

Requirements for a Disaggregated service

Document Details

Document type: Project Report
Author: Gareth Knight
Draft / Version: Version 1.1
Date of edition completion: 15/09/2005

Change history

1.0 29/06/2005. Internal draft
1.1. 15/09/2005. First version

Audience

This document is written for use by AHDS Executive staff and Partners within the SHERPA-DP project.

1. Introduction

To implement the SHERPA-DP disaggregated model, the E-Print Archive and Preservation Service must clarify the terms of their agreement. The greater organisational and technical interoperability that is intrinsic within the SHERPA DP project makes this stage more complex than it is for traditional full deposits, which can generally be completed without external consultation. It is envisaged that three parties will perform specific functions within the disaggregated model:

1. *E-Print Archive* – The Institutional archive responsible for managing the ingest process and provide access to digital content.
2. *Preservation Service* – An external organizational allocated to implement the technical infrastructure and preserve digital content in the long-term.
3. *Repository Advisory board* – A group, composed of advisors from different organizations that will provide advice and recommendations to the Preservation Service.

This report will outline preliminary responsibilities to be allocated to each party. It should be consulted in conjunction with the 'SHERPA-DP OAIS Report'.

2. Responsibilities of an OAIS-compliant repository

The SHERPA DP disaggregated model is relatively new and untested. Previous work examining the requirements for successful long-term digital archiving, notably the RLG/OCLC report *Trusted Digital Repositories: Attributes and Responsibilities* (RLG, 2002) and James et-al (2004), serves as a useful basis for examining the responsibilities that must be allocated to the e-print repository and preservation service.

To provide a trusted service, as defined by the RLG/OCLC study, the model must demonstrate it is able to provide the six OAIS mandatory responsibilities:

1. Negotiate for and accept appropriate information from information Producers.
2. Obtain sufficient control of the information provided to the level needed to ensure Long-Term Preservation.
3. Determine...which communities should become the Designated Community and, therefore, should be able to understand the information provided.
4. Make the preserved information available to the Designated Community.
5. Ensure that the information to be preserved is independently understandable to the Designated Community. In other words, the community should be able to understand the information without needing the assistance of the experts who produced the information.
6. Follow documented policies and procedures which ensure that the information is preserved against all reasonable contingencies, and which enable the information to be disseminated as authenticated copies of the original, or as traceable to the original.
(OAIS Reference model, page 3-1)

These responsibilities may be allotted to different parties within the disaggregated model. It is expected the e-print archive will have considered responsibilities 1-4 in the initial stages of set-up and have implemented relevant software to ingest e-prints and make them available to the relevant community. Additional procedures must be developed and refined to preserve e-prints during the project test-bed.

¹ The framework for identification of a trusted repository indicates actions that must be performed. The implementer must produce clear criteria to implement and monitor services.

2.1. E-Print Archive

Mandatory Responsibilities

To implement the disaggregated model the e-print archive MUST:

1. Implement appropriate repository software to accept, manage and disseminate e-prints;
2. Develop policies to identify preferred file formats for deposit and inform the Depositor (OAIS Producer) and Preservation Service of these requirements;
3. Create and implement a deposit licence that:
 - a. Establishes permission for the E-Print Archive to transform the submitted resource (e-print) for the purpose of preservation and accessibility.
 - b. Establish permission for the E-Print Archive to allocate responsibility for preservation to a third-party.
4. Create and/or refine discovery metadata necessary to locate the e-print within the E-Print Archive.
5. Assign an identifier to locate the e-print within the repository.
6. Accept responsibility for the creation of a dissemination copy of the e-print (Dissemination Information Package), derived from the submitted version created by the Depositor.
7. Provide facilities to monitor and transfer digital resources to the Preservation Service, by:
 - a. Providing a method for the Preservation Service to connect to the Repository and extract Information Packages.
 - b. Implement a mechanism to transfer Information Packages to the Preservation Service₂.
8. Provide a detailed description of the metadata schema implemented, including a list of elements and vocabulary.
9. Implement technical services, in cooperation with the Preservation Service, to identify recent submissions and extract relevant information

Elective Responsibilities

To implement the disaggregated model the e-print archive SHOULD:

10. Provide a full description of any actions performed on the e-print or associated metadata between the point of ingest and delivery (e.g. refinement of discovery metadata) and log such details in the relevant metadata or transmit the information to the Preservation Service for inclusion in the relevant metadata.

To implement the disaggregated model the e-print archive MAY:

11. Implement mechanisms and procedures to 'push' data and metadata held by the E-print archive to the Preservation Service.

2.2. Preservation Service

Mandatory Responsibilities

To implement the disaggregated model, the Preservation Service MUST:

1. Establish a preservation strategy that:
 - a. Identifies file formats held by the E-print Archive
 - b. Reviews recent developments and identifies file formats considered suitable for preservation
 - c. Establishes risk analysis procedures and identifies preventative measures that may be taken to ensure long-term access.
 - d. Establishes procedures for data migration.

² The AHDS is currently investigating methods to transfer metadata and data between the two parties and will work with the E-Print Archive to implement suitable mechanisms.

2. Review the preservation strategy on a regular basis to ensure the preservation format remains suitable for the e-prints submitted to the E-Print Archive.
3. Liaise with the e-print archive to establish a schedule for the transfer of Information Packages to the Preservation Service.
4. Provide details of the extended metadata scheme intended to describe and preserve the digital resource.
5. Perform integrity checks to ensure the validity of the data type.
6. Provide a full description of the actions performed on the e-print while in its custody.

Elective Responsibilities

To implement the disaggregated model the Preservation Service MAY:

1. Establish mechanisms and procedures to identify updates to metadata and data within the E-print archive.
2. Implement mechanisms and procedures to 'pull' data and metadata held by the E-print archive. Data transfer may be initiated according to a pre-determined timetable (e.g. every 7 days) or when notified by the E-print archive.

2.3. Repository Advisory Board

The Repository Advisory Board will comprise individuals from organizations familiar with e-prints and digital preservation. Possible partners may include the Digital Preservation Coalition (DPC) and Digital Curation Centre (DCC). The Repository Advisory Board should provide guidance to the Preservation Service on issues relating to the management of the disaggregated model and guidance on its implementation. This may be summarized to two recommended requirements.

The Repository Advisory SHOULD:

1. Maintain a technology watch of digital preservation issues, to refine existing documentation on relevant preservation formats, significant properties of an existing format, and other relevant areas.
2. Provide guidance on issues related to the sustainability, including recommendations to manage the technological and organizational aspects of the disaggregated service.