



The NERC Data Grid



The NERC Data Grid (NDG) provides the infrastructure which allows users to:

- Find data
- Explore what is known about datasets (including information about the observing or simulating tools)
- Access, manipulate and visualise data!

Like the web, the NDG has no “owner” or “central control”; data remain with data providers – be they managed data centres in the UK or abroad, or semi-managed data archives in large research groups. The location of the data can be transparent to the user, while still allowing data providers to maintain their intellectual investment by controlling access.

Although there is no “centre” to the NDG, the NDG team do provide two community services which all NDG participants may avail themselves of:

- **The “NDG Discovery Service”**: A database of discovery information, harvested from data providers around the world, together with both a web-site, which provides a portal to that discovery data, and a set of web-services, which allow NDG consumers to exploit that database without using the NDG web-site.
- **The “NDG Vocabulary Service”**: Consisting of databases of environmental thesauri and ontology tools to map between terms. The vocabulary service supports “machine assisted” browsing: which, for example, allows a search for **rainfall** to return datasets with **precipitation**.

NDG: the Future!

Currently focused on atmospheric and oceanographic data, but designed for extension to the full gamut of environmental sciences. Currently a research project, but elements of the NDG should receive ongoing, long-term NERC support as strategic infrastructure.

Competitors! How is NDG different from the rest?

e2edm, End-to-End Data Management: the NDG is based around mature OGC descriptions of environmental data, e2edm is based around a number of emerging applications. NDG and e2edm have similar aims but are using different technologies. Arguably NDG is more standards compliant, and will provide support for a wider range of data sources.

Thredds: provides catalogue services, primarily limited to file-based access. Within what it does, Thredds is significantly more mature than NDG, and is very suitable for data providers with limited metadata and/or data.

OPeNDAP: a protocol for accessing data, which has limited support for access control and requires the data user to know exactly what the data are before they can be used.

What Can't the NDG Do?

NDG can't generate the underlying metadata!

Automated manipulation and searching relies on comprehensive metadata, which require quality data management to create and maintain. However, NDG does provide a clear framework for metadata requirements.

NDG doesn't provide information services:- it provides data services! The distinction is important. However, the NDG does provide a framework within which it is possible to link data to the information services:- research publications, reports, derived information, policy advice ...

What Can NERC Data Grid Do For You?

As a Data User?

- **Discovery leads directly to data:** A simple desktop search – using a simple search page or “google toolbar” like interface – gives direct access to data, not just the metadata. Data can even be recovered directly into applications.
- **Comprehensive coverage:** The NERC Data Grid reports datasets from a wide range of sources from the UK, Europe and further afield. Even commercial datasets can be visible – if not directly accessed.
- **Seamless access:** Using NDG only requires a single-sign-on; no need to register with each data centre. NDG provides data access across multiple sources and even allows combination of data from multiple sources. This is achieved without compromising data security and protection of intellectual investment in data production.
- **Hides data format differences and provides generic visualisation tools:** No need to develop new reading routines or learn to use new tools for each new data set. Exploit the python I/O library to NDG data and integrate into your own software!
- **Enhances interdisciplinary research:** Investigation of metadata can allow discovery of previously unrecognised linkages.

As a Data Provider?

- **Provides a standards based metadata hierarchy:** Exploit the NDG metadata structures to either design a metadata system from the ground up, or use them as an interoperability tool. In the latter case, once you have a mapping from your information model to the NDG model, you can interoperate with NDG metadata browsing and discovery tools (and with other existing and emerging national and international regimes, such as INSPIRE). No need to run multiple security systems: You can leverage off an existing access control system, or develop a new one, and plug into the NDG single-sign on authentication and authorisation framework. Logging comes for free!
- **Increases visibility and usage of datasets:** Inclusion in NDG does not prevent inclusion in other data discovery or access initiatives. Even if your data centre or website is down, your data are discoverable through NDG discovery services.
- **Minimise user support load:** More users can find their data, and use them, without the need for direct support from data provider staff.

As a Developer?

- **Exploit NDG web services:** The architecture of the NDG is based on independent interacting web services. Build your own, extend the existing services, and deploy them in different ways!
- **Exploit NDG modular development:** The underlying NDG data manipulation software is written in python modules. Use `easy_install` to exploit the NDG software and build or port your own data manipulation tools. Add your favourite graphics package ...
- **Exploit NDG metadata standards:** Develop new services confident that they will be deployable and interoperable across a wide range of data types and data providers.
- **NDG developments are non-proprietary:** NDG does not require compromising on interacting with other projects. NDG exploits OGC interfaces to give more ways of interacting with data.
- **NDG standards compliance implies less future development:** You will be able to leverage off NDG compliance to OGC and ISO protocols to deploy more tools.
- **Exploit the NDG identifier conventions:** Lodge your copies of your data with a NERC designated data centre, and support persistent long-lived data citation.