

Managing scientific research assets for the long term: A structured method for the management of preservation processes

Esther Conway, Brian Mathews, Matthew Dunckley and David Giarretta

Science and Technology Facilities Council
Rutherford Appleton Laboratory, Chilton, Didcot, Oxfordshire OX11 0QX
esther.conway@stfc.ac.uk

Abstract

The preservation of scientific research data is a complex task data which involves decision making and the allocation of resources in an ever changing technical social and organizational environment. In this paper we propose a structured method for the preservation of scientific research data. This approach incorporates a number of key processes methods tools and techniques and permits the controlled management of changes in terms of investment and return on investment.

A business case for the preservation scientific data of captures the reasoning for initiating a preservation analysis or planning. Consideration should also be given to the option of doing nothing including the costs and risks of inactivity. The business case for the preservation of scientific data could exist for variety of financial, strategic or legislative reasons but should always be the driving force for any activities undertaken. A good business case should document expectations of achieving defined benefits. These benefits and prospect of them being realized can be reviewed. This checks that a valid case still exists for the preservation of data so that organizational resources can be appropriately directed and the required quality of data can be maintained for its lifespan,. We intend to show using illustrative examples from the atmospheric science community how a business case informs preservation analysis and other key processes. As a result communication between senior management and technical/scientific teams is supported and the desired research asset is created and maintained

The challenge of digital preservation of scientific data lies in the need to preserve not only the dataset itself but also the ability it has to deliver knowledge to a future user community. A true scientific research asset allows future users to reanalyze the data within new contexts. Thus, in order to carry out meaningful preservation we need to ensure that future users are equipped with the necessary information to re-use the data. This paper will present an overview of a preservation analysis methodology which was developed in response to that need on the CASPAR and DCC SCARP projects. We intend to place it in relation to other digital preservation processes discussing how they can interact to provide archives caring for scientific data sets with the full arsenal of tools and techniques necessary to rise to this challenge.

Good preservation analysis is essential in order to design a truly reusable asset. This methodology capitalizes on a community's expertise and knowledge by appreciating the nature of data use, its evolution and organizational environment. The methodology incorporates a number of analysis stages into an overall process capable of producing an actionable preservation plan for scientific data, which satisfies a well defined preservation objective. In this paper we will discuss how the creation of an archival information package using preservation actions selected on this basis ensures a measurable and testable solution. We aim to demonstrate with examples from the Digital Curation Centre SCARP project case studies how judicious analysis permits the design of archival information packages which deliver a greater return of investment by both improving the probability of the data being reused and potential outcome of that reuse .

The analysis method facilitates modeling of information networks based on the archival information package solution. Using illustrative examples from the CASPAR testbeds we intend to show how these network models are a representation of the digital objects, operations and relationships which allow a preservation objective to be met for a future designated community. The model provides a sharable, stable and organized structure for digital objects and their associated requirements. They expose the risks, dependencies and tolerances within an archival information package. This allows for the automation of event driven or the periodic review of archival holdings by knowledge management technologies. The clear definition of relationships also facilitates the identification of reusable solutions which can be deposited within registry repositories of representation information, thus sharing preservation efforts within and across communities.

Scientific Research assets are likely to be held by and transferred between institutional repositories. These repositories need to be designed planned and managed, competing for resources within complex organizational structures. In this paper we intend to conclude by touching upon how preservation analysis can inform audit and certification processes such as DRAMBORA and TRAC or planning activities for new repositories such as PLATTER. By doing this we allow preservation analysis at the data set level to be placed within the context of institutional planning and operations

In conclusion this approach seeks to apply the practices of quality assurance, risk/knowledge management, analysis, planning and modeling in a structured controlled manner to optimize beneficial re-use of an institutions scientific research assets for the community it serves.

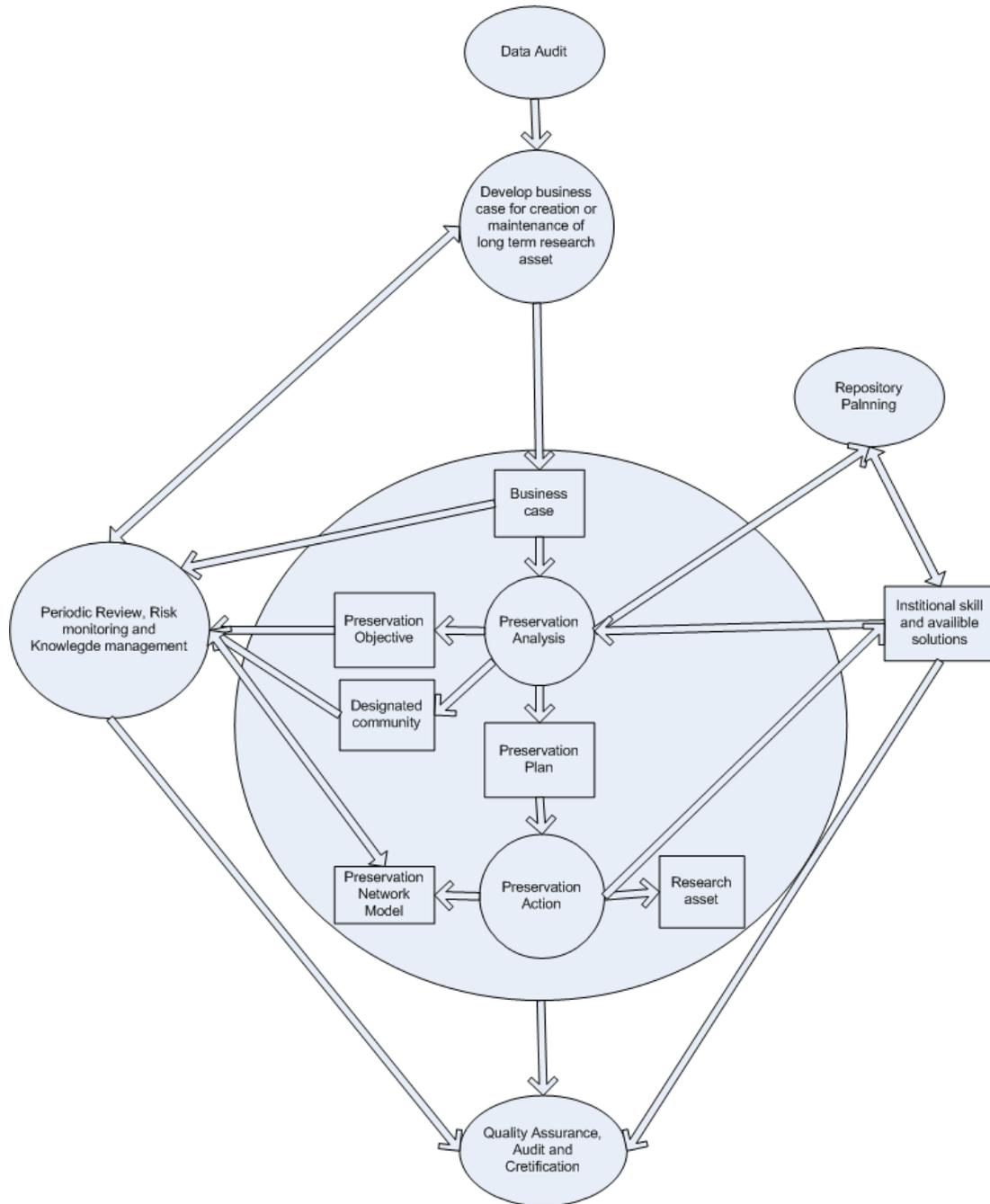


Fig1. Preservation process diagram