

# The NERC DataGrid (NDG)



Roy Lowry on behalf of the NDG, BADC  
and BODC.

Ray Cramer, Marta Gutierrez, Kerstin Kleese Van Dam, Venkatasiva  
Kondapalli, Susan Latham, Bryan Lawrence, Kevin O'Neill, Ag Stephens,  
Andrew Woolf



**British Oceanographic Data Centre**  
<http://www.bodc.ac.uk>





# Outline



- NDG Aims and Metadata Taxonomy
- Demonstration of NDG Discovery Service
- NDG Security Model
- Project Status



British Oceanographic Data Centre <http://www.bodc.ac.uk>  
British Atmospheric Data Centre <http://badc.nerc.ac.uk>



**NERC Centres for  
Atmospheric Science**  
NATURAL ENVIRONMENT RESEARCH COUNCIL



# Timelines



- 2002: e-Science arrives at NERC:
  - Legacy Systems with many files and existing access and authorisation systems that cannot easily be replaced.
  - Complex existing 'DISCOVERY' metadata systems.
  - No 'USE' metadata
  - Discovery based on Z39.50 (which never seems to work)
  - Utilisation based on file retrieval.
- 2004: NERC DataGrid ready to move forward
  - New metadata systems describe data as well as datasets.
  - OAI based harvesting supports scalable FAST data discovery.
  - New authorisation systems under development.
- 2005:
  - Moving towards utilisation based on metadata, on demand server side behaviours, grid-based back end parallelisation etc
  - NDG release rollout commences



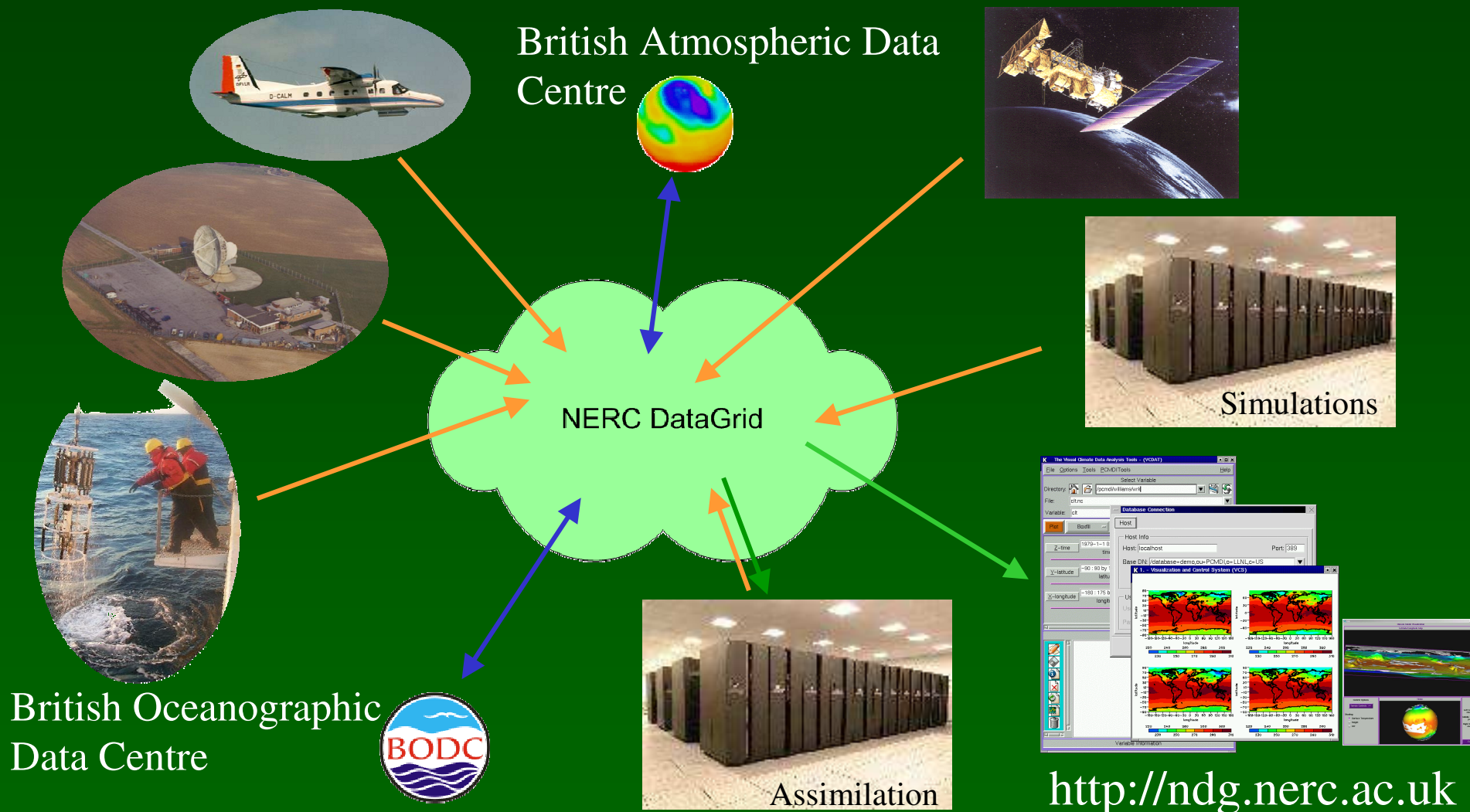
British Oceanographic Data Centre <http://www.bodc.ac.uk>  
British Atmospheric Data Centre <http://badc.nerc.ac.uk>



**NERC Centres for  
Atmospheric Science**  
NATURAL ENVIRONMENT RESEARCH COUNCIL



# Problem to be Addressed by NERC DataGrid



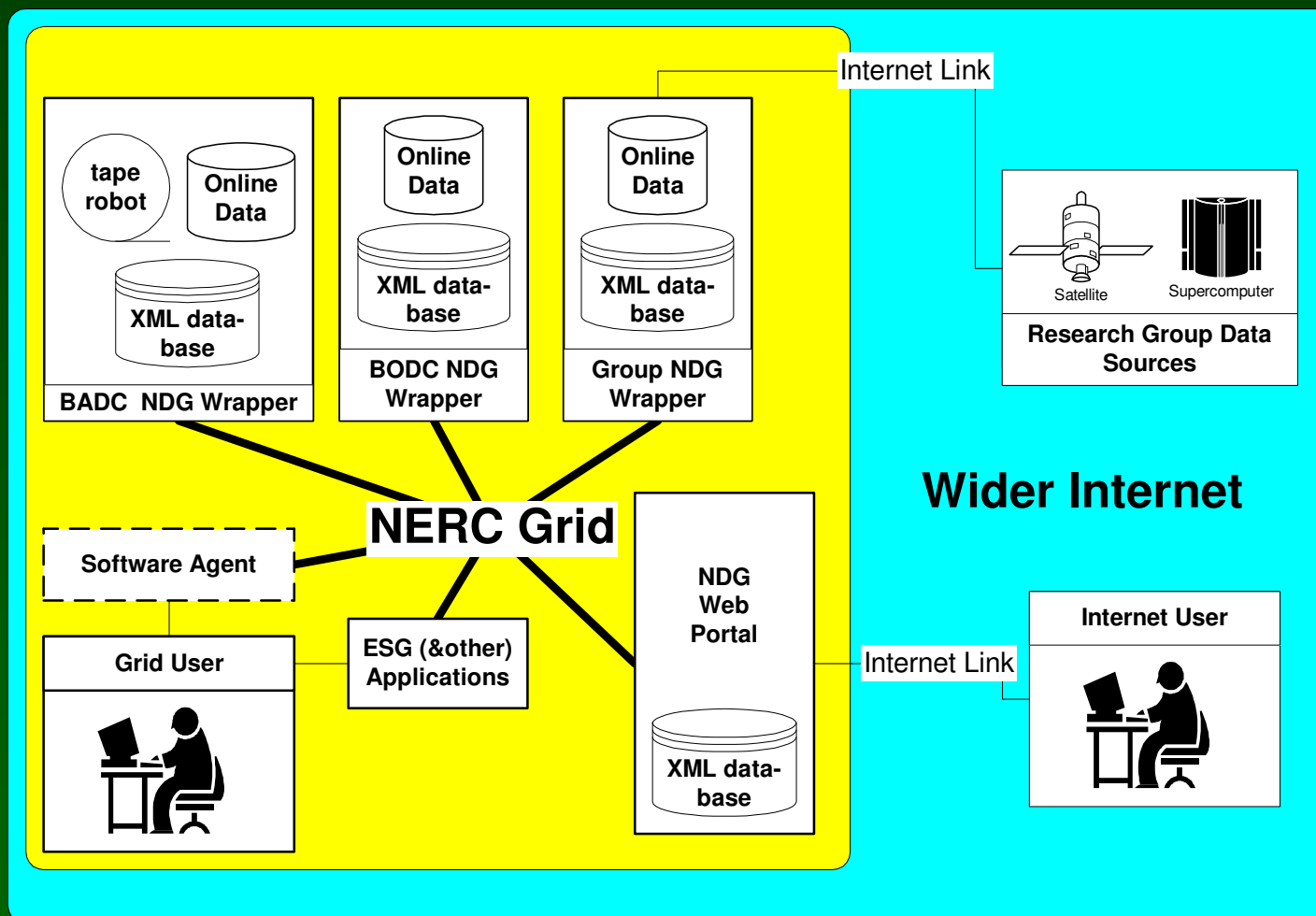
British Oceanographic Data Centre <http://www.bodc.ac.uk>  
British Atmospheric Data Centre <http://badc.nerc.ac.uk>



**NERC Centres for  
Atmospheric Science**  
NATURAL ENVIRONMENT RESEARCH COUNCIL



# NERC DataGrid Overview



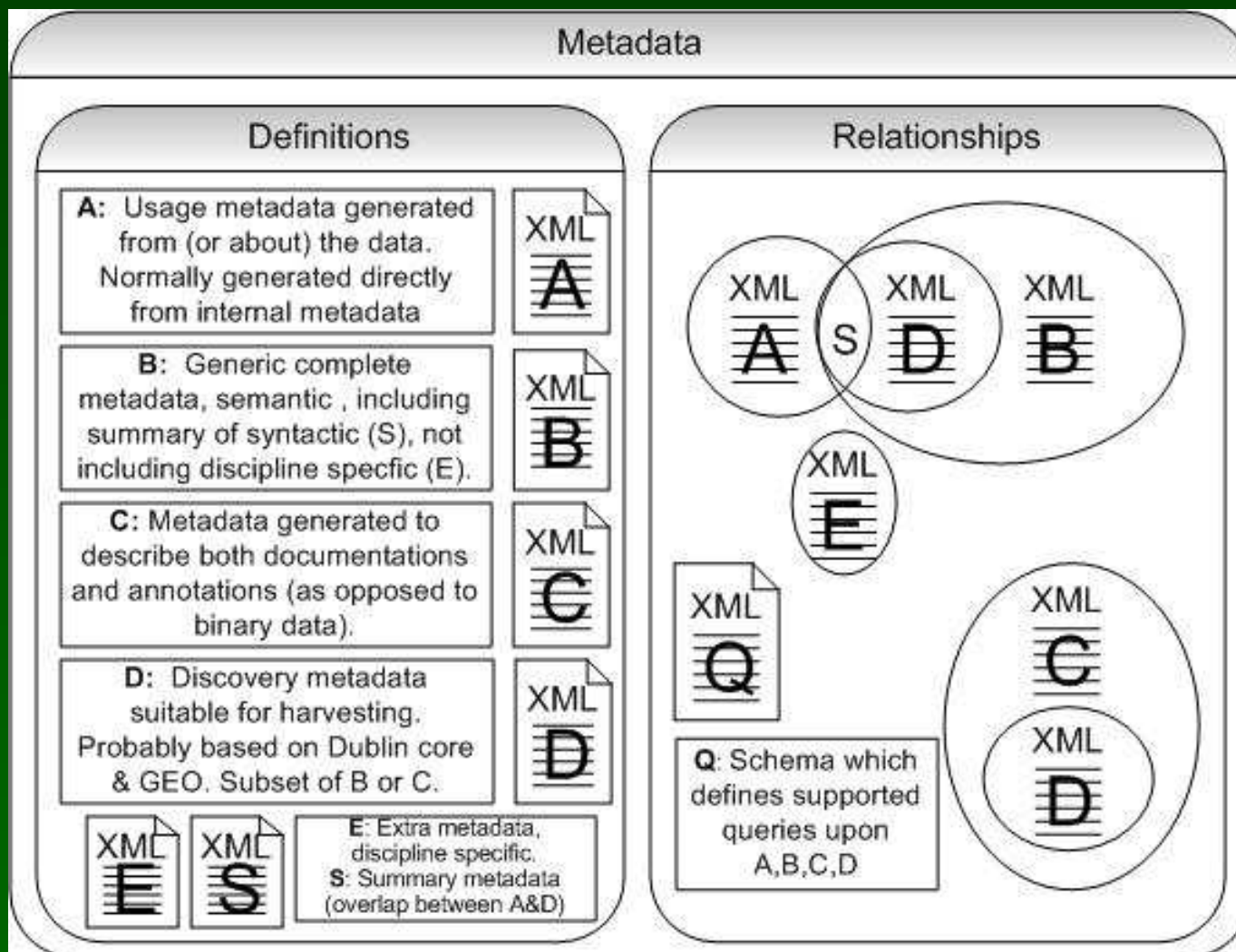
British Oceanographic Data Centre <http://www.bodc.ac.uk>  
British Atmospheric Data Centre <http://badc.nerc.ac.uk>



**NERC Centres for  
Atmospheric Science**  
NATURAL ENVIRONMENT RESEARCH COUNCIL



# NDG Metadata Taxonomy



British Oceanographic Data Centre <http://www.bodc.ac.uk>  
British Atmospheric Data Centre <http://badc.nerc.ac.uk>



NERC Centres for  
Atmospheric Science  
NATURAL ENVIRONMENT RESEARCH COUNCIL



# NDG Metadata Taxonomy



- Key Points
  - 'A' is 'Use' metadata built on GML, branded Climate System Modelling Language (CSML)
  - 'B' is a browsable network of 'Discovery' metadata, branded Metadata Objects for Links in Environmental Science (MOLES)
  - 'D' records are conventional dataset 'Discovery' records, currently GCMD DIF (but could be any suitable format such as ISO19115 profile)



British Oceanographic Data Centre <http://www.bodc.ac.uk>  
British Atmospheric Data Centre <http://badc.nerc.ac.uk>

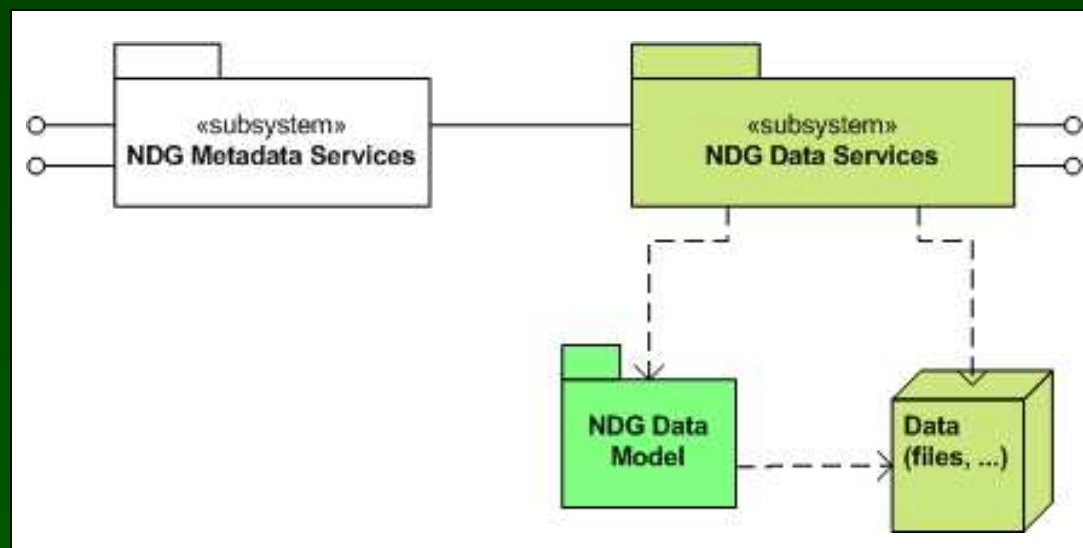


**NERC Centres for  
Atmospheric Science**  
NATURAL ENVIRONMENT RESEARCH COUNCIL





# NDG Metadata Architecture



Service based model:

- clear separation between discovery and use
- discovery service standards compliant and interoperable



British Oceanographic Data Centre <http://www.bodc.ac.uk>  
British Atmospheric Data Centre <http://badc.nerc.ac.uk>



NERC Centres for  
Atmospheric Science  
NATURAL ENVIRONMENT RESEARCH COUNCIL





# NDG Metadata Vocabularies



- Controlled vocabularies form an important part of NDG metadata
- Schemas support multiple vocabularies and can therefore include internal maps
- BODC are developing vanilla web service vocabulary support for NDG (and services will be public)
- A lot of work is required to rationalise vocabulary requirements across atmospheric and oceanographic domains



British Oceanographic Data Centre <http://www.bodc.ac.uk>  
British Atmospheric Data Centre <http://badc.nerc.ac.uk>



**NERC Centres for  
Atmospheric Science**  
NATURAL ENVIRONMENT RESEARCH COUNCIL



# NDG Discovery Service



- Data Providers each build a MOLES repository
- Discovery records (DIFs) generated by X-Query and XSLT and posted in a public OAI repository
- NDG harvests DIFs to build a central repository, which is queried by discovery web services
- Portal is one possible interface to these services, but they could equally well be used by software agents
- Data Providers with a stock of DIFs can post them without using MOLES for 'light' participation



British Oceanographic Data Centre <http://www.bodc.ac.uk>  
British Atmospheric Data Centre <http://badc.nerc.ac.uk>



**NERC Centres for  
Atmospheric Science**  
NATURAL ENVIRONMENT RESEARCH COUNCIL



# NDG Discovery Service





NDG Search Interface - Microsoft Internet Explorer provided by evesham.com

File Edit View Favorites Tools Help

Back Forward Stop Home Search Favorites Media

Address <http://glue.badc.rl.ac.uk/discovery/> Go Links



## NDG - Discovery Service

Full Text

Wildcards \*, ? Character ranges [a-zA-Z] support  Hits Displayed

[Advance Search](#)

Internet



British Oceanographic Data Centre <http://www.bodc.ac.uk>  
British Atmospheric Data Centre <http://badc.nerc.ac.uk>



NERC Centres for  
Atmospheric Science  
NATURAL ENVIRONMENT RESEARCH COUNCIL



# NDG Discovery Service





NDG Search Interface - Microsoft Internet Explorer provided by evesham.com

File Edit View Favorites Tools Help

Back Forward Stop Home Search Favorites Media

Address <http://glue.badc.rl.ac.uk/discovery/searchTextAdvance.jsp> Go Links



## NDG - Discovery Service

### Advance Search

**Text search**

**Full Text**

Wildcards \*, ? Character ranges [a-zA-Z] support

**Fields**

Search Fields

**Temporal Coverage**

		Date Input (yyyy-mm-dd)	
<b>Start Date</b>	after <input type="button" value="v"/>	<input type="text"/>	
<b>End Date</b>	before <input type="button" value="v"/>	<input type="text"/>	<input type="button" value="Search"/>

**Spatial Coverage**

	<b>North</b> <input type="text" value="90"/>		
<b>West</b> <input type="text" value="-180"/>	<input type="button" value="Select Geographic Area"/>	<input type="text" value="180"/> <b>East</b>	<input type="button" value="within v"/>
	<b>South</b> <input type="text" value="-90"/>		<input type="button" value="Search"/>

Hits Displayed

Done Internet



British Oceanographic Data Centre <http://www.bodc.ac.uk>  
British Atmospheric Data Centre <http://badc.nerc.ac.uk>



NERC Centres for  
Atmospheric Science  
NATURAL ENVIRONMENT RESEARCH COUNCIL



# NDG Discovery Service



NDG Search Interface - Microsoft Internet Explorer provided by evesham.com

Address: <http://glue.badc.rl.ac.uk/discovery/Task>

NDG - Discovery Service

Search results for query term: "humidity"

Check All <input type="checkbox"/>	Title	Temporal Coverage		Data Centre
		Start Date	End Date	
<input type="checkbox"/>	Met Office - UK and European Radiosonde data (1990 - June 1998) +	1990-01-01	1998-06-30	GB/IICAS/BADC
<input checked="" type="checkbox"/>	Atmospheric Chemistry studies in the Ocean Environment (ACSOE) +	1996-04-02	1998-12-30	GB/IICAS/BADC
<input checked="" type="checkbox"/>	VIRTEM Validation of IASI Radiative Transfer: Experiments and Modelling +	1999-01-01		GB/IICAS/BADC
<input checked="" type="checkbox"/>	Airborne Antarctic Ozone Experiment (AAOE-87) +	1987-08-12	1987-10-04	GB/IICAS/BADC
<input checked="" type="checkbox"/>	European Centre for Medium-Range Weather Forecasts (ECMWF) ERA40 Re-Analysis data ANALYSES on MODEL LEVELS. +	1958-01-01	2001-12-31	GB/IICAS/BADC
<input checked="" type="checkbox"/>	meteorology +	1996-07-06	1996-04-12	BADC
<input checked="" type="checkbox"/>	European Centre for Medium-Range Weather Forecasts (ECMWF) ERA40 Re-Analysis data ANALYSES on PRESSURE LEVELS. +	1958-01-01	2001-12-31	GB/IICAS/BADC
<input checked="" type="checkbox"/>	European Centre for Medium-Range Weather Forecasts (ECMWF) ERA40 Re-Analysis data ANALYSES on POTENTIAL TEMPERATURE (THETA) LEVELS. +	1958-01-01	2001-12-31	GB/IICAS/BADC
<input checked="" type="checkbox"/>	European Centre for Medium-Range Weather Forecasts (ECMWF) ERA40 Re-Analysis data ANALYSES on POTENTIAL VORTICITY LEVELS. +	1958-01-01	2001-12-31	GB/IICAS/BADC
<input checked="" type="checkbox"/>	Met Office - UK High Resolution Radiosonde Data +	1990-04-01		GB/IICAS/BADC

Displaying hits 11 to 20 (total: 43) Previous | Next

Browse Selected: ☐ Data ☒ Detailed Metadata

☒ HDG data service available  
☒ HDG metadata service available

[http://cdat.badc.nerc.ac.uk/cgi-bin/dxui.py?datasetURI\\_1=/badc/ecmwf-e40/metadata/e40-1.0-am-1958-2001.xml](http://cdat.badc.nerc.ac.uk/cgi-bin/dxui.py?datasetURI_1=/badc/ecmwf-e40/metadata/e40-1.0-am-1958-2001.xml)



British Oceanographic Data Centre <http://www.bodc.ac.uk>  
British Atmospheric Data Centre <http://badc.nerc.ac.uk>



NERC Centres for  
Atmospheric Science  
NATURAL ENVIRONMENT RESEARCH COUNCIL



# NDG Discovery Service



- Up to three types of service available for each dataset returned
  - Metadata browse (access to MOLES repository)
  - Data service (access to data through CSML)
  - Local service (anything the data host can deliver through a URL)
- Also possible to display DIF in HTML (human-readable) or XML (machine-readable)



British Oceanographic Data Centre <http://www.bodc.ac.uk>  
British Atmospheric Data Centre <http://badc.nerc.ac.uk>



**NERC Centres for  
Atmospheric Science**  
NATURAL ENVIRONMENT RESEARCH COUNCIL





# NDG Discovery Service



NDG Search Interface - Microsoft Internet Explorer provided by evesham.com

Address: <http://glue.badc.rl.ac.uk/discovery/Task>

NDG - Discovery Service

Search results for query term: "humidity"

Check All <input type="checkbox"/>	Title	Temporal Coverage		Data Centre
		Start Date	End Date	
<input type="checkbox"/>	Met Office - UK and European Radiosonde data (1990 - June 1998) +	1990-01-01	1998-06-30	GB/IICAS/BADC
<input checked="" type="checkbox"/>	Atmospheric Chemistry studies in the Ocean Environment (ACSOE) +	1996-04-02	1998-12-30	GB/IICAS/BADC
<input checked="" type="checkbox"/>	VIRTEM Validation of IASI Radiative Transfer: Experiments and Modelling +	1999-01-01		GB/IICAS/BADC
<input checked="" type="checkbox"/>	Airborne Antarctic Ozone Experiment (AAOE-87) +	1987-08-12	1987-10-04	GB/IICAS/BADC
<input checked="" type="checkbox"/>	European Centre for Medium-Range Weather Forecasts (ECMWF) ERA40 Re-Analysis data ANALYSES on MODEL LEVELS. +	1958-01-01	2001-12-31	GB/IICAS/BADC
<input checked="" type="checkbox"/>	meteorology +	1996-07-06	1996-04-12	BADC
<input checked="" type="checkbox"/>	European Centre for Medium-Range Weather Forecasts (ECMWF) ERA40 Re-Analysis data ANALYSES on PRESSURE LEVELS. +	1958-01-01	2001-12-31	GB/IICAS/BADC
<input checked="" type="checkbox"/>	European Centre for Medium-Range Weather Forecasts (ECMWF) ERA40 Re-Analysis data ANALYSES on POTENTIAL TEMPERATURE (THETA) LEVELS. +	1958-01-01	2001-12-31	GB/IICAS/BADC
<input checked="" type="checkbox"/>	European Centre for Medium-Range Weather Forecasts (ECMWF) ERA40 Re-Analysis data ANALYSES on POTENTIAL VORTICITY LEVELS. +	1958-01-01	2001-12-31	GB/IICAS/BADC
<input checked="" type="checkbox"/>	Met Office - UK High Resolution Radiosonde Data +	1990-04-01		GB/IICAS/BADC

Displaying hits 11 to 20 (total: 43) Previous | Next

Browse Selected: ☐ Data ☒ Detailed Metadata

☒ HDG data service available  
☒ HDG metadata service available

[http://cdat.badc.nerc.ac.uk/cgi-bin/dxui.py?datasetURI\\_1=/badc/ecmwf-e40/metadata/e40-1.0-am-1958-2001.xml](http://cdat.badc.nerc.ac.uk/cgi-bin/dxui.py?datasetURI_1=/badc/ecmwf-e40/metadata/e40-1.0-am-1958-2001.xml)



British Oceanographic Data Centre <http://www.bodc.ac.uk>  
British Atmospheric Data Centre <http://badc.nerc.ac.uk>



NERC Centres for  
Atmospheric Science  
NATURAL ENVIRONMENT RESEARCH COUNCIL





# NDG Data Service



Location Edit View Go Bookmarks Tools Settings Window Help

Location: [gi-bin/ndgDataAccess?datasetSource=dmgdev1.esc.rl.ac.uk&datasetID=acsoe\\_badc\\_eae96\\_macehead\\_mt1\\_1](http://gi-bin/ndgDataAccess?datasetSource=dmgdev1.esc.rl.ac.uk&datasetID=acsoe_badc_eae96_macehead_mt1_1)

BADC Testbeds Google BNL Dictionary Technology Places Personal Science Alexander the Grea... H >>

**NATURAL ENVIRONMENT RESEARCH COUNCIL**

**NDG - Data Service**

Select variables

☐ Wind speed  
(minimum in period) m/s

☐ Wind speed (mean  
value in period) m/s

☐ Wind speed  
(maximum in period)  
m/s

☐ Wind direction  
(degrees magnetic)

☐ Dry bulb  
temperature (degrees  
Celsius)

☐ Relative humidity  
(percent)

☐ Atmospheric  
pressure (millibars)

☐ Solar radiation  
(Watts)

☐ Rainfall (mm in  
period)

☐ Error flag

Get netCDF file

Get tar file



British Oceanographic Data Centre <http://www.bodc.ac.uk>  
British Atmospheric Data Centre <http://badc.nerc.ac.uk>



**NERC Centres for  
Atmospheric Science**  
NATURAL ENVIRONMENT RESEARCH COUNCIL





# NDG Browse Service



Location Edit View Go Bookmarks Tools Settings Window Help

Location: <http://glue.badc.rl.ac.uk/exist/Task>

BADC Testbeds Google BNL Dictionary Technology Places Personal Science Alexander the Grea... H >>

 **NDG - Detailed Metadata Browse** 

Displaying Detailed Metadata Records

```
</dgValidTermID>
<ListLevel>1</ListLevel>
</dgStdParameterMeasured>
<ParameterName>Wind speed (minimum in period)</ParameterName>
<ParameterAbbreviation>wind speed min</ParameterAbbreviation>
</dgParameterSummary>
<dgParameterSummary>
  <dgStandardUnit>
    <dgValidTerm>m/s</dgValidTerm>
    <dgValidTermID>
      <ParentListID>http://www.unidata.ucar.edu/content/software/uc
      <TermID>1</TermID>
    </dgValidTermID>
  </dgStandardUnit>
  <IsOutput>true</IsOutput>
  <dgRangeDataParameter>
    <HighValue>14.4</HighValue>
    <LowValue>0.4</LowValue>
  </dgRangeDataParameter>
  <dgStdParameterMeasured>
    <dgValidTerm>EARTH SCIENCE</dgValidTerm>
    <dgValidTermID>
      <ParentListID>http://gcmd.gsfc.nasa.gov/Resources/valids/gcmd
      <TermID>1</TermID>
    </dgValidTermID>
  </dgStdParameterMeasured>
</dgParameterSummary>
</dgParameterSummary>
</dgValidTermID>
```



British Oceanographic Data Centre <http://www.bodc.ac.uk>  
British Atmospheric Data Centre <http://badc.nerc.ac.uk>



NERC Centres for  
Atmospheric Science  
NATURAL ENVIRONMENT RESEARCH COUNCIL



# Example Local Service



Location Edit View Go Bookmarks Tools Settings Window Help

Location: <http://www.soc.soton.ac.uk/cgi-bin/seadog/seadog.pl>

BADC Testbeds Google BNL Dictionary Technology Places Personal Science

## SeaDog search

Welcome to the SeaDOG Database. This is a Deep Ocean Geophysics metadata database, and holds information on the data available within the archive. Requests for data are made via e-mail and data sent to users via CD-ROM. Use the search criteria pull-down menus and text boxes below to start the choice of data and then hit 'Submit Search'.

Category of Search	Choose one or more search criteria	
Cruise Name	<input type="text"/>	Leave blank to search all
General Data Category	<input type="text"/>	Leave blank to search all
Detailed Data Category	<input type="text"/>	Leave blank to search all
Geological Keyword	<input type="text"/>	Case Insensitive and partial words allowed
Area Search	Values	In degrees
Minimum Latitude	<input type="text"/>	Range -90 to 90
Maximum Latitude	<input type="text"/>	Range -90 to 90
Minimum Longitude	<input type="text"/>	Range -180 to 180
Maximum Longitude	<input type="text"/>	Range -180 to 180

(6.8 KB/s) 25%



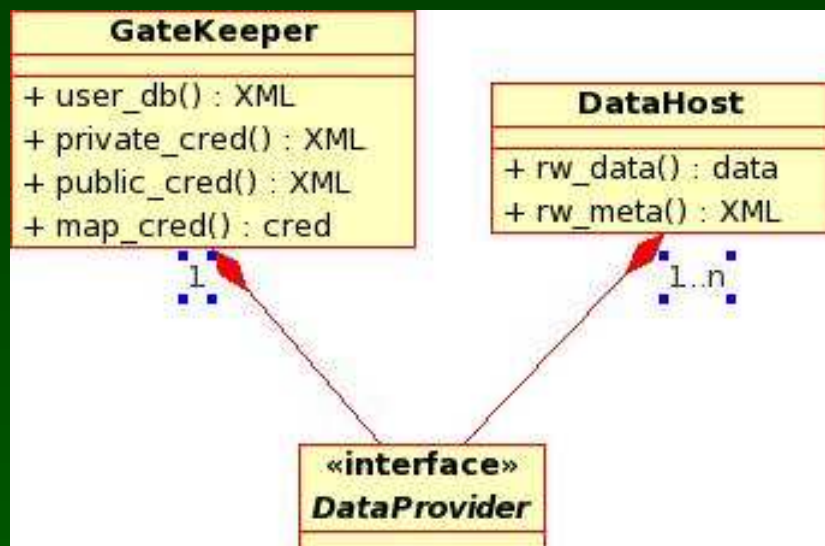
British Oceanographic Data Centre <http://www.bodc.ac.uk>  
British Atmospheric Data Centre <http://badc.nerc.ac.uk>



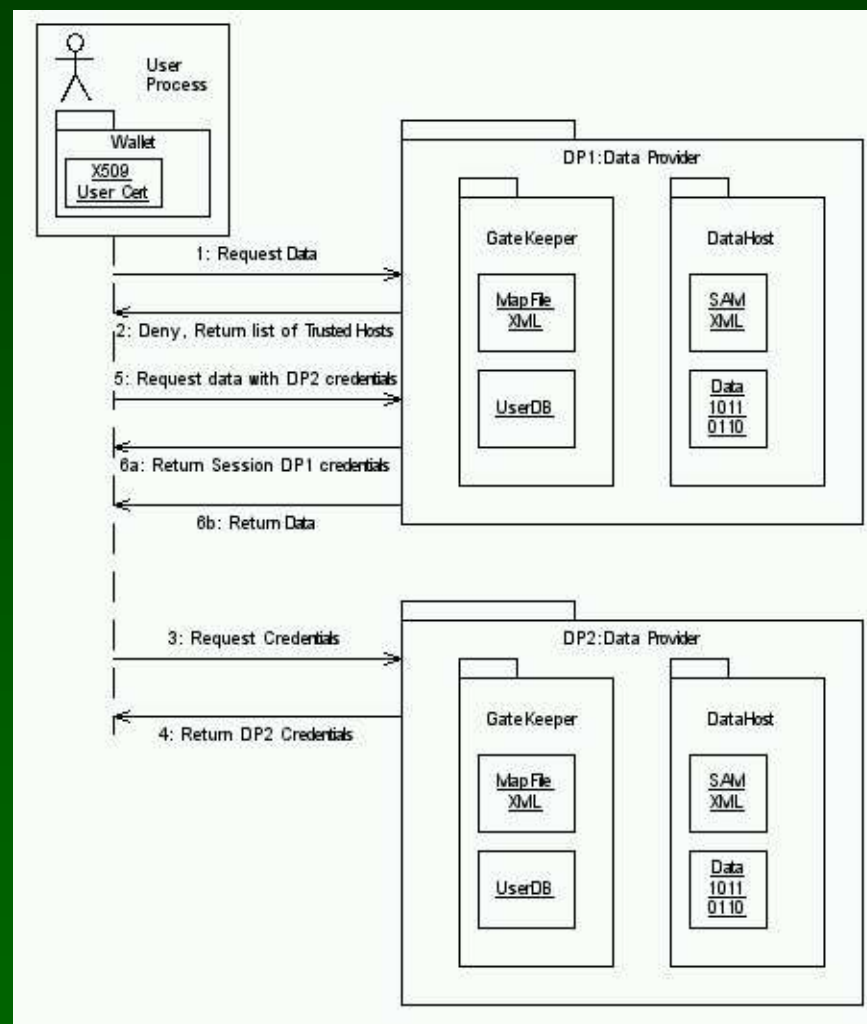
NERC Centres for  
Atmospheric Science  
NATURAL ENVIRONMENT RESEARCH COUNCIL



# NDG Security



Certificate based, pass encrypted credentials between user and gatekeeper.



British Oceanographic Data Centre <http://www.bodc.ac.uk>  
British Atmospheric Data Centre <http://badc.nerc.ac.uk>



**NERC Centres for  
Atmospheric Science**  
NATURAL ENVIRONMENT RESEARCH COUNCIL



# Authorisation



- Role-based access:

```
<dataset>
<host> badc.nerc.ac.uk </host>
<name>ukmo-obs </name>
<access-requires> researcher <access-requires>
<access-requires> ukmo-obs </access-requires>
<processing-requires> nerc </processing-requires>
</dataset>
```

Signed “conditions of use”  
form exists for this dataset

- Key concept: Only hosts that trust each other share data, even within a larger virtual organisation: e.g. at BADC:

```
<trusted>
<bodc>
<host>ndg.bodc.nerc.ac.uk</host>
<attribute remotename="nerc"> nerc </attribute>
<attribute remotename="ashoe"> ashoe </attribute>
<attribute remotename="staff"> nerc </attribute>
<other> bodc </other>
</bodc>
</trusted>
```



British Oceanographic Data Centre <http://www.bodc.ac.uk>  
British Atmospheric Data Centre <http://badc.nerc.ac.uk>



NERC Centres for  
Atmospheric Science  
NATURAL ENVIRONMENT RESEARCH COUNCIL



# Current Work



- Further service design underway but implementation details not yet obvious (e.g. GT4 etc).
- Deployment of CSML to describe observational and model data
- Building security infrastructure
- Ongoing MOLES development and population for:
  - Oceanographic data
  - Atmospheric Chemistry data
  - Numerical Modelling data
  - Remote Sensing Data



British Oceanographic Data Centre <http://www.bodc.ac.uk>  
British Atmospheric Data Centre <http://badc.nerc.ac.uk>



**NERC Centres for  
Atmospheric Science**  
NATURAL ENVIRONMENT RESEARCH COUNCIL



# Where are we?



- Release 0.1 (1<sup>st</sup> March 2005)
  - Discovery Service
  - Data extractor service
  - NASA Ames Python API
  - Data Provider tools and documentation
    - CSML documentation and schema
    - MOLES guide and schema
    - MOLES to DIF XQueries and XSLT
    - OAI guide
- Release 0.2 (23<sup>rd</sup> May 2005)
  - Support for DIF 9 (ISO-compliant)
  - Support for MOLES 1.02
  - Improved tooling (automation for operational use)
  - Improved documentation including new
    - Data Providers' Guide
    - Using Discovery Service as a web service
    - Guide on utilising eXist XML database
  - Improved and extended Discovery Service content



British Oceanographic Data Centre <http://www.bodc.ac.uk>  
British Atmospheric Data Centre <http://badc.nerc.ac.uk>



**NERC Centres for  
Atmospheric Science**  
NATURAL ENVIRONMENT RESEARCH COUNCIL





# Where are we?



- Release 0.3 (6<sup>th</sup> June 2005)
  - Security added
  - Support for DIF 9 extensions for model data
  - Dublin Core supported as additional discovery format
- Release 0.4 (July 2005)
  - Document handling library in python and java
  - Data-provider database package
  - WMS/WCS interface to NetCDF
  - CSML-based data delivery
  - Vocabulary server
  - Secure access guide
- Release 1.0 (Due September 2005)
  - Pre-operational release
- Second (operational) phase of project funded from October 2005 to September 2007



British Oceanographic Data Centre <http://www.bodc.ac.uk>  
British Atmospheric Data Centre <http://badc.nerc.ac.uk>



**NERC Centres for  
Atmospheric Science**  
NATURAL ENVIRONMENT RESEARCH COUNCIL