

# Metadata of the NERC DataGrid

Kevin O'Neill  
CCLRC e-Science Centre  
[k.d.o'neill@rl.ac.uk](mailto:k.d.o'neill@rl.ac.uk)



## Resources

- Funded by NERC and National e-Science Core programme
- Initially approx 15 sy over 3 years

## Investigators

- Bryan Lawrence (PI) – BADC
- Roy Lowry – BODC
- Kerstin Kleese van Dam – CCLRC e-Science Centre

Close co-operation with the Earth System Grid in the States



Starting with Atmospheric and Oceanographic data, held in the BADC and BODC, provide a framework for:

- Data Discovery
- Data Use with format transparency

within the environmental community funded by NERC.

It should be cross-disciplinary (big and small borders!)

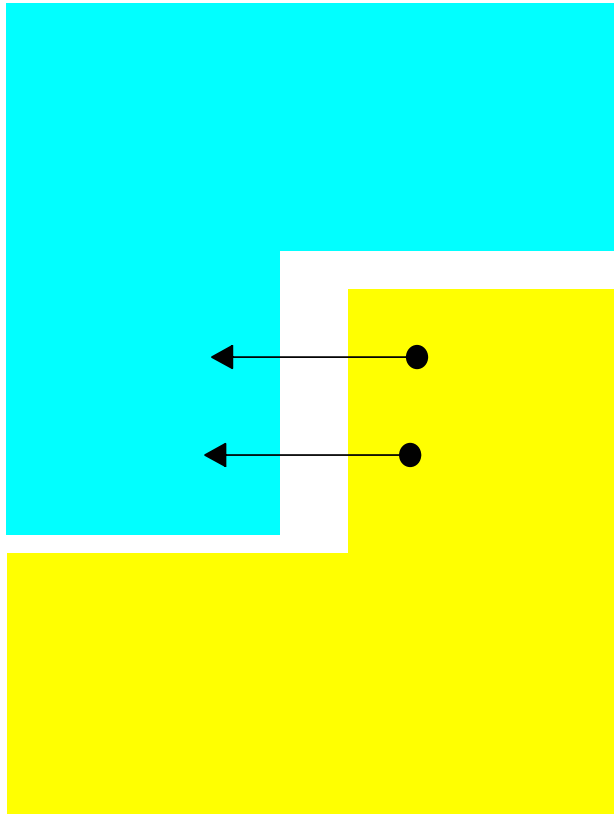
- Global warming and similar are no respecters of academic boundaries...

Build on existing initiatives and standards

- Not a green field site! Much existing work, albeit normally covering limited areas
- Particularly important are OGC and ISO programmes



# Relation of data and metadata models



## Provides clear separation of function

- Difference between data use and discovery etc.
- “Tuning” of metadata to include relevant detail

## Allows increased reuse of metadata model

- Avoids tie-in to details of a particular field’s data formats
- Can plug-in another data model
- (and the same applies to the data model...;-)

# A short taxonomy of the NDG metadata

## A - [Archive]

- Usage metadata, normally generated with the data, and always accompanying it. (For example, the metadata held in a NetCDF formatted file).

## B - [Browse]

- Complete set of metadata which covers both semantics, and syntax). This metadata is often built up over time.

## C -[Comment]

- Ancillary metadata, such as annotations and publications. Usually provided after the ingestion of data into an archive.

## D - [Discovery]

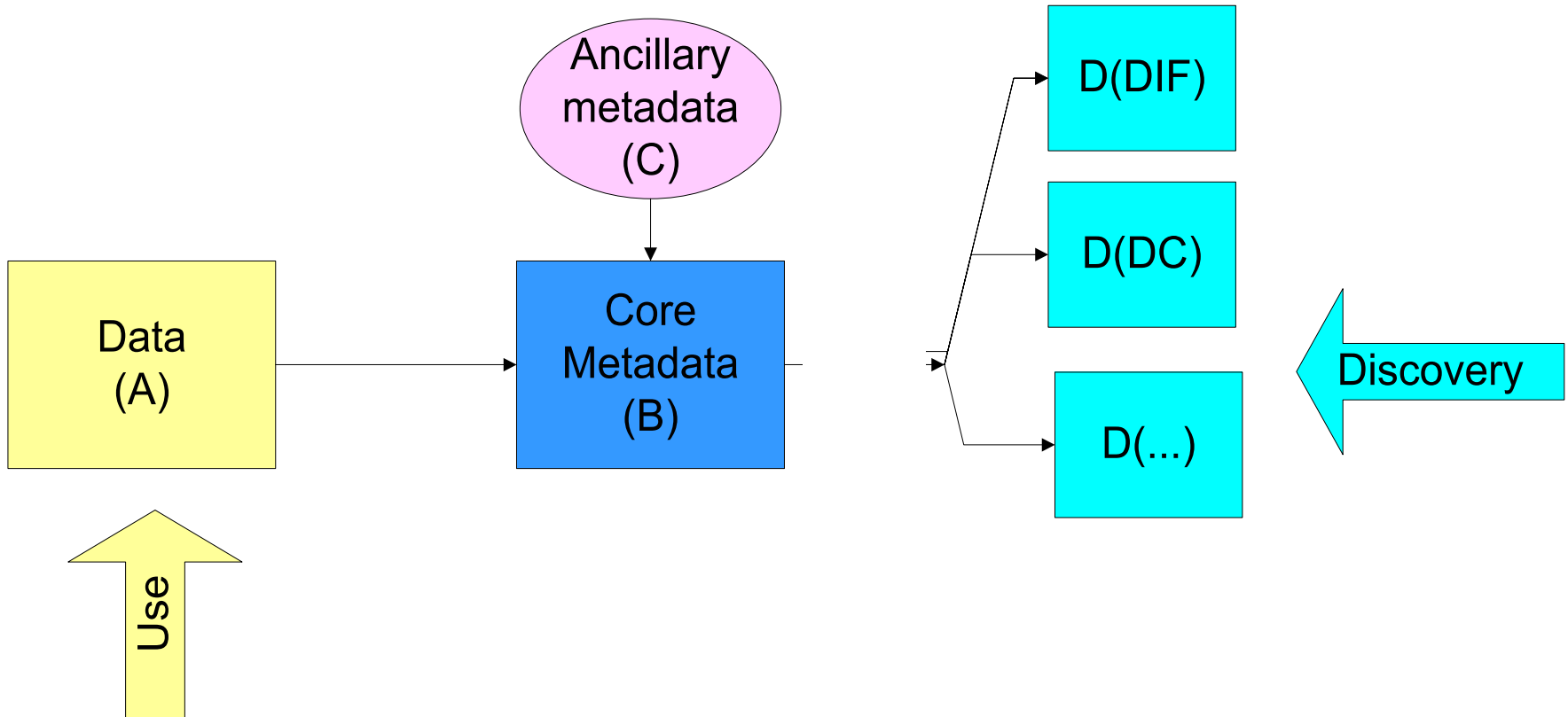
- The metadata needed to find datasets of interest. Usually produced by managed data centres, e.g. DIF, GEO, CLRC-MDF

## E – [Extra]

- Metadata specific to a discipline, which is not yet incorporated into the core



# Using metadata within the NDG

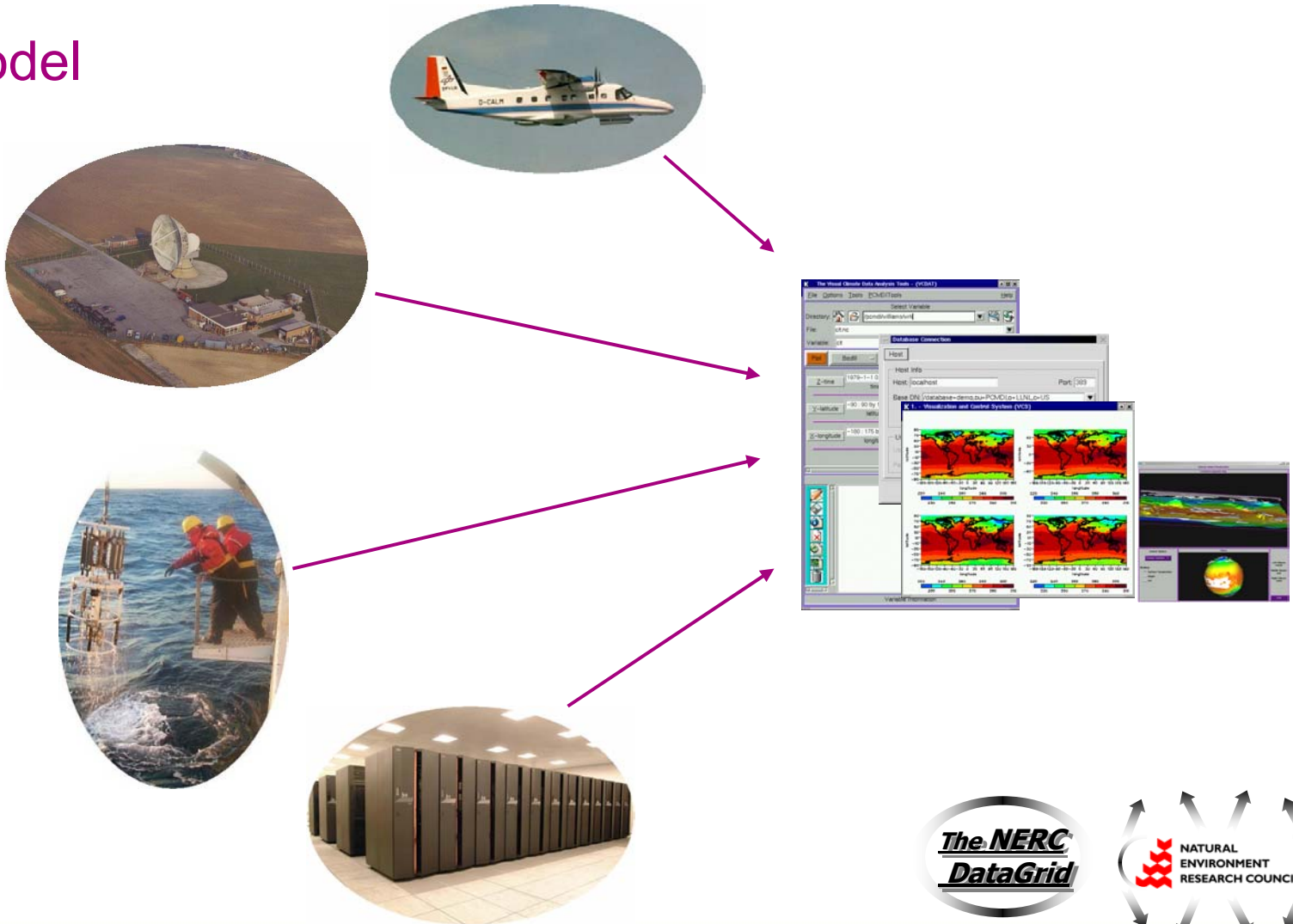


A & B overlap in data summary

D is usually a subset of B, but may be supplemented with E

# NERC DataGrid Data Model

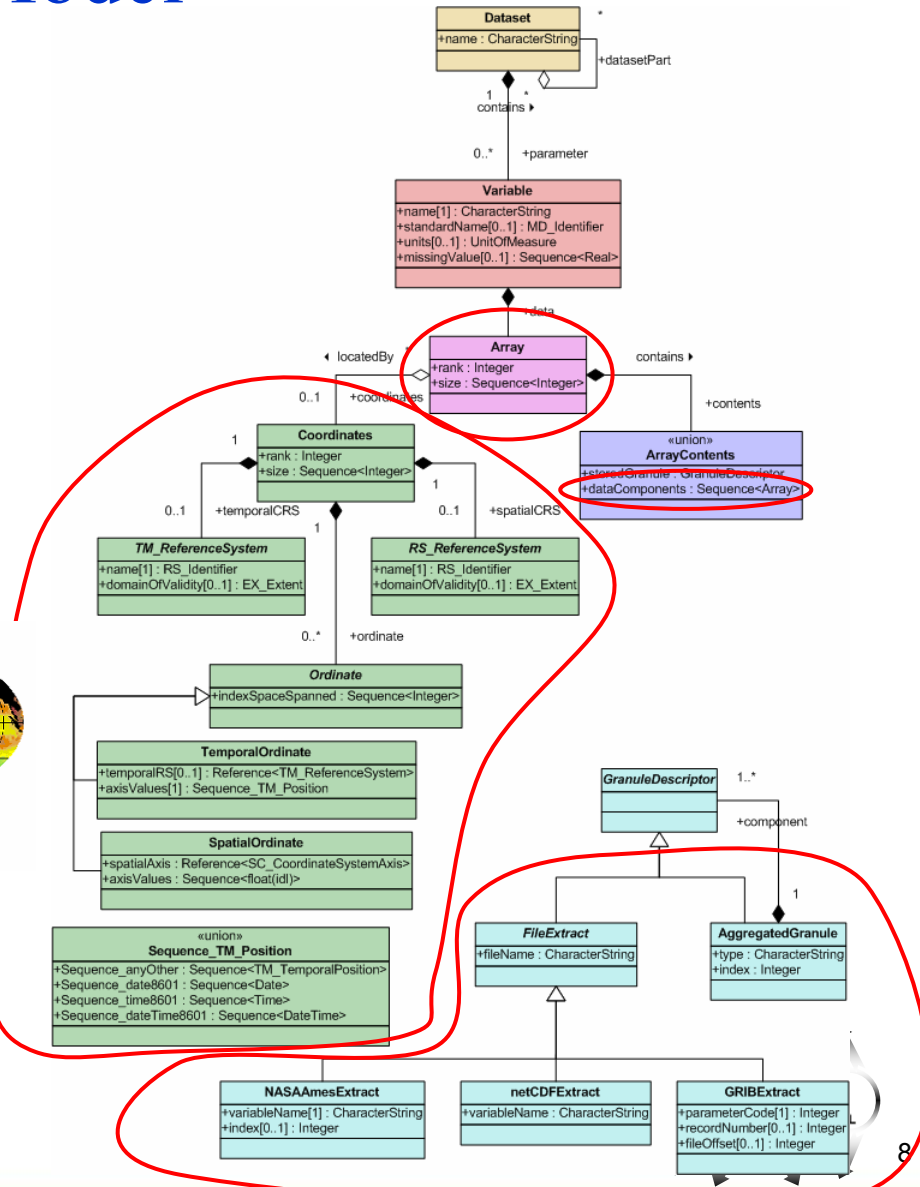
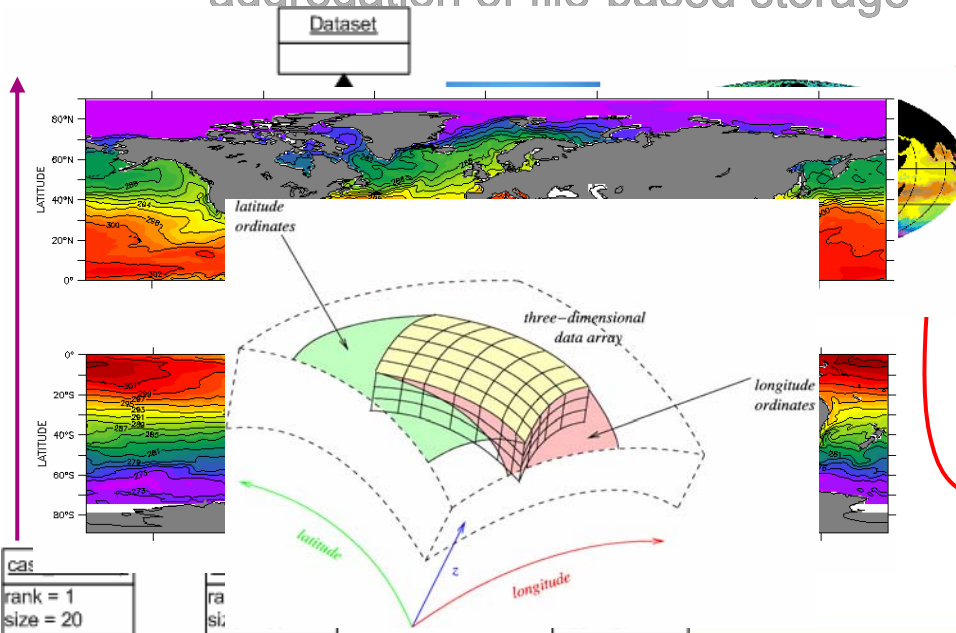
## Data Model



# NERC DataGrid Data Model

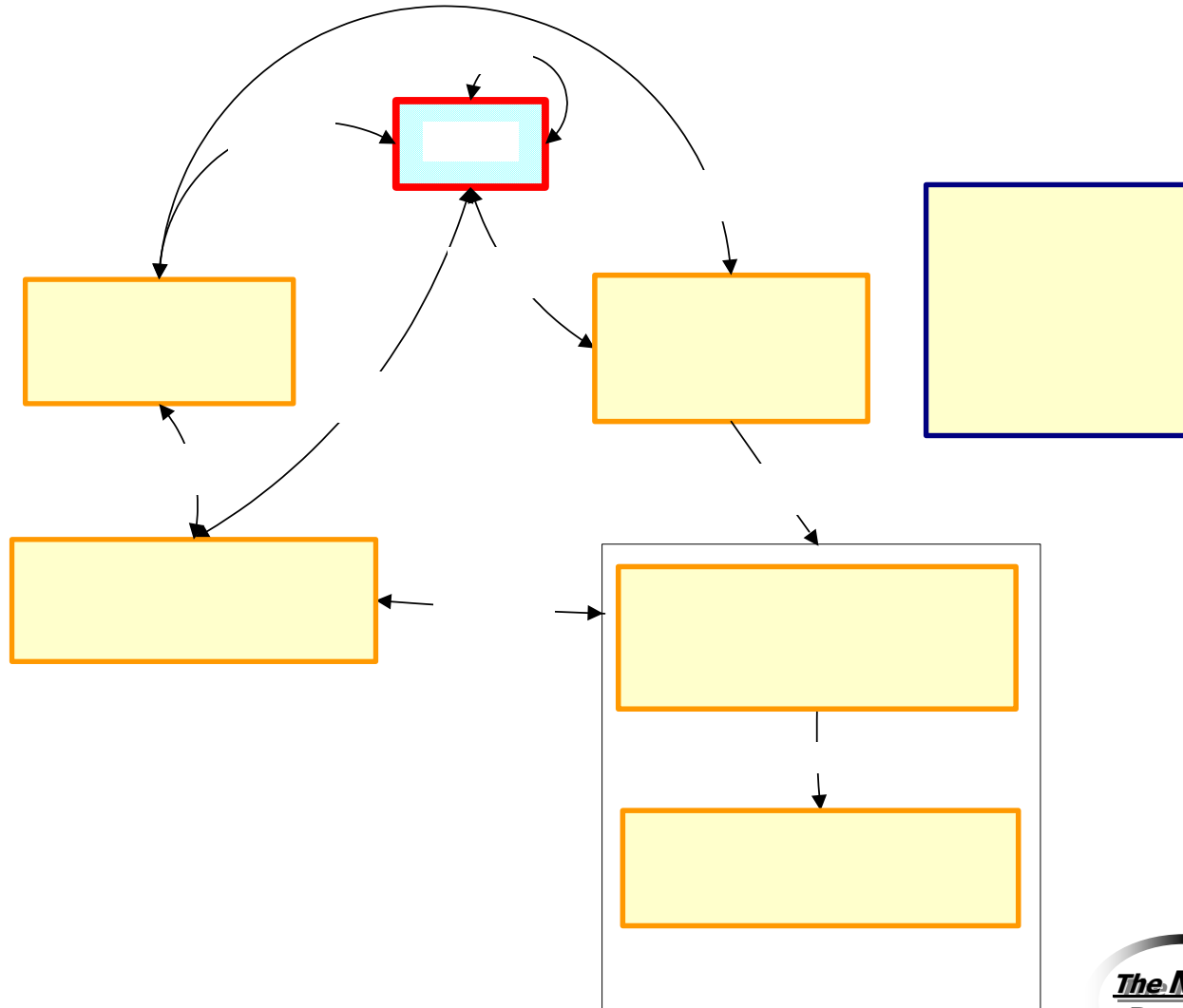
## Data Model

- nested hierarchies of multidimensional arrays
- aggregation of file-based storage

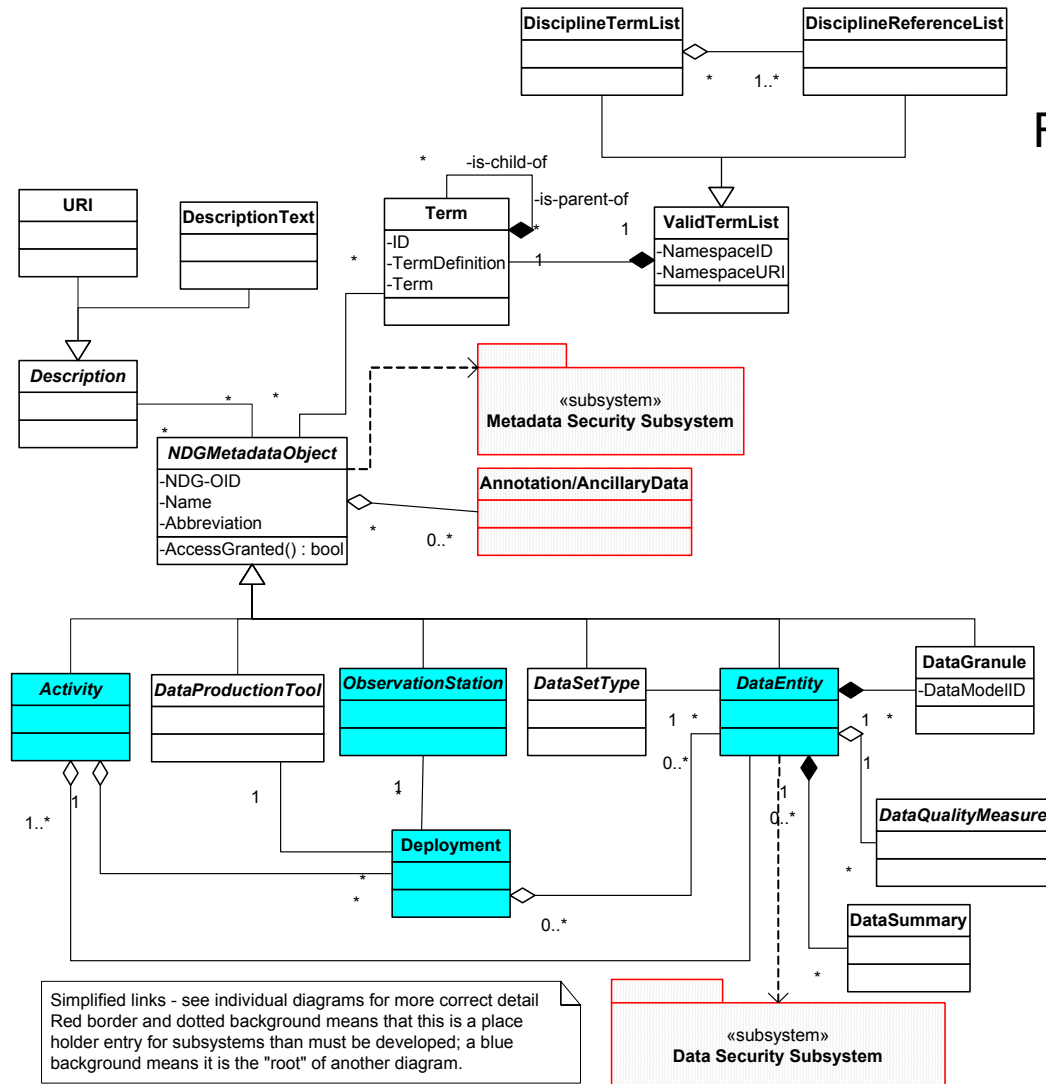




# Metadata model – conceptual view



## Top level – UML “sketch”



## Four types of metadata record:

- Activities
- Data production tools
- Observation stations
- Data entities

## Model expressed as an XML Schema

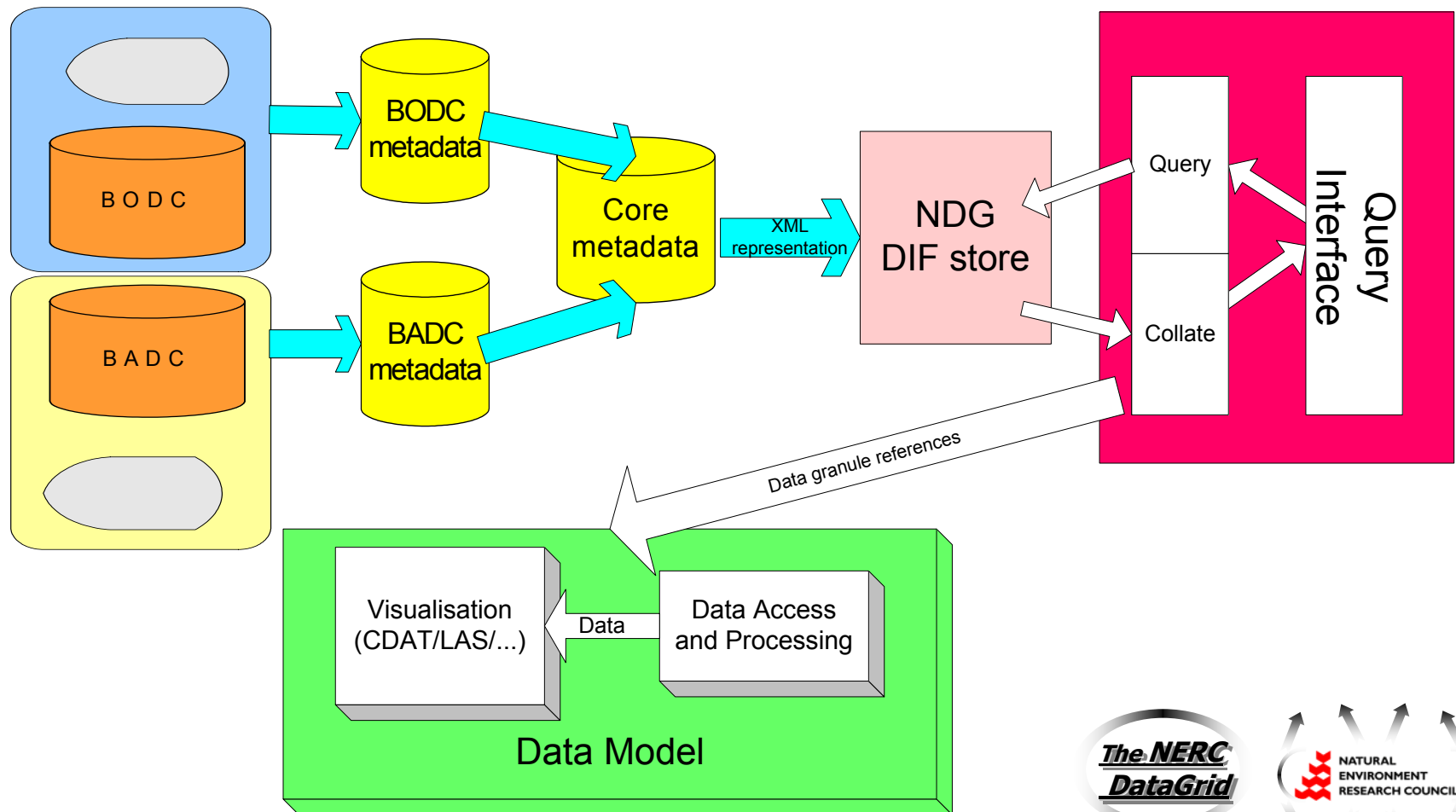
- Metadata records to be presented in XML, validated against the schema
- Wide range of data sources, XML provides a good intermediate format

## XSLT scripts being developed to map metadata into the various discovery formats

- GCMD's DIF format to be the first, then FGDC GEO profile



# End-to-end – currently...



Thank you

Questions?

# ISO TC211

- **Projects:**

- **19101 (15046-1): Geographic information - Reference model**
- 19102 (15046-2): Geographic information - Overview (Project deleted, see resolution 192 - Adelaide)
- **19103 (15046-3): Geographic information - Conceptual schema language**
- **19104 (15046-4): Geographic information - Terminology**
- **19105 (15046-5): Geographic information - Conformance and testing**
- **19106 (15046-6): Geographic information - Profiles**
- **19107 (15046-7): Geographic information - Spatial schema**
- **19108 (15046-8): Geographic information - Temporal schema**
- **19109 (15046-9): Geographic information - Rules for application schema**
- **19110 (15046-10): Geographic information - Feature cataloguing methodology**
- **19111 (15046-11): Geographic information - Spatial referencing by coordinates**
- **19112 (15046-12): Geographic information - Spatial referencing by geographic identifiers**
- **19113 (15046-13): Geographic information - Quality principles**
- **19114 (15046-14): Geographic information - Quality evaluation procedures**
- **19115 (15046-15): Geographic information - Metadata**
- **19116 (15046-16): Geographic information - Positioning Services**

Published IS

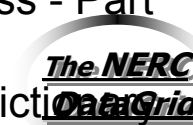
Draft

# ISO TC211

- **Projects:**

- **19117 (15046-17): Geographic information - Portrayal**
- **19118 (15046-18): Geographic information - Encoding**
- 19119 (15046-19): Geographic information - Services
- **19120 (15854): Geographic information - Functional standards**
- 19120/Amedment 1: Geographic information - Functional standards - Amendment 1
- **19121 (16569): Geographic information - Imagery and gridded data**
- 19122 (16822): Geographic information/Geomatics - Qualifications and Certification of Personnel
- 19123 (17753): Geographic information - Schema for coverage geometry and functions
- 19124 (17754): Geographic information - Imagery and gridded data components
- 19125-1: Geographic information - Simple feature access - Part 1: Common architecture
- 19125-2: Geographic information - Simple feature access - Part 2: SQL option
- 19125-3: Geographic information - Simple feature access - Part 3:COM/OLE option
- 19126: Geographic information - Profile - FACC Data Dictionary
- 19127: Geographic information - Geodetic codes and parameters

Technical Report



## ISO TC211 (last)

- **Projects:**

- 19128: Geographic information - Web Map server interface
- 19129: Geographic information - Imagery, gridded and coverage data framework
- 19130: Geographic information - Sensor and data models for imagery and gridded data
- 19131: Geographic information - Data product specifications
- 19132: Geographic information - Location based services possible standards
- 19133: Geographic information - Location based services tracking and navigation
- 19134: Geographic information - Multimodal location based services for routing and navigation
- 19135: Geographic information - Procedures for registration of geographical information items
- 19136: Geographic information - Geography Markup Language (GML)
- 19137: Geographic information - Generally used profiles of the spatial schema and of similar important other schemas
- 19138: Geographic information - Data quality measures
- **19139: Geographic information - Metadata - Implementation specifications**
- 19140: Geographic information - Amendment to the ISO 19115 Geographic information series of standards for harmonization and enhancements



## Many more areas to be considered

- Provenance, versioning: mostly seen as part of the “C” metadata
- Ontologies, taxonomies, and standard vocabularies

## Major problems at the moment are socio-cultural

- History of autonomy means that procedures and standards are diverse, and trust must be built from a low base