

JISC DEVELOPMENT PROGRAMMES

Project Document Cover Sheet

Progress Report 2

Project

Project Acronym	SHERPA-DP	Project ID	
Project Title	SHERPA Digital Preservation: Creating a Persistent Preservation Environment for Institutional Repositories		
Start Date	1st March 2005	End Date	28th February 2007
Lead Institution	Arts and Humanities Data Service, King's College London		
Project Director	Andrew Wilson		
Project Manager & contact details	Martha Brundin, martha.brundin@ahds.ac.uk 020 7848 1978		
Partner Institutions	University of Nottingham, University of Glasgow, University of Edinburgh, London LEAP Consortium, White Rose Consortium		
Project Web URL	http://www.ahds.ac.uk/about/projects/sherpa-dp/index.htm http://www.sherpa.ac.uk/		
Programme Name (and number)	Supporting Digital Preservation and Asset Management in Institutions. JISC Circular 4/04		
Programme Manager	Helen Hockx-Yu		

Document

Document Title	Project Plan		
Reporting Period	August 1 2005 – January 31 2006		
Author(s) & project role	Martha Brundin, Gareth Knight, Andrew Wilson - AHDS SHERPA DP Project staff		
Date	24th Feb 06	Filename	SherpaDPProgRep2-1a.doc
URL	In progress		
Access	x Project and JISC internal		<input type="checkbox"/> General dissemination

Document History

Version	Date	Comments
1a	24/02/06	Submitted to Helen Hockx-Yu Friday Feb 24th 2006

JISC Progress Report: Sherpa DP

1 Aug 2005 – 31 January 2006

Overview of Project

Grant Statement

The project is being conducted under the terms agreed in the letter of grant

SHERPA DP was granted a change of start and end date by Leona Carpenter. The revised project dates are March 1st 2005 to February 28th 2007.

2. Aims and Objectives

No changes have been made to the aims/objectives as outline in Project Plan Version 1b

The targets for this reporting period are as follows:

- Report on OAIS-SHERPA preservation structure now in its second draft and on schedule for completion as planned (WP 2.6).
- Proposal for the roles of the DCC and DPC complete and agreed by relevant organisations (WP 2.9)
- A lifecycle model for e-prints located in institutional repositories completed (WP 2.13) (Deliverable)
- Document summarising relevant preservation outputs of the SHERPA project completed (WP 3.1)
- A review of metadata application profiles in use by repositories completed (WP 4.1)
- A draft minimum set of elements for preservation and technical metadata on track for completion (WP 4.4)
- Liaison planning with the project partners is an on going activity. The technical specification and test environment setup has been discussed with the partners during the project partners meeting at Leeds (WP 5.1)
- DSpace software review completed as a part of (WP 5.2)
- EPrints software review completed (WP 5.3)
- Storage assessment for the preservation archive completed (WP 5.4)
- Methods of data synchronization between preservation archive and IR completed (WP 5.5)
- Fedora software review has been added to work packages to assess Fedora's capability for digital preservation. Review completed (WP 5.6)
- A specification of requirements for the use of METS as a mechanism for transferring data from repositories completed (WP 5.7)
- A review of file formats held by E-print repositories completed (WP 6.1)

3. Overall Approach

No changes have been made to the approach as outlined in the project plan

4. Project Outputs

- Report on OAIS-SHERPA preservation structure now in its second draft and on schedule for completion as planned (WP 2.11)
- Proposal for the roles of the Digital Curation Centre (DCC) and Digital Preservation Coalition (DPC) may play in commenting upon and promoting the work performed by Sherpa DP complete. Preliminary agreement in place between these organisations and the AHDS (WP2.9)
- A lifecycle model for e-prints located in institutional repositories complete (WP 2.13)
- Document summarising relevant preservation outputs of the SHERPA project complete (WP 3.1)
- A review of metadata application profiles and vocabularies in use by repositories participating in the SHERPA DP project complete (WP 4.1)
- A draft minimum set of elements and vocabulary for preservation and technical metadata on track for completion (WP 4.4)
- A review of the file formats held by E-print repositories, accompanied by an assessment of their suitability for preservation and problems that may arise during migration complete (WP 6.1).

5. Project Outcomes

Much of the significant technical investigation in the project has taken place in this reporting period. This work will be pulled together during the next 6 months with the development of the project's preferred solutions to such concerns as data synchronisation, preservation metadata, data and metadata transfer, etc.

Research performed for Work package 2.13 demonstrates that the e-print lifecycle model, developed by James et-al, may be refined to apply to different organisation models, such as the disaggregated structure of the Sherpa DP project. It also demonstrates areas for improvement that were not considered in the original lifecycle, such as OAI harvesting.

The metadata implemented by institutional repositories is diverse and cannot be refined to basic or qualified Dublin Core elements. To maintain e-print metadata, the metadata elements used by these services must be mapped to a merged schema and stored in a METS framework. It also indicates that current metadata schemes are insufficient to identify the source repository and ensure the integrity of the digital objects. Further metadata must be created by the institutional repository for harvest by the Preservation service.

6. Stakeholder Analysis

The SHERPA DP Technical Officer visited the PRESERV project at Southampton in September to discuss the different approaches and commonalities between SHERPA DP and the PRESERV project.

A meeting was held with project partners on November 4th in Leeds. The day comprised a review of work performed on the project, an outline of the services the AHDS is likely to provide and an indication of likely technical methods we will use. Discussion on the work required from project partners was held and a preliminary agreement made regarding staff and hardware availability.

7. Risk Analysis

No problems have occurred.

8. Standards

No changes currently planned

9. Technical Development

No changes currently planned

10. Intellectual Property Rights

No problems relating to third party rights have arisen to date.

Project Resources

11. Project Partners

No changes have been planned

12. Project Management

No changes have been made

13. Programme Support

- A Preservation Group mailing list has been started on JISCmail to coordinate cluster group activities and several groups have been invited to join.
- Links have been made with the PRESERV project and information is shared on a regular basis.
- The Programme Manager has been proactive in alerting all projects to relevant conferences and events and a SHERPA DP representative has attended many of these.

14. Budget

See Appendix A

Detailed Project Planning

15. Workpackages

The only deliverable for this reporting period is the Eprints Life-cycle model (WP2.13). This was completed to schedule.

Other documents scheduled for completion during the reporting period were completed as planned:

- Report on OAIS-SHERPA preservation structure now in its second draft and on schedule for completion as planned (WP 2.6).
- Proposal for the roles of the DCC and DPC complete and agreed by relevant organisations (WP 2.9)
- A lifecycle model for e-prints located in institutional repositories completed (WP 2.13) (Deliverable)
- Document summarising relevant preservation outputs of the SHERPA project completed (WP 3.1)
- A review of metadata application profiles in use by repositories complete (WP 4.1)
- A draft minimum set of elements for preservation and technical metadata on track for completion (WP 4.4)
- Liaison planning with the project partners is an on going activity. The technical specification and test environment setup has been discussed with the partners during the project partners meeting at Leeds (WP 5.1)
- DSpace software review completed as a part of (WP 5.2)

- EPrints software review completed (WP 5.3)
- Storage assessment for the preservation archive completed (WP 5.4)
- Methods of data synchronization between preservation archive and IR completed (WP 5.5)
- Fedora software review has been added to work packages to assess Fedora's capability for digital preservation. Review completed (WP 5.6)
- A specification of requirements for the use of METS as a mechanism for transferring data from repositories completed (WP 5.7)
- A review of file formats held by E-print repositories completed (WP 6.1)

WP 4.2 - Investigation of METS framework, originally scheduled for completion in this reporting period, has been rolled into WP 5.8 - Review of METS for data transfer, as both are being undertaken by the same external consultant. We expect that both work packages will be completed by 28 February as scheduled.

Since the last report WP 5.4 – Review of SRB was removed from the project. It was the opinion of the Project Director that the amount of work required to set up a test SRB client and server environment, and the time that would be needed for testing (based on experience with a separate SRB installation for AHDS), made it unfeasible to retain this particular work package in the project.

The end date of WP 5.1 – Liaison Planning has been extended as this as an ongoing activity across the timetable for WP5

Additional workpackages added since the last report are:

WP 3.2 – Handbook Specification

WP 5.6 – Fedora software review

WP 6.1 – File format review

WP 6.3 – File format typing and format registries assessment

Objectives for the next report period:

WP 3: Digital Repository Handbook

3.2 Handbook Specification

3.3 Establishment of Handbook Editorial Board

3.4 First Editorial Board Meeting

WP4: Metadata and METS

4.6 Final report on use of METS framework

4.7 Final report on minimum metadata element set

WP 5: Archiving of Institutional Repositories

5.8 Review of METS for data transfer

5.11 Final report on preferred solution

WP 6: Archive Management

6.2 Access review

6.3 File format typing and format registries review

6.4 Integrity, fixity and security review

6.5 Obsolescence and migration review

6.6 Draft report on preferred solution

WP 7: Implementation

7.1 Implementation plan

16. Evaluation Plan

No evaluation results for this reporting period.

During the next reporting period the metadata set will be evaluated for its suitability for purpose for the Sherpa DP partner repositories. This will be done using questionnaires and discussion with the partner sites as well as the DPC and DCC. A successful outcome would be the development of a metadata set that meets preservation requirements for the AHDS and the project partners.

17. Quality Assurance Plan

The Project Manager has signed off on one deliverable in this reporting period. The report setting out an e-prints lifecycle model was finalised and agreed as fitting its intended purpose as part of Work Package 2: Infrastructure.

Feb 2006- the Sherpa DP metadata set will be validated against the PREMIS data dictionary to ensure adherence to international standards for preservation metadata.

March 2006- adherence to international metadata standards for resource discovery will be checked by ensuring that the Sherpa DP metadata harvester complies with OAI-PMH.

18. Dissemination Plan

Report on Sherpa DP given at Sherpa Project Officers Meeting, Glasgow 30 November 2005
Sherpa DP is the first of the DPC's Featured Projects. Interview of Project Director available at <http://www.dpconline.org/graphics/join/sherpa.html>

Objectives for next reporting period:

- Meeting with PRESERV Project, Southampton 2 March
- Cluster Group meeting at JISC Programme Meeting, Warwick 27-28 March
- NDIIIP/JISC Digital Preservation Workshop, Washington DC 7-9 May
- Paper on Sherpa DP at Society for Imaging Science and Technology Archiving Conference, Ottawa Canada 23-26 May

19. Exit/Sustainability Plan

No project outputs listed in the plan have yet reached final completion date.

The draft report on the OAIS SHERPA Preservation Infrastructure (WP 2.11) develops a generic model for applying the OAIS model in a distributed structure. Copies of the draft report have been disseminated among the project partners and the DCC and DPC for comment.

In the next reporting period the following reports will be disseminated among the project partners and to the DCC and DPC:

- Final report on OAIS SHERPA preservation Infrastructure (WP 2.12)
- Report on the use of METS framework within the SHERPA preservation environment (WP 4.6 & 5.8)
- Minimum metadata set for SHERPA repositories (WP 4.7)

Report documenting investigation into archiving of institutional repositories (WP 5.11)

Appendix

Deliverables Summary		
Project Plan Version 1a WP1.1	30/04/05	✓
Reports on SHERPA OAIS model and roles WP2.12	31/08/06	
JISC Progress Report 1 WP1.6	31/08/05	✓
Eprints Life-cycle model WP2.13	31/12/05	✓
Report on use of METS Framework WP4.6	28/02/06	✓
JISC Progress Report 2 WP1.7	28/02/06	✓
Report on archiving of institutional repositories WP5.10	31/03/06	✓
Report on technical metadata elements set WP4.7	30/04/06	
Implementation Plan WP7.1	30/04/06	
JISC Progress Report 3 WP1.8	31/08/06	
Report on preservation archive management WP6.9	31/07/06	
Digital Preservation Handbook WP3.8	28/02/07	
JISC Completion Report WP1.9	28/02/07	