The International Journal of Digital Curation
Issue 2, Volume 6 | 2011

Curating Scientific Research Data for the Long Term: A Preservation Analysis Method in Context

Esther Conway, David Giaretta, Simon Lambert and Brian Matthews,

Science and Technology Facilities Council

Abstract

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 presents an overview of a preservation analysis methodology which was developed in response to that need on the CASPAR and Digital Curation Centre SCARP projects. We intend to place it in relation to other digital preservation practices, 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.

The *International Journal of Digital Curation* is an international journal committed to scholarly excellence and dedicated to the advancement of digital curation across a wide range of sectors. ISSN: 1746-8256 The IJDC is published by UKOLN at the University of Bath and is a publication of the Digital Curation Centre.

Introduction

This paper presents a brief overview of the preservation analysis methodology. After describing the main stages and purpose of the method, we intend to place it in relation to other digital preservation and existing archival practices, 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 the current challenges. We then intend to illustrate the benefits of preservation analysis with examples from the SCARP[1] case studies and solutions we have implemented on the CASPAR[2] project, with a focus on the following key areas:

Maximizing Return on Investment

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, evolution and organizational environment. When scientific data is used by a community it develops a history. During its lifetime, the custody of a data set may pass through several bodies, generating rich documentation which explains the scientific purpose of the dataset and how it has evolved over time. Organizations also develop around branches of science publishing or produce grey materials over time which prove to be important for the interpretation or analysis of data.

This method seeks to design the optimal research asset by capturing key information which facilitates reuse. We aim to demonstrate, with comparative examples from the Digital Curation Centre's SCARP project case studies, how judicious analysis permits the design of Archival Information Packages (AIP) which deliver a greater return of investment by both improving the probability of the data being reused and potential outcome of that reuse.

Provision of Measureable and Testable Solutions

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. Using an example from the SCARP Atmospheric Sciences case study (Conway, 2009) we will show how we developed such an objective, and how this facilitates a testable solution giving an archive the necessary assurance in any preservation action taken.

Management of Research Assets Through the Modelling of Preservation Networks

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.

---

[1] Digital Curation Centre SCARP Project: http://www.dcc.ac.uk/scarp/.
[2] The CASPAR Project: http://www.casparpreserves.eu/.

The models provide 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 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.

Informing Preservation Activities in the Wider Institutional Environment

Preservation analysis clearly exposes the requirements and issues associated with an individual dataset, thereby supplying audit, certification and repository planning activities with essential information. Archival Information Packages are likely to be held by, or 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[3] and TRAC (CRL & OCLC, 2007), or planning activities for new repositories, such as PLATTER[4]. By doing this, we allow preservation analysis at the dataset level to be placed within the context of institutional planning and operations.

Overview of the Preservation Analysis Method

The challenge of digitally preserving 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. This entails allowing future users to re-analyze 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.

The Digital Curation Centre's SCARP and CASPAR projects have a strong focus on the preservation and curation requirements for scientific data sets. These projects engaged with a number of archives based at the STFC[5] Rutherford Appleton Laboratory. In particular, we carried out extensive analysis work to consider the preservation requirements of the British Atmospheric Data Centre[6], the World Data Centre[7] and the European Incoherent Scatter Scientific Association (EISCAT)[8]. During these studies it became clear that there was a need for a consistent preservation analysis methodology.

[3] DRAMBORA: http://www.repositoryaudit.eu/.
[4] PLATTER: http://www.digitalpreservationeurope.eu/platter/.
[5] Science and Technology Facilities Council: http://www.stfc.ac.uk/.
[6] British Atmospheric Data Centre: http://badc.nerc.ac.uk/home/index.html.
[7] World Data Centre for Solar Terrestrial Physics: www.ukssdc.ac.uk/wdcc1/wdc_menu.html.
[8] European Incoherent Scatter Radar: http://www.eiscat.rl.ac.uk.

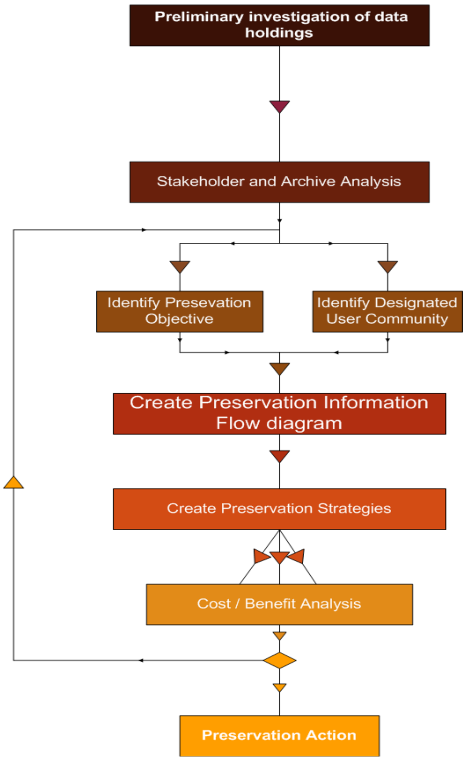

Figure 1. Preservation Analysis Workflow.

In the resulting methodology, we sought to incorporate a number of analysis techniques, tools and methods into an overall process capable of producing an actionable preservation plan for scientific data archives. Figure 1 illustrates the stages of this methodology. In the rest of this paper we shall discuss the stages in detail, illustrated with examples of work with the scientific archives.

Preliminary Investigation of Data Holdings

The initial step is to undertake a preliminary investigation of the data holdings of the target archive. The CASPAR project developed a questionnaire containing key questions which allowed the preservation analyst to initiate discussion with the archive. It critically allows the analyst to:

- Understand the information extracted by users from data;
- Identify Preservation Description and Representation information;
- Develop a clearer understanding of the data and what is necessary for its effective re-use;
- Understand relationships between data files and what constitutes a digital object within the archive.

While it is appreciated that this questionnaire is not an exhaustive list of questions which one may need to ask about a preservation target, it still provides sufficient information to commence the analysis process .The full questionnaire and results from the Ionosonde WDC holdings can be obtained from the CASPAR website (Conway et al., 2007).

Stakeholder and Archive Analysis

After carrying out the questionnaire process for each archive, it is necessary to carry out a stakeholder analysis. This is because:

- Stakeholders may hold different views of the knowledge a data set is capable of providing an end user;
- Stakeholders can identify different end users whose skill sets and knowledge base vary;
- Stakeholders may have produced, or be custodians of, information vital for re-use of data.

The stakeholder analysis classifies stakeholders into a number of categories, each with their own concerns. In addition to identifying the stakeholders from the different categories, it is also beneficial to understand how an archive has evolved and been managed. This process can be used to illuminate the different uses of data over time and highlight the existence of associated Representation Information.

Defining a Preservation Objective

The analysis carried out before this point may present one with a natural, easily defined preservation objective, or alternatively, there may be a greater number of options which overlap and are more difficult to define. It is important to note that this type of analysis cannot advise you as to which preservation option to choose, but merely clarifies the options available to you. Preservation objectives should be:

- Specific – well defined and clear to anyone with a basic knowledge of the domain;
- Actionable – the objective should be currently achievable;
- Measurable – it is critical to be able to know when the objective has been attained in order to assess if any preservation strategy developed is adequate;
- Realistic – based on findings from the previous stages of analysis.

Defining a Designated User Community

The Designated Community is defined in OAIS (CCSDS, 2002) as: "An identified group of potential Consumers who should be able to understand a particular set of information. The Designated Community may be composed of multiple user communities."

An archive defines the Designated Community for which it is guaranteeing to preserve some digitally encoded information, and must therefore create AIPs with appropriate Representation Information.

The Designated Community will possess a skills and knowledge base which allow them to successfully interact with a set of information stored within an AIP, in order to extract required knowledge or recreate the required performance or behavior. In common with the preservation objective, the analysis up to this point may present one with a range of community groups that the archive may chose serve.

The definition of the skill set is vital, as it limits the amount of information which must necessarily be contained within an AIP in order to satisfy a preservation objective. In order to do this, the definition of the designated community must be:

- Clear, with sufficient detail to permit meaningful decisions to made regarding information requirements for effective re-use of the data;
- Realistic and stable, in so far as there is reasonable confidence in the persistence of the knowledge base and skill set.

Preservation Information Flows

Once the objective and community have been identified and described, an analyst should be in position to determine the information required to achieve an objective for this community. An analyst proceeds by identifying risks which are to be addressed by preservation action. We advocate the creation of an OAIS preservation information flow diagram at this juncture.

An OAIS preservation information flow diagram is graphical representation and analysis tool, which is a hybrid of an information flow diagram and the OAIS information model. It provides a convenient format to facilitate group discussions over preservation plans and strategies. A preservation information flow diagram we created for the Mesosphere Troposphere Stratosphere (MST) data is shown below:

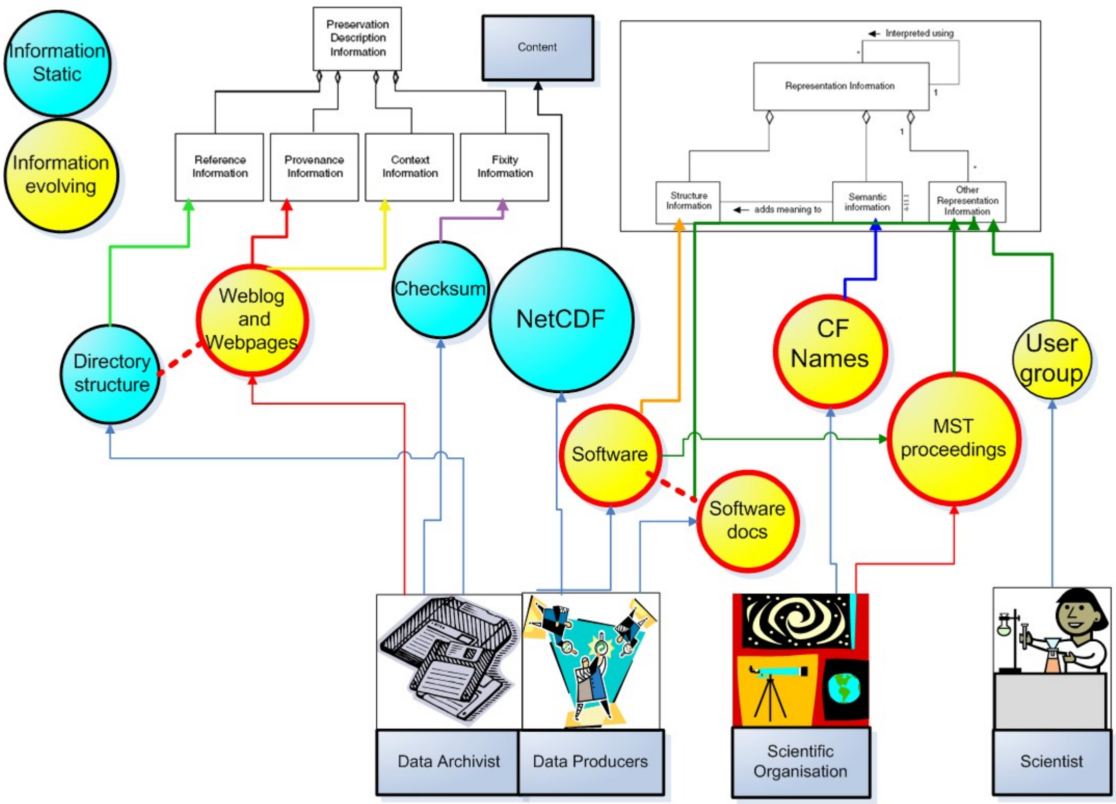

Figure 2. OAIS information flow diagram for the MST data set.

The OAIS reference model specifies that within an archival system, a data item has a number of different information items associated with it, each performing a different role in the preservation process. The preservation objective for a designated community should be is satisfied when each item of the OAIS information model has been adequately populated with sufficient information. The information model provides a checklist which ensures that the preservation objective can be met. All information objects must be mapped to at least one of the element of the OAIS information model.

In addition to information objects and the standard OAIS information model, the diagram contains a number of other components.

As multiple strategies can be developed, a number of competing preservation plans are available. A preservation plan should consists of a unique:

- Set of information objects,
- Set of supply relationships,
- Set of preservation strategies.

Each plan will allow an archive to carry out a series of clear preservation actions in order to create an AIP. The archive should now be in a position to take a number of plans to the cost/benefit/risk analysis stage, where they can be evaluated and a preferred option chosen.

Cost/Benefit/Risk Analysis

The final stage of the workflow is where plan options can then be assessed according to:

- Costs to the archive directly, as well as the resources knowledge and time of archive staff;
- Benefits to future users which ease and facilitate re-use of data;
- Risks inherent to the preservation strategies and accepted impact to the archive.

Once this analysis is complete, the optimal plan can be selected and progressed to preservation action. If no plans are deemed suitable then the process must begin again with an adjustment to the preservation objective and/or the designated community to be served.

Maximizing Return on Investment

When we examined different archives we discovered there was the potential to create different research assets for data, depending on how one might choose to support reuse of a dataset. We observed the following sorts of factors influencing the use and re-use of data over time:

- Birth and development of a science;
- Events which influence data use, such as the second world war or global warming;
- Development of technologies and the emergence of global networks;
- Publication of journals, technical manuals, interpretative handbooks, conference proceeding, minutes of user group meetings, software etc;
- Emergence of branches of science and associated organisations;
- Stewardship of data and the influence of different custodians.

This is not an exhaustive list, as many factors influencing data re-use are domain-specific, as is the categorization of the stakeholders. Naturally, most of these can only be expected to be dealt with in the most cursory way in any practical study. Nevertheless, even this can be extremely important in understanding the situation, as after this evaluation you should be in position to scope what types of reuse may be realistically achieved.

By comparing two datasets we can demonstrate how an appreciation of previous data use and stakeholder relationships can inform the design of an AIP, allowing an archivist to select the optimal set of information to realize return through improving the probability of the data being reused and the outcome of that use.

The Mesosphere Troposphere Stratosphere (MST)[9] dataset is extremely well documented and tightly managed. Access to the data is restricted, with end users required to report back on how they have used the data. The Archivist is the key manager of these data for a number of reasons:

1. He is the project scientist involved in production of the data;
2. He is a field expert and practicing scientist in close contact with relevant scientific organisations;
3. He provides support, runs and keeps records of user group meetings.

When we consider these factors, we can see that it is reasonable to try to capture information from current users which facilitate the re-use of data by future scientists. The information has been captured in user group minutes, conference proceedings and scientific papers resulting from the study of MST data. The capture and determination of value of these resources is possible as a result of the archivist's domain knowledge and close connection to users.

We now contrast the scenario above with that of the Ionosonde data holdings of the World Data Centre. The data archivist, whilst also being a skilled individual with domain knowledge, does not have the same strong connection with users. The data currently comes from 252 geographically diverse locations and current users are simply required to provide an e-mail address to gain access. As a result, it would be completely impractical to capture user generated information, even if it might facilitate re-use. We make the judgment that the archive would realize a return on the large investment required to source, capture and asses information from these user groups.

[9] The Natural Environment Research Council (NERC) Mesosphere-Stratosphere-Troposphere (MST) Radar at Aberystwyth: http://mst.nerc.ac.uk.

The added value of information which contributes to reuse must be assessed against the likely cost of its capture and maintenance. If creation of asset reliant on such information is deemed viable and cost effective, an archive may then proceed by developing a preservation objective, which accommodates the information in scope.

Measurable and Testable Solutions

We shall now examine how a preservation objective developed for the MST dataset, in accordance with the methodology, produces a testable solution.

The importance of the MST dataset lies in the fact that it is an irreplaceable earth observational record: once lost, these data cannot be replaced by the repetition of the experiment. The dataset is valuable because it:

- Contains data from the UK's most powerful and versatile wind-profiling instrument;
- Provides information about atmospheric stability, turbulence, humidity fields, precipitation and a variety of atmospheric phenomena;
- Contains measurements of winds up to many kilometres from the ground;
- Contains a record of winds sampled continuously, with a cycle time of a few minutes over a long period of time;
- Provides a record not only of the horizontal but also the vertical air velocity, which additionally has high temporal and spatial resolution.

If we accept that the challenge of digital preservation 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, we must therefore ensure that future users are equipped with the necessary information to re-use the data.

Use of atmospheric data may involve different activities, such as format conversion to make it interoperable with other datasets. Specialist visualisation of time related data may be needed to study some kinds of atmospheric behaviour. Interpolation, subsetting or different forms of statistical analysis may also be employed. Identification of atmospheric phenomena and behaviours is achieved through the creation or application of data to established models of dynamic atmospheric systems. The same data may be used to study different aspects of atmospheric behaviour; listed below are some which we identified in peer reviewed literature resulting from studies which employed the MST data.

- **Precipitation:** Clouds contain moisture; when the droplets in clouds coalesce they become sufficiently large to cause precipitation. A recent evolution of knowledge surrounding the MST radar data allows the data to be used to study precipitation.
- **Convection:** Convection is the transfer of heat by movement within a substance. The MST radar data permits you to study the convective circulation of air within the atmosphere.
- **Gravity Waves:** Gravity waves are generated in the troposphere by frontal systems or by airflow over mountains. The geographic position of the MST radar site is ideal for studying this phenomenon.

- **Rossby Waves:** Rossby waves are a subset of inertial waves. These atmospheric waves are large scale motions, with wavelengths of up to 6000 km. The continual monitoring of a discrete region of the atmosphere over a long period allows for the analysis of such waves.
- **Mesoscale and Microscale Structures:** The frequency of observation and the resolution of the MST radar also permit analysis to be carried on mesoscale (~50km) and microscale (lasting a matter of minutes) structures.
- **Fallstreak Clouds:** Cloud formations can be associated with atmospheric conditions, such as turbulence and waves, which the MST radar is capable of observing.
- **Ozone Layering:** Atmospheric dynamics can also be correlated with chemical composition of the atmosphere.

Mindful of these types of use, and with a desire to support them, we set the preservation objective as follows. A user from a future designated community should be able to extract a specific set of 11 parameters from data files for a given time and altitude. These include typical measurements, such as vertical wind shear and tropopause sharpness. In addition, we would want the data user to be able to correctly interpret the scientific parameter definitions and to be able access and read the following materials:

- Scientific output resulting from use of the data set,
- The MST international workshop conference proceedings,
- The MST user group meeting minutes.

This objective has the desired qualities of being specific, actionable, measureable and realistic.

While it could be tempting to try and specify a replication of current use, this may not be advisable. For example, if we had set the preservation objective as being the ability to study gravity waves or ozone layering occurring in the atmosphere above the MST site, we would rapidly discover that this is too vague an objective. This opens too many avenues of investigation when determining the skill and knowledge base needed to correctly interpret or analyze the data for these purposes. The unfortunate consequence would have been a time consuming analysis process and a lack of certainty that this objective had been achieved for future users.

An objective developed in accordance with methodology is heedful of the scope and preservation value of the dataset, whilst facilitating a measurable, testable solution.

Management of Research Assets Through the Modeling of Preservation Networks

Preservation analysis produces a solution whereby a number of digital interactions fulfill a preservation objective for a designated community. The preservation solution consists of a number of digital objects and sources of information, which will have been subjected to preservation action, such as format conversion or the addition of representation information. A future user may be required to interact with a number of unfamiliar digital objects in order to achieve meaningful reuse of data. As a result, an archivist will be confronted with the task of designing an information network which a future data user can navigate and effectively engage with. Solutions are also not permanent, with dependencies and associated risks which must be monitored by an archive. These risks must be monitored and managed by an archive, as the realization of these risks may result in a critical failure where the network cannot fulfill the defined objective.

A preservation information network model is a representation of the digital objects, operations and relationships, which allows 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. The model also directs the capture and description of digital objects which need to be packaged and stored within an OAIS compliant AIP.

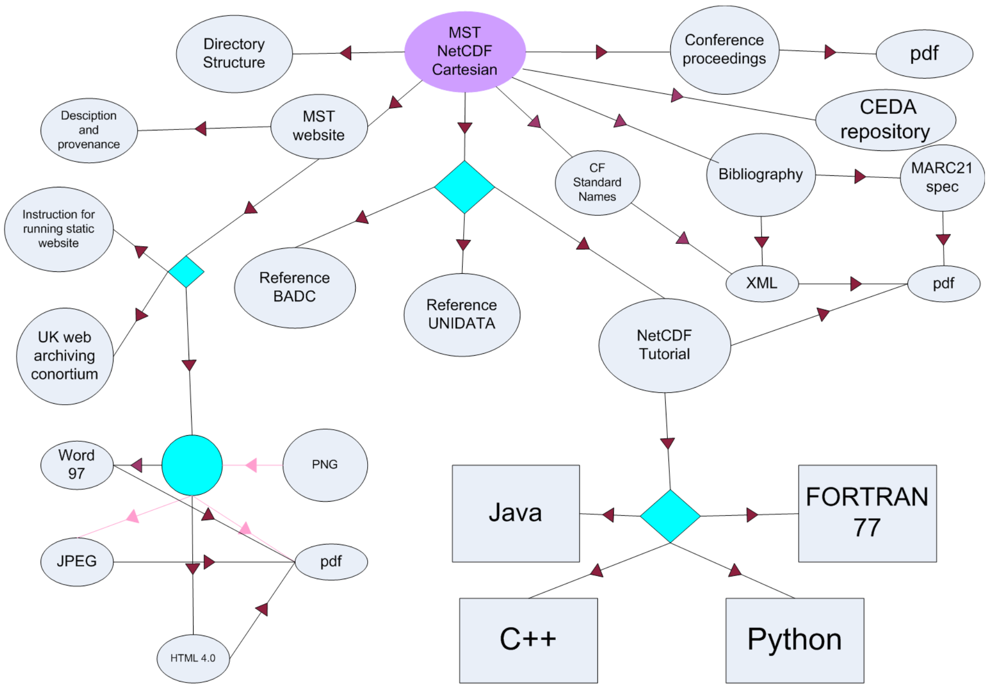

Figure 3. MST preservation network model. Some of the text in the smaller boxes is very difficult to read. Could you supply a clearer version of this diagram?

Figure 3 is preservation network model of a solution implemented on the CASPAR project. This model describes object and the functions which fulfill the preservation objective. Frequently, more than one strategy can fulfill a required function. This represented by the diamond symbol in the diagram above as an “or” relationship. Where more than one element is required to satisfy a prescribed function, this is represented by a circle as an “and” relationship. By modeling these functional relationships and attaching dependencies to the objects, monitoring and reviewing of solutions can be automated using knowledge management technologies, allowing critical failure of solution to be detected when risks are realized.

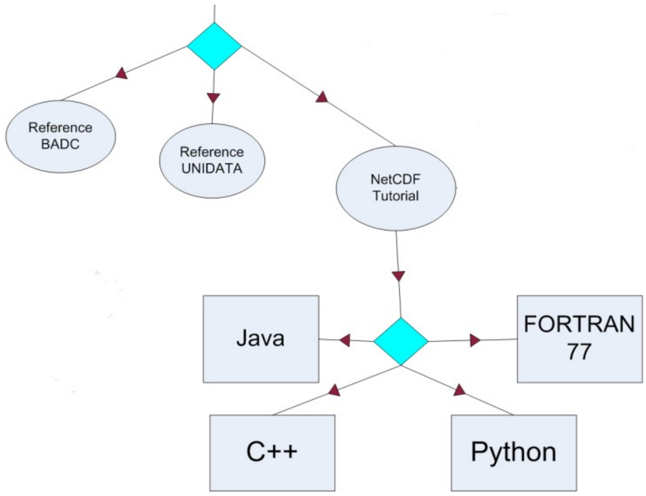

Figure 4. Preservation network model of a NetCDF data format (Network Common data form) reusable solution.

If we look at the network section above, which performs the function of allowing a user to extract the desired parameters from NetCDF formatted files, there are six different strategies a user can employ, all of which must fail before there is a critical failure of the solution. As this section of the network has a specific, well defined function, which is to allow a user to extract parameters form NetCDF formatted files, the solution can be deposited within a repository of representation information, such as RRORI. It can then be reused as part of wider solution for different atmospheric datasets which utilize the format.

Informing Preservation Activities in the Wider Institutional Environment

Preservation analysis clearly exposes the requirements and issues associated with an individual dataset, thereby supplying audit, certification and repository planning activities with vital information. 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 dataset level to be placed within the context of institutional planning and operations.

DRAMBORA

DRAMBORA describes digital curation as being characterized as a risk-management activity, where the job of digital curator is to rationalize the uncertainties and threats that inhibit efforts to maintain digital object's authenticity and understandability, transforming them into manageable risks. Six stages are implicit within the process. Initial stages require auditors to develop an organizational profile, describing and documenting the repository's mandate, objectives, activities and assets.

We view preservation analysis as being a complimentary activity, as it exposes the risks and objectives of a research assets created via the preservation analysis process. DRAMBORA can then proceed to assess these exposed risks in terms of their likelihood and potential impact, taking advantage of the clarity provided by preservation network modeling. Auditors are encouraged to conceive of appropriate risk management responses to the identified risk. The DRAMBORA process can then enable effective resource allocation, enabling repository administrators to identify and categorize the areas where shortcomings are most evident or have the greatest potential for disruption.

TRAC and the Expected ISO Standard

Work is proceeding under the auspices of the Consultative Committee on Space Data Systems (CCSDS), towards an ISO standard for audit and certification of trusted digital repositories[10]. Section 1.5 of the OAIS Reference Model [11]"Road Map for Development of Related Standards" included an item on standards for accreditation of archives, reflecting the long-standing demand for a standard against which repositories of digital information may be audited, and on which an international accreditation and certification process may be based.

The draft standard is based chiefly on the document "Trustworthy Repositories Audit and Certification: Criteria and Checklist" arising from the work of the RLG-NARA Task Force on Digital Repository Certification. The task force was formed between Research Libraries Group (RLG) and the National Archives and Records Administration (NARA). These experts were asked to define certification requirements, to delineate a process for certifications, and to identify a certifying body (or bodies) that can implement the process

These approaches are comprehensive, covering aspects of organizational commitment, reliability and practices, with a view to assessing all aspects of a repository's ability to preserve digital assets over the long term. The aim is that the eventual ISO standard will be usable for formal certification of repositories, as well as for repositories' own self-assessment.

The metrics are divided into three main sections, both in the TRAC version and the expected ISO standard:

1. Organizational infrastructure;
2. Digital object management:
3. Infrastructure and security risk management, or technologies, technical infrastructure and security.

By undertaking a preservation analysis, as described in this paper, evidence is provided for an audit according to the standard. The following are some examples of how the procedure of preservation analysis relates to the requirements of the audit. The metrics cited are from the draft standard, as submitted to the CCSDS editor. This is by no means an exhaustive list, but illustrative of the value of the preservation analysis approach from an audit and certification perspective.

[10] Digital Repository Audit and Certification Wiki: http://wiki.digitalrepositoryauditandcertification.org.
[11] OAIS Reference Model: http://public.ccsds.org/publications/archive/650x0b1.PDF.

Identifying Preservation Objectives provides evidence for:

> 3.1.3 The repository shall have a Collection Policy or other document that specifies the type of information it will preserve, retain, manage and provide access to.

Cost/Benefit Analysis - Resources and Staffing Aspects provides evidence for:

> 3.2.1 The repository shall have identified and established the duties that it needs to perform and shall have appointed staff with adequate skills and experience to fulfill these duties.

Information Flows provides evidence for:

> 4.1.2 The repository shall clearly specify the information that needs to be associated with specific Content Information at the time of its deposit.

Information Flows Diagram provides evidence for:

> 4.2.2 The repository shall have a description of how AIPs are constructed from SIPs.

Preservation Network Model provides evidence for:

> 4.3.3 The repository shall have mechanisms to change its preservation plans as a result of its monitoring activities.

PLATTER

The Planning Tool for Trusted Repositories (PLATTER) provides a basis for a digital repository to plan the development of its goals, objectives and performance targets over the course of its lifetime. PLATTER is designed to be complimentary to existing audit and certification tools.

The process is centered on a group of strategic objective plans. Preservation analysis can again inform this process by providing information on data complexity, data specialization, and acquisition through clearly identifying that which needs to be covered in a deposit agreement. It also provide information on specialization, the technical aspects by having highlighted the technical risks to the digital object within an AIP and, of course, the preservation plan.

Conclusions

This paper presented an overview of a preservation analysis methodology, which was developed on the CASPAR and DCC SCARP projects, placing it in relation to other digital preservation practices, and discussing how they can interact to provide archives caring for scientific datasets with the full arsenal of tools and techniques necessary to rise to the challenge of preserving a long term research asset. We have shown how the use of preservation analysis can provide greater return on investment, measurable solutions, assist the management if research assets and support audit/certification and repository planning activities.

Wider application, trialling and further development of the preservation analysis methodology outlined here would be desirable to test its validity in a broader range of disciplines and organisational settings. In addition, the production of training materials and support for archivists who wish to adopt our approach for data preservation would be of benefit.

Archives can find it difficult to articulate and specify reasons for the preservation of data. We additionally recommend that the organisations, such as the DCC, develop further guidance on setting preservation objectives and establishing valid business cases for the preservation of scientific data.

Acknowledgements

Work partially supported by European Community under the Information Society Technologies (IST) program of the 6th Framework program for project CASPAR and the Joint Information Systems Committee (JISC) for the Digital Curation Centre's SCARP project.

We would also like to thank our colleagues at STFC: David Hooper, Sam Pepler, Matthew Wild, Steve Crothers, Chris Davis, Rita Blake, Ruth Bamford, Simon Lambert, Stephen Rankin and Brian McIlwrath.

References

Consultative Committee for Space Data Systems (2002). Reference Model for an Open Archival Information System (OAIS). Recommendation for Space Data Systems Standard. CCSDS Blue Book. Retrieved July 8, 2011, from http://public.ccsds.org/publications/archive/650x0b1.pdf.

Conway, E. (2009). Curating atmospheric data for long term use: Infrastructure and preservation issues for the Atmospheric Sciences community. Retrieved July 13, 2011, from http://epubs.cclrc.ac.uk/work-details?w=50504.

Conway, E., Giaretta, D., Dunckley, M., Rankin, S., Wild, M., Davis, C., Lambert, S., Blake R. & Bamford, R. (2007) DCC Case Study: World Data Centre Ionosonde Holdings. Retrieved July 8, 2011, from http://www.casparpreserves.eu/other-caspar-products/other-caspar-products/ionosonde-case-study.pdf.

CRL & OCLC (2007). Trustworthy Repositories Audit & Certification (TRAC): Criteria and checklist. Retrieved July 8, 2011, from http://www.crl.edu/PDF/trac.pdf.