisfies the reasons for its development. This paper documents the SHAMAN Assessment Framework and explicates the relationship between assessment and software validation in the SHAMAN project.

Introduction

The Sustaining Heritage Access through Multivalent ArchiviNg (SHAMAN) Integrated project [1] examines the long-term preservation of digital heritage, with funding from the 7th Framework Programme for research and technological development of the European Commission. SHAMAN aims to develop a next generation Digital Preservation (DP) framework, which tackles the challenges of technological evolution and obsolescence and provides scalable, sustainable services for preserving increasingly complex objects and their relationships. The foundations of the SHAMAN approach lay on three distinct areas: data grids to support the creation of shared collections that are distributed across multiple institutions and locations; digital libraries to provide services for publishing, discovering, presenting and manipulating (often dynamic) data; and persistent archives to manage the long-term preservation of digital information.

In order to support the Digital Preservation framework, SHAMAN has developed a standardized Reference Architecture which provides the "architectural patterns and technical terminology" [2] to guide the development of the project's outputs. These outputs include reference implementations of "preservation tools for analysing, ingesting, managing, accessing and reusing information and digital objects across digital archives" [2]. The digital preservation framework and reference implementations form the basis for the creation of three functional research prototypes that exhibit, test and validate the principles, functionality, viability and usefulness of the SHAMAN solutions. Coordinated by the Integration and Demonstration Subprojects (ISPs), each prototype centres on a specific Domain of Focus: Memory Institutions (DoF1), Industrial Design & Engineering (DoF2) and e-Science (DoF3).

Under the lead of HATII at the University of Glasgow, the requirements for the functionality and mission of the prototypes were gathered through extensive interviews conducted with representatives from the three DoFs. The findings from the interviews were translated into use cases [3], which model the activities in the three domains and reflect the user expectations from the SHAMAN DP framework. The project team realised at an early stage the intricate nature of the relationships between user requirements and delivered research products. To this end, an assessment framework was developed to evaluate and validate the degree to which the delivered solutions extensively cover the identified requirements. The primary purpose of the assessment framework is to support the implementation of the SHAMAN project outputs. However, in taking a comprehensive approach to assessing and validating the preservation of digital libraries, archives and repositories, the application of the assessment framework can be extrapolated to the wider DP community.

This paper describes the fundamental aspects of the SHAMAN Assessment Framework [3] and the instruments that have been created to assess the DP framework as a whole and validate the software tools developed within it. To this end, we first outline the main characteristics of the SHAMAN persistent archive approach and its respective archive-centric information life-cycle. We then explain the assessment specifications, in terms of goals and outputs to be evaluated, assessment criteria and key performance indicators. In the last section, we report on an implementation of the assessment framework through a software validation methodology to verify that the SHAMAN software solutions satisfy the pre-defined use cases and requirements.

The SHAMAN Persistent Archive Environment

The SHAMAN approach to digital preservation systems is based on five interdependent constituents, which in turn provide the groundwork for designing and developing the project outputs. Specifically, the SHAMAN DP framework is perceived as a distributed, infrastructure-independent environment with policies in place to govern community goals and provisions for contextual characterisation and information discovery. The distributed environment allows for the management of very large-scale collections, through automation of administrative functions, shared custodianship of information across institutions and data integration. Data management necessitates the existence of regulatory forces for retention, disposition, distribution, replication, access, ingestion and long-term preservation of data. This necessity accentuates the organic role of policies in the SHAMAN DP framework. In particular, the preservation framework focuses on delivering automation of policies which – according to [4] - lies at the heart of long-term management of digital collections.

In parallel, the long-term management of collection properties must withstand the technological evolution and persist unrestricted from the computing environment. This is achieved by infrastructure independence, which enforces policies across the distributed network and enables the integration of technological advancements in the overall framework. Persistence over time further requires that digital objects are thoroughly documented, so that they can be understood in the future. The SHAMAN DP framework incorporates contextual metadata that provide provenance, procedural, descriptive and administrative information to characterise digital objects. Lastly, the framework supports information discovery systems which permit finding and accessing digital objects through context metadata and indexed annotations.

The SHAMAN DP framework suggests an information life-cycle that follows an archive-centric approach. This approach builds on the specifications of the Open Archival Information System (OAIS) [5] model and addresses all stages of the life-cycle depicted in Figure 1. From the above, it becomes evident that the SHAMAN DP framework conglomerates the synchronised action of organisational, technological and R&D mechanisms that together synthesise a complex matrix of interrelationships. Understanding this complexity, the development of an assessment methodology to evaluate the integrated components of the SHAMAN framework took into account the project goals, as well as extant assessing criteria and identification of key performance indicators. These issues are further discussed in the following section.

DP Framework Assessment Methodology

At the highest level of the Assessment Framework [6] lays the project's goals and outputs that will be subject to evaluation. The principal aim of SHAMAN is to furnish the community with a fresh and extensive theoretic underpinning for the development of digital preservation systems. The systems should promote augmented storage, access and discovery of digital information, through seamless integration of data management, archival and digital library technologies. Another aim is the implementation of a grid-based preservation system that will incorporate the preservation requirements identified within and across the three domains of focus for cultural heritage, engineering and e-Science data. Through this effort, SHAMAN intends to generate a knowledge base for dissemination that transcends project findings, by creating a network for sharing expertise and offering support in preservation and re-use of digital objects. These three non-orthogonal goals of the SHAMAN project dictate the hierarchy of tasks that must be evaluated and validated.

In particular, there are technical, conceptual, administrative and operational aspects that the assessment needs to consider. To facilitate development and evaluation, the SHAMAN goals and outcomes have been instantiated in demonstration scenarios that comprise of use cases and exhibit the SHAMAN outputs through demonstration applications. The content of the scenarios has been tailored to address exemplar cases within each domain of focus. The demonstrators represent working samples of the ISP prototypes, which proclaim the operations and benefits of the prototypes within each demonstration scenario. The demonstrators use the Integrated Rule-Oriented Data System (iRODS™) data grid technology as a storage substrate for digital preservation. Furthermore, the demonstrators represent the combined effort of individual work packages across the project. Hence, the assessment methodology must incorporate benchmarking tools and risk mitigation mechanisms, as well as evaluation methods directly stemming from the rule-oriented data system and from information systems success criteria.

In order to address these issues, the assessment methodology utilizes assessing criteria from extant efforts, specifically deriving from TRAC, DRAMBORA, iRODS rules and Information Systems models. The following sections provide more insight into these efforts and their role within the SHAMAN assessment framework.

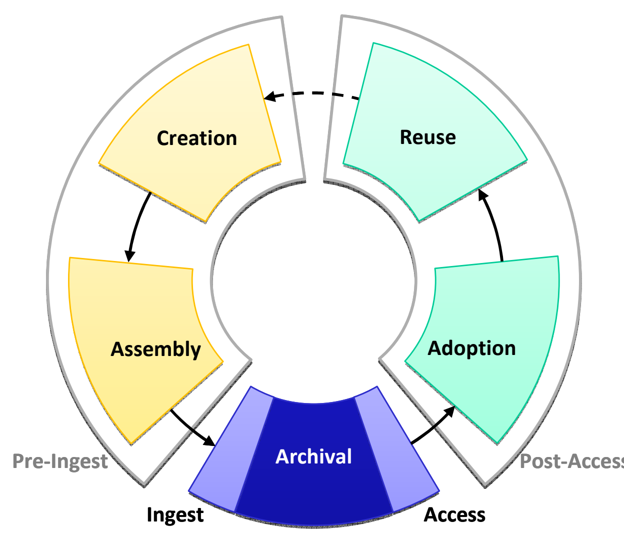

***Figure 1**. The SHAMAN Archive-centric Information life-cycle © SHAMAN project*

TRAC and DRAMBORA Criteria

For the assessment methodology to include benchmarking and risk mitigation mechanisms, the project team reviewed two well-established and recognized tools: TRAC and DRAMBORA. The Trustworthy Repositories Audit and Certification (TRAC) Criteria and Checklist [7] is an auditing tool, which aids in establishing the reliability, commitment and readiness of institutions to undertake long-term preservation responsibilities. Similarly to the SHAMAN archive-centric approach, TRAC is based on OAIS so as to provide benchmark criteria to certify the trustworthiness of repositories for digital preservation.

Along the same lines, the Digital Repository Audit Method Based on Risk Assessment (DRAMBORA) [8] is a methodology and interactive online toolkit that support the assessment of digital repositories through evidence-based risk management. DRAMBORA considers the institutional context and the organizational, technical and managerial structures pertaining to digital repository environments, in order to classify and evaluate the risks associated with ingest, curation and access to authentic digital information (Figure 2).

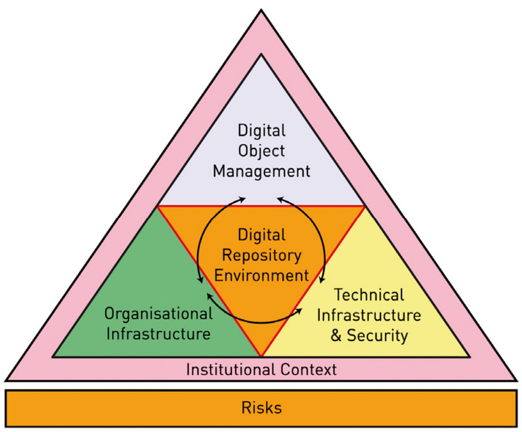

Figure 2. *DRAMBORA: Interrelationships within a digital repository environment © HATII at the University of Glasgow*

The analysis of TRAC and DRAMBORA revealed that the SHAMAN assessment methodology should provide the criteria to prove that a preservation system following the design and deployment specifications of the SHAMAN DP framework successfully supports the TRAC/ DRAMBORA rules. In this manner, it is possible to not only assess the outcomes of this project, but also provide a comprehensive methodology for future projects building on the SHAMAN approach to evaluate the affordances of a repository for digital preservation.

iRODS Rules Criteria

Since SHAMAN is employing the iRODS [9] data grid technology for data management and storage, it was decided that assessment of conformance with iRODS rules should be included in the methodology for evaluating the SHAMAN DP framework. The "iRODS Rule Engine" forms the core of the iRODS system, providing the means to manage, invoke and execute policies through automated services [10].

The purpose of including iRODS rules in the assessment of the SHAMAN framework is twofold: first, the capacity of individual framework elements to incorporate, showcase and support these rules must be evaluated. This also includes project outputs, such as software components and conceptual schemes that utilize the grid-based approach. Second, the results of this part of the assessment will shed light on the ability of preservation systems in general to function under these rules and benefit from their adoption.

Information Systems Success Criteria

The SHAMAN DP Framework effectively represents an abstract instance of a preservation system which in turn encapsulates a representation of an information system. Recognizing this analogy, the assessment working group has incorporated evaluation criteria in the methodology that derive from Information System (IS) Models. In particular, the success criteria introduced by DeLone and McLean [11-12] have been primarily considered, due to their wide acceptance for IS evaluation and their appropriateness for the SHAMAN case. The IS success model [11] specifies the dimensions, interrelationships and measures associated with IS success in seven distinct areas. Each area promotes system evaluation based on:

1. System quality criteria
2. Information quality criteria
3. Use criteria
4. User satisfaction
5. Individual impact
6. Organizational impact
7. Essential properties

The above criteria cover the entire range of functions, goals and outputs of the SHAMAN DP framework and are therefore essential in measuring its success in delivering the expected results. Further criteria were derived from the IEEE *Recommended practice for software requirements specifications* [13]. Although this document primarily offers guidance in delivering software specifications, the requirements can be projected to evaluating the SHAMAN technical outputs, particularly in terms of (1) Functionality; (2) External interfaces; (3) performance; (4) system attributes; and (5) Design constraints upon software implementation.

Definition of an Assessment Plan

The components of the assessment methodology presented in the previous section are not meant to be applied individually to the evaluation of SHAMAN. Instead, an assessment plan has been devised, which combines criteria and approaches in extant efforts so as to evaluate the SHAMAN efforts at all phases of the project's life-cycle. Front-end evaluation was conducted at the beginning of the project through the user needs assessment interviews and deriving use cases. The assessment plan further specifies formative design-and-evaluation cycles, which continually gauge the quality of the project outputs and designate their progress towards reaching the project's high-level goals.

In order to achieve this, the evaluation team encouraged the SHAMAN work package (WP) leaders to identify and explicate assessment criteria relevant to the areas of their WP activities. The process concluded with the definition of Key Performance Indicators (KPIs) for each work package, which provide: information to describe their content; measurement criteria for tasks within the WPs; and the respective targets to be achieved. The KPIs have been integrated with the technical, managerial and user-related criteria.

This integration is evident in the mapping of iRODS rules to TRAC and DRAMBORA. As explained previously, iRODS operates upon a set of rules expressing policies. In system terms, these rules are instantiated as small programmatic units (micro-services) that execute in a networked environment and trigger events that are based on pre-defined conditions. In the context of SHAMAN – and indeed preservation systems generally – iRODS rules are particularly apposite for defining policies regarding digital object ingest, access and preservation, and related data management operations. Efforts within SHAMAN have concentrated on creating a set of rules that reinforce conformance of IRODS-based repositories with the TRAC criteria. Building on these efforts, the generated rules have been further mapped to DRAMBORA specifications for risk mitigation by assignment of iRODS rule / TRAC tuples to relevant DRAMBORA risks.

In order to establish compliance of the SHAMAN DP framework with the iRODS / TRAC / DRAMBORA specifications, the assessment plan includes a workflow that can be used to verify adherence of outputs to the aforementioned matrix of rules. This assessment workflow (derived from [14]) incorporates a number of consecutive steps. In the first instance, the assessment criteria need to be defined. This process has been documented in the section on the DP Framework Assessment Methodology, where TRAC / DRAMBORA and Information System Models have been selected as the sources of evaluation measuring tools. The next step is to identify the relevant policies that govern and accomplish the assessment criteria. Furthermore, the rules that apply to the policies need to be examined. In effect, these steps involve the mappings between iRODS rules and TRAC/DRAMBORA criteria previously discussed. Once this matrix of rules has been defined, the workflow continues with the identification of a means to achieve the required functionality and the formation of preservation metadata to describe this functionality. In the IRODS-based environment, the implementation of required preservation functions is achieved through micro-services, while metadata such as persistent state information are recorded by the system. In the last step of the workflow, the entire set of recorded metadata is queried to evaluate whether the assessment criteria have been met.

The above assessment plan is meant to be initially applied to the demonstrator applications developed by the ISPs, as exemplary implementations of the SHAMAN DP framework. In the next section we present an application of the assessment plan on a validation methodology to verify the technical competence of the SHAMAN demonstrators for memory institutions.

Software Validation of ISP1 Demonstrators

Building on the Assessment Framework and plan, a methodology has been devised to validate the software components within the SHAMAN demonstrators. At present, software validation will be applied to the upcoming ISP1 Demonstrators for Domain of Focus 1, Memory Institutions. These demonstrators focus on four scenarios covering (1) indexing and archiving book-like publications; (2) indexing and archiving digitisations; (3) scientific publishing and archiving of heterogeneous interlinked material; and (4) processing and archiving web harvesting material.

Software validation investigates the correctness, completeness, accuracy, consistency and testability of software requirements, further determining the degree that the software satisfies its intended use and users through analysis, evaluation, review, inspection, assessment and testing of products and processes [15]. The purpose of software validation is to ensure that the software performs its intended functions, but also to eliminate unintended functions and measure its quality and reliability [16]. Validation engages in testing software and specifications at the end of the development effort to establish conformance with the overall system requirements (i.e. that the system does what it is supposed to). The criteria for validating the SHAMAN demonstrators at a software level have been drawn from the IEEE Standard for Software Verification and Validation [15]. The standard is widely used in software development and its framework can be adjusted to the specific needs of individual projects. The methodology is based on the Standard and utilises its practices, validation activities and measures to generate a bespoke validation plan for the SHAMAN ISP1 demonstrator.

The validation methodology addresses four activities within the development phase which relate to concept, requirements, design and implementation of the ISP1 demonstrators. Concept validation assesses whether the solutions that the demonstrators bring to the digital preservation problems are real and devoid of any false assumptions. Requirements validation ensures that the functional and performance requirements for the demonstrators accord with the objectives of the DP framework and provide a complete, accurate and consistent account of the use cases derived from the user study. Design validation evaluates whether software requirements have been correctly translated into a design specification to guide the implementation of the demonstrators. Lastly, Implementation validation verifies that the design specifications have been followed and that the final output accurately and entirely meets the software requirements.

A selection of tools specified in [15] will be used to validate the ISP1 demonstrators. Document evaluation ensures that the documentation of the demonstrators – from concept to implementation – satisfies the needs for re-use in the future. Traceability Analysis maps functional requirements with the user needs covered in the use cases. Software design validation confirms the existence and correctness of technical blueprints that regulate the implementation of software components. Implementation validation tests whether the final products perform the intended functions and reports on errors and / or omissions. From the above, it becomes evident that the software validation methodology inherits the properties of the assessment framework. However, by being a sub-part of the assessment framework, software validation only addresses system quality criteria.

As a first step towards employing the validation methodology, an exhaustive listing of functional requirements for the demonstrators has been generated, which derives directly from the use cases for DOF1. These functional requirements specify the entire range of functions and operations that the demonstrators need to incorporate, in order to fulfill their role as paradigmatic implementations of the ISP1 prototype and consequently the overall SHAMAN DP framework. This list of functional requirements has been translated into achievement indicators. Each requirement will be validated against one or more indicators, which have been formulated to precisely assess the demonstrator specifications. Validation at this level will be performed in a check-list manner through an online validation form. The form will be distributed to project partners and particularly demonstrator developers. The level of achievement for each indicator will be accompanied by appropriate justification. The format of this validation instrument has been based on previous efforts (e.g. the *Frescor* project [17]).

Although software validation is currently being applied to the ISP1 demonstrators, the specifications can be straightforwardly tailored to match the requirements of future software development within SHAMAN for the engineering and e-Science domains of reference.

Conclusions

This paper has presented an overview of a methodology devised to assess the SHAMAN Digital Preservation Framework. In doing so, we have described the organic components of the SHAMAN approach and infrastructure, as well as explicated the foundations for building an Assessment Framework. Inevitably, the assessment methodology focuses on data management and technical aspects, because of the nature of the SHAMAN project. There is a huge arena of organizational and financial characteristics, which – although outside the scope of SHAMAN – need to be considered. TRAC and DRAMBORA support the assessment of these repository facets and employment of the SHAMAN Assessment Framework in a real-life setting will require such organizational criteria to be addressed. Building on the process followed to define KPIs for the SHAMAN prototype, further indicators can be derived to assess the implementation of software components, risk mitigation strategies and organizational needs. The ultimate goal of SHAMAN is exactly to create a verifiable, open and extensible digital preservation framework for both current and future use. By presenting the SHAMAN Assessment Framework, we expect to contribute to the work of other preservation system development efforts and offer the community a working example of a comprehensive, validated methodology.

Acknowledgments

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

References

[1] Sustaining Heritage Access through Multivalent ArchiviNg (SHAMAN) Project website, 2009. http://www.shaman-ip.eu

[2] Dobreva, M., Nicholson, D., Hemmje, M., Borbinha, J., Hasan, A. 2009. Specification of the SHAMAN Reference Architecture, SHAMAN Project. http://shaman-ip.eu/shaman/document

[3] Innocenti, P. Aitken, B. Hasan, A. Ludwig, J. Maciuvite, E. Barateiro, J. Antunes, G. Mois, M. Jäschke, G. Pempe, W. Wilson, T. Hundsdoerfer, A. Krandstedt, A. Ross, S. 2009. SHAMAN Requirements Analysis Report (public version) and Specification of the SHAMAN Assessment Framework and Protocol, SHAMAN Project. http://shaman-ip.eu/shaman/sites/default/files/SHAMAN_D1.2_Requirements%20analysis%20report_0.pdf

[4] Ross, S. Digital Preservation, Archival Science and Methodological Foundations for Digital Libraries, Keynote Speech at the European Conference on Research and Advanced Technology for Digital Libraries (ECDL) 2007, Budapest, Hungary, 17 September 2007. http://www.ecdl2007.org/Keynote_ECDL2007_SROSS.pdf

[5] Consultative Committee for Data Space Systems, Reference Model for Open Archival Information Systems (OAIS), CCSDS, January 2002 http://public.ccsds.org/publications/archive/650x0b1.pdf

[6] Innocenti, P. Ross, S. Maceviciute, E., Wilson, T. Ludwig, J. Pempe, W. Assessing Digital Preservation Frameworks: the approach of the SHAMAN project, International ACM Conference on Management of Emergent Digital EcoSystems (MEDES'09), Proceedings of the International Conference on Management of Emergent Digital EcoSystems, France 2009

[7] CRL, 2008. Trustworthy Repositories Audit and Certification (TRAC) Criteria and Checklist, Centre for Research Libraries, Chicago. http://www.crl.edu/content.asp?l1=13&l2=58&l3=162&l4=9

[8] DigitalPreservationEurope & Digital Curation Centre, 2008. DRAMBORA Interactive: Digital Repository Audit Method Based on Risk Assessment. http://www.repositoryaudit.eu/

[9] iRODS Integrated Rule-Oriented Data System. https://www.irods.org/

[10] Introduction to iRODS. https://www.irods.org/index.php/Introduction_to_iRODS

[11] DeLone, W.H. & McLean, E.R. Information systems success: the quest for the dependent variable. Information Systems Research, 3(1), 60-95. (1992).

[12] DeLone, W.H. & McLean, E.R. The DeLone and McLean model of information systems success: a ten-year update. Journal of Management Information Systems, 19(4), 9-30 (2003).

[13] Institute of Electrical and Electronics Engineers, 1998. IEEE recommended practice for software requirements specifications, Institute of Electrical and Electronics Engineers. (IEEE Std. 830-1998), New York, NY.

[14] Moore, R. Towards a Theory of Digital Preservation. The International Journal of Digital Curation. 1(3), 63-75 (2008). http://www.ijdc.net/index.php/ijdc/article/view/63/42

[15] Institute of Electrical and Electronics Engineers, 2004. IEEE Standard for Software Verification and Validation. (IEEE Std 1012™-2004), New York, NY.

[16] Wallace, D. R. and Fujii, R. U. Software Verification and Validation: An Overview. IEEE Software, 6(3), 10-17 (1989).

[17] FRESCOR Project Evaluation Report II. FRESCOR Deliverable D-EVv2. http://www.frescor.org/index.php?mact=Uploads,cntnt01,getfile,0&cntnt01showtemplate=false&cntnt01upload_id=93&cntnt01returnid=54

Author Biographies

Perla Innocenti is co-Principal Investigator in the EU-funded projects SHAMAN and DL.org, and Research Associate at HATII, University of Glasgow. Her research interests include digital preservation methodologies, preservation of media art, audit and risk assessment for digital repositories, digital libraries design and usage models, digitization methodologies. She was involved in repository design, audit research as part of DigitalPreservationEurope (DPE) and Digital Curation Center (DCC), co-ordinating activities and development for the DRAMBORA) Toolkit, usage models research within the EU-funded project Planets, and the investigation of the potential application of the DRAMBORA toolkit in digital libraries within the DELOS project.

Dr. Leo Konstantelos is a Preservation Resources Officer in the Humanities Advanced Technology and Information Institute (HATII) at the University of Glasgow. He has conducted research into preservation of interactive and ephemeral digital content for the Planets project, and into a methodology for software validation for the SHAMAN Integrated Project. Leo holds a PhD in Humanities Computing in the area of user studies for Digital Art in Digital Libraries. He has delivered a number of seminars on digital libraries, user studies, statistical methods and the digital arts. He was a member of the DELOS Network of Excellence on Digital Libraries.

Seamus Ross is Dean and Professor, Faculty of Information, University of Toronto. Formerly, he was Professor of Humanities Informatics and Digital Curation and Founding Director of HATII (Humanities Advanced Technology and Information Institute) (1997-2009) at the University of Glasgow. He served as Associate Director of the Digital Curation Centre (2004-9) in the UK, and was Principal Director of ERPANET and DigitalPreservationEurope (DPE) and a co-principal investigator such projects as the DELOS Digital Libraries Network of Excellence and Planets. He recommends Digital Preservation and Nuclear Disaster: An Animation [video] (1999), and "Digital Archaeology" [PDF] (1999).

Dr. Elena Maceviciute is Professor, Swedish School of Library and Information Science, University of Boras, and holds a similar position at Vilnius University, Lithuania. Her research has covered a wide variety of topics in information management and digital libraries. She has worked on a number of European projects and is member of the editorial boards of a number of recognized information science journals.

Dr. Tom Wilson, is Professor Emeritus, University of Sheffield and Visiting Professor, University of Boras. His research has been chiefly in two areas: information management, where he has studied the impact of information technology in organizations, and information behaviour, a field in which is he one of the most cited authors. He publishes and acts as Editor in Chief of the electronic journal, Information Research.

Jens Ludwig studied philosophy and computer science, and since early 2005 is an employees of the Goettingen State and University Library. At the end of 2008 he also took a degree in Library and Information Science. He was a developer in Kopal project, in nestor II is Head of the AGs standards and is also engaged in the EU-funded project SHAMAN. His activities include the conceptual and technical work in the areas of long-term archiving and virtual research environments.

Wolfgang Pempe is research associate at the Research and Development Department of the Goettingen State and University Library and involved in the projects DARIAH, SHAMAN and TextGrid. After graduating in Assyriology he worked as software developer / XML expert at a typesetting company and an internet agency. He is co-founder and director of Saphor GmbH, an IT service provider.