

## The Problem

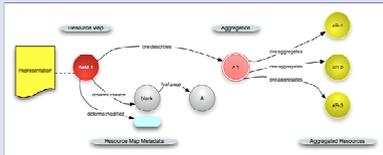
There are currently no broadly accepted mechanisms for recording, preserving, and making available the provenance and contextual information the science community needs to ensure usable data and replicable results in the Earth sciences and it is generally accepted that existing provenance models, such as OPM, are not sufficient

## The Question

Can the Open Archives Initiative's Object Re-use and Exchange Protocol (OAI-ORE) be used to describe and make available provenance and contextual information about digital data sets in the Earth Sciences?

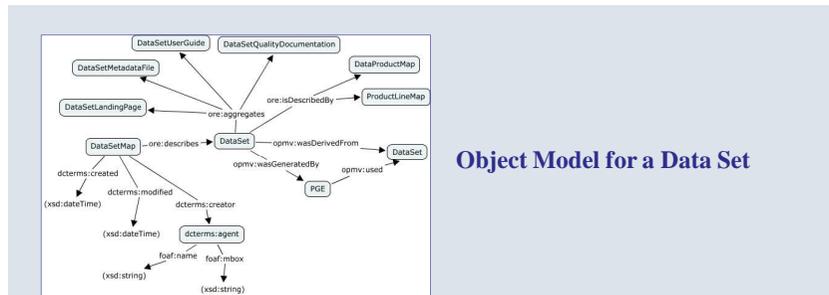
### What is OAI-ORE?

- A consistent methodology to describe the constituents or boundary of an aggregation
- Uses Resource Maps to communicate the relationships between members of a complex aggregation
- Resource Maps may be serialized as RDF XML, RDFa, or XML based ATOM
- URI's are used to identify:
  - individual components within an aggregation,
  - the aggregation itself (but this does not have to resolve to an actual object),
  - Resource Maps
- An example Resource Map:

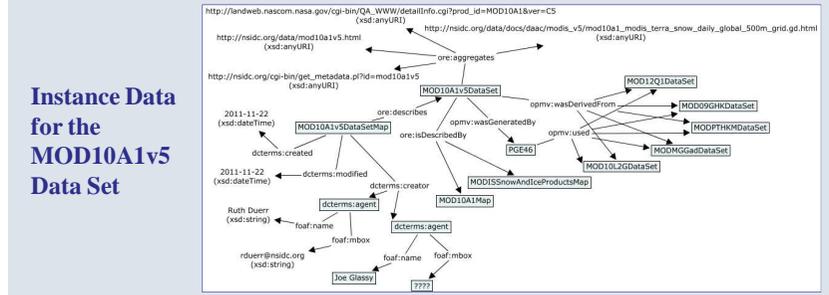


### Methods

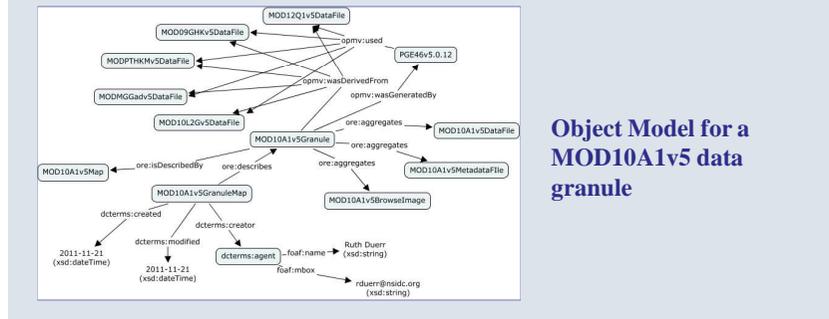
- Using MODIS/Terra Snow Cover Daily L3 Global 500m Grid (MOD10A1) as the test data set
  - Current data is from Version 5 of the data set
  - Data set has ~1.3 million granules of data, growing by ~ 300 granules per day of data
- Assessed existing MODIS information sources to determine types and levels of aggregations needed
- Assessed ESIP's developing Provenance and Context Content Standard (PCCS) to determine which information object belongs to which aggregation
- Develop object and instance models for each type of aggregation



Object Model for a Data Set



Instance Data for the MOD10A1v5 Data Set



Object Model for a MOD10A1v5 data granule

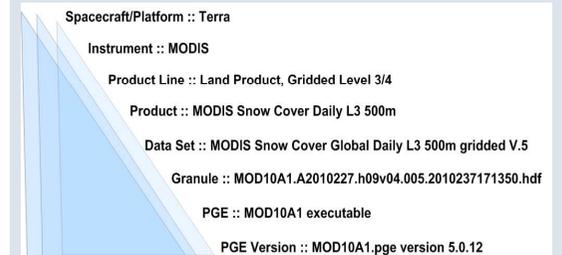
### Summary of PCCS Mapping to Aggregations

Number of PCCS Refs	Aggregation Category
10	data sets in their own right
7	data set aggregation
8	Partly in data set, partly in product or product line aggregation
2	Partly in granule and partly in data set aggregations
1	Split over granule, product, product line, and data set aggregations
1	Split over granule and PGE aggregations
3	Instrument aggregation
8	PGE aggregation
2	PGE version aggregation
2	Split over product and product line aggregations

### Provenance and context references:

- Reference Model for an Open Archives Information System (OAIS), Draft Recommended Standard, CCSDS 650.0-P-1.1 (Pink Book) Issue 1.1 August 2009
- U.S. Global Change Research Program (USGCRP). Global Science Requirements for Long-Term Archiving, Report of the Workshop, October 28-30, 1998, Boulder, Colo., March 1999.
- ESIP, Provenance and Context Content Standard, (in development)

### Types of Aggregations



### Issues

- It is not clear that mapping up from one aggregation to a higher level aggregation (e.g., from a granule to the data set to which it belongs) is acceptable OAI-ORE practice
- Provenance (and potentially other) relationships are used to relate individual objects within an aggregation to objects in other aggregations. From the OAI-ORE specification this seems to be acceptable practice.

### Early Results

- While OAI-ORE on its own is insufficient, in combination with other provenance models appears to be sufficient
- Developing a variety of aggregations at various levels appears to partition the provenance in a useful and maintainable way
- In particular, combining OAI-ORE with provenance models such as OPMV appears to be a fruitful path forward