UK Environmental Observation Framework

Sharing Environmental Observation Data – the Issues

UK-EOF Programme Team
September 2008

Environment Research
Funders' Forum

Sharing Environmental Observation Data – the Issues

Purpose of this paper

This briefing paper summarises the major issues affecting the sharing and reuse of environmental observation data. It suggests a vision of increased reuse in line with the strategic objectives of the UK Environmental Observations Framework (UK-EOF) and asks the Think Tank to debate what the next steps should be – if any.

Summary

a.   Environmental Observation data are a valuable resource, in which the UK invests up to £500m pa. The data that are generated are fundamental to improving our understanding of the environment and enhancing our decision making ability in the face of environmental change. The real significance and potential of the data may not be known initially until there is a time series or they are combined with other data.

b.   For optimal public and commercial benefit datasets need to be shared and reused. In principle, there is plenty of scope for reuse. Around four fifths of environmental observation data are collected at public expense and are generally subject to access legislation and legislation encouraging reuse. However, only an estimated 20% of all environmental observation data are currently held under a governance policy facilitating reuse. This is potentially a significant waste of resources and is a frequently voiced as a concern among any potential user of environmental observations.

c.   The process of ensuring effective sharing and reuse of environmental data has three main elements:
- recognition that problems exist
- agreement of what constitutes success (across a varied community)
- supporting and resourcing actions which produce the desired changes.

d   The main characteristic of success can be put fairly simply:

- **People/organisations in "UK plc" actually share and reuse environmental observation data to inform policy decisions, expand knowledge, contribute to international activities, stimulate markets for innovation, improve operational capabilities and improve their responses to the changing environment.**

e   Two further determinants of success support the main vision. These are that:
- **All data collectors adhere to best practice (as opposed to minimum compliance with legislation)**
- **Each organisation has a transparent and well publicised data policy which balances open access with economic and societal pressures.**

f.　　Examples of good practice exist in a piecemeal fashion across the environmental observation community. National and global initiatives are being developed which may result in better governance and wider data sharing and use. However, organisations still meet – or perceive - significant barriers to data sharing. These can be generally classified as:

- Barriers to awareness and discovery of data, including who owns it, where is it held and what conditions of reuse are.
- Technical barriers from storage and archiving incompatibilities.
- Cultural barriers where organisations do not value information governance (a prerequisite for data sharing) and do not include it as part of their "core business"
- Economic barriers arising from high prices and restrictive licence conditions, or from perceptions of these.

g.　　Some barriers will only be resolved as a result of debate and agreement at the highest level. There is a tension in Government between the desire for public good arising from more open access to data and the requirement that some organisations must make a financial return from their data. It is not likely that these economic pressures will weaken or that organisations will be willing to address them unilaterally.

h.　　However, given the will and sufficient resources, many of the other issues could be overcome by any organisation. Obligations and initiatives such as INSPIRE, SEIS and GEOSS will help in providing the technical knowledge to enable data sharing and there are examples of existing good practise which could be shared.

i　　Therefore, one of the main gaps between the vision and current practise concerns the cultural aspects. If every organisation is asked to examine their data policies and come to a clear statement of where they stand on charging and IPR issues, and specifically to which of their datasets licences and charges apply, we will be closer to success.

j　The think tank is invited to debate the proposed vision, the scope of the application and to offer guidance on appropriate working structures to achieve it.

Questions for discussion
The UK –EOF would like the Think Tank to discuss:

i.　　If we should seek to construct an "ideal" data policy
ii.　　If the scope of this work should cover all environmental data or be restricted to environmental observation data (including baseline surveys).
iii.　　Whether we should aim for minimum compliance by 2019 as required under INSPIRE – or whether we should be more proactive - such as 90% of data should be held under a transparent and effective data policy by 2013?
iv.　　if the environmental observation community (facilitated by UK-EOF) should seek clarification on where Government / Treasury intends the Trading Fund/Wider Markets Initiative model to apply, and where the "public good" model is preferred. How should charging policies be regulated? Are these issues better left to each funder of observations?
v.　　if we should seek to challenge and/or help enforce existing policies and best practice
vi.　　if we should be proactive in the international and European areas, setting data standards such that the UK is seen as leading the way?

vii.   If there should be a champion for this work at senior level – i.e. we establish an annual reporting process to show progress towards any agreed vision

1. Introduction – the UK Environmental Observation Framework

1.1   What is environmental observation? It is a very broad term used to describe the taking, on a reasonably regular basis, of any form of observations relative to the status of the environment. The observations may be made on land, in air, in fresh or salt water; from ships, satellites, or by using any technology including by hand or eye. They may measure physical, geological, chemical or biological properties of the environment, and be collected for a variety of purposes.

1.2   With all this variety, observations do however share three key characteristics. The wider significance they may have, or for what other purposes they may be needed in the future, may not initially be anticipated. The data generated by observation activities is unique – it is not possible to go back in time and recreate the circumstances of collection. Frequently this data reveals its full potential when combined with other data, allowing us to understand complex systems.

1.3   We are aware as never before of the impact of human activity on the environment – the many issues around climate change, depletion of resources, loss of species. Society needs observations to provide understanding and evidence of what is happening in the environment, why, and what can be done about it. Accordingly, the UK Environmental Observation Framework (UK-EOF) was launched in 2008 to co-ordinate environmental observation activities to achieve the following impacts:

- expand knowledge of the environment;
- inform policy decisions;
- improve societal response to changes in the environment;
- improve operational capability;
- contribute to international activities;
- stimulate a market for innovation.

1.4   Figure 1 shows how data collection, handling and use is at the heart of any observation programme and thus at the heart of UK-EOF. The Knowledge Base workstream of UK-EOF aims to:

- identify observation activities in a metadata catalogue, so that their existence, purpose and characteristics are known
- assess their suitability for use in achieving members' and UK-EOF strategic objectives

and lastly, but most importantly:

- facilitate that use, addressing barriers which may inhibit it.

Figure 1 The Concept of UK-EOF
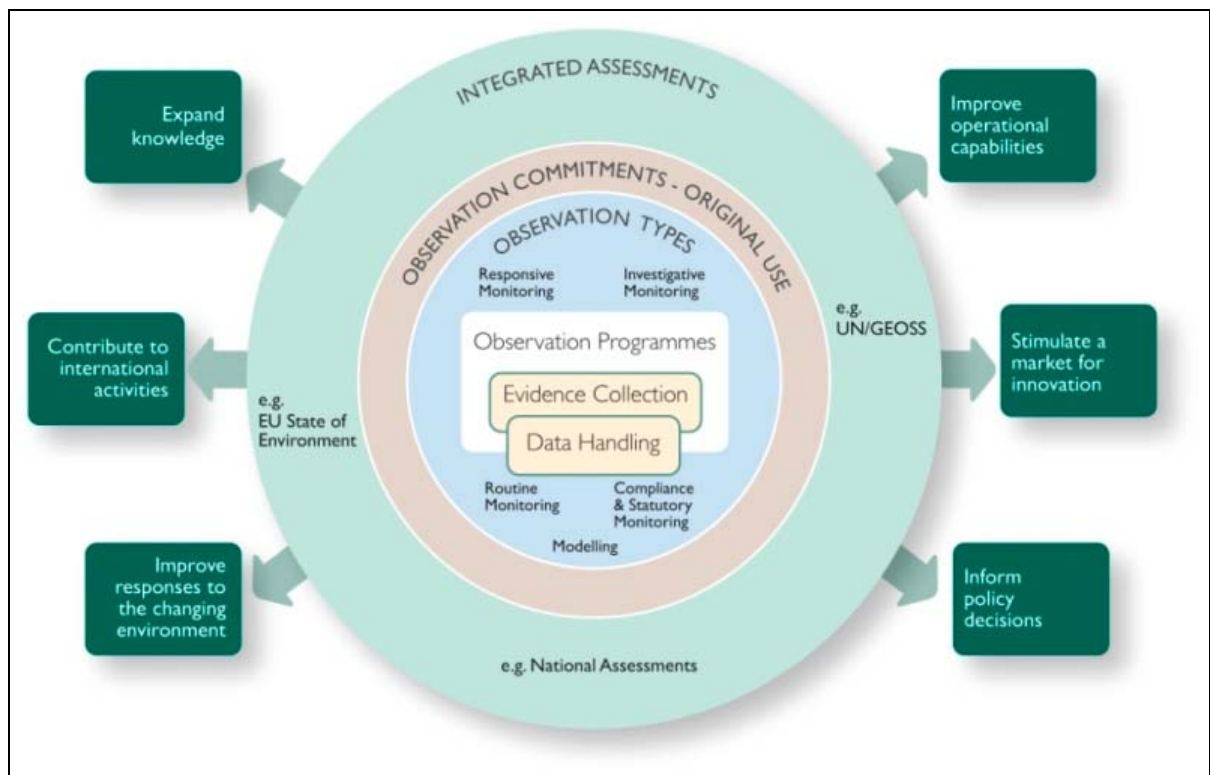
1.5   These barriers appear to have a significant effect. The Environment Research Funders' Forum database, created in 2006, contains information on over 600 environmental monitoring programmes. Only 20% of the data is held under a data policy allowing unconstrained reuse. As the cost of collecting this data could be as much as £500m, the 80% potentially underused could represent a significant waste of resources. Concerns over underuse of data, and frustration at the perceived causes, are regularly expressed by all sectors of the environmental observation community.

1.6   In order to facilitate reuse of environmental observation data, a set of several interrelated issues must be acknowledged. The purpose of this paper is to articulate the issues and to generate high level ownership of the challenges they pose. The paper examines the legal and economic framework under which data can be shared and the key national and international initiatives in data sharing, both current and planned. It then analyses the barriers to data sharing which remain to be addressed, and suggests some key activities for UK-EOF itself, and its stakeholder organisations, to progress.   It poses some controversial questions to stimulate debate.

2.   Environmental observation in context

2.1   Environmental issues are increasingly interesting to citizens and receive regular and high profile media coverage. The data produced by environmental observation activities are fundamental to increasing scientific understanding and to public decision-making in response to environmental change. This data is very varied in its nature. It may exist as numbers or images; be measurements of physical, chemical or biological properties of the environment; be produced by automated methods involving significant investment – for example satellite observations of the globe – or by volunteers manually recording the number of birds in their gardens. It may be produced by private, public or voluntary sector organisations, each responding to different economic and social drivers. The data produced by these diverse activities all contribute to the remit of UK-EOF.

2.2   The wide variety of bodies and individuals collecting environmental observation data also means that there are varying levels of understanding of the concepts involved in data management, and some differences in definitions. For the purposes of this paper we are proposing the working definitions and concepts detailed in Annex 1.

Legal context

The existing and future legal framework

2.3   There are few explicit legal requirements to share data. The main legislation covering data management is the Data Protection Act 1998 (DPA), and its primary concern is how data is processed - especially the safeguards given to personal information. DPA allows data sharing with precautions; it does not require it, and following some high profile losses of personal information, public concern is focussed more on data security than sharing. Under the Freedom of Information Acts (2000 in England, Wales and Northern Ireland; 2002 in Scotland) and Environmental Information Regulations (2004 in all parts of UK) (EIRs), public authorities must release information on request (subject to a range of exemptions). EIRs are particularly relevant to the environmental observation community; however, the Office of Public Sector Information (OPSI) notes that provision under access legislation does not include the right to reuse; intellectual property rights – copyright and database rights – are protected.

2.4   The Ministry of Justice advises that most public bodies, ministerial or non-ministerial, may share data using implicit powers, and many are encouraged to do so under the terms of the Reuse of Public Sector Information Regulations 2005 (ROPSI). This legislation aims to encourage the reuse of public sector information (PSI) by making it clear what is available and under what terms. It covers only information collected as part of a PSI holding organisation's "public task". (This legislation is currently under review, which may affect the sharing of data produced by public sector organisations).

2.5   Certain public bodies – the Trading Funds (executive agencies of Government under the Trading Funds Act 1973) – are required by HM Treasury to make an economic return on their operations beyond meeting their costs. Several major data owners are governed under this model, including Ordnance Survey, the Met Office and the UK Hydrographic Office, which between them cover the majority of the supporting or baseline datasets used by the environmental community.

2.6   Recent studies (CUPI, POIR, Cambridge) have examined the economic and societal effects of making public sector information more widely available. These have highlighted a dichotomy in government approaches to information. Broadly this arises from the Trading Fund legislation (and the non-mandatory Wider Markets Initiative), which aim to make a return on information "assets" for the organisation, and access/ROPSI legislation which has the aim of increasing public benefit by making information cheap and widely available. This dichotomy, and the tension it creates, is widely discussed in the environmental observation community.

2.7   Concerns over the effects of the Trading Fund business model on the wider information industry have prompted The Shareholder Executive and HM Treasury to carry out an assessment of Trading Funds. This is focusing on the price structure and effects on downstream markets. The assessment is expected to report in autumn 2008, and although not aimed primarily at the environmental community, changes to Trading Fund operations are likely to have a significant impact on such users.

2.8   Privately funded data (collected by commercial or voluntary organisations) is subject to relevant legislation on Intellectual Property Rights (IPR). (The provisions of DPA will also apply if the data fall within the definition of personal information –at present this is unlikely in the voluntary sector, which tends to focus on terrestrial ecology). About 20% of environmental observation data is collected by volunteers, and their efforts are very valuable, both in terms of the activities covered and the public engagement and media interest they generate. For example, organisations such as the RSPB continue to have significant political influence using data collected by their volunteers. Voluntary organisations' willingness to share their data depends on their policies and the existence of resources to do so. This is often complicated where the sole asset of the organisation lies in its data.

2.9   INSPIRE – Infrastructure for Spatial Information in the European Community – is a new EU Directive aiming to improve the sharing of spatial information between public authorities and improve public access to it. INSPIRE will mandate standards across all Member States for environmental spatial data via a set of Implementing Rules, currently being developed. Defra are in the process of transposing the Directive into domestic legislation, although full implementation is a staged process lasting until 2019. INSPIRE will allow for the possibility of charges to be levied for data.

2.10   SEIS - Shared Environmental Information System – is an EU proposal currently at Commission Working Group stage, and intended to cover all types of regulatory data collected in order to report under other EU Directives. It will be implemented in a similar way to INSPIRE, and public authorities who have developed data management systems which are compliant with INSPIRE will be able to apply this experience to SEIS. (More information is expected in November 2008.)

2.11   On a global scale, GEOSS and GMES are examining approaches to data policies that may ultimately result in changes to legislation. The trend of these will be in the direction of increased drivers for data sharing.

Major environmental data sharing initiatives

2.12   In addition to the legislation governing access to and reuse of data, there are a number of significant non-mandatory initiatives, ranging in scope from the global to the sub-national, which affect the landscape of data sharing. Compliance with legislation is essentially a wide scale but reactive process; contributing to an initiative in order to solve a problem is proactive. The key initiatives, from the point of view of the environmental observation community, are set out in the following sections.

(i)   Geographic Information

2.13   SPIRE – Shared Spatial Information Services – is a Defra initiative aimed at improving the use of Geographic Information (GI) in Defra, its agencies and associated bodies, such as the Environment Agency and Natural England. It is intended to be the national provider of services under INSPIRE. Issues such as how the underlying data is identified, quality assured and made available are key aspects of SPIRE.

2.14   The UK Location Strategy – entitled "Place Matters – the Location Strategy for the UK" – has been developed by the Geographic Information Panel to promote a coherent approach to the management and use of GI in the UK and to advise government on key medium to long term GI issues. It is not yet in the public domain but will provide a framework for the use of GI datasets, which are the main supporting data used by the environmental observation community.   There are five main strategic areas, intended to ensure that the UK exploits the full value of its Geographic Information:

  (1) we know what data we have, and avoid duplicating it;
  (2) we use common reference data so we know we are talking about
  the same places;
  (3) we can share location-related information easily through a
  common infrastructure of standards, technology and business
  relationships;
  (4) we have the appropriate skills, both among geographic
  professionals and among other groups who use location
  information or support its use;
  (5) we have strong leadership and governance to drive through
  change including the implementation of the Location Strategy and the
  implementation of INSPIRE.

(ii)   Marine

2.15   MEDIN – Marine Environmental Data and Information Network – is a partnership of public and private sector bodies with the objective of improving access to marine and coastal data and helping the UK better manage its marine data resources. It has begun to address the issues for the marine sector and has produced a straightforward protocol and established a set of distributed data centres linked via a web portal. It is however looking to UK-EOF for collaborative work and guidance on the more complex policy areas.

(iii)   Terrestrial

2.16   NBN – the National Biodiversity Network - is a partnership which involves many of the UK's wildlife conservation organisations, the government and country

agencies, environmental agencies, local records centres and also many voluntary groups, with the objective of collecting, storing and making available, a wide range of UK biodiversity data. It sets an example of collaborative working, especially in engaging the voluntary sector, and maintains a user-friendly web portal, the "NBN Gateway".

(iii)   Research/Academic

2.17   NERC - Data Centres and Grid. NERC maintains seven designated data centres, responsible for NERC's data assets. As well as undertaking long term storage (curation) and providing access to the data consistent with NERC's data policy, the centres provide guidance on data management issues. The NERC Data Grid allows potential data users to find, understand, manipulate and visualise data from many institutions (at present mostly holding oceanographic and atmospheric data) around the world. The Data Grid technology uses international standards to find data, converts it to a common format and enables use of it via a portal. This technology will also be integral to making more NERC data available under the provisions of INSPIRE.

(iv)   Economic

2.18   The Wider Markets Initiative (WMI) was established in 1998 by HM Treasury. It encourages organisations to create revenue from their datasets, although it is not mandatory on organisations and sets no financial targets.   A relaunch of WMI, as recommended by the National Audit Office in 2006, may intensify pressures on organisation to go down this more commercial route. At present, in the environmental observation community, the Environment Agency, Natural England and the British Geological Survey use this model to generate income from some or all of their data.

(v)   Global

2.19   GMES – Global Monitoring for Environment and Security – is a joint initiative set up in 2008 by the European Commission and the European Space Agency (ESA). It is a 1.2 bn euro initiative driving the future of the EU's satellite observations. GMES aims initially to produce services relating to land monitoring, marine and emergency management, with air and climate services also in scope. It requires data from both space observation and in-situ monitoring from Member States, together with spatial, hydrographic and meteorological reference data. It is working towards a data policy and common standards, in line with the data principles of INSPIRE and GEOSS. GMES has begun to face the barriers to sharing data and is looking to its membership for resources and guidance.

2.20   GEOSS – Global Earth Observation System of Systems – is an initiative of GEO, the Group on Earth Observations, an intergovernmental organisation including 74 countries, the EU and the UN. GMES is seen as the European contribution to GEOSS.   GEOSS aims to provide a very wide range of information for decision making in nine main areas of societal benefit (disasters, health, energy, climate, water, weather, ecosystems, agriculture, biodiversity), by combining existing and planned observation systems around the world, and developing new ones to fill gaps. It has a 10 year implementation plan running to 2015 and will create its own data policy and data architecture (storing and sharing mechanisms). GEOSS is currently consulting on a data suitability "tag" to aid its wide constituency of end users.

Other initiatives and links

2.21　　There are many other initiatives either aiming to promote sharing of data and information, or which require it to achieve their objectives. They may be addressed to the public at large, such as the Power of Information Taskforce, or at specific parts of the environmental data spectrum. Many are technically driven and aim to create standardised procedures for describing and accessing data.

Summary of current legal framework and initiatives

2.22　　Public bodies, which produce four fifths of UK environmental observation data, are generally allowed to share their data, and required to give access to their information, as a result of legislation or implied powers. The majority of the remaining observation data (up to 50% in the field of terrestrial ecology) is collected by voluntary bodies which are not affected directly by Government policy.　Sharing of this data will rely on the data policy of the organisation as owner of the IPR.

2.23　　The economic drivers of Treasury targets and the Wider Markets Initiative encourage data sets to be seen as assets and to be traded under an average cost (cost recovery) regime. There is a tension between the economic benefit to the data holding organisation, and the wider public good arising from innovative use of the data either by business or by the environmental sector.

3.   Barriers to data sharing and reuse

3.1   Despite (or maybe because of) the plethora of existing initiatives, a number of barriers to effective sharing and reuse still remain. While some sectors and organisations have found, or are developing, solutions, these are piecemeal and often still have broader funding, IPR and organisational challenges to deal with. UK-EOF has been tasked with articulating these barriers, identifying good practice and helping to replicate it.

3.2   The barriers described in this paper fall broadly into four categories: discovery of data; technical issues: economic and policy issues; and organisational/cultural attitudes. Each is discussed below as they relate to the environmental observation community.

Knowing what exists already – "discovery"

3.3   All players in environmental observation and the research community acknowledge problems with the discovery of data – in plain terms, knowing what's there. The wide variety of activities under the umbrella of environmental observation means that data is collected and stored by a range of organisations, often without a common approach to data management (archiving or storing the data and describing it so it can be reused). Very few organisations have a "shop window" through which third parties can see what data is held.

3.4   The result is decreased efficiency in sourcing useful data; at worst, decisions are made on sub optimal information or the data is re-collected at unnecessary cost. These problems are most marked where potential users are looking for data outside their own speciality – arguably where cross cutting or system approaches will deliver new insights, or where information is being sought on a policy issue. It is estimated by the marine data community that 75% of the costs of reusing data arise from searching for that which already exists, with only the remaining 25% spent on analysis. Moreover, these searching or "discovery" costs arise each time the data is accessed – no efficiencies arise from the experience.

3.5   Some parts of the community are responding to this challenge by creating information gateways to facilitate the discovery of, and access to, separate datasets on a common theme. The National Biodiversity Network (NBN) currently has over 30m species records drawn from 264 datasets; its web portal also offers advice on searching, transparent metadata and a data quality assessment. MEDIN is creating a similar gateway for marine topic areas. Unfortunately, this "umbrella" approach has not yet reached a number of global sector based initiatives with no oversight in the UK, so a potential user still has to search for them all separately.

Technical Issues

3.6   Technical issues, related to the variety of data management procedures and the lack of common standards, act as a barrier to data sharing. Interoperability (the ability of a system or dataset to interact and exchange data with one or more other system or datasets) is the goal, and mechanisms are being developed to progress this. INSPIRE will mandate standards for environmental spatial data via a set of Implementing Rules (currently being developed); SEIS is expected to do so. The need to become INSPIRE compliant should drive an improvement in interoperability among public authorities, although full implementation is not until 2019, and SEIS will

be later.   In the meantime, technical problems will remain unless organisations make it a matter of policy to adopt common standards and develop the capability – by attracting and retaining suitably qualified staff – to do so. In essence, technical issues arise not because the technology can't cope, but because the people who operate don't – or are not encouraged to – talk to each other.

3.7   The community view is that larger data holders have from necessity developed procedures for data management that stand them in good stead, not least because they maintain staff expertise and archiving infrastructure as a corporate resource. The large international initiatives will - eventually – impose common standards on certain datasets, and national programmes such as SPIRE will encourage interoperability between the datasets accepted as SPIRE data layers. However, many of the smaller and less sustainably funded (or unfunded) organisations, especially those not in the scope of legislation, may struggle to achieve this. The rate of progress will depend on the level of resources available and the organisation's willingness to engage with the technical requirements.

Economic and policy issues

3.8   Once a dataset is identified, the potential user must gain access to it and secure permission for reuse. While access to information may be available under legislation (EIRs or FoI), there are often a number of obstacles to reuse:
- high cost (relative to the user's budget)
- the holding organisation can't price it
- the holding organisation isn't resourced to provide it, even for payment
- licence conditions prevent full use (eg no web publishing or sharing with stakeholders)

The result is that less than 20% of raw environmental data is freely available, with 70% available only under a range of constraints, and 10% apparently unavailable for reuse.

3.9   There are of course legitimate reasons for an organisation not releasing its data, such as protection of vulnerable species, owing a duty of confidence, and security/counter terrorism considerations. However, there are cases where these reasons do not appear to apply and yet the data remains hard to access, either because there is no clear data policy or because of the number of conditions. There may also be some data which is collected for a defined purpose and for which there is (at present) no demand for reuse. (Though this may not always be the case – the Continuous Plankton Recorder (SAHFOS) is no longer needed to support the herring fleet as it was in the nineteen thirties, but it is now recognised as a valuable environmental and climatic indicator).

3.10   There is a tension between the desire for openness and access to data, in line with policy and legislation, and an organisation's desire, or need, to maintain control of its data for policy or financial reasons. A reduction in income may affect an organisation's ability to reinvest in data quality or place greater pressure on other income streams, with consequent effects on other parts of its public task. While this disbenefit may be balanced by an increase in the overall public good arising from more open access to data (as indicated by economic analysis such as the Cambridge Study), it is perhaps unreasonable to expect any organisation to address this tension unilaterally.

3.11   There is evidence from several sources that such tension affects other organisations, which may miss opportunities to use relevant data. The most obvious result is that some organisations may be unable to afford access to desired datasets. Uncertainty, or a perception of expense or difficult licence conditions, may be equally inhibiting to effective data access. For example, the MOD submitted to the Commons Communities and Local Government Committee in 2007 that the license costs imposed by OS are particularly high and where data is not covered by the advantageous Pan Government Agreement, public bodies in general are paying very high prices or even finding the data to be unaffordable.

3.12   A more specific example relates to the reuse of marine data. The Department for Transport and the Maritime and Coastguard Agency pay £5.4m pa to UKHO to generate hydrographic charts required under international obligations to the Convention on Safety of Life at Sea (SOLAS). This data product has to be recast in order to provide a product needed by the environmental sector. UKHO set up a subsidiary company (Seazone) to create and licence a hydrographic chart for environmental purposes, using the same base survey data. The cost to Defra (and other public bodies) for this recast data was proposed at £1m pa, a significant increase on the £60k pa of the previous licence.

3.13   On the international front, problems with access to datasets owned by Trading Funds/WMI organisations could become a major barrier to fulfilling the UK's commitments to GMES. In order to make the best use of the billion-euro investment in satellite observations, these need to be processed with in-situ (earth-based) data. The complications of ownership and cost allocation for access to baseline data held by OS, UKHO, BGS, EA etc will give rise to problems. As will the UK's inability to deliver a complete set of in-situ data – a key contribution to GMES.

3.14   Further potential barriers arise from the question of "downstream" IPR – the derived copyright issue, where reuse of data is only allowed subject to a condition that reuse stops with the first reuser. While it is understandable that the original supplier of data would wish to ensure the widest range of potential customers for their data (and hence maximize revenue), the restrictions further down the chain generally require all the organizations concerned to spend time working out the implications for their operations. This imposes an opportunity cost and adds to the perceived barriers, a problem noted by Natural England in respect of their use of OS data to share maps with their stakeholders.

3.15   To summarise, this is a complex area governed by powerful economic, societal and environmental drivers. Significant change to the overall data sharing landscape will only occur if there is a debate and consensus at the highest policy level. However, this should not deter organizations from examining their data policies and coming to a clear statement of where they stand on charging and IPR issues, and specifically to which of their datasets licences and charges apply.

Organisational/cultural issues

3.16   Many barriers to effective data sharing result from the attitude of an organisation (or of key individuals) to data. The organisation's culture is critical. Data – and particularly the processes of good data management – may not be high on senior management's priority list, despite the reliance of many organisations on data to fuel the "core business". This attitude may extend to the staff engaged in data management activities, who as a consequence may feel undervalued, or to the

non-enforcement of such data policies as do exist – a vicious circle resulting in demotivation and eventual loss of capability.

3.17   Some organisations and individuals, especially those close to the research sector, may have a tendency to hang on to results with a view to future publication, or as material for future research contracts. Under the HEFCE model, a scientist's professional advancement depends on publication, so it would be unrealistic to expect change at individual level without support from the organisation and its funders and clients. However, the marine sector has seen a significant change when those producing data have accepted that devoting resources to data handling is just as valid as resourcing the collection activities themselves. This results in data producers thinking through all the issues associated with their data before collection occurs, with consequent increases in internal consistency of approach and external transparency.

3.18   In bringing about cultural change in organisations, it is generally established that a shift in attitudes and behaviours will only occur if senior management articulate a clear message of support for change and a robust plan for implementation at all levels in the organisation. This is not a simple undertaking, and will only happen if the organisation and its funding partners and clients are persuaded that data management is a key competence for the whole organisation, and one deserving of senior oversight and championing.

3.19   It is in this area of strategic data management, or "information governance" that UK-EOF would like to see significant change, as without it, the technical, financial and management issues will not be resolved. A further advantage of the information governance approach is that it is a credible mechanism to achieve transparency of the organisation's data policy and has sufficient power in the organisation to ensure enforcement of the policy.

4.   What does success look like? – and some possible steps to get there

4.1   The process of ensuring effective sharing and reuse of environmental data has three main elements:
- recognition that problems exist
- agreement of what constitutes success (across a varied community)
- supporting and resourcing actions which produce the desired changes.

4.2   The previous section has discussed some of the problems and issues which organisations and individuals are facing in managing and sharing data. This section begins to articulate what successful data sharing might look like in the UK as a whole and from the point of view of organisations and individuals.

4.3   The main characteristic of success can be put fairly simply:

People/organisations in "UK plc" actually share and reuse environmental observation data to inform policy decisions, expand knowledge, contribute to international activities, stimulate markets for innovation, improve operational capabilities and improve their responses to the changing environment.

4.4   To carry out their core business, organisations obviously create and maintain their own data, and have been doing so for many years. Some data sharing has always taken place, especially between operational partners, and – as detailed in section 2 – wider initiatives are being developed. However, this does not mean that success is close at hand. The barriers identified in section 3 and the concerns expressed by the community at the UK-EOF's Launch and Workshop events indicate that there is much more to be done. Significantly, as concerns over the environment become more intense, the questions become more difficult – scientifically and societally. To achieve sound outcomes in a complex future, all relevant data must be accessed for the purpose in hand. The processes that characterise the key causal relationships, and hence guide policy and operational decisions, must not be weakened or constrained by inadequate access to data.

4.5   The main vision is one of "market pull". It is externally focused on the needs of organisations which must deliver effective policy, respond to emergencies, innovate etc. Use of shared data is not an end in itself, but a means to achieve outcomes.

4.6   Two further determinants of success will support the main vision. These are that:
- **All data collectors adhere to best practice (as opposed to minimum compliance with legislation)**
- **Each organisation has a transparent and well publicised data policy which balances open access with economic and societal pressures.**

These success factors relate more to supply side "push" – not merely of technology, although "best practice" will eventually encompass solutions to the technical barriers identified in section 3. The factors encapsulate what data collectors and owners will have to do to meet the needs of the "market" – using the term in the widest sense.

4.7   The following paragraphs set out some aspects of best practice, though by no means an exhaustive list. Achieving these would require a number of different

activities, undertaken on different timescales; most will need the active partnership of UK-EOF and of organisations across the environmental observation community.

Best practice – helping people to know what's already available

4.8   The creation of an Environmental Observation Catalogue is a key deliverable of UK-EOF and its members. This in itself should go a long way to ensure that decisions – whether policy, operational or business related – are based on data which is not only easy and practical to discover, but is accompanied by metadata which establishes its suitability and relevance. As such, the creation of the Catalogue will be a joint undertaking. Organisations contributing metadata will have to build a process into their core business to ensure that metadata is kept up to date (estimated as a minimum 0.1FTE per organisation). The Catalogue will benefit from the example set by UKDMOS and the ERFF Research database.

Best practice - technical

4.9   The major EU obligations such as INSPIRE and SEIS will provide a mandatory framework for sharing GI and regulatory data. In preparation for this, environmental observation organisations will have to commit resources to identifying and resolving technical problems. There is an opportunity to make progress in advance of the 2019 implementation date – i.e sooner rather than later. However, these major initiatives do not apply to all types of observation data or to all types of organisation, so there must be a process of extending the "lessons learned" and best practice from INSPIRE and other initiatives.

Best practice – organisational

4.10   Cultural change in organisations is a vast subject. However, commonly accepted factors in successful change include leadership at the highest level and implementation throughout the organisation by empowered people who understand the reasons for the change. Within the environmental observations community, culture change in organisations could be assisted by a "toolkit" of measures, including: standard data clauses for contracts; a citation scheme for reused datasets; other recognition mechanisms, training in data management (which could be technically focused or in the context of managerial skills); data management skills included in post graduate programmes.

Data policy and economics

4.11   Some organisations will already have an effective data policy, while for others this may involve a journey. Best practice in terms of data policy will involve gaining agreement on what makes an effective data policy, finding examples of model wording, scope etc, and setting out a process for replication. Successful replication of existing good practice will depend on each organisation being clear about what benefits they are hoping for, and on their willingness to own the solution.

4.12   Certain organisations will be recognised as beacons of effective data management. They will be well placed to influence others, either in their own "supply chain" of data, or simply by example.

4.13   Policies must be implemented, and if necessary, enforced. Organisations should consider how they enforce the provisions of their data policy and what sanctions they are prepared to impose for non-compliance.

4.14   It is unrealistic to expect economic pressures to weaken. Assessing the financial imperative to make a return on assets against societal expectations of greater public good arising from making those assets more freely available is not likely to be an easy or attractive process. However, organisations should be prepared to do this assessment as part of arriving at a sustainable, and above all, transparent data policy, which they can present to the public.

5.   Conclusions….and what are the next steps?

5.1   This paper summarises the legislative landscape relating to data sharing, pointing out the dichotomy between those measures which encourage sharing (at marginal or no cost) and those which impose income generating targets on the data provider.

5.2   There is a suite of initiatives on sectoral, national, European and international scales which will help with some of the aspects of sharing and reusing data.

5.3   The paper also outlines the issues that still exist, and will continue to do so even if we comply fully with all the forthcoming EU obligations.

5.4   A vision of what success in data sharing might look like is proposed. In order to reach the vision there are a number of actions at all levels of the organisation ranging from technical to cultural, financial and political

5.5   The UK –EOF would like the think tank to discuss:

    i.   If we should seek to construct an "ideal" data policy
    ii.   If the scope of this work should cover all environmental data or be restricted to environmental observation data (including baseline surveys).
    iii.   Whether we should aim for minimum compliance by 2019 as required under INSPIRE – or whether we should be more proactive - such as 90% of data should be held under a transparent and effective data policy by 2013?
    iv.   if the environmental observation community (facilitated by UK-EOF) should seek clarification on where Government / Treasury intends the Trading Fund/Wider Markets Initiative model to apply, and where the "public good" model is preferred. How should charging policies be regulated? Are these issues better left to each funder of observations?
    v.   if we should seek to challenge and/or help enforce existing policies and best practice
    vi.   if we should be proactive in the international and European areas, setting data standards such that the UK is seen as leading the way?
    vii.   If there should be a champion for this work at senior level – i.e. we establish an annual reporting process to show progress towards any agreed vision

    And to give guidance on:

    viii.   which organisations should be asked to take forward the actions identified in the discussion. Should this work be facilitated by UK-EOF? Should individual organisations lead on each action?

Annex 1 DEFINITIONS

These definitions are not intended as the last word in characterising a subject which is changing rapidly, nor is this list exhaustive. However, the main concepts may benefit from some explanation, particularly where some parts of the community use them differently.

Data

Data as a word is often used generically and imprecisely. In environmental observation, it tends to include (at a minimum) digital observations, scientific monitoring, data from sensors, and qualitative or observed behavioural data. Data may be held in either analogue or digital form and be stored either on paper or a variety of computer-compatible media.

Basic or "raw" data is fundamentally a set of measurements (though some organisations will include a checking or validation step). This raw data can be analysed or manipulated to form "derived" or "value added" data. For example, The Office of Fair Trading distinguishes between "unrefined" and "refined" information; the marine community use a four stage model (raw data – value added data – information – knowledge) to describe the process of using data to its full value; the UKHO acknowledge seven steps to describe their value-adding process – from "raw survey" to "hydrographic user application".

Data access

The processes and characteristics of retrieving and distribution of data. Terms commonly associated with data access are:
'Free and unrestricted' means non-discriminatory and without charge.
'Without charge', means at no more than the cost of production and delivery, without charge for the data and products themselves. (Not "free" in everyday sense of the word.)
'Non-commercial' means not conducted for profit, cost-recovery or re-sale.
'Product' means a value-added enhancement of data applied to a particular application

Data architecture

This can mean different things to different people. It could refer to
- the actual structure of the data products – how they are physically coded up in bits and bytes,
- the way data sets are structured inside a data base (for instance the tables in an oracle data base,
- a hierarchical description of the data according to various categories
   e.g.   environmental - > marine -> biological -> fauna ->flora - >fish

Database

A database is defined – very generally - in IPR legislation as "a collection of independent works, data or other materials which are arranged in a systematic or methodical way and are individually accessible by electronic or other means." This definition would also apply to a catalogue of metadata.

In the scientific community it is commonly used to describe a collection of data gathered in a particular project or for a particular project – it does not imply a long term secure archive where data are properly managed and quality controlled.

Data sharing
This term is often used in the same sense as data access. It does not carry the implication of mutuality (ie each party offers data in return for the other's) but more usually one party may demand payment for access to and use of the data by the other. Also, depending on the policies of the sharing organisations, a set of conditions or protocols is attached to the agreement to share data.

Data Licensing

Where the owner of data allows use of the data for a specified period, under specified conditions, often for payment but also to protect Intellectual Property Rights (IPR). The price of a licence can vary from the marginal cost of making the data available to a significant sum intended to provide a return on the organisation's investment. The UK has many models of licence agreements in place: bilateral, pan-Government Agreement, click use licensing for PSI.

Discovery

The concept of discovery is used in three main senses.
Used simply, discovery tends to mean the process involved in finding whether or not data on a certain topic exists, for example, fluoride in stream water.
Discovery metadata allows the user to answer the question and to find out who provides it and how he/she may get it. (The answers in this example are : yes; British Geological Survey; data sold under licence.)
Knowledge discovery, or data mining, is a more complex process usually defined as 'the extraction of implicit, previously unknown and potentially useful information from data'.

Geographic Information (GI)
Geographic Information (GI) constitutes any information that is referenced to a specific location, defined by a grid reference, postal address or an agricultural field or holding number, for example. All GI is capable of being represented on a map and analysed spatially.

Geographical Information Systems (GIS) are software suites which permit the visualization of data which has spatial attributes, ie they represent the spatial relationship between data.

Information
'Information' - at any rate in the scientific community - means a value-added enhancement or synthesis of data, interpretation of data or conclusions obtained from data. The term is also used generally as in everyday language.
'Information Governance' is used in the sense of a strategic approach to data management, resulting in increased organisational capability and clarity in policy terms.

Interoperability
Interoperability is used in general sense as the ability of computer systems to communicate and work with each other. It has three main aspects:
■ Technical, which is concerned with the technical issues of linking up computer systems, the definition of open interfaces, data formats and protocols;
■ Semantic, the objective of which is ensuring the precise meaning of exchanged information is understandable by any application involved;
■ Organisational, which deals with modelling organisational processes, aligning information architectures with organisational goals, and helping these processes to co-operate.

Intellectual Property Rights (IPR)

Intellectual property rights (IPR), very broadly, are rights granted to creators and owners of works that are the result of human intellectual creativity.

Databases (as defined in the legislation) have protection under the Copyright, Designs and Patents Act 1977 and also as a result of "database rights" based on the Copyright and Rights in Databases Regulations 1997. Copyright protection lasts for 70 years and database right for 15 years, which may in future affect databases containing long time-series data – often significant in environmental monitoring.

The development of interactive technologies such as Web 2.0 and the growth of "data mashing" (a mashup is a website or application that uses content from more than one source to create a completely new service) may blur the position on IPR.

<u>Metadata</u>

Often described as "data about data", it is contextual information which describes the data set (eg, type of data, owner, method of collection etc). A fundamental of effective data management is to have adequate metadata associated with each dataset.

For example, the UK-EOF Environmental Observation Catalogue (what observations the UK undertakes) will contain metadata only. A user will not see or use the underlying data itself, rather that it exists, who owns it, what it costs and if it is available.

<u>Portal</u>
A web-based tool that provides links to data sources, and access to a range of further capabilities depending on the specific portal (e.g. map based searching for data, tools to identify, select, and view data,)

<u>Spatial</u>
 An attribute implying connection to a fixed location (and possibly date / time).

ANNEX 2 BACKGROUND MATERIAL

This briefing paper has not been formally referenced, but the following main documents have been consulted in its preparation:

Data Sharing Review; Thomas and Walport; July 2008.
Confines analysis to personal data but identifies many common organisational issues

Models of Public Sector Information provision via Trading Funds; Newbery, Bentley and Pollock, February 2008.
The "Cambridge Study" – concentrates on six largest TFs, recommends moving to marginal cost charging for unrefined data and improving governance

Terms of Reference, Trading Fund Assessment, BERR/HM Treasury, July 2008.
Aims to clarify the business model for each TF and report in detail on six of the largest TFs, including OS, UKHO, Met Office

The UK Implementation of the European Directive on the Reuse of Public Sector Information (the first two years); Office of Public Sector Information (OPSI); July 2007.

UK Government response to European Commission on review of Directive; OPSI; July 2008.
Notes progress; detailed assessment of benefits needed if universities, research bodies etc were to be brought into scope of Directive

The United Kingdom Report on the Reuse of Public Sector Information 2008 (Unlocking PSI potential); July 2008.
Reviews position on PSI, legal and governance issues, agrees the need for increased clarity, notes the probable impact of new technology eg Web 2.0 "data mashing"

The impact and success of legislation on the reuse of PSI; UK Advisory Panel on Public Sector Information (APPSI); July 2008
APPSI is an independent advisory body to Government, with members drawn from public and private organisations. Recommend making PSI reuse mandatory, greater clarity over public task, generally favour a marginal cost approach

House of Commons Communities and Local Government Committee; Ordnance Survey - Fifth report (no 268) Session 2007/8; January 2008
Report of Committee and written evidence submitted to Committee

The Power of Information Report; Steinberg and Mayo, June 2007.
Analysis encourages reuse of information across society

Commercial Use of Public Information, Office of Fair Trading, December 2006.
The " CUPI report" – the first main report to highlight the dichotomy between the TFs/Wider Markets Initiative aim to generate income and the ROPSI/FOI/EIRs aim to make PSI openly available

The Wider Markets Initiative, National Audit Office, January 2006
Reviews the working of the WMI since 1998 and recommends re-energising it

To Share or Not to Share: Publishing Research Data; Research Information Network; June 2008
Although aimed at research community, their definition of data includes long time series observations so conclusions are relevant to UK-EOF; identifies common organisational problems

Annex 3 Acronyms and abbreviations

BGS	British Geological Survey
CUPI	Commercial Use of Public Information (Office of Fair Trading)
DPA	Data Protection Act 1998
EA	Environment Agency
EIRs	Environmental Information Regulations 2004
ESA	European Space Agency
ERFF	Environment Research Funders' Forum
FoI	Freedom of Information
GEOSS	Global Earth Observation System of Systems
GI	Geographic Information
GMES	Global Monitoring for Environment and Security
HEFCE	Higher Education Funding Council for England
INSPIRE	Infrastructure for Spatial Information in the European Community
IPR	Intellectual Property Rights
MEDIN	Marine Environmental Data and Information Network
MOD	Ministry of Defence
NBN	National Biodiversity Network
NE	Natural England
NERC	Natural Environment Research Council
OPSI	Office of Public Sector Information
OS	Ordnance Survey
POIR	Power of Information Review (Steinberg and Mayo)
PSI	Public Sector Information
ROPSI	Reuse of Public Sector Information Regulations (2005)
RSPB	Royal Society for the Protection of Birds
SAHFOS	Sir Alistair Hardy Foundation for Ocean Science
SEIS	Shared Environmental Information System
SPIRE	Shared Spatial Information Services
UKDMOS	UK Directory of Marine Observing Systems
UK-EOF	United Kingdom Environmental Observation Framework
UKHO	United Kingdom Hydrographic Office
WMI	Wider Markets Initiative