Information Modelling

MOLES
Metadata Objects for Linking Environmental Sciences

S. Ventouras
Rutherford Appleton Laboratory
OUTLINE

• Introduction
  - What MOLES is about
  - MOLES v2

• MOLES v3.4
MOLES model refers to B-Browse Type Metadata understanding the context of data and choosing between similar data sets

MOLES model was originally developed within the Natural Environment Research Council DataGrid (NDG) project to fill a missing part of the ‘metadata spectrum’

[Lawrence et al. doi:10.1098/rsta.2008.0237]

Currently MOLES v2.0 is deployed in CEDA
Key Entities

**DataEntity**: which consists of aggregations of more Data entities

**DataProductionTool**: which characterizes the instruments and/or processes available for producing data

**Observation Station**: which characterizes the location(s) (and observers) of data production

**Activity**: which characterizes the projects and campaigns etc., associated with data production

These four key entities are related by the **Deployment**, which binds a production tool deployed at an observation station on behalf of an activity to produce data.
Conceptual modelling is one of the important components in the development of any information system.

Owing to the lack of structured approaches to system design:

- the time or resources required for a system development are typically underestimated
- the system is inefficient to meet the demands of the intended end-users
- documentation is limited
- maintenance is difficult.
MOLES 3.4 is rooted in the ISO 19100 series of standards from the ISO/TC 211 “Harmonised Model”. In particular, it has been created:

a) following the guidance provided by ISO/TC 211 (i.e. ISO 19101, 19106 and 19109) concept of feature (type and instance)

a) has been formalised in the Unified Modelling Language (UML) ISO/IEC 19501, following the guidance of ISO/TS 19103.

c) by integrating reusable modules of conceptual schemas defined within ISO 19100 series e.g. temporal schema, metadata, Observations & Measurements etc.

User Requirements

MOLES v3.4 has been influenced by a range of environmental sciences (e.g. Earth Observation, Meteorology, Geochemistry, etc)
The general structure of the model gives a central place to the concept of **observation**.

**ISO 19156**

An **observation** is an act that results in the estimation of the **value** of a feature **property**, and involves application of a specified **procedure**, such as a sensor, instrument, algorithm or process chain.

An **observation** is an act associated with a **discrete time instant or period** through which a number, term or other symbol is assigned to a phenomenon. The **result** of an observation is an estimate of the value of a **property** of some **feature**, so the details of the observation are metadata concerning the value of the feature property.
MOLES – v3.4

Base Standard: ISO 19156 Observations & Measurements

```
«FeatureType»
OM_Observation
+ parameter: NamedValue [0..*]
+ phenomenonTime: TM_Object
+ resultQuality: DQ_Element [0..*]
+ resultTime: TM_Instant
+ validTime: TM_Period [0..1]

constraints
{observedProperties shall be a phenomenon associated with the feature of interest}
{procedure shall be suitable for observedProperty}
{result type shall be suitable for observedProperty}
(a parameter.name shall not appear more than once)
```
Key components of MOLES v3.4 include:

- **Project** descriptions
- The **observation** event itself (ISO 19156), and
- The **processes** used to acquire or generate the observation (ISO 19156, 19115 and 19115-2)
MOLES – V3.4 Structure

- **Observation**
  - Feature Of Interest (Subject of the Observation)
    - Specimen
  - Observed Property
  - Result
  - Process
    - Observation Process
      - Acquisition
      - Computation
    - Specimen Process
      - Specimen Acquisition

- **Project**

- **Observation Collection**
Leaf packages of the MOLES3.4 schema
Observed Property

```plaintext
«metaclass»
General Feature Model::
GF_PropertyType
{root}
+ definition: CharacterString
+ memberName: LocalName

instanceOf

«dataType»
MO_AbstractObservableProperty
+ description: CharacterString [0..1]
+ label: CharacterString [0..*]

instanceOf

«dataType»
MO_ObservablePropertyConstraint
+ description: CharacterString [0..1]
+ label: CharacterString

«CodeList»
MO_BasePhenomenonValue

+ subConstraint 0..*

«CodeList»
MO_CompositeObservableProperty

+ component 2..*

«dataType»
MO_SimpleObservableProperty
+ basePhenomenon: MO_BasePhenomenonValue
+ constraint: MO_ObservablePropertyConstraint [0..1]
```
‘Acquisition’ and ‘Computation’ can serve as a basis for describing any observation chain process.

**Acquisition**
A process step which interacts with the feature of interest (e.g. an atmospheric column, a specimen etc) to provide a result.

**Computation**
A process step which involves only numerical computation.
Process Computation

Data quality information - Imagery::LE_Source
  + processedLevel: MD_Identifier [0..1]
  + resolution: LE_NominalResolution [0..1]

Data quality information - Imagery::LI_Source
  + description: CharacterString [0..1]
  + scaleDenominator: MD_RepresentativeFraction [0..1]
  + sourceCitation: CI_Citation [0..1]
  + sourceExtent: EX_Extent [0..*]
  + sourceReferenceSystem: MD_ReferenceSystem [0..1]

MO_Computation
  «featureType»

MO_Process
  «featureType»

MO_Processing
  «featureType»
  + description: CharacterString [0..1]
  «voidable»
  + documentation: MO_Citation [0..*]
  + identifier: MD_Identifier [0..*]
  + processingInput: MO_InputOutput [0..*]
  + processingOutput: MO_InputOutput
  + relatedParty: MO_ResponsibelPartyInfo [1..*]
  + softwareReference: MO_Citation [0..*]

MO_InputOutput
  «dataType»
  + description: LE_Source
  + name: CharacterString [0..1]

Data quality information - Imagery::LE_Algorithm
  + citation: CI_Citation
  + description: CharacterString

MO_ResponsibelPartyInfo
  + relatedEntity: MO_ResponsibelPartyInfo [1..*]

MO_Citation
  + citation: CI_Citation
  + description: CharacterString
**Project**

**Definition**
An identifiable activity designed to accomplish a set of objectives.

**Description**
A typical sequence of data capturing involves one or more projects under which a number of further activities are undertaken, using appropriate tools and methods to produce the datasets.

EXAMPLE 1 The UK Surface Ocean /Lower Atmosphere Study (UK SOLAS) was a Directed Mode programme of the Natural Environment Research Council (NERC). The overall aim of UK SOLAS was to advance understanding of environmentally significant interactions between the atmosphere and ocean, focusing on material exchanges that involve ocean productivity, atmospheric composition and climate.

NOTE A project may be a member of a complex of parent- and sub-projects.
**Observation Collection**

**Definition**
A collection of existing observations which has a title and a reason for existence.

**Description**
NOTE  The observation results can be appeared in collections which are organised with significantly more flexibility than would be done if one used the original project alone.
Implementation

- From UML diagrams to XMI (From Enterprise Architect) then
- XML/GML
  - FullMoon (CSIRO), NewMoon (CEDA)
- DB
  - CEDA activity, ongoing
- OWL-DL
  - ISO 19152?