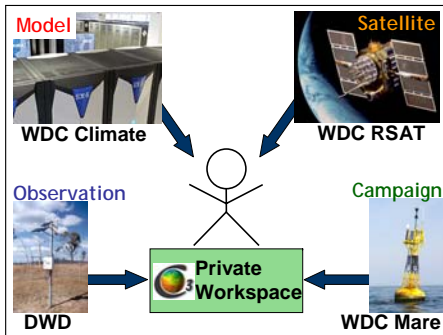


Introduction

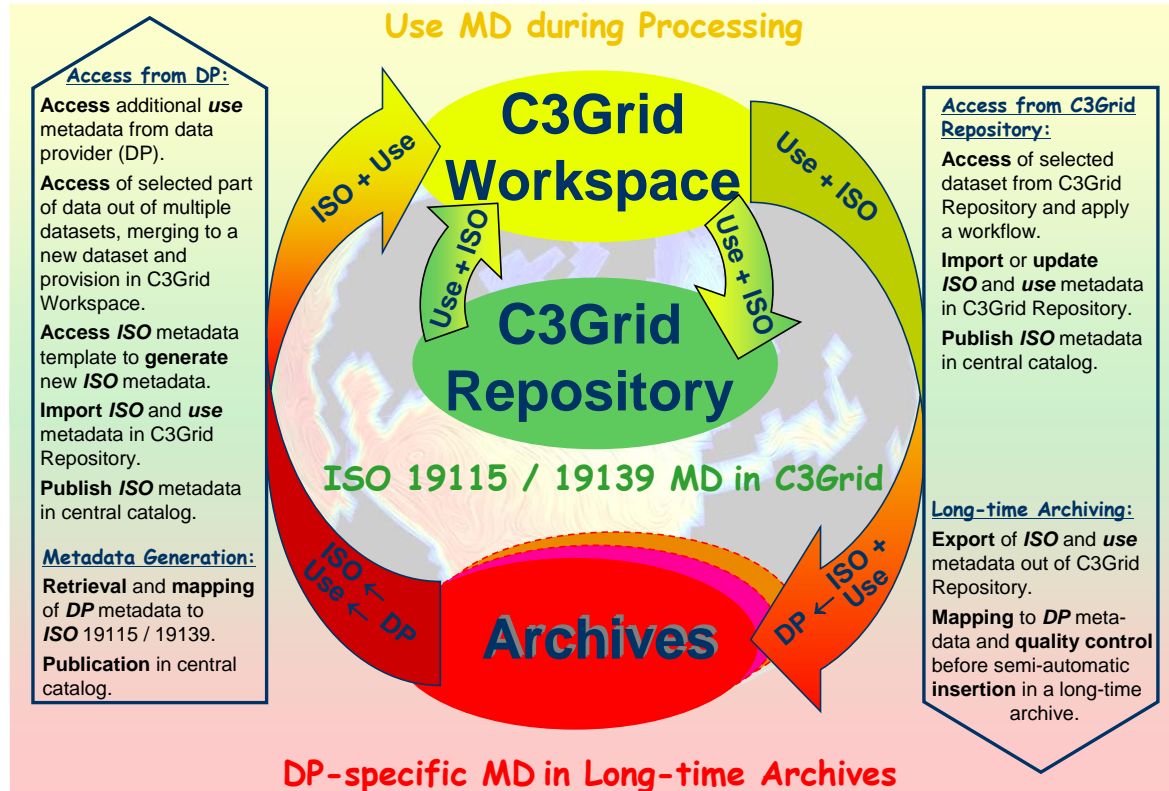
Earth System Sciences strongly depend on data comparison. These data, resulting from observations or modeling studies, are distributed over many archives, and vary tremendously with respect to description, access and quality.

In order to enable scientific work on a grid platform like C3Grid (Collaborate Climate Community Data and Processing Grid), thorough and continuous handling and management of the accompanied metadata as well as an intelligent organization of these derived metadata products are essential.



Metadata Situation in Earth Sciences

- none or highly specific metadata at the different institutes
- groups of datasets homogeneously describable, e.g. model run or satellite data
- no automatic metadata handling throughout scientific workflows
- creation of metadata is done for sharing data in a specific stage of the overall workflow



Access from DP:
Access additional *use* metadata from data provider (DP).
Access of selected part of data out of multiple datasets, merging to a new dataset and provision in C3Grid Workspace.
Access ISO metadata template to **generate** new *ISO* metadata.
Import ISO and *use* metadata in C3Grid Repository.
Publish ISO metadata in central catalog.

Metadata Generation:
Retrieval and mapping of *DP* metadata to *ISO* 19115 / 19139.
Publication in central catalog.

Access from C3Grid Repository:
Access of selected dataset from C3Grid Repository and apply a workflow.
Import or update ISO and *use* metadata in C3Grid Repository.
Publish ISO metadata in central catalog.

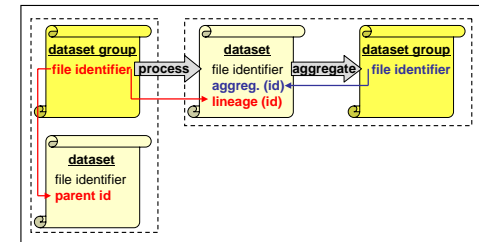
Long-time Archiving:
Export of ISO and *use* metadata out of C3Grid Repository.
Mapping to DP metadata and **quality control insertion** in a long-time archive.

Organization Principles for Metadata

The granularity of the discovery metadata should reflect the logical organization of the data repository at a sufficiently coarse grained level [1].

Grouping aspects of ISO 19139 used by C3Grid:

- **historical – lineage information:** for data provenance information
- **hierarchical – parent-child concept:** for collections of homogeneous datasets, i.e. belonging to one product, together accessible
- **topical – aggregation information:** for collections of inhomogeneous datasets belonging to different data products



Outlook

- Shibboleth based AA infrastructure
- sustainable distributed C3Grid Repository
- integration of automatic metadata quality assurance procedures
- metadata tool framework integration (rf. poster XY0265 on 17.04.2008)
- usage of C3Grid infrastructure for IPCC-AR5 data access and preprocessing from WDC Climate DB is assigned (stress test)

References:
 [1] INSPIRE: DT Metadata - Draft Implementing Rules for Metadata (Version 2, 02/02/2007).
www.ec-gis.org/inspire/reports/ImplementingRules/draftINSPIREMetadataRv2_20070202.pdf

Metadata in C3Grid

- uniform common metadata descriptions: C3Grid profile of ISO 19115 / 19139 used for discovery
- availability of accompanied ISO 19139 metadata to every dataset in C3Grid during its whole lifecycle
- data specific additional use metadata for data processing stored in C3Grid repository

C3Grid Support Components

- publication interface for metadata: OAI / PHM
- central metadata catalog: lucene index
- GUI for metadata creation: XForms (Oberon)
- metadata management: XML DB (eXist), OGSA-DAI
- shared community resources: Workspace and Repository, and Compute resources integrated by C3Grid middleware