UK Environmental Observation Framework

UK-EOF Data Solutions Workshop

Breakout Session C:
National Infrastructure

David Lister &
Liz Fox
Breakout Session C: National Infrastructure

• Contents:
  – What do we mean by National Infrastructure?
  – Why are we looking at this?
  – What is a Data Warehouse?
  – Future requirements from UK-EOF – can the infrastructure cope?
  – How does this relate to the Data Status Table?
  – Are we SEIS ready?
Breakout Session C: National Infrastructure

• What do we mean by National Infrastructure?
  – Primarily this session is addressing the technology required to enable metadata and the data itself to be accessible to users
  – In this session we will not be discussing the additional effort that might be required to carry out the conditioning of the data. This will be addressed in Session B
  – We will though discuss the technology required for enabling data to be transferred consistently
Data Service Providers e.g. NBN, EA, NERC, etc

Observations recorded in the field → Observations digitised - data created → Data and Metadata transferred → Data conditioning, curation & storage → Data Publication → Access agreements

~ 250 Collectors

Users

Activity Catalogue as part of the UK-Location Programme

From the Catalogue users are able to connect directly with the collectors or data warehouses to access and retrieve data.

Data stored close to source with easy to use licence conditions.

Users requiring a strategic overview of activities e.g. are we investing in the right areas of science to address the big societal issues?

Users requiring information about specific activities e.g. to join up resources and share platforms.

Users requiring access to the data the activity generates.

Some users continue to connect directly with collectors to access and retrieve data.

PRIVATE SECTOR

CONTRACTS & STANDARDS (via Location Council)

Observations

Data Service Providers e.g. NBN, EA, NERC, etc

Data Publication

Users

Access agreements

Data Discovery and use

Collectors

Observations recorded in the field

Observations digitised - data created

Data and Metadata transferred

Data conditioning, curation & storage

Data

Metadata

Digitised -> data created

Observations

Observations

digitised

Observations

digitised

Observations

digitised

Observations

digitised

Observations

digitised
Breakout Session C: National Infrastructure

- So what does the national infrastructure cover?
  - Physical hardware of the ‘pipes and buckets’
  - The schema required to enable different elements to talk to each other
  - The standards required to ensure greater consistency and to enable interoperability
Breakout Session C: National Infrastructure

• Why are we looking at this?
  – We need to make the UK-EOF DI work and if additional infrastructure is required, this needs to be provided
  – The Business Case that the UK-EOF team is putting together will identify the additional requirement and seek to justify the investment
• What is a ‘Data Warehouse’?
  – In this context a data warehouse is the repository for observation data and metadata from the programmes and activities undertaken by or on behalf of the UK
  – Different descriptions:
    – Data Warehouse
    – Data Archive Centre
    – Data Grid (NERC)
    – Data Service Providers (Location Programme)
  – Are there any more?
Breakout Session C: National Infrastructure

• What is a ‘Data Warehouse’? (cont)
  – Examples of key facilities managed by:
    – NERC
    – Environment Agency
    – British Library
    – UKHO
    – INSPIRE/ Location Programme
    – NBN
    – Defra Network e.g. SPIRE
    – Ordnance Survey
  – What others exist e.g. for storing data from smaller/ diverse collectors? (group 1)
  – Do they cover all the domains/ observation collectors (over 250)? (group 2)
  – What additional infrastructure is required to enable all data to be stored? (plenary)
Breakout Session C: National Infrastructure

- Current capability and future requirements?
  - Increasing demands from achieving the UK-EOF vision:
    - Discovery Catalogue
    - Requirement to publish programme/ activity metadata
    - Improving access to the data
    - Future increase in standards for collecting, storing and publishing data
  - What will be the impact on the current infrastructure? (Group 1)
  - How will this need to be addressed? (Group 2)
  - What are others e.g. Location Programme doing to address these issues? (Group 1 & 2)
Breakout Session C: National Infrastructure

- How does this relate to the Data Status Table?
  - will provide the means to evaluate performance and
  - report on the ability to store and transfer data and metadata effectively from one place to another
  - the resulting table will enable UK-EOF to highlight where resources need to be directed
<table>
<thead>
<tr>
<th>Data set e.g.</th>
<th>Data and Metadata Generation</th>
<th>Storing &amp; Archiving</th>
<th>Publishing/ Sharing</th>
<th>Access Agreements</th>
<th>Overall Ability to re-use</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Collection Standards</td>
<td>Data entry/ digitisation</td>
<td>Quality Assurance i.e. checking &amp; Cleaning</td>
<td>Creation of data documentation &amp; Metadata</td>
<td>Place to store physical samples</td>
</tr>
<tr>
<td>Atmosphere</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Example 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Example 2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Example 3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Biosphere</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Example 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Example 2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Example 3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lithosphere</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Example 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Example 2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Example 3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cryosphere</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Example 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Example 2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Example 3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fresh Water</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Example 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Example 2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Example 3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Marine</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Example 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Example 2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Example 3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

This is an example, for illustrative purposes only.
• How does this relate to the Data Status Table?
  – Place to store electronic data
  – Long term preservation strategy
  – Publication place e.g. portal
  – Publication standards e.g. GI or INSPIRE

• From the perspective of National Infrastructure, what is required to ensure that these are green?
• Are we SEIS ready?
  – How does this relate to SEIS?
  – What area is SEIS going to impact?
  – Is there anything else required for SEIS or any other initiative?
• Wrap up prior to plenary
  – What are the key messages we wish to pass back to plenary?
  – Are there any other issues we have not covered?