Geophysical Research Abstracts
Vol. 17, EGU2015-4022-1, 2015
EGU General Assembly 2015

Dynamic Data Management Based on Archival Process Integration at the Centre for Environmental Data Archival

Esther Conway, Alison Waterfall, Sam Pepler, and Charles Newey
CEDA, STFC, Didcot, United Kingdom (esther.conway@stfc.ac.uk)

In this paper we decribe a business process modelling approach to the integration of exisiting archival activities. We provide a high level overview of existing practice and discuss how procedures can be extended and supported through the description of preservation state. The aim of which is to faciliate the dynamic controlled management of scientific data through its lifecycle. The main types of archival processes considered are:

• Management processes that govern the operation of an archive. These management processes include archival governance (preservation state management, selection of archival candidates and strategic management) .
• Operational processes that constitute the core activities of the archive which maintain the value of research assets. These operational processes are the acquisition, ingestion, deletion, generation of metadata and preservation actvities,
• Supporting processes, which include planning, risk analysis and monitoring of the community/preservation environment.

We then proceed by describing the feasability testing of extended risk management and planning procedures which integrate current practices. This was done through the CEDA Archival Format Audit which inspected British Atmospherics Data Centre and National Earth Observation Data Centre Archival holdings. These holdings are extensive, comprising of around 2PB of data and 137 million individual files which were analysed and characterised in terms of format based risk. We are then able to present an overview of the risk burden faced by a large scale archive attempting to maintain the usability of heterogeneous environmental data sets.

We conclude by presenting a dynamic data management information model that is capable of describing the preservation state of archival holdings throughout the data lifecycle. We provide discussion of the following core model entities and their relationships:

• Aspirational entities, which include Data Entity definitions and their associated Preservation Objectives.
• Risk entities, which act as drivers for change within the data lifecycle. These include Acquisitional Risks, Technical Risks, Strategic Risks and External Risks
• Plan entities, which detail the actions to bring about change within an archive. These include Acquisition Plans, Preservation Plans and Monitoring plans
• The Result entities describe the successful outcomes of the executed plans. These include Acquisitions, Mitigations and Accepted Risks.