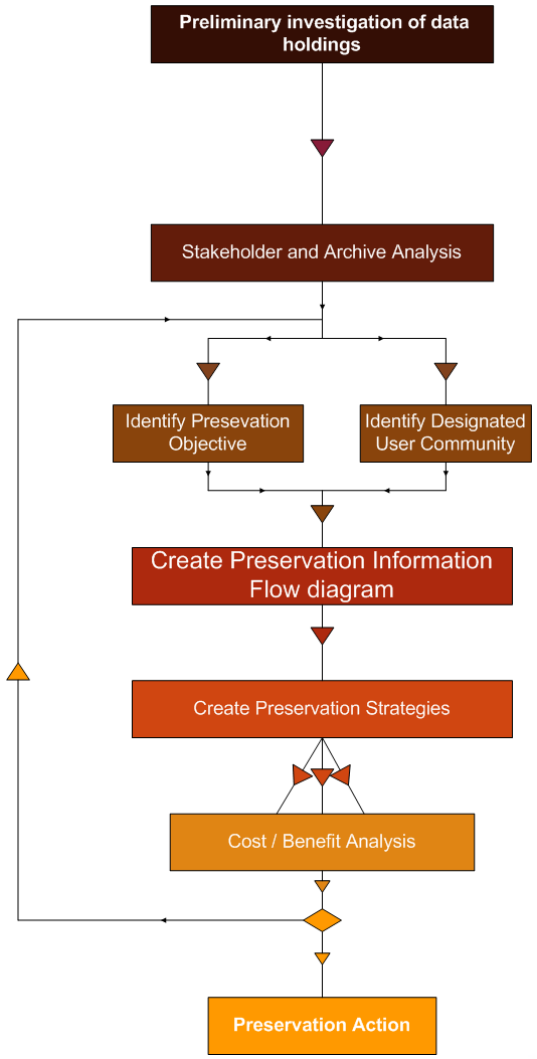
Curating Scientific Research Data for the Long Term: A preservation analysis method in context

Esther Conway, Matthew Dunckley, David Giaretta

12/05/2009

Science & Technology Facilities Council
e-Science

Overview of the Preservation Analysis Methodology



Preliminary investigation of data holdings

- 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

Stakeholder and Archive Analysis

- Stakeholders may hold different views of the knowledge a data set was 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

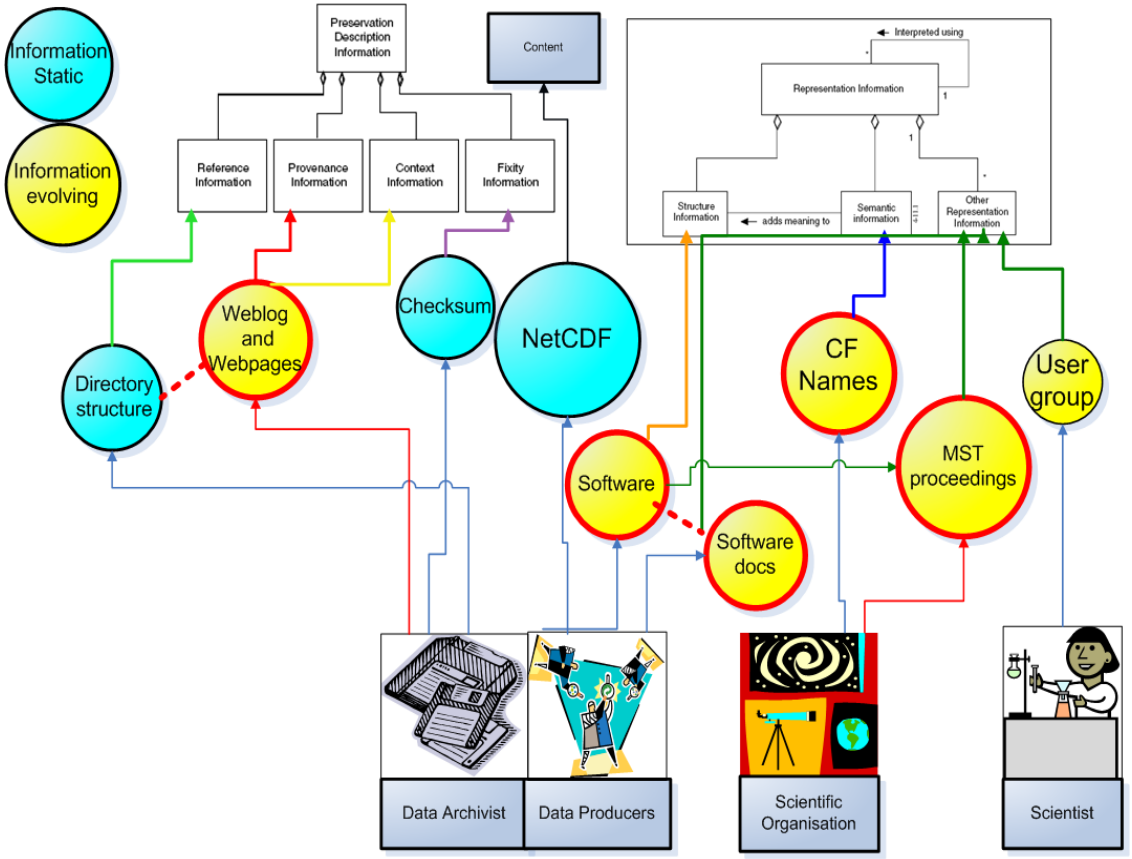
Defining a preservation objective

- 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

- 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.

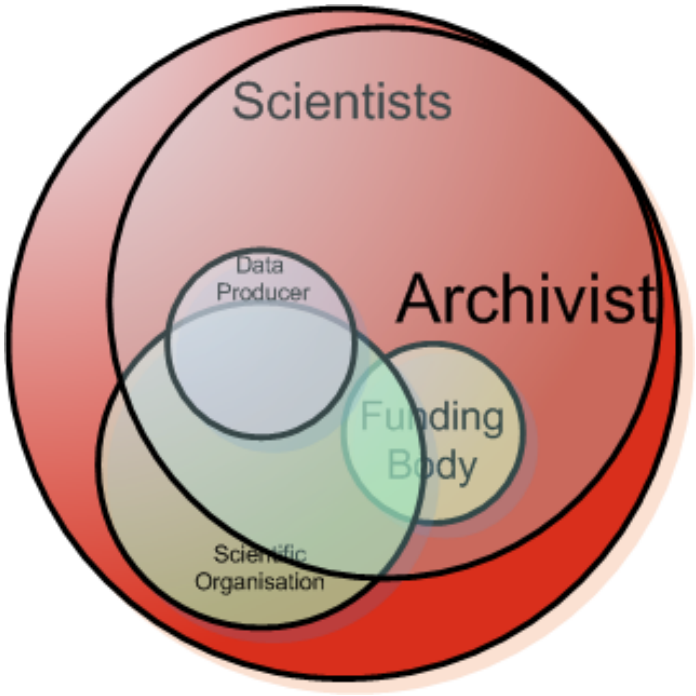
Preservation information flows

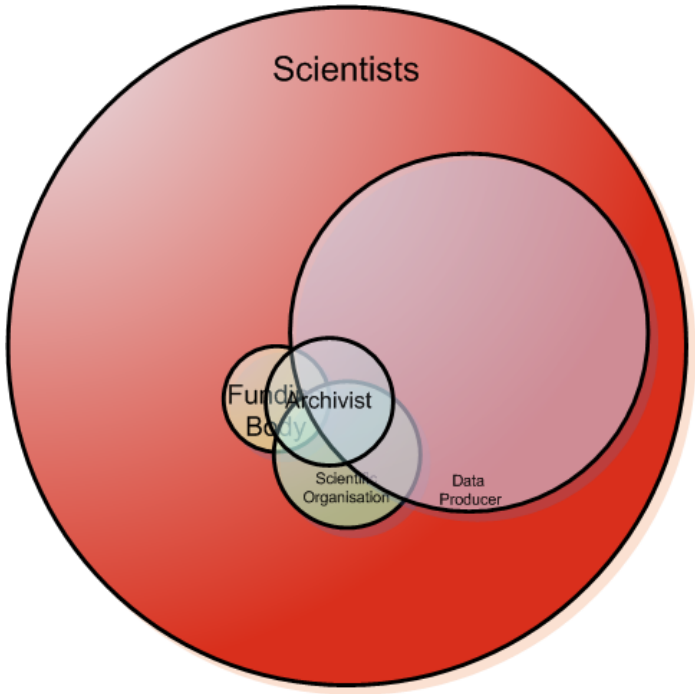


Cost/Benefit/Risk Analysis

- 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.

Maximizing return on investment





MST Data

- 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.

What can scientists understand using this data ?

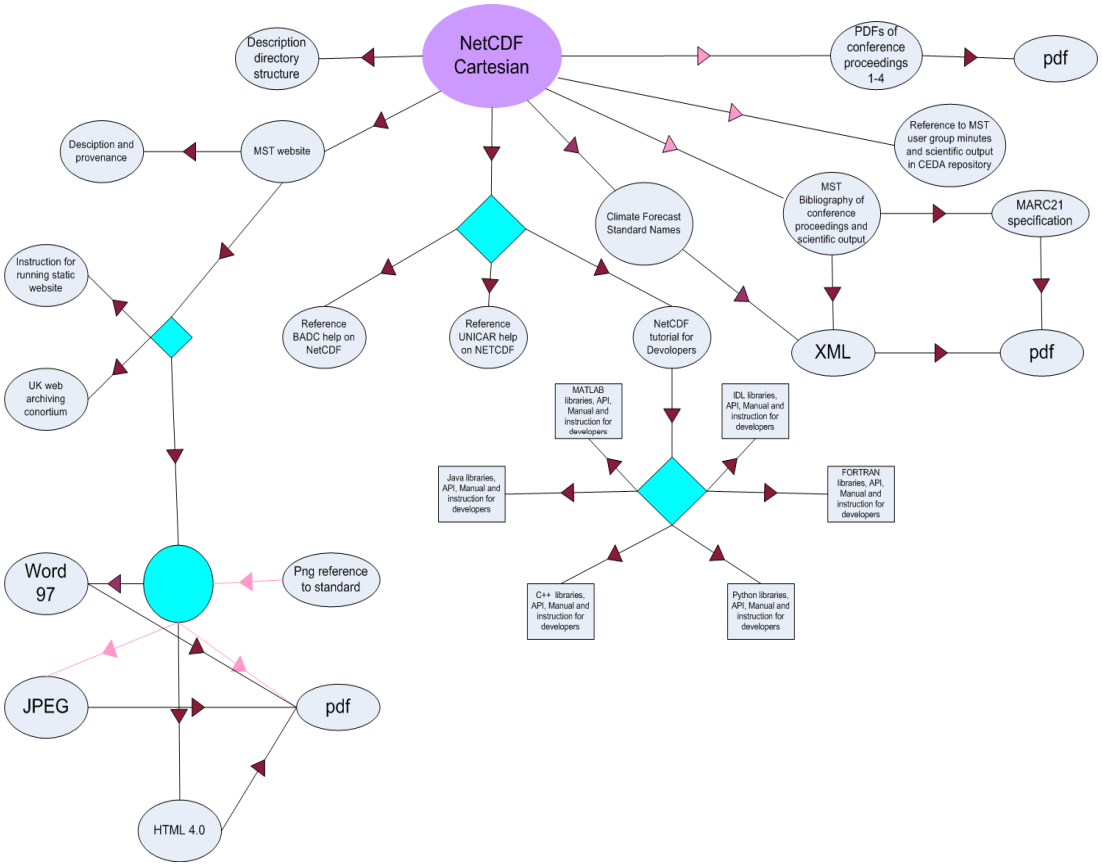
- **Convection.** Convection is the transfer of heat by movement within a substance. The MST radar data permits you to study the convective circulation of air within the atmosphere.
- **Gravity Waves.** Gravity waves are generated in the troposphere by frontal systems or by airflow over mountains. The 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 (atmospheric 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

The MST preservation objective

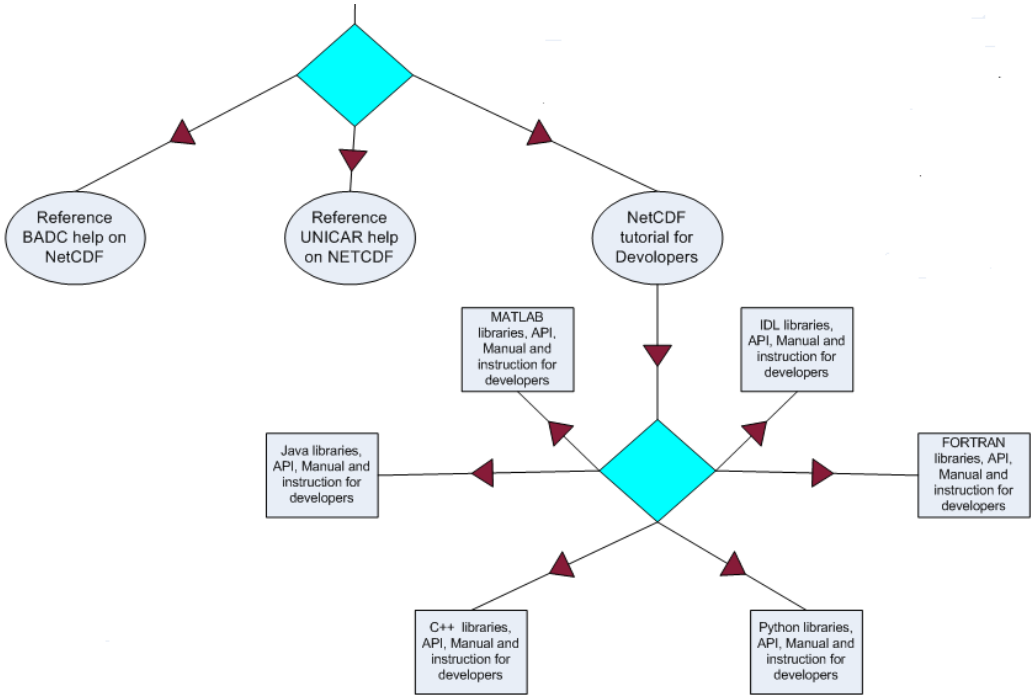
A user from a future designated community should be able to extract a specific set of 11 parameters from data files for a given time and altitude. These include typical measurements such as vertical wind shear and tropopause sharpness. In addition we would want the data user to be able to correctly interpret 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

Management of research assets through the modeling of preservation networks



Re-usable solutions



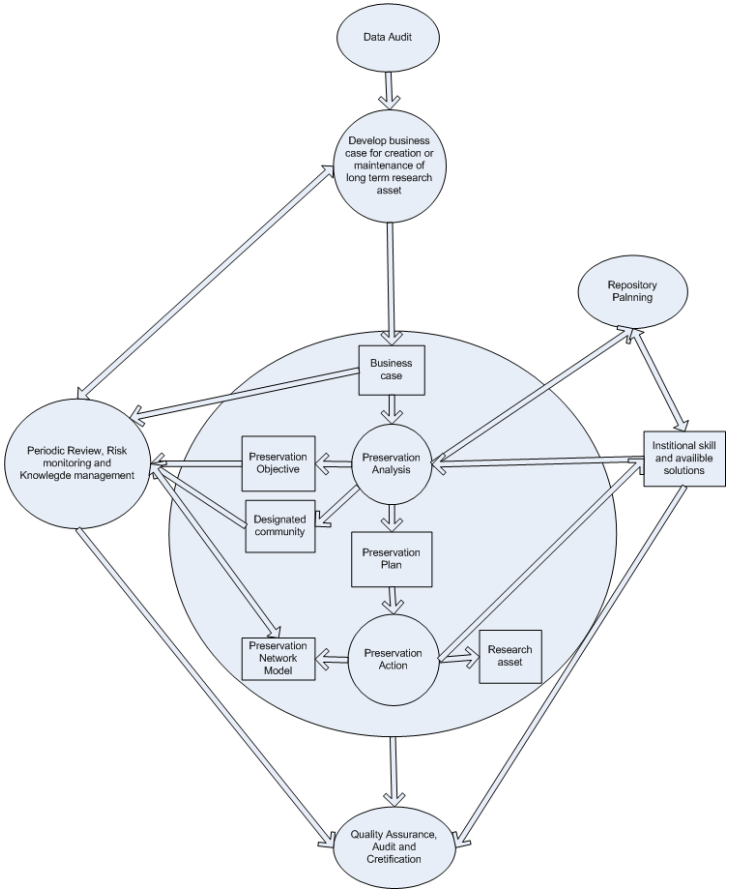
Informing preservation activities in the wider institutional environment

- Drambora
- TRAC
- PLATTER

TRAC

- By establishing duties the repository needs to perform,
- By identifying skills that staff require proving an adequate professional development program is in place
- Providing a definition of a repositories designated community
- Through preservation network modelling supporting periodic review
- By preservation analysis providing transparency.
- Repository will have identified properties it needs to preserve for digital objects,
- The repository will have clearly specified the information that needs to have associated with digital material at the time of deposit
- The repository will have documented preservation strategies it has employed

Moving towards a structured method for preservation processes



Summary

- Preservation Analysis Methodology
- Preservation Network Models
- Scarp project case studies
- Implemented Solutions using Digital Curation Centre and CASPAR software

Questions?

http://www.dcc.ac.uk/scarp

http://www.casparpreserves.eu