



Project no. 033572

CASPAR

Cultural, Artistic and Scientific knowledge for Preservation, Access and Retrieval

Instrument: Information Society Technologies

Thematic Priority: 2.5.10 Access to and preservation of cultural and scientific resources

REPORT ON EVALUATION CRITERIA



| | |
|----------------------|--------------------------------|
| Document identifier: | CASPAR-4103-RP-0101-1_0 |
| Submission Date: | 15-01-2009 |
| Due date: | 15-01-2009 |
| Work package: | 4100 |
| Partners: | All Partners |
| WP Lead Partner: | STFC/ERC |
| Document status | FINAL |

Abstract: This document provides the implementation plan for the CASPAR testbed.



Delivery Type Report
Author(s) CASPAR Consortium
Approval David Giaretta
Summary
Keyword List
Availability PUBLIC

Document Status Sheet

| Issue | Date | Comment | Author |
|-------|--------------|---------------------------|---|
| 0_0 | 30 June 2008 | Material gathered on Wiki | Whole consortium |
| 1_0 | 15 Jan 2009 | ERC report | ERC with formatting changes by David Giaretta |
| | | | |
| | | | |
| | | | |
| | | | |
| | | | |
| | | | |





Project information

| | |
|------------------------|---|
| Project acronym: | CASPAR |
| Project full title: | Cultural, Artistic and Scientific knowledge for Preservation, Access and Retrieval |
| Proposal/Contract no.: | IST-2006-033572 |

Project Officer: Carlos Oliveira

| | |
|----------|---|
| Address: | INFSO-E3 Information Society and Media Directorate General Content - Learning and Cultural Heritage Postal mail: Bâtiment Jean Monnet (EUFO 1167) Rue Alcide De Gasperi / L-2920 Luxembourg Office address: EUROFORUM Building - EUFO 1167 10, rue Robert Stumper / L-2557 Gasperich / Luxembourg |
| Phone: | +352 4301 33052 |
| Fax: | +352 4301 33190 |
| Mobile: | |
| E-mail: | Carlos.Oliveira@ec.europa.eu |

Project Co-ordinator: David Giarretta

| | |
|----------|---|
| Address: | STFC (formerly CCLRC), Rutherford Appleton Laboratory Chilton, Didcot, Oxon OX11 0QX, UK |
| Phone: | +44 1235 446235 |
| Fax: | +44 1235 446362 |
| Mobile: | +44 (0) 7770326304 |
| E-mail: | d.i.giarretta@rl.ac.uk |





CONTENT

| | | |
|----------|---|-----------|
| 1 | INTRODUCTION | 5 |
| 2 | DETAILED MATERIAL | 13 |
| 2.1 | PEER REVIEWED PAPERS IN JOURNALS AND AT CONFERENCES | 13 |
| 2.2 | STANDARDS ACTIVITIES | 16 |





1 INTRODUCTION

The CASPAR project has defined a set metrics and evaluation criteria in Table 1 in the CASPAR Description of Work [Ref 1].

The External Review Committee (ERC) was tasked to:

1. Evaluate the metrics and suggest other or more detailed metrics
2. Periodically evaluate these metrics

The ERC review of the metrics and evaluation criteria found them to be complete with no significant revisions needed.

This report evaluates the progress on each metrics for the period before the EU review on January 30-31,2009. The Table below is an extension of Table 1 in the CASPAR Description Of Work with a column added for a brief evaluation comment. Additional detailed material will be found in sections of this document referenced from the brief assessment column.





Table 1 CASPAR Measurable Objectives

| OBJECTIVE | PROGRESS ASSESSMENT METHODOLOGY | CURRENT ERC ASSESMENT |
|--|--|---|
| DIGITAL PRESERVATION METRICS | | |
| 1) Demonstrate a sound theoretical basis for the approach taken, including | | |
| a. the compatibility with the OAIS Reference Model and related standards – all of which have been peer reviewed extensively in the standards process itself and also by practitioners of digital preservation in a great number of areas | Review of CASPAR architecture and OAIS components with the External Review Committee – showing the relationship to OAIS and the relevance to digital preservation. | EXCELLENT: The CASPAR architecture and design documents have sections mapping the CASPAR Components into the OAIS RM CASPAR has also given several presentations and published several papers demonstrating the implementation of the needed core functionality within the OAIS RM framework. They have also developed formal requests to the OAIS Five Year Review team for clarifying the OAIS RM |
| b. the development of a peer-reviewed theoretical under-pinning for the specifics of the CASPAR approach | Paper(s) submitted to peer-reviewed journal, and presented at international conference(s) | EXCELLENT: See section 2.1 of this document for details of papers |
| 2) Provide a practical demonstration by means of what may be regarded as “accelerated lifetime” tests. These should involve demonstrating the ability of the Framework and digital information to survive: | | |
| a. environment (including software, hardware) changes | Demonstration to the External Review Committee of usability of a variety of digitally encoded information despite changes in hardware and software of user systems, and such processes as format migration for, for example, digital science data, documents and music | INCOMPLETE: The scenarios for demonstrating this preservation threat are complete in however the implementation of these scenarios is not yet complete |





| | | |
|---|--|--|
| b. changes in the Designated Communities and their Knowledge Bases | Demonstration to the External Review Committee of usability of a variety of digitally encoded information by users of different disciplines | INCOMPLETE: The scenarios for demonstrating are complete in however the implementation of these scenarios is not yet complete |
| 3) Show improved trustworthiness of repositories | | |
| | Use (draft) certification standard at various stages of application of CASPAR techniques to a number of repositories. The repositories should improve in trustworthiness. | GOOD: CASPAR chairs the CCSDS/ISO Registry Certification and Audit Working Group is generating a draft of two CCSDS/ISO Recommendations that will provide the Requirements for certification of "long-term Repositories" and the requirements and processes for an audit and certification organization. CASPAR has also produced documents explaining "digital trust" and has emphasized preservation threats in test scenarios. CASPAR should provide an exemplar for other repositories |
| 4) Review of evaluation methodology for: | | |
| a. CASPAR Framework evaluation | External Review Committee report on proposed validation techniques for CASPAR Framework | OK: The documentation in D3301 adequately describes the Framework testing approach. A separate volume for performance testing is probably not needed since factors such as scalability would be unlikely to be testable in the timeframe of this project |
| b. CASPAR Testbed evaluation | External Review Committee report on proposed validation techniques for methodology for validation of the Testbeds | UNKNOWN: The scenarios for the testbeds testing are complete. The Testbed Implementation Plan is reasonably complete. However the implementation of these scenarios is not yet complete. |
| 5) Demonstrate the applicability of the CASPAR virtualisation techniques for digital information in the following areas (all of which also include | | |





| | | |
|---|--|---|
| normal documents of various kinds), and their maintainability over time. | | |
| a. Science, Cultural Heritage, Contemporary Arts | Demonstrate to External Review Committee some specific examples virtualisation of data from a number of disciplines and show usability. | INCOMPLETE: The scenarios for demonstrating are complete in however the implementation of these scenarios is not yet complete |
| b. Digital Rights Management | Demonstrate to External Review Committee the use of the DRM description for a variety of types of data and rights | OK: There are complete scenarios, requirements, interfaces, design and some code. However no significant testing has been done in the testbeds |
| c. Preservation Description Information | | |
| i. Provenance | Demonstrate to External Review Committee the capture of provenance information with change of ownership, events, processing etc and can allow for evolution of these over time. | INCOMPLETE: The scenarios for demonstrating are complete in however the implementation of these scenarios is not yet complete |
| ii. Persistent identification | Demonstrate to External Review Committee the use of a system of Persistent Identifiers together with an analysis of the socio-economic survivability of the infrastructure supporting this identifier. | INCOMPLETE: The documentation for defining the persistent ID issues were identified in D2301 however there is no specific implementation defined |
| iii. Authenticity | Demonstrate to External Review Committee the system for proving authenticity of a digital object, together with an analysis of the maintainability of this authenticity system over time and also the issues arising from the transformation of the digital encoding of the information. | INCOMPLETE: There is a large amount of documentation on both the requirements and design of an authentication mechanism |
| PROJECT IMPACT METRICS | | |
| Show contribution to standards | Production of evidence of contributions to the development of some or all of the standards in section 5.1 | EXCELLENT: CASPAR has made significant contributions in many of the standards areas listed in 5.1. Section 2.2 of this report discusses some of these contributions |
| Show adoption of CASPAR results in the | Demonstrate effective adoption of the CASPAR by | INCOMPLETE: Premature The |





| | | |
|--|---|--|
| preservation community | users involved in test beds and number of entities external to the project, interested in the CASPAR framework and endorsing it in some way | dissemination and training activities have not yet occurred in the science and However, the performing arts community is redeveloping existing tool based on CASPAR concepts and components |
| Show that effectiveness of dissemination processes | Indicate a number of providers of tools and services of digital preservation interested in CASPAR framework and developing complementary services or tools | EXCELLENT: The CASPAR Preservation Data Store is the basis of an IBM Preservation Data Store that is being heavily publicized by IBM. It is very rare to see a commercial vendor develop a product based on a research project prior to the end of the research effort. |
| OBJECTIVE | PROGRESS ASSESSMENT METHODOLOGY | CURRENT ERC ASSESMENT |
| DIGITAL PRESERVATION METRICS | | |
| 1) Demonstrate a sound theoretical basis for the approach taken, including | | |
| c. the compatibility with the OAIS Reference Model and related standards – all of which have been peer reviewed extensively in the standards process itself and also by practitioners of digital preservation in a great number of areas | Review of CASPAR architecture and OAIS components with the External Review Committee – showing the relationship to OAIS and the relevance to digital preservation. | EXCELLENT: This group has published several papers demonstrating the implementation of the needed core functionality within the OAIS RM framework. See section 2.1 for a list of those documents. They have also sent formal requests to the OAIS Five Year Review team. |
| d. the development of a peer-reviewed theoretical under-pinning for the specifics of the CASPAR approach | Paper(s) submitted to peer-reviewed journal, and presented at international conference(s) | EXCELLENT: See section 2.1 of this document for details of papers |
| 2) Provide a practical demonstration by means of what may be regarded as “accelerated lifetime” tests. These should involve demonstrating the ability of the Framework and digital information to survive: | | |





| | | |
|---|--|--|
| c. environment (including software, hardware) changes | Demonstration to the External Review Committee of usability of a variety of digitally encoded information despite changes in hardware and software of user systems, and such processes as format migration for, for example, digital science data, documents and music | INCOMPLETE: The scenarios for demonstrating are complete in however the implementation of these scenarios is not yet complete |
| d. changes in the Designated Communities and their Knowledge Bases | Demonstration to the External Review Committee of usability of a variety of digitally encoded information by users of different disciplines | INCOMPLETE: The scenarios for demonstrating are complete in however the implementation of these scenarios is not yet complete |
| 3) Show improved trustworthiness of repositories | | |
| | Use (draft) certification standard at various stages of application of CASPAR techniques to a number of repositories. The repositories should improve in trustworthiness. | GOOD: The CASPAR leadership of the CCSDS/ISO Registry Certification and Audit Working Group and the emphasis on preservation threats in test scenarios should provide an exemplar for other repositories |
| 4) Review of evaluation methodology for: | | |
| c. CASPAR Framework evaluation | External Review Committee report on proposed validation techniques for CASPAR Framework | OK: The documentation in D3301 adequately describes the Framework testing approach. A separate volume for performance testing is probably not needed since factors such as scalability would be unlikely to be testable in the timeframe of this project |
| d. CASPAR Testbed evaluation | External Review Committee report on proposed validation techniques for methodology for validation of the Testbeds | UNKNOWN: The scenarios for the testbeds testing are complete. The Testbed Implementation Plan is reasonably complete. However the implementation of these scenarios is not yet complete. |
| 5) Demonstrate the applicability of the CASPAR virtualisation techniques for digital information in the following areas (all of which also include | | |





| | | |
|---|--|--|
| normal documents of various kinds), and their maintainability over time. | | |
| a. Science, Cultural Heritage, Contemporary Arts | Demonstrate to External Review Committee some specific examples virtualisation of data from a number of disciplines and show usability. | INCOMPLETE: The scenarios for demonstrating are complete in however the implementation of these scenarios is not yet complete |
| b. Digital Rights Management | Demonstrate to External Review Committee the use of the DRM description for a variety of types of data and rights | OK: There are complete scenarios, requirements, interfaces, design and some code. However no significant testing had been done in the testbeds |
| c. Preservation Description Information | | |
| i. Provenance | Demonstrate to External Review Committee the capture of provenance information with change of ownership, events, processing etc and can allow for evolution of these over time. | INCOMPLETE: The scenarios for demonstrating are complete in however the implementation of these scenarios is not yet complete |
| ii. Persistent identification | Demonstrate to External Review Committee the use of a system of Persistent Identifiers together with an analysis of the socio-economic survivability of the infrastructure supporting this identifier. | INCOMPLETE: The documentation for defining the persistent ID issues were identified in D2301 however there is no specific implementation defined |
| iii. Authenticity | Demonstrate to External Review Committee the system for proving authenticity of a digital object, together with an analysis of the maintainability of this authenticity system over time and also the issues arising from the transformation of the digital encoding of the information. | INCOMPLETE: There is a large amount of documentation on both the requirements and design of an authentication mechanism |
| PROJECT IMPACT METRICS | | |
| Show contribution to standards | Production of evidence of contributions to the development of some or all of the standards in section 5.1 | EXCELLENT |
| Show adoption of CASPAR results in the preservation community | Demonstrate effective adoption of the CASPAR by users involved in test beds and number of entities external to the project, interested in the CASPAR | INCOMPLETE: Premature |





| | | |
|--|---|---|
| | framework and endorsing it in some way | |
| Show that effectiveness of dissemination processes | Indicate a number of providers of tools and services of digital preservation interested in CASPAR framework and developing complementary services or tools | EXCELLENT: The CASPAR Preservation Data Store is the basis of an IBM commercial Preservation Data Storage |





2 DETAILED MATERIAL

2.1 PEER REVIEWED PAPERS IN JOURNALS AND AT CONFERENCES

The following tables, which were adapted from the CASPAR Dissemination Report (D5202) provide a list of most of the peer reviewed papers written by CASPAR personnel. In addition the CASPAR Dissemination report lists over 60 presentation at different professional conferences

| Title | Relevant Conference | Author/s |
|---|--|---|
| Preservation Workflows, Strategies and Infrastructure | Papers accepted for DigCCurr 2009: Digital Curation Practice, Promise and Prospects | David Giaretta et al |
| Data access and long term data & knowledge preservation for Earth Science: An overview of some ESA initiatives | | Sergio Albani, Vincenzo Beruti, Luigi Fusco and David Giaretta |
| Spreading the burden infrastructure and tools for preservation and use of digital resources | | David Giaretta |
| Preservation Infrastructure in Context | Paper accepted for the African Digital Scholarship and Curation, Pretoria, South Africa, 12-14 May 2009. | David Giaretta |
| How does one know which repository is worth its salt? | | David Giaretta |
| Authenticity and Provenance in Long Term Digital Preservation: Modeling and Implementation in Preservation Aware Storage | Paper accepted for 1st Workshop on the Theory and Practice of Provenance (TaPP '09) February 23, 2009, San Francisco, CA | <i>Michael Factor, Ealan Henis, Dalit Naor, Simona Rabinovici-Cohen, Petra Reshef, Shahar Ronen, Giovanni Michetti, Maria Guercio</i> |
| Ontology-Based Temporal Modeling of Provenance Information | MELECON 08 - The 14th IEEE Mediterranean Electrotechnical Conference | <i>A.Mikroyannidis, B. Ong, K. Ng and D. Giaretta</i> |
| On Storage Policies for Semantic Web Repositories that Support Versioning | ESWC 2008 - 5th European Semantic Web Conference | <i>Y. Tzitzikas, Y. Theoharis and D. Andreou</i> |
| A Formal Approach for RDF/S Ontology Evolution (accepted for publication) | ECAI 2008 - 18th Conference on Artificial Intelligence | <i>G. Konstantinides, G. Flouris, G. Antoniou and V. Christophides</i> |
| (Semantic Web) Evolution through Change Logs: Problems and Solutions, Artificial Intelligence and Applications | AIA'2007, Innsbruck, Austria, February 2007 | <i>Yannis Tzitzikas and Dimitris Kotzinos</i> |





| | | |
|---|--|--|
| Ontology for Preservation of Interactive Multimedia Performances | MTSR '07 - 2nd International Conference on Metadata and Semantics Research | <i>K. Ng, T.V. Pham, B. Ong, A. Mikroyannidis and D. Giaretta</i> |
| Dependency Management for the Preservation of Digital Information | DEXA 07 - 18th International Conference on Database and Expert Systems Applications | <i>Y. Tzitzikas</i> |
| Mind the (Intelligibility) Gap | ECDL 2007 - 11th European Conference on Research and Advanced Technology for Digital Libraries | <i>Y. Tzitzikas and G. Flouris</i> |
| Terminology and Wish List for a Formal Theory of Preservation | PV 2007 - Ensuring Long-term Preservation and Value Adding to Scientific and Technical Data | <i>Giorgios Flouris and Carlo Meghini</i> |
| CASPAR project and the validation of digital preservation technique | | <i>C. Meghini, D. Giaretta, L. Fusco, S. Ross, M. Guercio, M. Hernandez, U. Di Giammatteo, Z. Bjelogrljic, D. Naor, S. Boi, D. Terrugi, K. Ng, L. Briguglio, V. Christophides, B. Bachimont, H. Vinetand and P. Sedlak</i> |
| Dependency Management for the Preservation of Digital Information | | <i>Y. Tzitzikas</i> |
| Some preliminary ideas towards a theory of digital preservation | | <i>Giorgos Flouris and Carlo Meghini</i> |
| Preservation DataStores: Architecture for Preservation Aware Storage | | SST '07 - IEEE Conference on Mass Storage Systems and Technologies/ Published in the conference proceedings (Proceedings of the IEEE Conference on Mass Storage Systems and Technologies (MSST), September 24-27, 2007, San Diego, USA). |
| Steps towards a theory of information preservation | PRESDB07 - International Workshop | <i>Giorgos Flouris and Carlo Meghini</i> |





| | | |
|--|--|---|
| | on Database Preservation | |
| On the Foundations of Computing Deltas between RDF models | ISWC 07 - The 6th International Semantic Web Conference and the 2nd Asian Semantic Web Conference | <i>D. Zeginis, Y. Tzitzikas and V. Christophides - this paper won the Best Paper Award</i> |
| Ontology-Driven Digital Preservation of Interactive Multimedia Performances | Proceedings of the 3rd International Conference on Automated Production of Cross Media Content for Multi-channel Distribution. | <i>Alexander Mikroyannidis, Bee Ong, Kia Ng and David Giaretta</i> |
| On Graph Features of Semantic Web Schemas | Hellenic Data Management Symposium, 2006 | <i>Yannis Theoharis, Vassilis Christophides, Yannis Tzitzikas,</i> |

Table 2:PAPERS PRESENTED AT CONFERENCE

| TITLE | JOURNAL | AUTHORS |
|--|--|--|
| CASPAR and a European Infrastructure for Digital Preservation | European Research Consortium for Informatics and Mathematics Journal special issue on European Digital Library | <i>David Giaretta</i> |
| The CASPAR Approach to Digital Preservation." | International Journal of Digital Curation 2, no. 1 (2007) | <i>Giaretta, David.</i> |
| The Need for Preservation Aware Storage - A Position Paper | ACM SIGOPS Operating Systems Review, Special Issue on File and Storage Systems, Volume 41, Issue 1 (January 2007), pages 19-23 | <i>M. Factor, D. Naor, S. Rabinovici-Cohen, L. Ramati, P. Reshef and J. Satran</i> |
| Preservation DataStores: New storage paradigm for preservation environments | IBM Journal of Research and Development on Storage Technologies and Systems, Volume 52, Number 4/5, 2008 | <i>M. Factor, D. Naor, S. Rabinovici-Cohen, L. Ramati, P. Reshef, S. Ronen, J. Satran, and D. Giaretta</i> |
| Digital Preservation of Interactive Multimedia Performing Arts | | <i>Kia Ng, Tran Vu Pham, Bee Ong, Alain Bonardi, Jerome Barthelemy, and David Giaretta</i> |
| Towards OAIS-Based Preservation Aware Storage - A White Paper | IBM Internal White Paper, November 2006 | <i>S. Cohen, D. Naor, L. Ramati, and P. Reshef</i> |

Table 3 PAPERS IN PEER REVIEWED JOURNALS





2.2 STANDARDS ACTIVITIES

The following material lists contributions to some of these Standards activities listed in Section 5.1 of the CASPAR Description Of Work:

1. The evolution of the OAIS Reference Model:

- 1) David Giaretta is the major editor of the OAIS Five Year Review
- 2) CASPAR personnel input many change requests to the five year review process

2. Extensions to the Object Storage Device (OSD) standard to support an implementation of the OAIS concepts:

- 1) CASPAR/IBM personnel participate in the Storage Networking Industry Association (SNIA) LTACSI - Long Term Archive & Compliance Storage Initiative and have presented Preservation Data Store to that group

3. Contribution to the definition of the national guidelines and recommendations for preservation

- 1) CASPAR and PARSE.insight (Permanent Access to the Records of Science in Europe) jointly developed a survey to better understand the views of researchers on long-term preservation and access to scientific data
- 2) CASPAR and PARSE.insight will hold several workshops to discuss the survey and refine the results
- 3) PARSE.insight will then develop a roadmap and recommendations for developing the e-infrastructure in order to maintain the long-term accessibility and usability of scientific digital information in Europe

4. Certification standard for trusted repositories

- 1) David Giaretta chairs the CCSDS/ISO Registry Certification and Audit Working Group that is developing a draft of two CCSDS/ISO Recommendations that will provide the Requirements for certification of "long-term Repositories" and the requirements and processes for becoming an audit and certification organization.
- 2) CASPAR personnel have participated as members of this Working Group

5. Contributions to AIP-type packaging techniques

- 1) CASPAR personnel were members on the CCSDS Information Packaging and Registry Working Group that developed the XFDU standard

