



Enhanced Access to Environmental Science Information

Scoping Workshop

Final Report for

Environment Research Funders' Forum
The British Library

February 2010

1 Introduction

1.1 Background

The digital environment is changing constantly and rapidly and, as the availability of services, data and information increases, so do user expectations. Managing the collection, storage and accessibility of information has become both increasingly complex and increasingly necessary.

The Environment Research Funders' Forum (ERFF) hosted a workshop on 14 January 2010 at The British Library to explore emerging issues around the collection, storage and accessibility of environmental science information. Participants - drawn from ERFF's member organisations, from The British Library and from other potential stakeholder groups such as the local government and business communities - discussed whether the current information infrastructure is working well enough to meet funder and user needs for information in today's complex digital environment or whether it needs to be updated to enhance access. Participants were particularly keen to explore whether the time is now right to develop a coordinated approach to gathering and accessing environmental science information.

The specific aims of the discussion were therefore to

- Consider enhanced access to environmental science information from a number of different perspectives
- Explore the extent to which enhanced access already exists and what gaps might need to be filled to create a more effective system
- Get some practical insights into what works and what doesn't
- Determine whether there is a business case to build enhanced access to environmental science information; and, if there is
- To suggest what needs to be done to get going

1.2 Workshop structure

The workshop was structured around 5 main discussion sessions:

- Why we need this conversation
- What enhanced access to environmental science information looks like from the perspective of
 - Users
 - Researchers/institutions
 - Funders
 - Information brokers
- What exists already that can we built on and what gaps need to be plugged
- Confirmation – or not – of the business case
- What needs to be done to get started

In addition, there were three presentations – from The Wellcome Trust, RCUK and SNIFFER – offering some practical lessons and thoughts about the way forward.

1.3 This report

This report documents the main findings and outcomes from the workshop. The outputs of each discussion session are recorded in sections 2 – 6 and we have added a brief commentary in most sections which summarises our thoughts on the discussion.

2 Why we need this conversation

2.1 Introduction

We asked participants to work in groups of 6-8 to discuss why they believed it important to consider a shared approach to enhancing access to environmental science information. Answers to the specific question fell into three broad categories: cost, efficiency and accessibility. In addition, participants highlighted a number of operational issues that any effective system would need to address.

2.2 Cost

- We can't afford to do this alone
- We need best value

2.3 Efficiency

- It will minimize duplication and optimize the impact of research funding
- There are too many organisations involved at the moment
- The onus is on evidence based policy making - but which evidence should be used?
- We need information finding to be more efficient and to increase discoverability
- We need to get the right information to the right people at the right time
- Discoverability vs relevance

2.4 Accessibility

- We are constantly looking for the best research and this will help us
- We need to bring in-house research out and make it accessible
- Its existence will help producers to process data and make information available
- It will increase transparency of information – and this will be very helpful for public communications and debate around contentious issues such as climate data
- It will help us move towards an 'open source' data model
- The information needs to be disseminated differently for different user groups (researchers, policy makers, public)
- There is a question about the skills of users and their ability to use the information – any approach needs to support efforts to increase information literacy amongst users

2.5 Operational issues

- We need to think about what kind of information we need
- We need to define the boundaries: interdisciplinary, broad definition of 'Environment', which data sets to use
- To make it work, we need to address issues around IP, interoperability, standards
- Issues about trust (worthiness)
- Important to avoid apparent bias – information gathered must be balanced in quality and quantity
- Do we need a European dimension? Euro data, euro information, European standards
- Need to be realistic about what this can achieve
- ...and how it replaces expert/face to face communication

2.6 Commentary

Overall, we noted strong enthusiasm from participants for taking a collaborative approach. We did not detect – at this or any other stage of the workshop – any unwillingness to work together or any significant divergence of opinion.

3 What enhanced access looks like

3.1 Introduction

For this discussion, participants worked in groups that reflected their principal interest - users, researchers/institutions, brokers (two groups) or funders – and discussed the key characteristics of a system to enhance access to environmental science information.

Characteristics fell into one of four broad categories: system design, inputs, outputs, use.

3.2 System design

- A single portal
- Technologically up to date
- A joined up (connected, hyperlinked) information landscape
- Use of appropriate metadata
- Semantic linkages
- Different levels of information available
- Wholly open? Appropriate access?
- Well indexed
- Excellent user support
- Follow the best systems elsewhere (e.g. health)
- Future proofed to take account of emerging technology

3.3 Inputs

- Broad/wide, but relevant
- Quality assured: must meet strict scientific criteria? In other areas (such as observational data) the level of quality assurance must be clear
- Co-ordinated
- Published at the right time
- Horizontally and vertically integrated
- Includes grey literature
- Researchers need incentives to put material in
- Researcher biographies

3.4 Outputs

- Must be relevant and discoverable
- “if you searched for this, you might like...”
- a peer review system that gives ‘Amazon-like’ approval ratings

3.5 Use

- A magnifying glass on what’s going on and what we need to fund (supporting funding decisions)
- Affordable
- Free
- Some support for those with limited scientific literacy to access and interpret information?
- Trusted
- Access to historical data is essential

3.6 Commentary

The purpose of this session was to build a shared sense of the characteristics that an enhanced access system might have, rather than to highlight or debate differences. Nevertheless, the responses show - we suggest - a high degree of alignment between different groups.

Two characteristics which might deserve further discussion are

- The level of access (wholly open or appropriate)
- Whether the information in the system is that produced by ERFF members alone or whether it is information produced by all environmental science researchers. This point was not explicitly discussed during the workshop.

4 Building on what already exists

4.1 Introduction

Participants continued working in their interest groups and addressed two questions:

- What do we have currently that we can build on?
- What are the main gaps that need to be filled?

Each group's responses are set out in the tables in this section. We have noted two categories of response: ones that can be loosely defined as 'information sources' and ones that can be loosely defined as 'management issues' which encompass aspects of system design and use.

At the end of the exercise, we asked groups to identify which elements can be built on and which gaps seemed most important. Not everyone carried out this step; where groups did, their responses are presented in bold.

4.2 Users

| To build on | Gaps to fill |
|--|--|
| <u>Information sources</u> <ul style="list-style-type: none">• Existing technology such as Wiki, google• Reviews – expand, better questions, better topics, wider publication• Biomed experts (Collexis)• Network of experts, papers• Text mining – enhance searches, better discovery | <u>Information sources</u> <ul style="list-style-type: none">• Access to older data and information• Specialist information such as maps, photos, images <u>Management issues</u> <ul style="list-style-type: none">• Resources and funding |

4.3 Researchers/institutions

| To build on | Gaps to fill |
|---|--|
| <p><u>Information sources</u></p> <ul style="list-style-type: none"> • DataCite • Google • Scientifically based briefing and guidance notes • ERFF metadata catalogue • Institutional support : subscriptions, repositories • Databases: eg web of knowledge, conference proceedings, cabi, institution websites <p><u>Management issues</u></p> <ul style="list-style-type: none"> • Layered simple interface with levels of detail • QAd information discovery (cf Wikipedia) | <p><u>Management issues</u></p> <ul style="list-style-type: none"> • Information needs to be free at point of use • Policy people are not disciplined about citing services • Text mining tools for environmental science • Lack of broad approach to quality assurance and labelling v peer review • Information literacy skills • Information on the people providing information • Discovering relevant electronic communities |

4.4 Brokers (i)

| To build on | Gaps to fill |
|---|--|
| <p><u>Information sources</u></p> <ul style="list-style-type: none"> • DataCite • Data discovery service • NERC data grid • GCMO • Go Geo • UK spatial infrastructure • ERFF research database • ERFF Environmental Observation catalogue <p><u>Management issues</u></p> <ul style="list-style-type: none"> • Technical mechanism for discovery • Metadata standards: Dublin Gore, MOLES, EMF-Serif, ISO 19139 | <p><u>Management issues</u></p> <ul style="list-style-type: none"> • Mechanisms for measuring impact • Common useful classification for environmental sciences • Time, money • Silo mentality and 'ownership' • A way to resolve the tension between specialism and inter-disciplinary expertise (maximise shared standard/shared language) |

4.5 Brokers (ii)

| To build on | Gaps to fill |
|--|--|
| <p><u>Information sources</u></p> <ul style="list-style-type: none"> • ERFF research database and classification scheme • UK environmental observation framework • Material in institutional repositories • Libraries • NERC data centres • Publisher databases • Centre for Environmental Evidence work on systematic reviews • Pubmed, Cochrane model • Social networking <p><u>Management issues</u></p> <ul style="list-style-type: none"> • Drive for evidence based policy making • Exponential growth of knowledge to be managed (is this a gap?) • Knowledge Council work on knowledge management, skills • Funders' influence on researchers and institutions (eg mandates) • REF – incentives for researchers • RCUK work on open access publishing • European legislation (eg INSPIRE/SEIS) and standards database • Use capability reviews to drive change in Other Government Departments • Use the Operational Efficiency Programme to drive change in government • British Library expertise • European/global activities and partners • Requirements for digital continuity | <p><u>Information sources</u></p> <ul style="list-style-type: none"> • Internet: everybody's records are part of the archive <p><u>Management issues</u></p> <ul style="list-style-type: none"> • Leadership • Long term funding • Incentives • Common data and metadata standards to allow interoperability • Links between data and publications • Rewards for use/citation of data • Protocols for data citation • Persistent identifiers for data • Impact of machinery of government changes? • Funders' reluctance to collaborate? • Prioritisation: which actions do we need to do first? What is the critical path analysis? • Funders' shared interest and influence. Funding/pressure/leadership to make this happen • Ownership for long term continuation • Clarity in impact of the Data Protection Act on data shoring • Peer review of data • Metrics on value, usage of data • Exponential growth of the data to be accessed • Cost – who pays? • Powerful and developing data indexing and search capability • Free and timely access to results |

| | |
|--------------------------------|--|
| coming out of the Dacre Review | |
|--------------------------------|--|

4.6 Funders

| To build on | Gaps to fill |
|--|---|
| <p><u>Information sources</u></p> <ul style="list-style-type: none"> • ERFF products: EOF, Research database, website portal to member websites • NERC Open Research Archive • Science Impact database • Grants on the web (GOTW) • NERC datagrid • NERC centre websites • NEG TAP • National biodiversity network • BRAG • ERANETS (Centre for Evidence based Conservation, Air Pollution Information) • ESDS social science • EST ESRC database • Defra SIS • SIBSTA Info • Scottishairquality.co.uk brings together air quality data from Scotland's LAs, summarises and presents map-based current data and interprets it in health terms <p><u>Management issues</u></p> <ul style="list-style-type: none"> • Build on networks – formal, informal, UK, international • Good practice elsewhere (eg biomed) • International collaboration information (eg peer, marine RA, Heads Geological Services) • Developing Scottish and Welsh portals for provision of and access to Scotland's or Wales' main sources of environmental monitoring data | <p><u>Information sources</u></p> <ul style="list-style-type: none"> • Finding the grey literature <p><u>Management issues</u></p> <ul style="list-style-type: none"> • Multiple stakeholders have different priorities • Synthesis for users (policy, public, industry) • Funding – more from industry needed • Legislative barriers and constraints • Interdisciplinary barriers • Geographical issues and barriers • IP issues • The challenge of bridging the cultural gap between specialist scientists and generalist policy makers • Careers and skills • Data protection and privacy |

4.7 Commentary

There are clearly a lot of information sources with which to get started. The challenge in creating an enhanced access system, therefore, is perhaps more likely to relate to knowing where to start; we can easily imagine that the perfect could become the enemy of the good. The most important thing to do, we suggest, is to begin building a system organically, using fewer elements in the first instance to test the approach and then scaling up as practical issues are resolved.

The gaps that need to be filled are almost exclusively management issues. Many of them relate to the design of the system and are therefore in the control of whoever sets it up, but we are particularly struck by the range of points made about the motivation and skills within government departments and policy teams. We imagine that proving the usefulness of any system (and therefore ensuring its sustainability) must require Other Government Departments – (OGDs) to be engaged by – not just aware of – it.

We have no argument with the prioritisation of elements that can be built on:

- ERFF products: EOF, Research database, website portal to member websites
- Build on networks – formal, informal, UK, international
- Good practice elsewhere (eg biomed)

...nor with the prioritised gaps to be filled:

- Leadership
- Long term funding
- Incentives
- Common data and metadata standards to allow interoperability
- Multiple stakeholders' have different priorities
- Synthesis for users (policy, public, industry)

5 The business case

5.1 Making the case

Participants were in agreement that a system to enhance access to environmental science information is needed and can be viable. It was noted that setting the boundaries that define 'environmental science' would require skill and thought.

5.2 Commentary: some critical strategic questions to address

At this stage of the workshop, we noted a number of strategic questions relating to the design and implementation of an enhanced access system. These emerged from the discussions and presentations:

- Is there a tension between using this to convey information and using it to demonstrate environmental, economic and societal impact? To support funding decisions?
- *Can* it be used to support funding decisions if compliance is below 50%?
- Outputs: what format and type (web pages, word documents, PDF?)
- Inputs: organisation or individual? Data, metadata, database of data?
- Where does the intelligence lie: the uploader, the downloader, or the system...?
- Which business model: Wiki, Google or Amazon? UKPMC? Something new and improved over these models?
- Can the system be made so good that it is the only game in town for everyone? Compliance, design, value?
- ...and how many phases will it take to get there?

6 What needs to be done to get started

6.1 Introduction

In the final session of the day, we asked participants to identify what needs to be done to begin building a shared approach to enhancing access to environmental science information.

We asked participants to identify no brainers, creative and heroic ideas¹ for putting the system in place.

6.2 No brainers

Leadership

- Secure leadership for and commitment to the idea
- Articulate the business case and secure funding
- Design a business model for a sustainable resource

Build on what exists

- Create federated searches of databases: join up existing information with a google search tool
- Ensure grey literature is a core component
- Explore how to digitise the archive/repository and make it discoverable

System design

- Design in interoperability
- Design a single portal approach
- Work towards a single system in the longer term that is easy for people to lodge information in
- Use the system to signpost a range of resources
- Design in links between publications and the data
- Develop metadata standards
 - Programmes
 - Projects
 - Data
 - Skills

¹ A 'no brainer' is something that is easily identified, even obvious, although it may not be easy to deliver. A 'heroic' idea is one that carries some reputational or operational risk – but which will make things happen quicker if it is successful.

Incentives

- Make it a condition of funding regimes that people store information in the system
- Identify the carrots and sticks that will make researchers provide their research

6.3 Creative ideas

Funding

- Get funders from other places (eg Vodafone Foundation; Tesco)

Build on what exists

- Use the DPSIR - Driving forces, Pressures, States, Impacts, Responses - system
- Link the database to valuing ecosystem services
- Put the system on a government secure intranet

System design

- Impacts with Digital Object Identifiers (DOIs): system of DOIs to all components of a research project (people, inputs, outputs)

Create new forms of value

- Use the service to synthesise research and perform new systematic reviews
- Create a virtual research environment

6.4 Heroic ideas

Funding

- Ask Google to help set up and fund the system

System design

- Establish an environmental version of PMC

Create new forms of value

- Create a classification system that is universally accepted – convert all subjects to ISN
- Open the system beyond ERFF membership
- Taxonomy of taxonomies: universal interoperability
- Citation index as standard for a wider range of information (including data sets)

6.5 Next steps

Participants identified what must be done now to take the project forward:

Secure agreement

- Put a paper to ERFF

Map what exists in the environmental information landscape

- List current databases
- User requirements in different sectors
- Services and content
- Stakeholders, use cases, approaches
- Size and motivations of the user base
- Use this as a driver to improve individual databases in organisations and define standards

Confirm users

- Scope who would use it nationally and at European level
- Agree who it is for: government departments? Business?
- Talk to publishers to test ideas

Clarify costs and benefits

- Clarify costs and benefits for all audiences/users

Engage funders

- Create a funders register to collate all grants and contract information in the UK
 - RCs and other organisations
 - All subject areas
- Harvest information from organisations' websites

Develop operating principles

- Develop mandates on open access and how to enforce compliance
- Work towards a GEOSS approach globally

Build commitment

- Engage NGOs and champion (Prince Charles?) to persuade government to do this
- Brief HOC/Scottish Parliament Select Committees
- Engage with EEA, OECD

6.6 Commentary

We have little to add to this section other than to re-iterate some earlier comments

- Throughout the workshop, we noted strong enthusiasm from participants for taking a collaborative approach to enhancing access to environmental science information.
- We did not detect any significant divergence of opinion about what should be done.
- The challenge in creating an enhanced access system is perhaps knowing where to start. The perfect could easily become the enemy of the good, however, and we recommend that, if funding can be secured, the system is built organically using fewer elements in the first instance to test the approach and then scaling up as practical issues are resolved.