# UKEOF

**ENVIRONMENTAL OBSERVATION FRAMEWORK**

<div align="center">

**UKEOF catalogue supplier workshop: Database alignment**
**Workshop report**
12th June 2013, 10.30-15.00
Austin Court, 80 Cambridge St, Birmingham, B12NP

</div>

## 1. Background – engaging catalogue suppliers

The UKEOF catalogue upgrade is driven by the need to provide up to date information in an easily accessible database which informs users of what observations are being collected where, why and how much it costs. In additional to the existing metadata, this means that UKEOF also needs to hold metadata about geo-location so that the catalogue can display 'where' observations are being taken. The upgrade therefore requires suppliers to provide additional and updated information.

The UKEOF upgrade coincides with the establishment of the EU INSPIRE Annex III Environmental Monitoring Facilities (EF) legislation which seeks to standardise environmental metadata across Europe. This means that all UK public sector bodies will be required to update their monitoring facilities metadata to this standard. The UKEOF catalogue will align with this standard and therefore reduce the burden of providing metadata in different formats to different systems.

## 2. Workshop aim

The aim of the workshop was to establish how organisations will comply with INSPIRE EF and the best way of working together to ensure metadata can be shared as easily as possible. One of the key benefits of the UKEOF catalogue is that it allows the whole environmental observation landscape to be viewed in one place. UKEOF therefore want to ensure that suppliers are able to provide updated information with as little effort as possible to ensure that users review the most up to date information possible through the catalogue.

## 3. Workshop presentations

### 3.a INSPIRE Annex III EF theme

The CEH project team, leading on the catalogue upgrade for UKEOF, have created a schema for the UKEOF catalogue based on that of the INSPIRE Annex III EF theme. A detailed presentation of the INSPIRE EF theme was provided[1].

***Main messages:***

- **Terminology** – the INSPIRE terminology differs from that commonly used in the UKEOF and other communities. This will initially be a barrier that organisations will have to address. See Annex 1 for definitions.

- **Hierarchy** – the hierarchy of how programmes, activities, facilities and networks fit together was also raised as a potential issue that may require UKEOF and others to amend their current ways of working and descriptions within databases.

- **Genealogy** – another powerful tool which will allow organisation to trace if facilities have been replaced over time or if other observations are using the same sensors, for example.

---

[1] Slides can be provided on request by emailing office@ukeof.org.uk or for UKEOF members on
https://lwec.centraldesktop.com/uk-eofcatalogue/file/24582846/

- **Observing capability** – INSPIRE EF will also allow users to explore the full capability of what a programme is and could measure using descriptions of the instruments.

- Linkages between the programmes/activities/activities/networks are complex but this will eventually allow users to drill down to the data itself.

- UKEOF will not be asking for organisations to provide **genealogy and capability** information as the Management Group has agreed this would be too burdensome. UKEOF will build these fields into the schema so there will always be potential to populate this information if organisation have it and want to use it in the future.

### 3.b Developing the new catalogue specification

The new database will be a built around storing xml document to give a flexible approach. This provides good version control and a record of updates made to the data. The new catalogue will be populated with the data from the existing database and modified as organisations prepare and share their INSPIRE compliant data. Additional information such as geographic location will need to be added to existing records. This is potentially difficult for suppliers to provide.

Large organisations will be able to supply xml format data for UKEOF to bulk upload. Suppliers with fewer records will be able to add and modify records through a web form.

Once added, the data will be stored in a new schema created by CEH as the INSPIRE schema is complex. Having control of this schema allows us to update it in the future if requirements change.

### 3.c Making efficient use of data and resources

CEH have investigated how data from one organisation's database could be transformed for use in the new catalogue. Examples using UKDMOS and CEH data were presented, demonstrating the use of three transformation tools. The open source tool from HALE (http://community.esdi-humboldt.eu/projects/hale/wiki) was of particular interest as it would be free of charge to organisations. Participants were enthused and reassured by the ease of transformation.

CEH have also developed an application programing interface (API) for the catalogue which will allow users to develop their own programs to use with the UKEOF catalogue.

### 4. How do organisations currently store their information?

Attendees were asked to provide a short summary of how they currently hold their data and how they intend to comply with INSPIRE Annex III. Organisations are at different stages of compliance, mainly depending on whether they have had to already undertake work to comply with Annex I or II.

### 5. Discussion session Q&A

The following questions were discussed during the workshop:

### When will the new catalogue be available?

A "beta" version of the catalogue will be completed by November 2013. There will then be a period of testing to iron out any problems and bugs ready for a launch in April 2014.

**How will I get my records into the catalogue?**

The records that are in the existing catalogue will be moved to the new catalogue when it is available. The new catalogue will also have a web interface which will enable you to add new records (and edit existing ones). There will also be a facility to upload new records.

Currently, to update and add new records, data providers supply information to UKEOF. The data are then reformatted for the catalogue and added by the UKEOF team. With the new catalogue data providers will be able to add/edit records directly.

**What types of information will the catalogue manage?**

There are four types of information that the catalogue will handle:

**Programmes -** A programme is a related group of environmental monitoring campaigns which is policy-led and which is normally run over a long period (years or decades).

**Activities -** An activity is an environmental monitoring campaign that is carried out with specific equipment for a specific period of time.

**Facilities -** A facility is a specific area or object which directly collects or processes data about environmental properties (physical, chemical, biological). Facilities can be considered as:

- sites – e.g. A nature reserve
- stations – e.g. a weather station inside the nature reserve
- sensors – e.g. instruments on the station which are generating data

**Networks -** Networks are groups of *facilities* that are managed for a specific purpose, targeting a specific area. They can be at an organisational, regional, national or international level.

*These definitions are best illustrated with an example:*

*Programme* - National weather monitoring programme

    *Activity* – National rainfall monitoring

        *Facility* – Site number  1, Penrith

        *Facility* – Rainfall gauge no.A7RB4A

    *Activity* – National windspeed monitoring

        *Facility* – Site number 1, Penrith

        *Facility* – Wind vane no. 283847

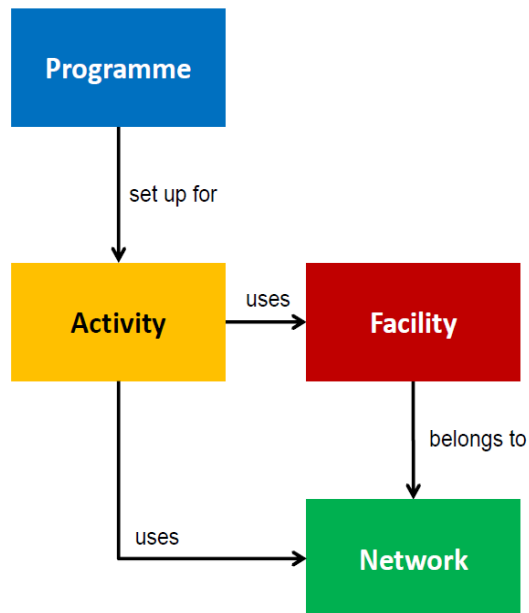        *Facility* – Anemometer no.2011/06/AQZXG

    *Network* – Cumbrian weather sites

        *Facility* – Site number 1, Penrith

*Facility* – Site number  2, Brampton

*Facility* – Site number  3, Drumburgh

## How are these information types related?



The relationships between the four data types can get complex but to try to simplify it:

- A *programme* normally consists of multiple *activities*
  However, an *activity* can also contribute to multiple *programmes*
- A *network* is a collection of *facilities*
  However, a *facility* can belong to more than one *network*
- An *activity* uses one or more *facilities* (or *networks*) to collect data
  But, a *facility* (or *network*) can generate data for more than one *activity*

NB - See Annex 1 for definitions

## Can I have a programme without an activity?

Yes.  Environmental monitoring programmes can provide a very generic access point to environmental monitoring information with a broad view.

## What level of detail will we be expected to provide?

The existing data in the catalogue concerns *programmes* and *activities* – they describe what is monitored, why and by whom.  However, a significant drawback currently is the lack of data on <u>where</u> the monitoring is taking place.  We will be asking data providers for this additional information during the upgrade process.

The existing catalogue does not hold information on related *facilities*.  The new catalogue will be able to manage such information but it will not be mandatory – if the data providers want to use that functionality they can do so.

## What format will I be asked to provide the data in?

UKEOF will be shortly publishing a schema document which will define the structure and format for new data. There are a number of tools that can help transform data from the format organisations currently hold it in to the UKEOF format. For example, <u>FME</u>, <u>Altova MapForce</u> and <u>Humbolt Alignment Editor</u>.

**How does the UKEOF schema differ from the INSPIRE EF schema?**

The UKEOF schema is a simplified translation of the INSPIRE schema which removes some of the complexities of the INSPIRE data structure. Translating the UKEOF schema into the INSPIRE schema will make tit clearer and easier to use.

The only additional information the UKEOF schema requires surrounds financial investments. This enables UKEOF to provide an estimate of the total investment made in observing the environment by the UK.

**Who will be able to use the catalogue?**

The catalogue will be on a public website and so details of the activities, programmes, etc. will be publicly available.  Members of the UKEOF community will be provided with a login which will provide them with additional functionality such as the ability to add/edit records.

**Will the public be able to see ALL the information in the catalogue?**

No. Certain information will be kept private. For example, financial data about the costs of maintaining/running a programme will be visible only to the "owner" organisation and the UKEOF management group.

Our current thinking is that there will be a publication "workflow" which will allow data providers to decide whether the information they are adding to the catalogue is:

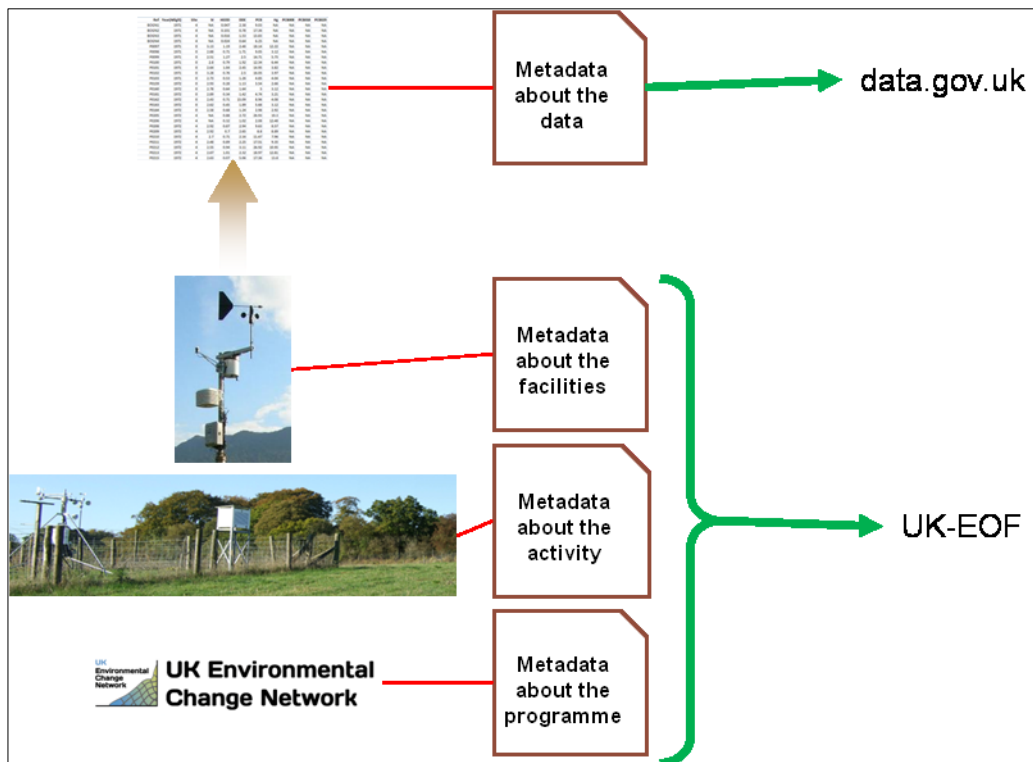| | |
|---|---|
| **Private** | visible only to themselves and the UKEOF management group |
| **Sensitive** | visible to everyone in the UKEOF community (i.e. to logged-in users) |
| **Public** | visible to the public |

**What if the location of my monitoring site/equipment is sensitive?**

We are currently considering the best option for ensuring that sensitive information is not made public.  For example:

- Omitting sensitive information from the published record altogether
- Generalising location information (e.g. "Cumbria" rather than a map reference)
- Publishing location information but setting those fields to a restricted audience (e.g. data provider and UKEOF only)

**How does the UKEOF catalogue fit in with data.gov.uk?**

Currently, data.gov.uk only maintains information on datasets.  The UKEOF catalogue is concerned with the monitoring activities and facilities that produce that data.

However, there will be overlap between metadata in the UKEOF catalogue and data.gov.uk. We are working with UKLP to define these relationships and to ensure that data providers are not duplicating effort.

## What about information I already provide to UK DMOS?

We will continue to harvest information from UKDMOS on behalf of the marine community. UKEOF will therefore only need to approach the marine community to ask for information about investments in marine observations.

## 6.    Workshop achievements and conclusion

- An in-depth description of the INSPIRE EF theme schema was successfully presented by the project team, detailing some of the difficulties and solutions that have emerged through the UKEOF catalogue work. The project team have 'translated' the INSPIRE schema and presented this in a manageable manner that could be more easily understood by the wider community.
- Important information was made available to participants regarding transformation of metadata. This process has been made relatively easy by the availability of transforming tools which will make it easier for organisations to provide their data in the UKEOF schema format. The use of Open Source software was particularly appealing.
- UKEOF will be able to assist many of its partner organisations as they develop their own internal databases to ensure they are INSPIRE and UKEOF compatible.
- Additional information such as geographic location will need to be added to existing records – this is potentially the most resource intensive process in terms of updating the existing metadata records.
- Although many organisations are already working on compliance, some are further on in development than others and this may mean that updating information in the UKEOF catalogue will be delayed. In the meantime, it was clarified that UKEOF will be able to

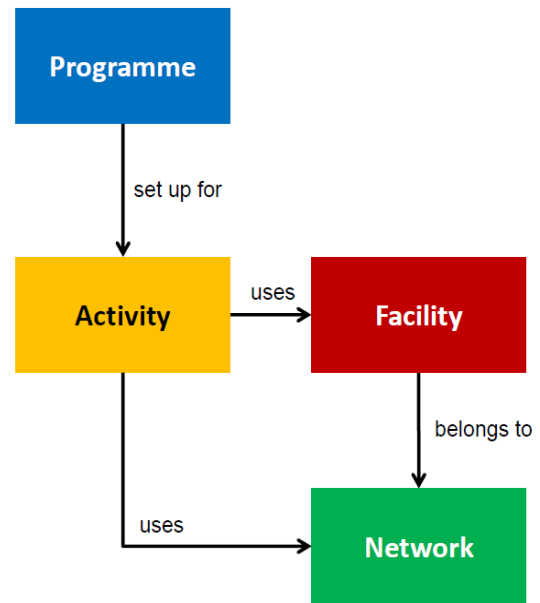use existing records as a placeholder for updated records to help organisations deliver their updated information.

## Annex 1 – INSPIRE terminology

**Programme** - Framework based on policy-relevant documents defining the target of collections.

**Activity** - Set of monitoring features used in a coherent and concise timeframe, area and purpose. Or a campaign carried out with specific equipment for a specific period of time. These are commonly treated as one time step in a programme.

**Facility** - A geo-referenced feature directly collecting or processing data about environmental properties (physical, chemical, biological). These can be fixed or mobile, sites, stations or Sensors for example.

**Network** - Aggregation/grouping of facilities managed in the same way for a specific purpose, targeting a specific area. A network can be at the organisation/regional/national/international level.



## Annex 2 – List of attendees

### Attendees

Helen Beadman (UKEOF); Paul Willis (Ricardo-AEA); Dave Chambers (JNCC); Simon Coleman (NE); Jess Durkota (EA); Claire Edwards (NE); Gaynor Evans (MEDIN); Monica Jones (NRW); Amber Vater (UKEOF).

### Project team (NERC CEH)

Mike Brown; Sharon Hayton; Chris Johnson; Phil Trembath.

### Apologies

Mark Hallard (SEPA); Keiran Millard (Seazone); Phil Northing (FERA); James Penman (Met Office).

**Annex 3 – Workshop agenda**

| Time | Item | Lead |
| --- | --- | --- |
| 10.30 | **Welcome & workshop aims** | Mike Brown (Chair) |
| 10.40 | **UKEOF catalogue overview** | Amber Vater |
| 10:50 | **Benefits of the new UKEOF catalogue** | Mike Brown |
| 11:10 | **UKEOF catalogue upgrade – technical aspect** | Phil Trembath |
| 11:30 | **Data providers (part 1)** How data providers currently manage their information. | Phil Trembath |
| 11:50 | Refreshment break | |
| 12:05 | **Data providers (part 2)** Session continues... | Phil Trembath |
| 12:25 | **Case study**: how have UKEOF aligned with existing databases? | Phil Trembath |
| 13.15 | Lunch & networking | |
| 14.00 | **Discussion session** | Mike Brown |
| 14.40 | **Next steps** | Phil Trembath |
| 15.00 | **Close** | Mike Brown |

**Annex 4 – Pre-meeting request for information**

Delegates were asked to provide the following information to discuss at the workshop. The aim of this was to judge how organisations currently store information and their future plans.

- What application do you use to manage the information (e.g. Access database, spreadsheet, Oracle database)?
- How is the data structured?
  - If the information is stored in a relational database, what are the tables and their relationships
  - If the information is stored in a spreadsheet, what are the column headings
- Do you use controlled vocabularies/codelists to describe your data, if so which ones?
- Who in your organisation is responsible for managing the information?
- Is the information used purely for internal use or is the information already publically available?
  - If so how/in what format?
- What steps has your organisation taken to meet INSPIRE requirements for the Environmental Monitoring Facilities theme?
- When do you expect to have compliant metadata ready within yor internal systems?