Geophysical Research Abstracts
Vol. 13, EGU2011-3368, 2011
EGU General Assembly 2011
© Author(s) 2011

Streamlining the Metadata Objects for Linking Environmental Sciences

Spiros Ventouras and Bryan N. Lawrence
Rutherford Appleton Laboratory - STFC, United Kingdom (spiros.ventouras@stfc.ac.uk)

The Metadata Objects for Linking Environmental Sciences (MOLES) model was originally developed within the Natural Environment Research Council DataGrid (NDG) project to fill a missing part of the 'metadata spectrum' (see Lawrence et. al. 2009). It is a framework within which to encode the relationships between the tools used to obtain data, the activities which organised their use, and the datasets produced. With an emphasis on the relationships between entities which descibe these things, it has a similar focus to the upcoming ISO19156 Observations and Measurements (O&M) standard (Cox 2010, personal communication), and work over the last few years has focussed on harmonsing MOLES and O&M. In this presentation we summarise the characteristics of V3.4 of MOLES, highlighting both the changes, and the reasons for the changes, from previous versions, and discuss techniques for migrating information beween different versions of the underyling information models.

MOLES is primarily of use to consumers of data, especially in an interdisciplinary context, to allow them to establish details of provenance, and to compare and contrast such information without recourse to discipline-specific metadata or private communications with the original investigators. MOLES can also be of use to the custodians of data, providing an organising paradigm for the data and metadata (which is how it is being deployed in the Centre for Environmental Data Archival, CEDA). The concepts of MOLES v3.4 are rooted in the harmonised ISO model, exploiting both the metadata standards (ISO 19115, ISO 19115-2) and O&M. Where possible MOLES exploits existing concepts and relationships, via specialisation and extension. MOLES is intended to complement other profiles of O&M (such as GeoSciML and CSML), not replace them.

The key structures of MOLES 3.4 revolve around specialisations of ISO19115-2 and ISO19156, aimed at capturing the important provenance elements associated with observations (and simulations) and the processes which produce them. MOLES is not particularly concerned with the details of the result structures themselves for which other metadata is used. Key components of MOLES include:
• Project descriptions
• The observation event itself, and
• The processes used to acquire or generate the observation
Where MOLES 3.4 differs from previous versions is in a much clearer delineation of key process components: distinguishing between acqusition (which inherits from the MI_AcquisitionInformation classes of ISO 19115-2) and computations (which inherit from the Processing parts of the Data Quality extensions in ISO19115-2). MOLES also addresses clearly the distinction between the acquisition of an observation, and the acquision of a specimen, from which an observation might be made.

The evolution from MOLES 3 has thus far been mainly influenced by constructing UML instances from a range of environmental sciences, however, recently MOLES development is being influenced by considering how we will do the migration of existing information instances. Currently MOLES 2.0 is deployed in the Centre for Environmental Data Archival (CEDA), but it is intended that MOLES 3.4 will be deployed in CEDA during 2011. Moving from the previous version where the key entities are "data production tool", "project", "data entity" and "deployment" (see Lawrence et al op cit), to the new version as described above, has a number of challenges – clearly new software infrastructure is required, but perhaps the larger challenge is migrating the existing information in a managed way.

CEDA is currently trialing using RDFa to markup important concepts foreshadowed for MOLES 3.4 in MOLES 2.0 entity XHTML attributes, so that the information migration can be carried out independently of the software evolution. This exercise appears both useful in helping stage the information migration and in informing the information modelling itself, with some key attributes in MOLES 3.4 being informed by the experiences of trying to establish information in the MOLES 2.0 context. We believe this technique for managing the information evolution (as opposed to the underlying information model) is an important new innovation, and we will report

experiences with this activity.

B.N. Lawrence,R. Lowry, P. Miller, H. Snaith, and A. Woolf (2009): Information in environmental data grids. Phil. Trans. R. Soc. A, 367, 1003 – 1014. doi:10.1098/rsta.2008.0237.