Metadata and Coupling

Rupert Ford
Overview

- What is Metadata?
- Coupled model configuration Metadata
- Aside: Metadata/code consistency
- Provenance Metadata
- Configuration vs. Provenance Metadata
- Vision
What is Metadata?

- U.S. trademark “The Metadata Company” 1986
- “data about data”
- "Metadata is used to facilitate the understanding, use and management of data."
- One persons data is another persons Metadata
Simple illustration

- We have the value 9.8
- Some additional information
  - “gravity (on earth at sea level)”
  - “m/s2”
  - “accurate to 2 decimal places”
  - “used in simulation X”
- Common vocabularies essential (e.g. CF, udunits, CIM)
Coupled model configuration

Metadata

- Metadata input to Coupled Models
- coupling systems (at least) separate composition information from component code
- Some information may be coded
Coupled model configuration

Metadata

- OASIS3 NAMCOUPLE
  - source/target fields
  - source/target grid
  - Transformations
  - coupling or I/O
  - coupling rate
Coupled model configuration

Metadata

- OASIS4 XML
- AD, PMIOD (Application and component model description)
- SMIOC, SCC (coupled configuration) xml files
Coupled model configuration

Metadata

• METADATA in ESMF
  • Coded
  • Rocky’s configuration approach
Coupled model configuration

Metadata

- Metadata in BFG
  - Model/component interface description
  - Coupled model composition information
    - Connections
    - Schedule
  - Coupled model deployment information
    - Mapping to executables
    - Mapping to threads
    - Choice of framework
    - [Partition descriptions]
- Grid description
Aside: Metadata/code consistency

- Where to maintain component metadata?
- Code is master or description is master?
- ESMF, OASIS - mostly code
- BFG - description
- Consistency between the two
  - analyse code: Code->(OFP)->XML->(XSL)->CIM
Provenance Metadata

CIM development

“The main objective of METAFOR is to develop a Common Information Model (CIM) to describe climate data and the models that produce it in a standard way, and to ensure the wide adoption of the CIM”

Goal:

• One normative artefact – the UML model
• Derived XSD generated automatically
An essential aim of **Metafor** is that the conceptual model is not changed by the manor in which it is used or applied.
Metafor Infrastructure

Develop, deploy, and evaluate a prototype infrastructure that will allow key data and models to be discovered and compared between distributed digital repositories.

- Single sign-on services to populate and manipulate, the CIM metadata
- Services exploit NDG CSML to provide a common Geographic Markup Language interface to climate data
- Centralized CIM content harvested from individual repositories using OAI-PMH (Open Archives Initiative Protocol for Metadata Harvesting).
CIM for Configuration Metadata

- 2 Metafor deliverables
  - CIM enabled OASIS4
  - CIM enabled BFG

- Metafor developing the CIM to drive the configuration of OASIS4 and BFG. Hoping to liaise with ESMF developers
Coupling Technologies for Earth System Modelling: Today and Tomorrow. Toulouse 15th-17th December 2010
Configuration vs. Provenance Metadata

- Configuration – what will happen (need to know options)
- Provenance – what happened
- component model interface – all possible component inputs/outputs vs. only those that were connected
Configuration vs. Provenance Metadata

Scientific Provenance – don't care about how produced if results are the same, so deployment information possibly not required

Coupler independent description?
Configuration vs. Provenance Metadata

Configuration – some values may be coded, some may be separate Metadata

Provenance – all is Metadata

e.g. ESMF uses a coded API. Therefore there is no configuration information.
Vision

- Standard configuration descriptions using the CIM
- Standard (Scientific?) coupling provenance information using the CIM
- CIM (or some future derivative) as the "lingua franca" for both provenance and configuration Metadata.
Thanks!