

# NERC SCIENCE INFORMATION STRATEGY

<b>1. Introduction .....</b>	<b>2</b>
<b>2. Background .....</b>	<b>5</b>
<b>3. Strategic Assumptions and Actions.....</b>	<b>7</b>
<b>4. Implementation Plan .....</b>	<b>18</b>
<b>5. Data Centre Funding and Governance.....</b>	<b>29</b>
<i>Appendix A: Implementation Plan &amp; Key Dependencies.....</i>	<i>33</i>
<i>Appendix B: Summary of Communication &amp; Change Management Activities ...</i>	<i>34</i>
<i>Appendix C: Definition of Terms.....</i>	<i>36</i>

*The NERC Science Information Strategy was approved by Council in July 2009. The NERC Executive Board approved the Implementation Plan for the Strategy in February 2010.*

*For further information please contact:  
Mark Thorley, NERC Data Management Coordinator  
[mrt@nerc.ac.uk](mailto:mrt@nerc.ac.uk)*

# 1. Introduction

## VISION

*NERC will work with its communities to deliver the data and information services in support of Next Generation Science for Planet Earth, enabling a step-change in science delivery and appropriate support for wider stakeholders. This will require NERC to work with and provide leadership to other suppliers of environmental information at both a national and international level.*

- 1.1. Information is pervasive in the NERC community and fundamental to the advancement of science. The full achievement of the strategic goals, outlined in *Next Generation Science for Planet Earth*, the NERC strategy for 2007 to 2012, will not be possible without continuing development of NERC's Information Management policies and methods. This Science Information Strategy sets the context within which all of NERC's scientific data and information activities, especially those of its network of data centres, will be carried out.
- 1.2. Efficient and effective data and information management is essential to the delivery of Next Generation Science for Planet Earth. This not only requires greater access to data, for novel reuse and repurposing by scientists, but also underpins significant National Good activities, for example, the delivery to UK plc and significant elements of the science in society agenda. While it is generally agreed that NERC already has examples of world class data management within its six data centres there is a desire from within and without NERC for the data agenda to move on significantly further. A significant driver for NERC to take on a greater leadership role is new European legislation concerning environmental and geo-spatial data. Any response to these challenges must see NERC being internally more harmonised, and externally unified in communication.

### **Why does the NERC community collect scientific data?**

- 1.3. NERC collects data primarily to support research investigations, in the short, medium and long-term. NERC data collection also supports wider National Good requirements (often supported via external income).

### **Why does NERC invest in information management?**

- 1.4. Many of the data collected by NERC provide a unique and irreplaceable record of the environment. Measurements accumulated over decades and historical records dating back beyond the 19th century provide a valuable resource to support research, survey and monitoring activities; and for users in academia, government, the public sector, industry and commerce. These data are of high importance and utility in understanding and predicting the behaviour of the earth system and man's impact on it.
- 1.5. The scientific method demands that data are available to support the scientific record and the repeatability of research. Data underpinning research publications are a key part of the scientific record.

- 1.6. Systematic survey of a large part of the environment can form an efficient basis upon which multiple responsive or directed science projects can be based – which otherwise may have to repeat data collection, or would prove financially unattractive.
- 1.7. NERC will only curate data that will be potentially valuable for reuse or repurposing (on their own or when further combined), or of significant evidential benefit in the scientific record or elsewhere. The data once collected can be exploited by a range of stakeholders, primarily scientists, but also policy makers, industry, charities and the general public. The extent of this reuse determines and demonstrates the value of NERC's data assets. Other organisations also collect environmental data for a variety of reasons, including the support of regulation. There are no clear boundaries between NERC's reasons to curate environmental data and theirs.

#### **EFFICIENCY THROUGH GOOD INFORMATION MANAGEMENT**

An internal audit undertaken by Shell International in 2002 revealed that their exploration geoscientists working in 'new frontiers' areas spent as much as 53% of their time finding the relevant data, 24% of their time archiving and documenting the data, and 23% of their time interpreting and adding-value to those data. Shell responded by setting targets aimed at increasing time spent on interpretation and adding-value to 46% by reducing time spent on finding data to 30%.

Source: Alan Dobson, Shell International, The Hague, Netherlands.

#### **What is the purpose of this strategy?**

- 1.8. The strategy outlines the policies, methods of working and technical infrastructure that NERC will need to implement in order to deliver Next Generation Science for Planet Earth (NGSPE).
- 1.9. It sets the strategic objectives and priorities for the management of science data and information within NERC in support of the implementation of NGSPE, including NERC's 'national good' activities. It also aims to guide the NERC science information management (IM) community as to how to increase the efficiency and effectiveness of their activities.

#### **What subject areas does it cover?**

- 1.10. This strategy covers all aspects of scientific data and information management; however, it initially expresses these in relation to digital data. In the final sections, it addresses how the management of analogue data and physical samples, and the work of libraries and archives integrates with the principles and practices for digital data.

#### **What time period?**

- 1.11. This strategy looks ahead some five years, but due to the rapidly evolving information management environment, will need to be refreshed in two years.

#### **Funding and Governance**

- 1.12. The NERC Executive Board (NEB) considers that a sector based governance model is the most appropriate structure to drive forward change within the science information domain. NEB perceives benefits from the vertical integration of data management within the NERC research and collaborative centres; *i.e.* data collection, data curation and community wide data modeling are focused within the research and collaborative centres.

1.13. The Science Information Strategy is thus written from this sector based governance assumption and broadly from the position of constant real terms funding (used more effectively) but it lays out the actions necessary to better define the user-driven vision we should aspire to, regardless of funding level, and methods to better understand our current position and how to improve it. Thus helping to support Council’s aim to decrease National Capability significantly as a proportion of total spend over the next 5-10 years.

## Implementation

1.14. The strategy identifies a large number of actions to address over time. The initial drive will be on the following:

- a. Understanding and prioritising aspirational customer demand: Review by consultation with current and potential customer groups, as to how NERC could improve service levels;
- b. Creating ‘one voice’ within NERC and working with other suppliers to meet demand: Responding to the consultation above, new legislation (INSPIRE and the UK Location Strategy) and the need for clearer UK leadership in environmental data management, by developing a clear single NERC ‘voice’ and combined technology infrastructure;
- c. Driving efficiency and best practice in data centres and the wider community: Through, for example, refinement of the data ingestion policy and assessment process, to ensure that data management is planned early in the research process and that data centre effort is targeted to add value at the outset of grants rather than to clear up any issue at the end.

1.15. Underpinning these overtly positive actions are a number of foundational actions, where NERC will better understand and define the activities, costs and role of its data centres.

### DEFINITIONS

The terms data, information and knowledge have specific meanings in the field of information science. So as to bring clarity to this strategy these terms will be used, as far as possible, in line with the following definitions:

**Data** are individual items or records (numeric or other) usually obtained by measurement, simulation or observation of the natural world and human impacts upon it;

**Information** is created by interpreting and/or processing data so that their context and relationships are expressed and understood. In contrast to data, information has value in decision-making;

**Knowledge** emerges from the expert analysis and synthesis of multiple sources of information, and is used as a basis for decision-making.

It is recognised that data, information and knowledge are a continuum, and in some instances their boundaries are fuzzy. **Information Management** spans this continuum.

The strategy makes reference to physical **Samples**. These are samples, specimens and materials collected during NERC’s research, survey and monitoring activities or deposited with NERC by third parties (sometimes under statute). The term **Data-set** is used commonly to describe aggregations of both analogue/digital data and analogue/digital information.

## 2. Background

### NERC – AN INTERNATIONAL EXEMPLAR

NERC'S data centres and information management programmes have a highly successful track record in acquiring, managing and disseminating science data and information generated by publicly and privately-funded research and other investigative activities. NERC information management activities are well respected both nationally and internationally.

At its last SMA review (2003) BGS was described as a world leader in the field of geoscience information management and delivery. The National Geoscience Data Centre is a key element in BGS's information management activities. British Atmospheric Data Centre was rated alpha-5 at its last SMA. The British Oceanographic Data Centre was reviewed in 2006 by an international group of data management experts receiving an overall alpha-4(high). There was clear recognition that BODC provides a world-class service when benchmarked against similar activities in other countries, and is probably best in world in terms of working with the community for the collection and quality control of data.

Many of the NERC Data Centres have international links and collaborations, for example, BODC provides the chairs for expert groups under the banner of UNESCO's International Oceanographic Commission.

One of the reasons for NERC's international reputation in this area is that its data centres do not just provide data management services. NERC is world-leading because its data centres also undertake informatics research and development to support future operational requirements, for example, the development of vocabulary services for data set discovery, or mark-up languages used to provide machine readable descriptions of data sets.

### Current Position

- 2.1. NERC has an established Data Policy (see <http://data.nerc.ac.uk>) which was introduced in 1992. The policy outlines the roles and responsibilities of the key players in data management (including NERC funded researchers and the NERC data centres) and how NERC will licence and charge for data. One of the key principles within the policy is that recipients of NERC grants must offer to deposit with NERC a copy of datasets resulting from the research supported, for use by other researchers, but without prejudice to intellectual property rights.
- 2.2. Other areas of NERC's scientific information management activities (libraries, archives and sample curation) are guided by best practice (including codes of practice) and established procedures.
- 2.3. NERC currently supports six data centres. These are the:
  - British Atmospheric Data Centre (BADC);
  - British Oceanographic Data Centre (BODC);
  - Environmental Information Data Centre (EIDC);
  - NERC Earth Observation Data Centre (NEODC);
  - National Geoscience Data Centre (NGDC);
  - Polar Data Centre (PDC).
- 2.4. NERC also has an arrangement with the AHRC (Arts and Humanities Research Council) supported Archaeology Data Service at the University of York to manage data resulting from NERC funded activities in science based archaeology. In addition, during 2010 NERC will be taking on responsibility for the 'Earth oriented' element of the UK Solar System Data Centre. This is as part of the transfer of support for the 'Earth oriented' elements of solar terrestrial physics from STFC to NERC.

- 2.5. NERC runs a Data Discovery Service providing an over view of NERC’s data centre holdings through an integrated, searchable catalogue (see <http://ndg.nerc.ac.uk/discovery>). It can be used to find information on what data the NERC data centres hold and how to access these data.
- 2.6. In total the NERC data centres manage some 2600 digital data sets, with a total data volume approaching 600 TBytes. The headline figure for the number of digital data sets does not reflect the scale and complexity of these data holdings. The number of data sets managed depends very much how you define a data set within a database. For example, the Biological Records Centre contains 20 million records of 1200 species occurrence across 1500 UK sites and is defined as one ‘data set’.
- 2.7. In addition to digital data, some data centres have large holdings of analogue materials, primarily NGDC with 20M geological samples and paper records and archives amounting to 8km of shelving.
- 2.8. A summary of the data centres, including their National Capability sector area of responsibility, is presented at Table 1.

	<b>British Atmospheric Data Centre</b>	<b>British Oceanographic Data Centre</b>	<b>Environmental Information Data Centre</b>	<b>NERC Earth Observation Data Centre</b>	<b>National Geoscience Data Centre</b>	<b>Polar Data Centre</b>
<b>NERC National Capability Sector</b> (primary responsibility)	Atmospheric Sciences	Marine Sciences	Terrestrial & Freshwater Sciences	Earth Observation	Earth Sciences	Polar Sciences (working with BODC for marine and BADC for Atmospheric)
<b>Host Centre</b>	NCAS. Run under contract by STFC at Rutherford Appleton Lab.	Hosted by NOC, though responsive to all the NERC marine centres.	CEH	NCEO. Run under contract by STFC at Rutherford Appleton Lab	BGS	BAS
<b>TByte of digital data held</b>	201	7	28	189	45	110
<b>Physical samples and hard copy data</b>		Archive of tide gauge records			Samples: 20M Archives & records: 8km	

**Table 1:** Summary of NC sectors, host centres, and information holdings of the NERC Data Centres.

### Usage of Data Centres and Data Holdings

- 2.9. Being able to understand what data sets have been used for, especially the research papers arising, is an issue for all data centres and not just those run by NERC. This is especially true where data sets are directly accessed via the web, and data centres do not develop an interactive relationship with the user.
- 2.10. The NERC data centres require users to report back to the data centre on publications arising from NERC data sets, however, users do not often meet this requirement. Tracking of usage, through data set citation, is an area of active research, to which the NERC data centres are contributing. The Digital Curation Centre and Research Information Network are also working to develop metrics for measuring data centre performance and tracking of data set usage. There are activities within the strategy focusing on data set citation and publishing, and the development of better usage metrics.

### 3. Strategic Assumptions and Actions

#### NERC IS AN INFORMATION ORGANISATION

NERC is an information and knowledge driven organisation. Over time programmes change, people change, skill requirements change and funding changes. However, NERC's data and information assets remain.

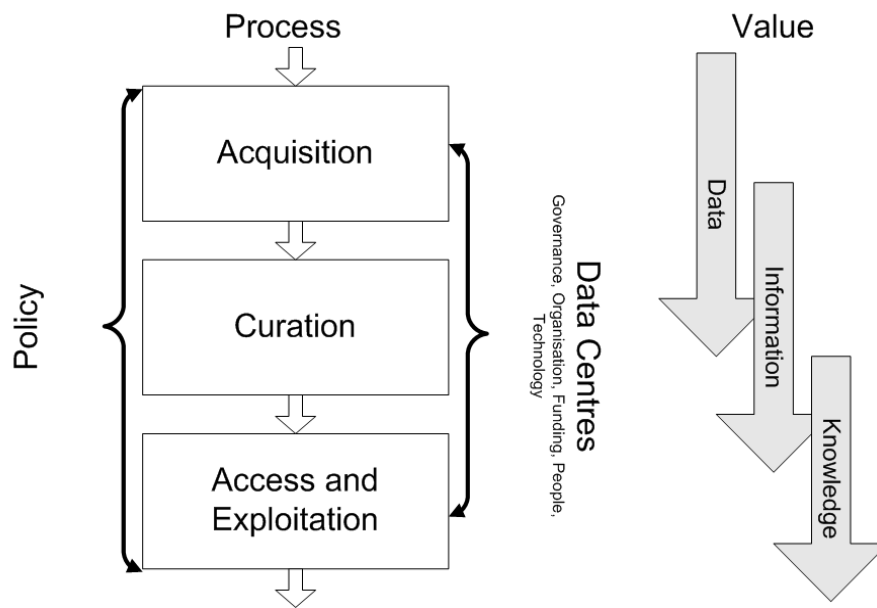
#### Strategic Assumptions

3.1. The activities outlined in this strategy are predicated on the following assumptions for the next five year period:

- NERC will continue to undertake long-term environmental data management through a network of data centres;
- The structure of the NERC Data Centre network will remain aligned to NERC's defined National Capability sectors;
- NERC will continue to fund the data management activities of its data centres and provide funds for project and programme data management activities for research grants;
- New instrumentation, technologies and models means that volumes of data will continue to grow. Volume alone is not a problem, though it comes with a cost. However, heterogeneity of data and a need to support more interdisciplinary working will be major drivers of data centre activities;
- There are currently data fundamental to NERC's mission that are not under the controlled management of the NERC data centres;
- Strategic delivery partners in data management will continue to exist. Specifically, the Archaeology Data Service at the University of York, The Economic and Social Data Service at the University of Essex and the JISC supported EDINA-Digimap facility at the University of Edinburgh;
- The Research Information Network will continue in its key role of developing the evidence base on researcher behaviour in relation to the management and sharing of research data. The Digital Curation Centre will continue in its role in providing both advice on research data management and as a forum to bring together the research data management community;
- The British Library will continue to play a major role in the management and dissemination of research information, and will move to clarify and understand the role of libraries in the management and dissemination of research data;
- Institutional repositories within the HEI sector will continue to develop and will have an increasing role in the dissemination of research information;
- User expectations of how they can access and use public sector information will continue to develop, and will have the potential to challenge how NERC makes its scientific information available. This reflects the growing 'Power of Information' agenda<sup>1</sup>;
- Current legislative drivers, including European and UK policy on charging and access to public sector information, will remain largely unchanged. Government will continue to review the charging and licensing policies and guidelines for public sector information, with growing expectations of more innovative re-use and exploitation licensing regimes.

## Change Drivers

3.2. The information management activities that NERC undertakes will need to respond to internal developments in strategy and operations, as well as external developments in legislation and other drivers. These change drivers will evolve over time, but key current influences are listed in Table 2. The sections that follow describe how NERC will adapt its information management activities in response to these drivers. The recommendations follow the flow of the information management activities value chain shown in Figure 1.



**Figure 1:** The information management value chain

Driver	What this means for Information Management within NERC
NGSPE Science activities	<p>Larger scale, more integrated and cross-disciplinary science activities require new approaches to IM. Thus the actions within the Information Strategy focus on:</p> <ul style="list-style-type: none"> <li>• Supporting researchers in the location and evaluation of the data to maximise its re-use and re-purpose;</li> <li>• Providing better facilities to manage and interpret the increased data heterogeneity and volume;</li> <li>• Equipping information scientists and scientists with key skills to work with the large scale data that the proposed science requires;</li> <li>• Providing the informatics tools and techniques to support the development of integrated research communities;</li> <li>• Working with discipline-based and cross-disciplinary bodies, both nationally and internationally, to influence the development and adoption of information science and technology standards.</li> </ul>
Knowledge Exchange and Science in Society and User Behaviours	<p>The Government requires NERC to demonstrate a step-change in economic impact, and needs more extensive engagement on science issues with society at large. User communities' information access expectations are changing, as are their reasons for access. Therefore the actions within the Information Strategy focus on:</p> <ul style="list-style-type: none"> <li>• Flexibility and clarity of licencing, allowing 3<sup>rd</sup> parties to develop more innovative, user focused applications;</li> <li>• Clarifying the data charging approach and removing of unnecessary barriers to access;</li> <li>• Developing online solutions, including portals, to enable search and download across the spectrum of NERC's data and information assets;</li> <li>• Improving engagement with non-research users of NERC data;</li> <li>• Supporting the publication and citation of data sets to reward researcher effort expended on good data management.</li> </ul>
Legislation and Government Policy	<p>UK and EU legislation and policy are increasingly influencing NERC's IM activities. Therefore actions within the Information Strategy focus on:</p> <ul style="list-style-type: none"> <li>• Implementation of the Inspire Directive, and influencing the development of a UK Spatial Data Infrastructure;</li> <li>• Implementation of the UK Location Strategy and influencing the UK Location Council;</li> <li>• Developing opportunities to sell environmental information management facilities to other public sector organisations covered by Inspire;</li> <li>• Continuing to prioritise the compliance with Freedom of Information legislation, Environmental Information Regulations, Data Protection and Information Security.</li> </ul>
Technology	<p>Rapid technological developments are providing new technology for measurement, modelling, managing information and disseminating it. Therefore the actions within the Information Strategy focus on:</p> <ul style="list-style-type: none"> <li>• Being able to deal with the increasing volumes and complexity of data arising from HPC, new instruments and monitoring systems;</li> <li>• Developing a range of innovative web-services and portals which will help support increasingly sophisticated user requirements for data discovery, visualisation and cross-linking;</li> </ul>
Efficiency and Effectiveness	<p>National Capability cost pressures will require greater IM efficiency and effectiveness. Therefore the actions within the Information Strategy focus on:</p> <ul style="list-style-type: none"> <li>• Closer working between all NERC Data Centres and greater sharing of skills and resources;</li> <li>• Aligning Data Centre governance and funding, implementing common performance measurement where appropriate;</li> <li>• Identifying the data NERC requires, raising the quality of that it acquires and discarding what it doesn't.</li> </ul>

**Table 2:** Key drivers of change for NERC's Information Management Activities

## **Strategic Actions**

### **Clarifying policy and role of Information Management in NERC**

3.3. NERC must ensure relevant stakeholders are aware of why NERC does IM, the activities it undertakes, the value that these add and the obligations that they bring.

3.4. To achieve this NERC will:

- With input from key stakeholders, refine the NERC data policy and its supporting guidance to clearly explain the roles, rights and responsibilities of all stakeholders, including creators, managers and users of data, the data centres and NERC's own science managers<sup>ii</sup>, and develop mechanisms for assessing compliance with the NERC Data Policy;
- Actively engage with creators and users of data to ensure they understand their rights and responsibilities in relation to the NERC Data Policy.

3.5. Implementation implications:

- Straightforward;
- Cost: Small consultation costs, no change ongoing.

### **Clarifying Governance, Organisation, Funding and People**

3.6. NERC must ensure that its Data Centre network effectively serves the widest UK community. Data centre activities must reflect the priorities placed on it by NGSPE and outlined elsewhere within this strategy.

3.7. Maintaining and enhancing the skills of NERC's information management staff is central to delivery of the activities outlined within this strategy. Successful Information Management is also heavily influenced by the skills of all those active within the research process, who have a responsibility to ensure the data they collect is effectively managed from the time of collection onwards.

3.8. In order to address the needs of the strategy, communities and people NERC will:

- Produce a clear definition of the role of a NERC data centre, covering all aspects of a data centre's work. This definition will also cover the boundaries between a host research centre and the data centre, ensuring appropriate focus on all data providers and users;
- Define a clear and robust funding model for information management activities (across RP, RM & NC) which absorbs the varying short term demand on data centres as NERC priorities shift in the context of longer term requirements. The funding model will also reflect the differing costs of information management between disciplines at the same time as recognising the proportional spend between NERC disciplines, and being responsive to changing long-term requirements;
- Establish a data centre governance structure which provides leadership and clear accountability for activities and resources. Each data centre will report to the relevant NEB member representing the national capability sectors. There will be clearly defined boundaries between the research centres and the data centres, especially the information management programmes of the host centres;
- Ensure research centre information management activities are judged by the same criteria as data management of NERC funded researchers in the HEI community;

- Ensure the data centres can deliver better value for money through active sharing of expertise and resources in both operational and research activities;
- Develop a fuller understanding of the range of informatics (information science, processing and technology engineering) skills needed to deliver successful Information Management and the skills gaps that currently exist;
- Produce proposals as to how the informatics skills base is developed and the gaps filled. This will address the skills needed both within data centres and research teams, and also consider issues around the recruitment and retention of informatics personnel within NERC and in research teams within the HEI sector;
- Continue with underlying investment in Informatics Research to ensure future requirements can be met effectively.

### 3.9. Implementation implications:

- Challenging – developing proposals is straightforward. Implementation will be more challenging;
- Cost: small cost in consultation, though there is likely to be a longer-term requirement for sustained investment in development of the informatics skills amongst NERC Data Centre staff and, through education, of the wider research base.

## **Improving the Efficiency and Effectiveness of Data Acquisition**

3.10. Currently some data are not offered to NERC that we would wish to collect and others are offered that we do not. Other data are collected and offered to NERC in a suboptimal format. These practices reduce the effectiveness of NERC's data holdings and cause inefficiency to the projects / programmes and data centres.

3.11. Activities to quality assure and document data sets, so that they are in a form suitable for others to re-use them, are not always seen as 'productive' science. Rather they are undertaken as a 'moral' obligation, for which there is little direct reward or recognition.

3.12. There also appears to be a lack of understanding in the community on issues surrounding ownership and rights to exploit data, so that some data creators are unwilling to offer copies of the data they create to NERC as they are concerned that they will lose all the rights to use these data.

3.13. NERC also acquires data from third parties to facilitate the efficient delivery of the science programme, and because NERC has a recognised national capability for environmental data curation. There are only limited criteria to guide data centres in this acquisition process.

3.14. To improve the efficiency and effectiveness of its acquisition of data of long-term value NERC will:

- In consultation with key stakeholders develop criteria to identify data of long-term value to NERC and the standards and formats which will maximise the opportunities for efficient re-use and re-purposing. This will be in a form that can be understood from the proposal stage of any project;
- Put the onus on researchers to use these criteria to identify which data they collect are of long-term value to NERC, and in these circumstances, write appropriate data management plans in proposals. This would be facilitated by data centre support depending on the size and nature of the proposal (described in guidance documents);

- Enhance the peer-review and moderation processes to include adjudication upon the need for and quality of data management plans where a proposal meets the NERC criteria. The adequacy of the data management plan would form one of the assessment criteria used;
- On appropriate grants develop a process to enable data centres to engage with and support funded researchers to ensure efficient data management from the outset;
- Enforce data submission in projects meeting the long-term data criteria, by making delivery of the data and necessary supporting documentation (in a defined format) a condition of receiving the final settlement payment;
- Produce clear guidance on issues of rights and ownership of data, and explain how NERC supports a right of first use of data by PIs and research teams through use of reasonable embargo periods;
- Actively work to develop mechanisms to allow the peer review and publication of data sets through data ‘journals’, enabling the tracking of usage of data sets through citation. It is hoped this will result in peer recognition and reward of the efforts put in to data management activities by research teams and others. This will be done in conjunction with the learned societies and scholarly publishers, and using the data centres as the recognised places of deposit of the published data sets;
- Continue to identify third party data which are either of significant benefit to the science programme, or which fits within a national curation remit, and provide appropriate archival services (appropriate to the business model sustaining the acquisition and/or the scientific goals of acquiring the data);
- Ensure third party data holdings have license conditions which enable NERC to exploit them as required, regardless of how these third-party data are acquired;
- Implementation of the INSPIRE directive will require non-specialist organisations to manage and disseminate environmental information. NERC could offer, at cost, an expert service to these organisations.

### 3.15. Implementation implications:

- Implementation: straightforward, but will require a programme of active communication and engagement with stakeholders, and a change to peer review and moderating panel processes;
- Cost: Neutral to the data centres, with proactive data centre support replacing current management of poorly conditioned or ultimately unwanted data sets.

## **Ensuring data curation standards are consistent, fit for purpose and represent good value for money**

3.16. Variations in the type of data and materials retained by the data centres, and different ways of working, has resulted in differences in approach and quality of information management activities. Currently NERC has many areas of world class curation, however we need to ensure better consistency of activities and standards across our data centre network.

3.17. To achieve quality, consistency and value for money, NERC will:

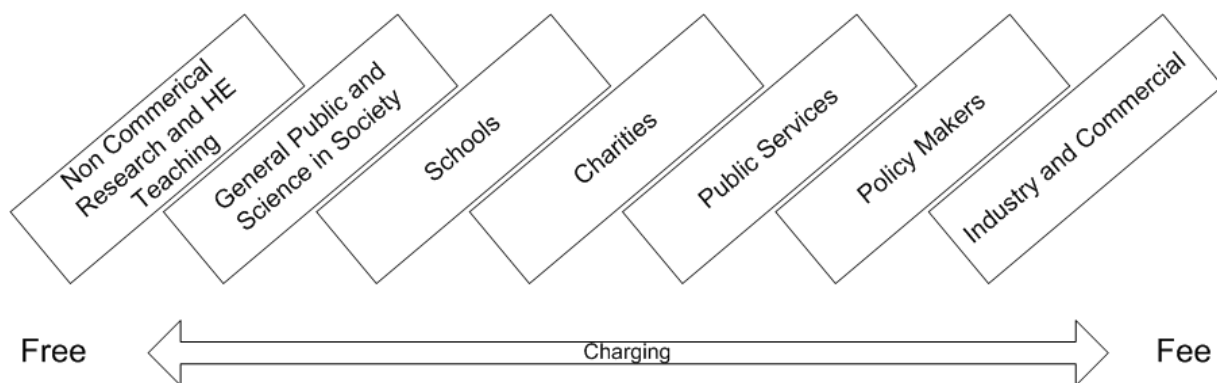
- Understand and challenge the differences in approach to information management between the data centres;
- Use this understanding to ensure a closer working relationship between the data centres, and a greater sharing of skills and resources. For example, in the areas of data conditioning and the development of metadata to support discovery and re-use;
- Develop and promulgate clear rules for retention and disposal of data.

3.18. Implementation implications:

- There are several difficulties here. The costs of different data centres vary significantly, as does the level and nature of service. The necessary analysis, discussion and agreement would probably take a year. Given that acquisition of data is the main cost driver rules on retention will be tackled later;
- Cost: Small cost in consultation.

### Improving Access and Exploitation

3.19. NERC is occasionally criticised for having a heterogeneous approach to charging for data and there is no clear policy to guide behaviour or investment levels to different extant and potential user communities (see Figure 2). A key principle underpinning data exploitation is that the NERC data centres act as gateways to data and information products, however, they will not normally develop value-added information products. This is a role for NERC's research centres and others, applying their knowledge and expertise to data from the data centres. Note: issues relating to technology and access are addressed in the next section.



**Figure 2:** User communities for NERC information

3.20. To improve access, whilst demonstrating clarity and fairness of approach, NERC will:

- Understand its current level of investment in exploitation for each user group. Form a range of objectives to better serve the needs of the identified user segments and associated costs of execution, and present these to Council by the end of 2009/10;
- Supply NERC data free of charge for teaching and non-commercial research purposes<sup>iii</sup>, except for 'large' requests (to be clearly defined) where NERC may recover the marginal costs of supply;

- Demonstrate transparency, fairness and clarity for access and charging, by ensuring that the relevant NERC research centres follow the lead set by BGS in gaining the Office of Public Sector Information's 'Information Fair Trader Scheme' accreditation. In so doing NERC will demonstrate its adherence to UK government policy and guidelines on charging for data and information as defined by HM Treasury;
- Co-ordinate the activities of NERC Data Centre help desks to ensure consistency of advice and service levels;
- Work to make selected data holdings available for innovative methods of access and linking with other data by third parties<sup>iv</sup> and provide 'innovation' licences to encourage re-use and development of its data holdings;
- Provide a generic, portal based data discovery, view and download service which meets NERC's INSPIRE obligations;
- Implement technologies upon which others will be able to develop a range of data access portals. NERC will encourage user communities to develop portals targeted at their specific communities, by providing a funding mechanism to co-fund community portal developments. For certain non-technical communities Knowledge Exchange or Science and Society funding will be used to provide technical input driven by user requirements;
- Consider working with other key user sectors, for example, Policy Makers and regulators, to develop portals targeted at their specific requirements. These portals may well link NERC data with data from other providers;
- Provide alternative channels of access to NERC data through working with 3<sup>rd</sup> party delivery agents, such as JISC or other non-NERC data centres;
- Wherever possible, provide leadership within the ERFF Environmental Observation Framework initiative and the UK Location Strategy.

#### 3.21. Implementation implications:

- Challenging – getting agreement on prioritisation will be difficult. All other actions are achievable;
- Costs of technology and data conditioning are heavily dependent on the outcome of the prioritisation exercise and the rate of change required. Gradual change can be absorbed within current baselines, step-change will require investment.

### **Developing technology and delivery methods**

3.22. NERC needs to deliver data to users within the NERC community, and to support other user segments. To enable this, an appropriate 'spatial data infrastructure' needs to exist which can expose NERC data via a range of services and portals.

3.23. NERC is already a world leader in data interoperability, and can base its infrastructure developments on the work of major existing programmes (including the NERC DataGrid and GeoSciML).

3.24. It is recognised that technology is both an enabler of information management and delivery as well as a driver of change and that investments in technology need to be balanced between these two outcomes.

3.25. To achieve this NERC will:

- Deploy a NERC Spatial Data Infrastructure which will link major NERC data resources using standards-based protocols which support both legislative requirements (INSPIRE) and community requirements;

- Provide one centrally supported INSPIRE compliant entry point to that infrastructure, and continue to support multiple community specific portals where necessary;
- Continue to develop and deploy innovative solutions to supporting the data interoperability required by evolving science and legislative requirements;
- Resource a small team tasked to engage with national and international activities in spatial data infrastructures and coordinate appropriate NERC service deployments.

3.26. Implementation implications:

- Although these are challenging objectives, it will be important to have a robust operational infrastructure and clear procedures for supporting the alignment of the NERC infrastructure to externally changing infrastructures (and the evolution from research practices into the operational systems). The NERC Technology Theme Action Plan will also need to be responsive to the national capability requirements as they involve;
- Cost: support for small technical coordination and service management teams will be required. Note that the physical cost of the central infrastructure will be marginal<sup>v</sup> and the service deployment costs should fall within existing budgets.

**What does successful information management look like?**

3.27. The success indicators for NERC's information management activities include:

3.28. Data acquisition:

- NERC and its community will understand what data NERC wishes to curate for the long-term and why;
- Clear guidelines will exist to guide a NERC funded PI as to:
  - how to successfully address data management in a bid to NERC;
  - how to manage data through the life of the grant, and;
  - how NERC Data Centres will support them at each stage.

3.29. Data curation:

- Data curation and conditioning practices across NERC's Data Centres will be optimised to enable simple information access and exploitation for stakeholders requiring data from across a range of NERC's Data Centres, without significantly compromising the needs of the stakeholder communities served from only one;
- NERC will identify (or set) the appropriate international standards for information management and demonstrate it meets them.

3.30. Information exploitation and access:

- NERC and its data centres will lead the UK environmental science community in complying fully with the requirements of the INSPIRE directive;
- NERC will provide appropriate routes of access and consistent charging regimes for its range of data holdings to all stakeholders communities (i.e. science, government, regulators, private, public and voluntary sectors);
- NERC will work with other UK and International bodies to provide users with the ability to access location based data across a range of providers.

### 3.31. Organisation and funding:

- NERC Data Centres are clearly established to serve the integrated NERC communities described in Next Generation Science for Planet Earth, and will be funded to a defined service level;
- NERC Data Centres are funded transparently in relation to the NC (including National Good), RP and RM support that they give;
- NERC will be compliant with all legislation, and actively influence future UK and EU policy concerning data management.

### 3.32. Metrics and indicators:

- Appropriate key performance indicators (KPIs) and user feedback mechanisms will be used to demonstrate Data Centre excellence, assess policy compliance, support benchmarking and drive continuous improvement; within financial constraints;
- Appropriate management information will be generated to show the totality of the Data Centre's activities where they are broadly like-for-like, particularly in Data Acquisition and Information Access/Exploitation. These will be combined with meaningful metrics concerning more specialised and localised activities to demonstrate the full impact of the data centres.

### 3.33. Implementation implications:

- Challenging – defining robust KPIs, and performing benchmarking is a non-trivial task;
- Cost – no additional costs.

## **Analogue Data and Physical Samples**

3.34. Though the general principles covering acquisition, curation, access and exploitation described above apply principally to digital data, they also apply to analogue data and physical samples, all-be-it with some practical differences. It should be noted that some of the NERC Data Centres, for example the National Geoscience Data Centre, house unique and very extensive collections of physical samples and other analogue materials.

3.35. The practical differences between the management of digital data and the management of analogue data and physical samples, relate principally to physical storage and access. The infrastructural costs and challenges of housing and providing access to analogue materials are greater than those for purely digital materials.

3.36. The actions proposed for digital data in previous sections will also be applied, as appropriate, to analogue data and physical samples. It is recognised that cost efficiencies might be realised through rationalisation of storage sites across NERC.

3.37. Some of NERC's facilities for sample storage and management, for example, the British Ocean Sediment Core Research Facility, are currently outside the purview of the NERC Information Strategy Group and this strategy. Such relationships will be reviewed to ensure that there are no inconsistencies in the acquisition, curation, access and exploitation of sample and specimen material across all NERC funded facilities.

## Science information held in libraries and archives

- 3.38. Archives services across NERC hold many types of information, ranging from original records and data covering scientific, business and heritage matters to material formally published by the host centre. Information and data are held on a wide range of formats - digital data, analogue data, physical records and objects including paper, photographs, moving film, sound recordings and maps. Given the broad range of information held, the services are targeted at different user groups, who are often only interested in a subset of the entire holdings. NERC archives services are classified as 'Places of Deposit' under the Public Records Act and are managed accordingly, conforming to appropriate national standards.
- 3.39. NERC recognises that effective library services are an essential underpinning for research, curating and managing the supply of high quality published information for scientists. The NERC Library Services Review in 2008 confirmed the value of library services and recommended more effective service delivery mechanisms and on-going evaluation and measurement to ensure that services remain user focused and relevant. The potential for libraries to play a role in initiatives such as open access publishing has also been recognised.
- 3.40. NERC will ensure that archives and library services are effective and efficient through a more joined up approach by:
- Delivering the service modernisation agenda outlined in the Library Services Review to provide high quality services to NERC staff;
  - Ensuring the archives services work together more closely to share best practice and knowledge;
  - Linking archives and library catalogues with data discovery services as appropriate, and conforming to appropriate standards to link to national archives portals;
  - Maximise usage by ensuring that collections and services are exploited more widely.
- 3.41. Implementation implications:
- Implementation - straightforward;
  - Cost - Implementation of the NERC Libraries Restructuring Project and the associated budget will be subject to approval by NEB.

## 4. Implementation Plan

### Scope

4.1. This plan covers the implementation of actions detailed within the NERC Science Information Strategy. It covers the activities of its data centres, the information management programmes of its research centres and the data management activities that NERC requires those it funds within the HEI sector to undertake. The plan has a primary focus on digital data and the technologies for their dissemination via web-based services, however, implementation activities will cover all NERC data, including analogue data, physical specimens and samples.

### Vision

4.2. The Science Information Strategy sets the context within which all of NERC's scientific data and information activities will be carried out. The vision for implementation of the strategy is that NERC will work with its stakeholders to deliver the necessary data and information services required for successfully meeting the objectives outlined within Next Generation Science for Planet Earth. Central to this is a 'single face' and 'single voice' for NERC data management supported by an integrated NERC information infrastructure, providing users with a way of identifying and accessing the data and tools they need. This step-change in delivery and support for wider stakeholders will identify NERC as a leader in the provision of environmental information at both the national and international level.

### Benefits

4.3. Key benefits which arising from the implementation of the strategy are:

- A clear understanding of how science information activities are aligned to the delivery of NGSPE and an improved understanding of the costs and benefits of science information activities in NERC, to provide a framework for better targeting of investments in the future;
- A more user-focused approach, with a priority on developing services and delivering the interoperable data that users need;
- A closer working together and sharing of skills across the data centres, with a focus on sharing best practice, reducing duplication of effort and ensuring better value for money;
- A re-engagement with researchers in the HEI sector, through the 'single voice' of NERC data, providing them with clearer information on which data NERC identifies as having a long-term value and their roles and responsibilities in ensuring the data are available for the long-term;
- A 'single face' of NERC data, providing a single route into NERC's data holdings, for users who are not members of existing data communities;
- A clear plan and schedule for NERC meeting its obligations under the INSPIRE Directive<sup>vi</sup> for the discovery, view and download of environmental data.

## **Implementation**

4.4. Central to all the implementation activities is the core concept of delivering a net benefit to NERC as a whole. The plan recognises that there are real differences in local practice and procedure in data management across NERC's data and research centres. These differences will be retained where it can be established they clearly deliver benefit. The implementation governance structure will provide a mechanism to moderate priority demands to ensure an overall net benefit to NERC.

4.5. The Implementation Plan is structured around:

- What is going to be done – covering implementation activities and outcomes, and communication and change management;
- How this will be achieved – covering the management of the implementation, governance, the resources required for implementation and the assumptions, risks, issues and dependencies which underlie the implementation activities.

## **Implementation Activities and Outcomes**

4.6. The implementation of the strategy will be undertaken through three phases of work, each of approximately one year's duration, though the exact length of the second and third phases will not be known until phase one is substantially completed. After the end of each phase the Information Strategy Group will undertake a formal review, to allow the implementation plan to be refreshed and revised, including any necessary re-scoping of resources. The information management activities that NERC undertakes will need to respond to internal developments in strategy and operations, as well as external developments in legislation and other drivers.

## **Implementation Phase 1 (2010/11)**

4.7. This phase will cover first-order decisions and enabling activities, policy changes, quick wins, base-lining of costs, activities, requirements & roles, and data licencing. It will put in place the foundations from which the programme of change will be built. For example, understanding roles & responsibilities and funding models and challenging centres to explain and justify differences in practice.

4.8. *User perspectives:*

- Key stakeholders will be consulted to improve understanding of their requirements for data (including non-digital data) and systems from NERC, resulting in a clear statement of what users can expect from NERC data centres;
- NERC will start to implement a 'single voice' and 'single face' for data management activities by the provision of a comprehensive and easy to use directory of NERC's data holdings, including information on how to access the data for reuse;
- A clear statement on how and when NERC will charge for access to data and information and the scale of these charges, to be applied consistently across NERC<sup>vii</sup>;
- Guidance to researchers, especially within HEIs, on how NERC will identify data of long-term value and the processes it will follow to ensure these data are ingested into the data centres for reuse and repurposing;
- Active engagement with the community on issues around the citation, peer-review and publication of data sets, with a clear statement of what NERC intends to progress in this important area, and by when it will do this.

#### 4.9. *Structural activities:*

- Agreement on ways to remove duplication and unnecessary difference, for example, by the adoption of consistent standards;
- An agreed NERC information architecture;
- A shared understanding of data centre cost bases and processes, to enable removal of duplication and identify efficiency savings where appropriate;
- NERC data centres will have access to up-to-date and consistent information on all awards made, to enable the identification of data of long-term value being generated through NERC funded activities;
- In order to measure the impact of the strategy a set of common performance metrics for data management activities will be developed, and a baseline for measuring performance will be established;
- Using the criteria for identification of data of long-term value, NERC will identify its key legacy data sets necessary for the successful delivery of NGSPE and to be INSPIRE compliant. These will have been prioritised these for any necessary data conditioning activities;
- INSPIRE data ‘discovery’ compliance.

#### **Implementation Phase 2 (2011/12)**

4.10. This phase will cover the implementation of high priority transformational changes and efficiency savings, along with second-order decisions and activities. It is the substantive building phase, where activities related to systems and data are consolidated to deliver the ‘single voice’ and ‘single face’ of NERC data management.

4.11. Some of the unforeseen consequences of the strategy, as well as clarification and a better understanding of the data centre framework will also start to emerge. The second and third phases will be directly influenced and guided by the essential work undertaken within the first phase.

#### 4.12. *User perspectives:*

- A clear statement of ‘NERC data’ as an identifiable and quality ‘brand’ that can be valued for environmental data and can be used to help support delivery of NGSPE;
- Continuing implementation of the ‘single face’ of NERC data, delivered by the developing NERC Information Infrastructure, providing a common set of services for web access to data;
- An increased understanding of user requirements, developed through more effective and targeted engagement, will be used to add value to key data sets. This will enable them to be used more effectively in support of NGSPE questions, and to ensure that productive and innovative data solutions are engineered to meet evolving requirements;

4.13. Through deployment of a revised NERC Data Policy, stakeholders will become aware of their rights and their responsibilities regarding the creation, management and use of data funded or held by NERC.

#### 4.14. *Structural activities:*

- Delivery of the ‘single voice’ and ‘single face’ of NERC data management will be supported through the closer working together of data centres, with rationalisation and harmonisation of activities where appropriate. NERC’s data holdings will be managed as a corporate asset;
- NERC will consult with stakeholders to identify which non-NERC data are going to be crucial to delivery of NGSPE, and then work with third-party data management partners to secure access to these data, in order to be able to include them in the ‘single face’ of NERC data;
- Based on an understanding of the efficiency savings that can be made within the data centres, and on the resources that are thus available for redeployment, the first stages of transformational change will be undertaken;
- A clear funding model for all data and science information activities within NERC’s data and research centres will be developed. This will help ensure that the finite effort of data centres and NERC’s science information activities are focused onto those groups and communities of most strategic importance to NERC;
- A commitment to informatics research and development will be demonstrated by the data centres through the production of appropriate data science papers and conference posters and an engagement with the wider environmental informatics community;
- Through changes to operations, performance metrics and service level agreements the activities of the data centres will be clearly structured and prioritised to meet NGSPE requirements;
- INSPIRE compliance for ‘newly collected and extensively restructured’ data sets.

#### **Implementation Phase 3 (2012/13)**

- 4.15. It is envisaged that this phase will focus on a targeting of effort to ensure the services being delivered to NERC’s stakeholders are maximised. The cost-benefits to NERC of key information management activities will be available and the value of the ‘single voice’ and ‘single face’ of NERC data to stakeholders will have been established.
- 4.16. This phase will implement a second wave of transformational change and third-order decisions and activities, informed by a review of the impact of the changes in phases one & two. There is an expectation that this phase will also establish processes for the effective management of data arising from Responsive Mode. This phase will work to maximise the services being delivered to NERC’s stakeholders, especially in terms of discovery, view and download of NERC’s data holdings, and will result in full INSPIRE compliance.
- 4.17. It is planned that proposed activities for this phase will be reviewed and assessed as the programme progresses. It is likely that outcomes and resource requirements identified at the start of the implementation process will have been revised before this phase is started.
- 4.18. *User perspectives:*
- Priority data will be conditioned to enable interoperability, and to meet the requirements of INSPIRE;
  - Users will see NERC as a more even and transparent organisation to deal with for the supply of data and information, part of which will have been achieved through CEH & NOC gaining Information Fair Trader accreditation;

- NERC will have clarified its approach and support for Open Access publishing, in light of the increased support of OA by the Research Councils and the changing scholarly communications landscape.

4.19. *Structural activities:*

- To ensure successful delivery of strategy outcomes the NERC data centres will continue to work in close co-operation, jointly setting priorities for activities and ensuring outputs still meet users' key requirements. Infrastructure developments will reflect and support these requirements;
- Using the improved understanding of our stakeholder requirements, NERC will have improved the level and structure of feedback routinely received.

4.20. A detailed list of implementation actions arising from the strategy has been developed. From this a set of implementation projects has been defined, with an estimate of resources for each project. An implementation plan based on this set of projects and showing key dependencies is presented at Appendix A.

### **Communication and Change Management**

4.21. Communication about the Science Information Strategy will be required on several levels; communicating the existence and benefits of the strategy, reporting progress on the actual implementation work and at a more detailed level within each of the significant projects. Parts of the plan, in particular responsibilities, will be dependent upon the final decision on the governance model.

4.22. It is recognised that the implementation of the strategy will require significant change in the operations of the Data Centres and how they interact with their stakeholders. Change management will be considered as part of each of the individual projects within the implementation plan. The NERC Change Management Framework and the network of Change Facilitators will be utilised where appropriate. This is especially important as the programme moves into phase two, which has major change elements within it.

4.23. A summary of planned communication and change management activities is presented at Appendix B.

### **Management of the Implementation**

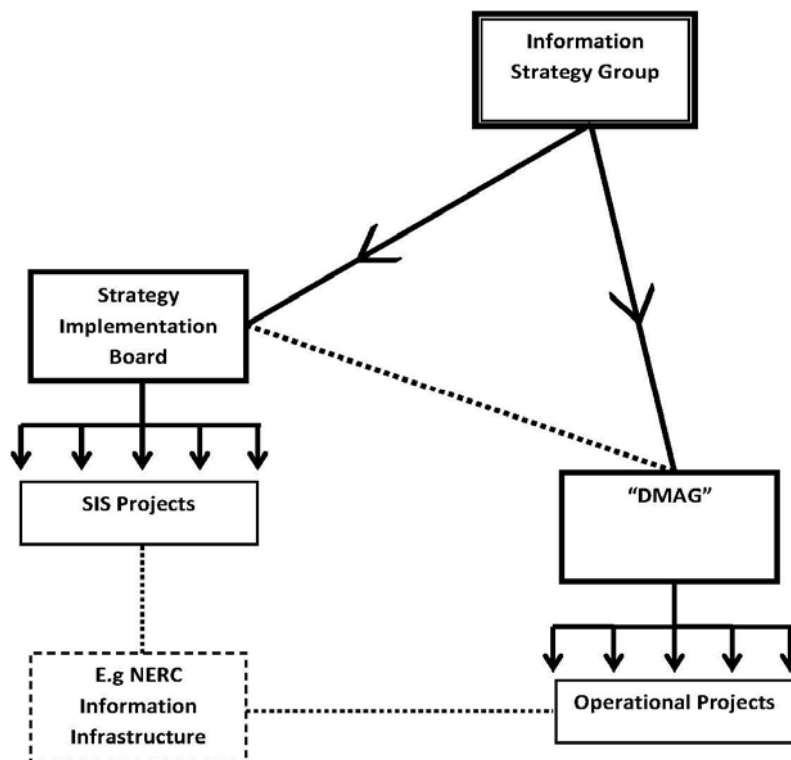
4.24. The implementation of the Science Information Strategy will be run as a programme with a portfolio of larger projects using the standard NERC project management practices. A rolling plan of smaller activities will also be required. The NERC Information Strategy Group will be responsible for overseeing the strategy's objectives and outcomes and a Strategy Implementation Board will manage the overall engagement of staff on the portfolio of projects and activities.

4.25. Standard Programme Management practices will be used to manage the strategy's progress including management of communications, costs, benefits, resources, risks, issues and projects' scope and timescales. Programme management will require approximately 1 full-time equivalent (FTE) centrally and 0.1 FTE per data centre for the duration. An annual progress report will be provided to the NERC Executive Board (NEB) and to The National Capability Advisory Group (NCAG). The scope of projects will be fully defined before being initiated and a single Project Executive will always be appointed supported by representation from across NERC, the Data Centres and key stakeholders as appropriate.

- 4.26. The detailed portfolio of projects will be modified as the programme progresses and a rolling plan of actions will be managed. A key driver for this will be the outcomes from the consultation activity in phase one and the target dates that NERC must meet to achieve INSPIRE compliance. The Science Information Strategy, and thus the Implementation Plan, are likely to require update during 2012/13.
- 4.27. Research centres will have a level of autonomy as to how they implement their information management programmes and data centre functions, provided strategy requirements are met.

## **Governance**

- 4.28. The NERC Director, Finance and Operations will be the NEB Director responsible for delivery of the NERC Science Information Strategy and will be accountable to NEB. NCAG will be actively consulted in major strategic decisions and planning of activities. Progress updates will be provided as required to the NERC Science & Innovation Strategy Board and to Council.
- 4.29. Governance of the Science Information Strategy will be undertaken at two levels:
- Governance of the Science Information Strategy's objectives and outcomes; including approval of changes and substantive programme and design issues;
  - Governance of the programme of projects and activities that deliver the strategy. Responsibilities include ensuring that significant local initiatives are aligned with the strategy and this will require for some to be expanded, some modified and some stopped.
- 4.30. The role of the NERC Information Strategy Group (ISG) will be adjusted to perform this first level of governance and a new body, a 'Strategy Implementation Board', will be formed for the second. Terms of reference for ISG and the Strategy Board will be developed and submitted to NEB for approval. Figure 3 provides a schematic of the governance model.
- 4.31. NERC Data Centres will be represented on both ISG and the Strategy Board, and appropriate external representation will also be sought. An extensive consultation exercise within the first phase of work, with resultant stakeholder engagement activities, will also ensure the views of all key external stakeholders and communities are adequately represented.
- 4.32. The role of the Data Management Advisory Group (DMAG) will be adjusted. The group will assure the design of changes emerging from the programme, ensuring integrity of the organisation, process, informatics and technology solutions. It will advise both ISG and the Implementation Board, and provide a focus for best practice between data centres to minimise duplication of effort. DMAG will also continue to have responsibility for coordination of delivery of operational data management by the data centres, through the 'single voice' and 'single face' of NERC data management.



**Figure 3:** Science Information Strategy Implementation Governance

4.33. Existing specialist DMAG subgroups will be reviewed and fixed-life groups to tackle specific issues will be established as required. For example, the information architecture community will need to work together to agree informatics standards and drive the sharing of skills and technologies.

4.34. Financial benefits accruing as a direct result of strategic activities will normally be retained within the budgets of the data centres and research centre science information programmes, and reinvested to help deliver further activities within the strategy. However, there may be instances where strategic priorities will require the reapportionment of benefits across the data centres and/or the research centres. This will be guided by ISG reporting to Director, Finance and Operations.

**Resources Required for Implementation**

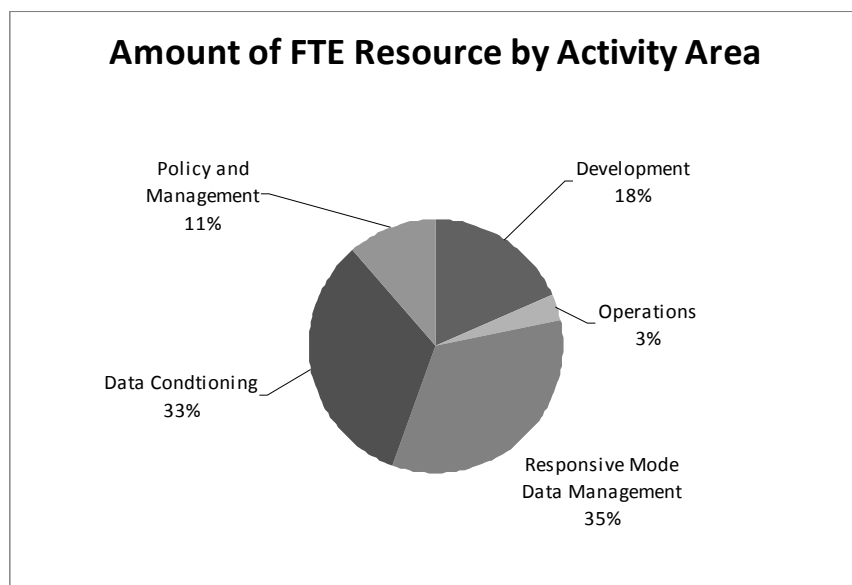
4.35. The implementation plan assumes constant real term funding for Science Information activities. Staff working on the programme will be funded from local budgets up to a level of 2 FTE per data centre (see Table 3). Staff engagement costs above this will be met centrally. Non-labour expenditure will also be met centrally.

Centre	Phase 1 FTE		Phase 2 FTE		Phase 3 FTE	
	Existing <sup>1</sup>	New <sup>2</sup>	Existing	New	Existing	New
BADC / NEODC	1.2	1.7	1.2	1.8	2.2	2.2
BODC	2.3	1.4	2.2	1.8	3.3	2.2
EIDC	3.3	1.5	3.2	1.8	4.3	2.2
NGDC	1.2	1.5	1.2	1.8	2.2	2.2
PDC	1.6	1.5	1.2	1.8	2.2	2.2
Swindon	0.2	1.7	0.2	2.1	0.3	1.7
<i>Total FTE</i>	9.7	9.2	9.2	10.9	14.6	12.9
Contingency <sup>3</sup>	2.9	2.8	3.7	4.3	7.3	6.4
<b>TOTAL</b>	<b>12.6</b>	<b>12.0</b>	<b>12.9</b>	<b>15.2</b>	<b>21.9</b>	<b>19.3</b>

**Table 3:** Overall resources estimated for the implementation of the strategy.

4.36. The resource estimates includes programme management support and, for phases two & three, the effort required to support effective management of Responsive Mode (RM) data. For phases two and three some of the required resources will be found from efficiency savings elsewhere within the data centres. However, the amount is not yet known.

4.37. The amount of FTE resource expended by activity area is shown in Figure 4. Over the three phases of the programme, approximately one-third of effort will be on data conditioning, one-third on responsive mode data management and support, and one-third on other activities.



**Figure 4:** Data centre resource allocation by activity area, over all phases.

<sup>1</sup> **Existing** – this refers to work that the data centres already plan to undertake to meet their current commitments and which are recognised as required for strategy implementation.

<sup>2</sup> **New** – refers to the resources required to meet additional activities arising from the implementation of the strategy.

<sup>3</sup> **Contingency** - is calculated as follows: Phase 1 30% (risk 15%, change 15%); Phase 2 40% (risk 20%, change 20%); Phase 3 50% (risk 25%, change 25%).

- 4.38. In addition, non-labour spend of £375k (excluding contingency) is required for supporting technical infrastructure, stakeholder consultation and engagement and community workshops. This is broken down as phase one £75k, phase two £125k, and phase three £175k.
- 4.39. There is a requirement for ongoing and sustained, long-term funding of some £250k a year to support the development and maintenance of core data discovery, view and download services (the NERC Spatial Data Infrastructure). In phase one this will be targeted at ongoing support for the NERC Data Grid. And, subject to agreement on a common information and technical architecture, will in subsequent phases provide support for common services used by all the data centres.
- 4.40. Informatics research activities by the data centres will also require support, and it is not unreasonable to aim for 5% of the data centre resource to be aligned to informatics research and information architecture activities. This 5% represents approximately £500k to £700k funding, some of which is currently provided centrally and some of which is explicit to centre's information management programmes.

### **Consequences of Resource Allocation**

- 4.41. The diversion of resources into the implementation of the Science Information Strategy will result in some current activities being either rescoped, delayed or stopped. Choices will have to be made, however, implementation activities are not yet specified sufficiently for detailed consequences to be assessed. The necessary information will only become available towards the end of phase one.
- 4.42. Successful implementation of the strategy will require the balancing of three key areas: ongoing operations, data conditioning to support responsive mode and the 'single face' of NERC, and new work to meet other requirements of the strategy.
- 4.43. The governance structure provided by the Information Strategy Group and the Strategy Implementation Board will be responsible for overseeing the process of making the necessary resource allocation choices and understanding and communicating the consequences.

### **Assumptions, Risks, Issues and Dependencies**

- 4.44. Key assumptions, risks, issues and dependencies impacting on implementation of the strategy have been identified. The Science Information Strategy (SIS) Implementation Plan and its associated assumptions, risks, issues and dependencies needs to be understood and owned by the appropriate senior managers across NERC. This collective understanding must reflect the variations in NERC's Data Centres and how this broad skills base strengthens the value of the 'single voice' and 'single face' of NERC data management.
- 4.45. *Key Assumptions:*
- The implementation plan assumes constant real term funding for Science Information activities and a long-term commitment by NERC to the maintenance of its environmental data centres as a key element of the UK National Capability in the environmental sciences;
  - The existing network of data centres, co-located with the appropriate NERC centres of scientific excellence, will be maintained. This is to allow local flexibility to meet specific community requirements;

- NEB and SIS governance groups will ensure that the necessary resources are allocated for effective implementation of the programme, which will of necessity, include input from other activity areas within NERC (for example, the Science Delivery function within Swindon Office);
- Implementation activities will be prioritised – so not all stakeholders will see their requirements met.

#### 4.46. Key Risks:

- Solutions delivered by the strategy may not meet the expectations of stakeholders, in terms of both vision and scale. What looks good in a demonstration may not successfully scale up to a production strength system.

*Mitigated through stakeholder engagement and expectation management, with explicit reference to the scope of the strategy. This will need to be in terms of the range of information services that the data centres can provide within the available resources.*

- Efficiency savings within the data centres will not deliver all the resources the data centres will need for new activities under the strategy.

*Mitigated through an explicit reprioritisation of existing data centre activities and a clear understanding of the scope and scale of efficiency savings that are achievable.*

- Process becomes more important than outcome, with too much focus on inward looking activities and too little focus on what stakeholders require.

*Mitigated through the Strategy Implementation Board focusing on delivery of outputs, with a key focus on improved access to data for reuse and repurposing. ISG and the Strategy Implementation Board must constantly remember that this strategy is fundamental to the successful delivery of NGSPE and that the criteria used to value the impact of NERC science depend implicitly on NERC's data, information and derived value added products.*

- There is a complex relationship between the data centres and their host research centres and/or research communities. Changes to this relationship have the potential to impact on the activities of the centres and the research process.

*Mitigated through the data centres actively managing this relationship and ensuring the Strategy Implementation Board are made aware of any likely adverse consequences. The board will be tasked to keep these adverse consequences to a minimum.*

#### 4.47. Key Issues:

- There is a large amount of work to be undertaken for data conditioning and management of responsive mode data. Initial estimates of resources for this and all other subsequent activities will be refined after the completion of phase 1.
- SIS implementation will be a sustained, long-term activity, which will require on-going management focus from the NEB level downwards. Senior management must understand and identify with the long-term benefits of the strategy.
- Other major change projects in NERC, for example implementation of the Shared Service Centre, will impact on the ability to drive through change. Strategy implementation activities must be coordinated in light of these other change initiatives to avoid delays and other unforeseen consequences.

4.48. *Key Dependencies:*

- IT infrastructure will support the needs of the SIS.

4.49. A comprehensive assumptions, issues and dependencies analysis and risk-register is being developed as a primary resource to underpin all implementation activities.

**Definition of terms**

4.50. A definition of terms used within the implementation plan, and elsewhere within the strategy, is presented at Appendix C.

## 5. Data Centre Funding and Governance

5.1. This chapter provides information on the proposed funding and management models to support future data centre activities.

### Data Centre Funding

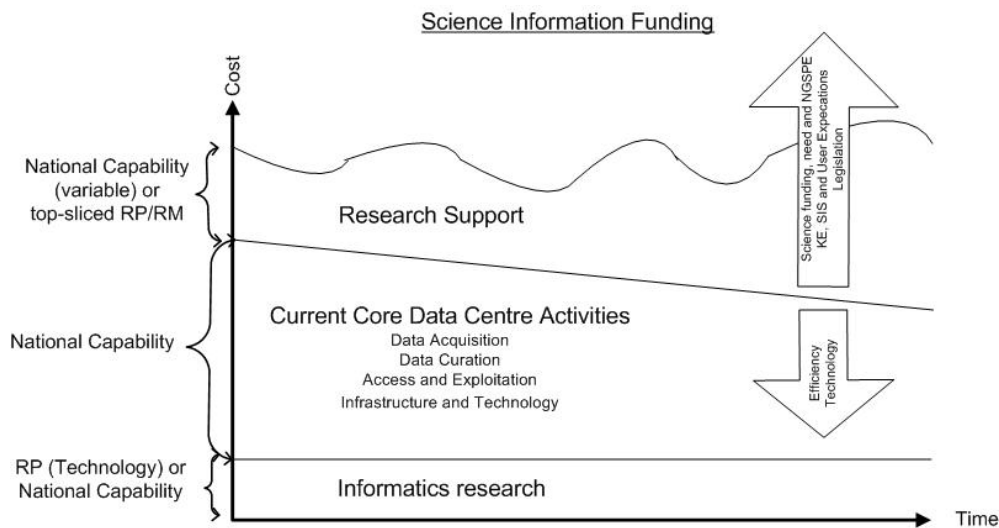
5.2. The network of data centres is there to provide support and guidance in data management to all those funded by NERC. However, this support is often provided on a best endeavors basis, and the level of support provided varies between data centres, depending on the governance model and funding allocation. Historically, data centres embedded within the NERC Research Centres have been less effective at offering support to researchers within the HEI sector. In addition, there is no formally agreed level of support that data centres are expected to offer. Competing priorities and resource has lead to problems where data centres are not always able to offer the necessary support to ingest data or on occasion provide comprehensive access to data sets.

5.3. As the primary cost in long-term data management is the ‘accession’ cost of getting data into the data centres, the more support in data management activities that the data centres can provide to researchers collecting data (especially in documenting and quality assuring data sets to enable reuse) the better the long-term quality of the data sets and the lower the accession costs. The evidence is that early engagement of data management specialists in the research process, leads to better long-term outcomes, both in the long-term usability of data sets and the overall costs of long-term data management.

5.4. Hence, the requirement of the Science Information Strategy for clarity of funding for data centre activities, clear governance and agreed levels of service.

### New Data Centre Funding Model

5.5. To achieve clarity of funding, the strategy will work to implement a common model for data centre funding. This is summarised in Figure 5.



**Figure 5:** Proposed data centre funding model

5.6. The funding model has three different funding elements, to cover the three main activity areas of the data centres.

- a) Long-term National Capability (NC) funding for core data centre activities;
- b) A variable Research Support funding element that reflects current NERC expenditure on research activities - Responsive Mode (RM) & Research Programmes (RP);
- c) An Informatics Research funding stream, which would be supported from Research Programmes (technology), Responsive Mode or via the National Capability interdependent research stream.

5.7. The long-term NC funding for core data centre activities will support on-going, data management and dissemination outside of any one project or programme, and will support NERC funded researchers in development of data management plans at the proposal writing stage.

5.8. Part of the process of determining allocations to individual data centres will involve an understanding why data management costs vary between data centres and disciplines. There is a requirement to understand if cost differences which exist are based on differing needs between research communities, or on custom and practice that should be challenged. If differences are based on need there is a requirement to understand if these are needs which NERC wants to support.

5.9. The variable Research Support funding element will reflect current NERC expenditure on research activities (RM & RP). This would be used to provide within project data management support to award holders, and crucially, cover the accession costs of getting data into long-term storage within the data centres. The allocation of this element would vary over time – both in terms of the overall level of NERC funding going into research activities and in terms of the amount that any one data centre would receive. The amount a data centre receives would reflect the differing costs of data management between disciplines at the same time as recognising the variation in proportional spend between NERC disciplines, and would be profiled over time to reflect the differing schedule of data management activities as opposed to research activities.

- 5.10. How this variable Research Support element is funded needs to be resolved in a pragmatic fashion. The viable options are either to cover this directly from NC or through top-slicing of RP/RM. The NC route is the simplest, however, it risks losing the direct relationship between the amount and type of research supported which the RP/RM route offers. However, it is considered that it is neither practical nor cost effective to directly charge individual award holders for the majority of data management services provided by the data centres.
- 5.11. If specific research projects require additional data management services above and beyond those covered by the Research Support element, these would be commissioned by projects using their own funds. This reflects current practice.
- 5.12. In addition, data centres will undertake an element of Informatics Research<sup>viii</sup>, which would be funded from RP (technology), RM or via the NC interdependent research stream. It is likely that most of the informatics research carried out by data centres would be directed at keeping data centre capability at the cutting edge, and hence would come within the NC interdependent research category.
- 5.13. Individual data centres also receive income for commissioned activities from external sources.

### **Data Centre Governance**

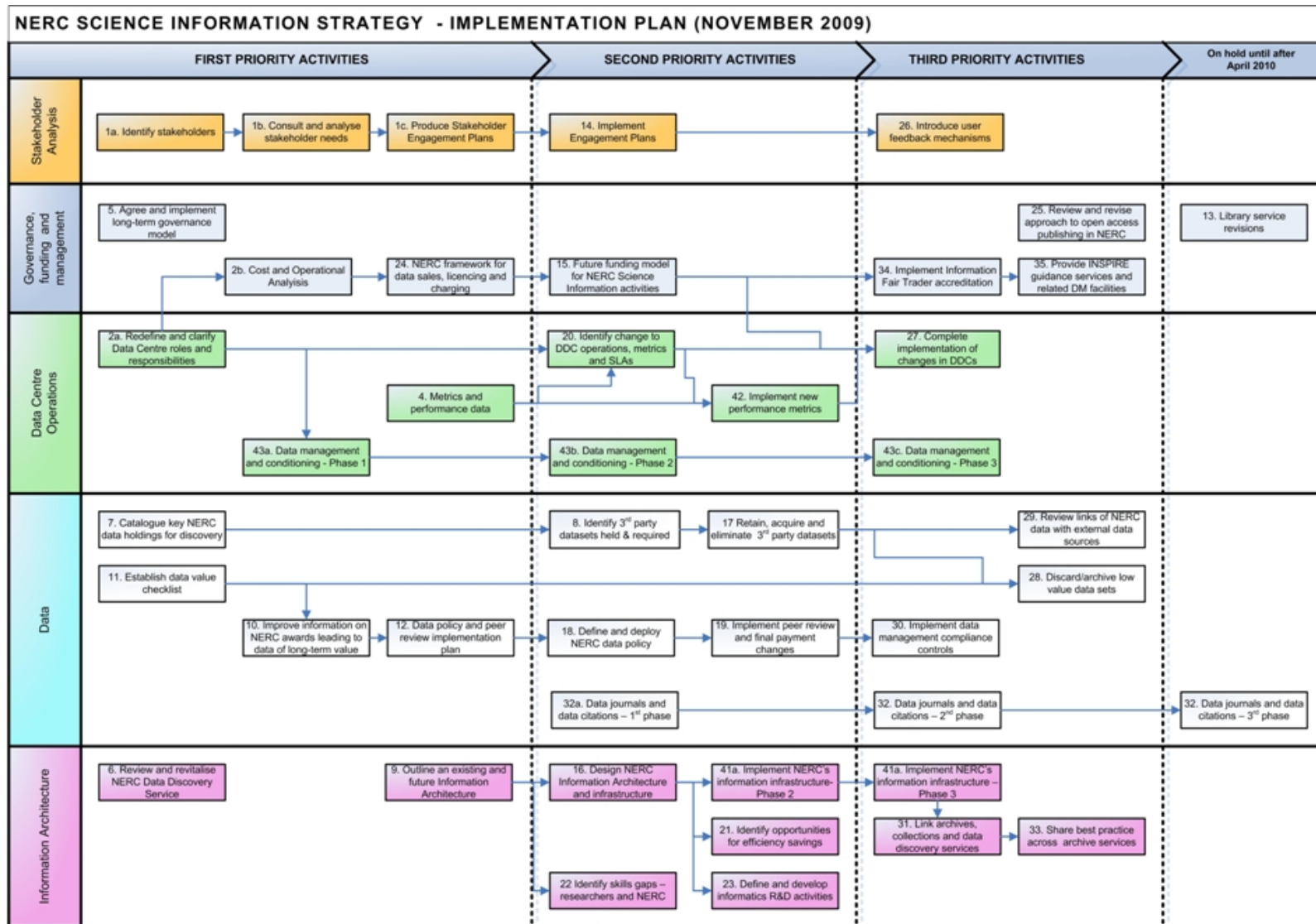
- 5.14. Presently, three of the data centres (EIDC, NGDC and PDC) form an integral part of their host centres information infrastructure. The other three (BODC, BADC and NEODC) are run as separate entities with clear funding lines and unambiguously act for the wider community. In the case of the former, it is not clear how the data centres relate to the wider community, whilst for the latter concerns have been expressed as to where the data centres report.
- 5.15. The present model of six sector orientated designated data centres requires a strong framework to ensure close co-ordination between data centres; closer consistency of service and activities across the data centres; efficient use of resources and rationalisation of activities where duplication exists. To help achieve this, the NERC Information Strategy Group (ISG) has considered the following three governance structures for the data centres:
- i. The creation of a single physical NERC Data Centre with a single Director;
  - ii. The appointment of a single NERC Data Centre Director, to which the network of data centres will report;
  - iii. Sector based management with strong central coordination and oversight.
- 5.16. A brief summary of the pros/cons of each of these three governance structures is outlined in Table 4.

Management model	Pros	Cons
Physically unified NERC Data Centre with single Director	Provide consistency of service; Maximise efficiencies; Enable balance of spend across disciplines; Will provide a 'one stop shop' for all users.	Isolated from the community and less driven by / responsive to community requirements; Seen by users as monolithic and unresponsive; Would remove data and informatics skills from the research centres. Real potential to impact on RC strategies which are strongly built on management and exploitation of environmental information and require access to data and informatics skills to do this.
Single NERC Data Centre Director managing a network of data centres.	Provide consistency of service; Enable balance of spend across disciplines; Data and information skills remain within the research centres.	Direction is isolated from the community and less driven by / responsive to community requirements; Could be seen by users as monolithic and unresponsive; Unclear as to how data and information staff would work for the data centre and for the research centre supporting research centres own IM programmes.
Sector based management with strong coordination and oversight.	Potentially more linked to and responsive to community and RP requirements; Data and information skills remain within the research centres.	More difficult to achieve consistency of service and to less efficient; Will need strong, clear coordination and oversight.

**Table 4:** Summary of data centre governance structures.

- 5.17. ISG recognised merits in all three models, but recommend to NEB that a sector based management model be maintained, with the data centres supporting the interests of their appropriate NC sector, and reporting to the relevant NEB member representing that NC sector. The NERC Director of Finance and Operations, supported by ISG, will provide strategic leadership and direction, within a framework of activities arising from the NERC Science Information Strategy.
- 5.18. In addition, ISG considered if there is scope for efficiency savings through merging back office functions / technologies for data storage and dissemination. Whilst there are efficiencies to be gained through merging data storage and dissemination technologies, there are also risks. For example, the reliance on high capacity, high resilience network connections to enable all NERC sites to access the centralised data holdings.
- 5.19. It is ISG's opinion that closer working between the data centres, with adoption of more standard ways of working and cross cutting, interoperable technologies will deliver many of the benefits that could accrue from merging of data storage and dissemination technologies. ISG therefore recommended that merging of data storage and dissemination technologies is not explored further at this time.

## Appendix A: Implementation Plan & Key Dependencies



## Appendix B: Summary of Communication & Change Management Activities

### Communication Plan

1. Communication about the Science Information Strategy will be required on several levels; communicating the existence and benefits of the strategy, reporting progress on the actual implementation work and at a more detailed level within each of the significant projects. Parts of the plan, in particular responsibilities, will be dependent upon the final decision on the governance model.
2. Two-way communication with NERC's key stakeholders is essential for success. This is especially true of communications with NERC funded researchers within HEIs.
3. The primary objectives of the communication plan are:

Objective	Action
Champion	Lead the communications and the change process. Take responsibility for leading the strategy implementation.
Engage	Two-way engagement of key stakeholders.
Inform and consult	Stakeholders of the changing processes for the operations of the data centres, the benefits to them and any resulting obligations / expectations for each stakeholder group.
Change	Change in operations and functions of the data centres.

4. The key groups and individuals involved in carrying out the communication plan include:

Group	Task / Responsibility
ISG	Active communication on detail and change linked to the development and implementation of the strategy. Enable change within the data centres.
Individual ISG Reps	Change Facilitator and Local Champion for their relevant data centres.
Chair of ISG	Act as SIS Champion.
Strategy Implementation Board and DMAG	Lead the active communication within specific stakeholder groups and the data centres. Monitor and report on progress against the implementation plan.
Implementation Project Leaders	Report on progress of their specific project to the Strategy Implementation Board.

5. Focusing on the communication activities associated with the actual implementation of the strategy and monitoring progress against the implementation plan, the primary methods of communication will be:

<b>Method</b>	<b>Frequency</b>	<b>Responsibility</b>	<b>Aim / Outcome</b>
ISG meetings	Quarterly	ISG	Monitor and support progress against the implementation plan.
Strategy Implementation Board and DMAG meetings	At least quarterly	Strategy Implementation Board	Monitor and maintain progress against the implementation plan
SIS Implementation Plan	Created prior to implementation, reviewed regularly at SIS Implementation Board meetings	Strategy Implementation Board	Document and define the actions, timescales and resources required to deliver the SIS
Project Progress reports	As dictated by the implementation group in response to scale of project	Implementation Project Leaders	Monitor progress at individual project level, to be fed up to Strategy Implementation Board
Project Plans	Project start. Level of detail dependent upon scale of project	Implementation Project Leaders	Document and define project with associated deliverables, resources and timescales
Implementation Plan Progress Report	Annually – for NEB and NCAG	ISG with input from the Strategy Implementation Board	Inform senior management of progress on the SIS Implementation

### **Change Management**

6. It is recognised that the implementation of the strategy will require significant change in the operations of the Data Centres and how they interact with their stakeholders. Change management will be considered as part of each of the individual projects within the implementation plan. The NERC Change Management Framework and the network of Change Facilitators will be utilised where appropriate. This is especially important as the programme moves into phase two, which has major change elements within it.
7. Crucially, all key stakeholders must accept the benefits that the strategy will deliver to ensure their support for the on-going change activities and to manage expectations about what can be achieved

## Appendix C: Definition of Terms

This glossary lists definitions for key words in the Science Information Strategy and this implementation plan. It concentrates on words and phrases that are poorly understood outside of the data management community and where there is ambiguity.

Term	Definition
<b>BADC</b>	The British Atmospheric Data Centre, based at Rutherford Appleton Laboratory
<b>BAS</b>	British Antarctic Survey – a NERC Research Centre
<b>BGS</b>	British Geological Survey – a NERC Research Centre
<b>BODC</b>	The British Oceanographic Data Centre, hosted by NOC, Liverpool
<b>CEH</b>	The Centre for Ecology and Hydrology – a NERC Research Centre
<b>Data</b>	Data are individual items or records (numeric or other) obtained by measurement, simulation or observation of the natural world and human impacts upon it.
<b>Data centre</b>	An organisation that collects, ingests, archives and distributes data and information. The NERC data centres are those specifically tasked with managing the data within the discipline areas: BADC, BODC, EIDC, NEODC, NGDC and PDC.
<b>Data conditioning</b>	<b>Data conditioning</b> is the process of working up data into a standard form suitable for archive. This is included in the <b>data ingestion</b> process of taking data from a data provider, data conditioning, data checking and finally depositing the data in an archive.
<b>Data set</b>	An aggregation of data and information with a common theme. Frequent themes used are data collected by a particular instrument, measurements from a particular project, or data from co-located instruments.
<b>Discovery</b>	The process by which data is discovered using tools such as the NERC data discovery service and Google. Discovery of raw data files is of very limited use; instead the items found are metadata records designed define and describe data sets. A data set catalogue is a searchable system of metadata records.
<b>EDINA</b>	EDINA is a JISC National Data Centre based at the University of Edinburgh
<b>EIDC</b>	Environmental Information Data Centre, based at CEH Wallingford
<b>ERFF</b>	The Environmental Research Funders’ Forum
<b>EU</b>	European Union
<b>GeoSciML</b>	A mark-up language for the Geosciences
<b>HEI</b>	UK Higher Education Institutions
<b>HMG</b>	Her Majesty’s Government (ie the UK Government)
<b>IM</b>	Information Management
<b>Informatics</b>	Informatics is the science of information, the practice of information processing, and the engineering of information systems. Informatics studies the structure, algorithms, behaviour, and interactions of natural and artificial systems that store, process, access and communicate information. It also develops its own conceptual and theoretical

	foundations and utilises foundations developed in other fields. Since the advent of computers, individuals and organisations increasingly process information digitally. This has led to the study of informatics that has computational, cognitive and social aspects, including study of the social impact of information technologies.
<b>Information</b>	Information is created by interpreting and/or processing data so that their context and relationships are expressed and understood.
<b>Information Architecture</b>	The structural design of shared information environments. For NERC's data and science information this includes common data models, metadata standards and communal web services.
<b>Information Management</b>	It is recognised that data, information and knowledge are a continuum, and in some instances their boundaries are fuzzy. Information management spans this continuum.
<b>INSPIRE</b>	INSPIRE is an EU directive that requires public bodies like NERC to make its data discoverable, viewable and downloadable. The directive comes into force piece by piece over the next few years.
<b>INSPIRE</b>	Directive 2007/2/EC of the European Parliament and of the Council of 14 March 2007 establishing an Infrastructure for Spatial Information in the European Community (INSPIRE) was published in the official Journal on the 25th April 2007. The INSPIRE Directive entered into force on the 15th May 2007.
<b>Interoperability</b>	The ability to exchange and use information. This requires: 1. Tools to exchange information (and the ability to use those tools). Where those tools consist of systems that integrate information from a variety of systems, without special effort, then we can talk about <b>service interoperability</b> . 2. The ability to interpret and use the information when it is acquired. When that use involves integrating information from disparate sources, we want <b>data interoperability</b> - that is, we want to be able to integrate the data/information to a common format/view without special effort. This last requires standardised formats, standardised metadata structures, and standardised vocabularies.
<b>JISC</b>	The Joint Information Systems Committee. Funded by the UK higher and further education funding bodies to provide world-class leadership in the innovative use of information and communication technologies to support education and research.
<b>KPI</b>	Key performance indicator
<b>Long-term</b>	Data management is a long-term activity – some data sets are extended over many decades. Critically data management is longer term than the funding cycles of programmes and centres.
<b>NC</b>	National Capability for environmental science – one of the three main NERC funding streams which enables the UK to deliver world-leading environmental science, support national strategic needs and respond to emergencies. Long-term observation, survey, mapping, scientific collections, community Earth-system models and data management are a key activity funded under NC.
<b>NEB</b>	The NERC Executive Board

<b>NEODC</b>	The NERC Earth Observation Data Centre, based at Rutherford Appleton Laboratory
<b>NERC Data Grid</b>	Data discovery and delivery are inherent components of many aspects of science. They can be considered part of a processing chain that starts with raw data from a variety of sources, and ends with the graphical production of information that is directly used in scientific research. The NERC Data Grid provides tools and information products that can be used to facilitate this process.
<b>NGDC</b>	The National Geoscience Data Centre, based at BGS Keyworth.
<b>NGSPE</b>	Next Generation Science for Planet Earth. The NERC Strategy for 2007 – 2012.
<b>NOC</b>	The National Oceanography Centre, a NERC Research Centre, with sites at Southampton and Liverpool.
<b>PDC</b>	The Polar Data Centre, based at BAS, Cambridge
<b>PI</b>	Principal investigator within a research team
<b>Publication</b>	Publication in the research context is understood as the public release of books, reports and papers in academic journals that give the authors credibility with funders and within their discipline. <b>Publication of data</b> is thus the public release of data where academic credit should be given for the creation of the data. A <b>data journal</b> is a journal which sets out to give academic credit for data creation
<b>RAE</b>	The UK Research Assessment Exercise, in future to be replaced by the Research Excellence Framework.
<b>RM</b>	Responsive Mode research for environmental science – one of the three main NERC funding streams which supports original investigation and training to gain, advance or expand knowledge and understanding. Funding is provided in response to applications for research proposals in any area relevant to NERC's remit. This includes the dissemination of that knowledge and understanding, and the research infrastructure required to deliver that research.
<b>RP</b>	Research Programme for environmental science – one of the three main NERC funding streams which delivers world-leading, strategically themed environmental research.
<b>SISB</b>	The NERC Science and Innovation Strategy Board
<b>SMA</b>	NERC Science and Management Audit
<b>Spatial data infrastructure</b>	A spatial data infrastructure is a framework of spatial data, metadata, and tools that enable the interoperable use of spatial data. <b>Spatial data</b> is data with a direct or indirect reference to a specific location or geographic area.
<b>Standards</b>	The words <b>standards</b> and <b>standardised</b> are meant to imply protocols and implementations which conform to something which has either been through a de facto or de jure process of community definition. The issues for the applicability of standards resolve to "fitness for purpose" and "community acceptance". Standards for data and metadata allow interoperability.
<b>Third-party data</b>	Data where the rights to exploit the data rest outside the control of NERC.

<b>UNESCO</b>	The United Nations Educational, Scientific and Cultural Organisation
<b>Value added products</b>	Data and information produced by analysis, aggregated and re-structuring existing data.
<b>Web-based services</b>	Web-based services are both portals for human usage and interfaces for computer to computer use. A <b>Portal</b> is a web based access point bringing together resources for a particular community.

## End notes

<sup>i</sup> The Power of Information: An independent review by Ed Mayo and Tom Steinberg (2007), available at [www.cabinetoffice.gov.uk/publications/reports/power\\_information/power\\_information.pdf](http://www.cabinetoffice.gov.uk/publications/reports/power_information/power_information.pdf).

<sup>ii</sup> Science managers are those responsible for developing and managing NERC's own process and procedures for evaluation of proposals and activities, running funding rounds, and all other activities that NERC undertakes in developing and delivering its programmes of work. They are largely, but not exclusively, based in NERC Swindon.

<sup>iii</sup> Where the results of the research are freely accessible in the public domain, normally via publication in peer-reviewed publications.

<sup>iv</sup> Innovative web-based applications that provide access to information often bring together data from multiple sources on-the-fly. This is known as producing a data 'mash-up'. To have data available for others to 'mash-up' requires a level of data conditioning and documentation above and beyond that currently often undertaken.

<sup>v</sup> Although network bandwidth into NERC institutions will need to be regularly reviewed.

<sup>vi</sup> The INSPIRE directive creates a framework for spatial information across Europe to be collected, stored, manipulated and made available in a more standardised electronic environment to facilitate the sharing of information between public sector organisations and the citizen.

<sup>vii</sup> All NERC activities in this area will utilise and make reference to the framework of 'default' licences being developed by the Office of Public Sector Information.

<sup>viii</sup> Informatics is the application of information science, processing and technology engineering skills to help further environmental sciences.