



**NERC Data Centres – future services: A forward look
Town Hall Meeting, 13th October 2016**

Executive Summary

This document provides an overview and presents a summary of the information gained at a brokerage event, held on the 13 October 2016, at the Victory Services Club (VSC) in London. The event, facilitated by Jamie McDonald Coaching and Development, brought together 55 stakeholders from a diversity of organisations. Specific attendees and their organisations are listed in Appendix 1.

The aim of the day was to help NERC better understand users' changing requirements for the NERC data centres and services they provide and so to inform the commissioning process for a NERC Data Service, as well as NERC's information strategy. These aims were met by providing an interactive, lively and engaging environment for gathering users' views and opinion. Information was elicited in a variety of ways through a number of different sessions. An outline of the event is provided in Appendix 2.

At the event participants felt strongly about the need to maintain the provision of discipline specific expertise and support that is currently offered by the data centres. There was some confusion over what was meant by a single data service and concern that this may result in a loss of the valued discipline specific expertise. There was a feeling that one single data centre was not desirable as the data and use of current data centres is very diverse. Despite this there could be benefits of a single data service including improved access to data and standardisation.

Future considerations for a data service included a requirement for more interoperability between different data including that with an international emphasis and the need to be able to adapt to or incorporate new technological developments. It was felt that any future data service would also need to operate within funding and resource constraints (in terms of both technological and financial).

It was agreed that in the future an ideal data service would provide a wide, international scope; high quality data; sound tools and methods to allow access use and interchange; it would also provide ongoing engagement with users and would be sustainably resourced and funded.

These findings and information gathered should be considered with the caveats that the group was made up of 55 self-selected stakeholders, some of whom worked for, or have close ties with, the data centres themselves. During the event participants were asked for their first ideas and opinions and may not therefore reflect well considered positions. Finally in collating the information there may be some loss of resolution / specificity relative to the raw data.

Session 1: Initial activity

With a dual purpose of getting people interacting and starting the discussion on the data centres, this activity asked people to note each other's answers to the following questions:

- Which data centres do you use?
- What do you use the data centre for?
- What is done well?
- What is not done so well?
- What would enhance their services?

The responses to the questions are provided in the tables below. For the initial question, "Which data centre do you use?" Sixty-seven responses were gained (some participants noted more than one centre) and the following table (1) summarises that information. Where a response was given for, for example BGS, it was allocated to the appropriate data centre, in this case the NGDC. In this question only, data centre staff were excluded from the analysis.

Table 1 – responses to the question, "Which data centre(s) do you use"

Data centre	Users	
BODC	12	
CEDA		
Unspecified	8	
BADC	4	
NEODC	2	
USSDC	0	
EIDC	3	
NGDC / BGS	7	
PDC	3	
OTHERS	14	
Unclear response or/ none	6	
DC personnel	8	

These data suggest that the BODC (12) and CEDA (14) are the most used individual data centres, by the group present. The NGDC is next (7) and then the EIDC and PDC (3 each). Fourteen responses indicated other sources of data and these ranged from International to UK and very local sources. This data may reflect groupings present within the participant cohort.

Table 2 – responses to the question, “What do you use the data centre for?”

Activity	Count
Research activities	8
Modelling	4
Commercial applications	7
Data management activities	
Managing / overview	5
Providing / storing data	7
Discovery / sharing data	5
Combining data	4
Unclear / ambiguous responses	5

Research activities top the list, particularly if modelling is included within that. And commercial applications are also shown (engineering projects, for example).

Table 3 – responses to the question, “What is done well?”

Activity	Count
Accessibility / dissemination / availability	10
Data support and advice	9
Staff responsiveness / service	8
Archiving and curation	8
Scale of data and connection to other datasets	7
Data delivery	5
Data discovery	4

Stakeholders rate highly (10 counts) the fact that data can be accessed and is available. The support and advice offered is also highly rated along with the responsiveness and service of the staff (17, combined). The curation of data (8), the scale of the dataset available and its connection to wider sets are also rated well (7) as are the processes of delivery and discovery of data.

The importance of expert staff in offering support, advice and responding to user needs is evident.

Table 4 – responses to the question, “What is not done so well?”

Activity	Count	
Searching / finding data	10	
Technological issues	10	
Funding and resources	9	Perhaps the lower count for data discovery and delivery above (table 3) is echoed in the challenges of the top category in this table (4). There are also clear concerns about technological constraints and funding/resource constraints. The table also shows the desirability of data harmonisation and human contact in the DCs.
Data sharing and connectivity to other data	6	
Standardisation	5	
Usability	4	
Communication / identity	4	
Lack of a human interface	2	
Others (security/hacking and contribution of data not being statutory)	2	

Table 5. Responses to the question “What would enhance their services”?

Activity	Count	
Data discovery, searching and capture	13	This table confirms the importance to stakeholders of the ease of finding the right information through improved interfaces and tools, as well as ensuring the connection with a bigger community of centres and stakeholders. The availability of funding and resources are again seen as a key constraint.
Funding and resources (inc staff)	10	
Connection between centres and elsewhere	9	
Enhanced user interfaces / tools	7	
User focus	4	
Others	5	

Overall, and given that these tables (3-5) present only a brief snapshot of information, the key messages emerging were:

- The provision of data centre expertise and support is of great importance
- Ease of access to data is important, and a challenge
- Connection between different data centres and to a wider community and dataset is important
- There are concerns about technological matters and resources / funding

Session 2: Exploring the possibility of ‘one data service’

This activity saw 12 groups of 4-5 participants each, collating as many plus (P), minus (M) and interesting (I) points as they could in 2 minute slots (6 minutes in total) in relation to the proposal that there be a single, integrated data service. The form, governance and structure of any such service was deliberately not specified. The resulting P, M and I answers from each small group were collated and clustered and are presented in table 6, below.

In summary, the Plus column highlights the possibility of an integrated approach and ‘one-stop-shop’ identity, as well as money saving through efficiencies of scale, shared practice and improved data access and standardisation. This might lead to further connections between different disciplines and increased awareness of NERC as a brand and enhanced influence/reach. Other possible benefits were a single approach to data security and a new, perhaps more resilient, funding model emerging. The ‘others’ category included developing a career structure within any service and building a community of users.

These Ps are somewhat offset by the Ms that show concerns about losing the (highly valued) specialism, expertise, diversity and identity of the individual data centres (and datasets). There were concerns about the costs of such a process, the challenges presented by the complexity and unwieldiness of a single service and adding a layer of administration. There was also a theme of a single point of failure and cyber-attack. Also of concern was the potential to lose the valuable contact with end users or to have that obscured through a less personal identity (one size does not fit all). The ‘others’ category includes job security, poor history (?), duplication of other data services (World data centre) and lack of suitability for analogue data.

Table 6 – Summary of Plus, Minus and Interesting points in response to the proposal of a ‘single, integrated data service’.

Plus	Count	Minus	Count	Interesting	Count
One stop shop	13	Loss of identity	9	New stakeholders	7
Money saving	13	Costs	13	Commercial options	6
Integrated working	13	Complexity / Unwieldiness	12	Fit with UK research context	6
Efficiencies	9	Administration	9	Form / governance	6
Sharing practice	8	Single point of failure	7	Change and innovation	7
Enhanced data standardisation	8	Loss of data	5	New data practices	9
Ease of access	8	Loss of expertise	7	International connections	6
Cross discipline working	7	Loss of diversity	8	Cross-disciplinary work	7
Brand and influence	7	Loss of contact with users / one size does not fit all	12	Image	5
Funding model	3	Single point of failure	7	Funding	2
Data security	2	Cyber attack	3	Motivation behind this	3
Others	3	Others	4	Others	6

The Interesting possibilities included similar themes, such as the potential for new users, for enhanced connections between disciplines and international links. It would also be interesting to see the form (centralised or distributed) and governance of such a service and how this would affect its image, its funding and the commercial opportunities open to it. Several people remarked on how this might fit with the wider UK research agenda, and on the possible emergence of new data practices and wider change / innovation. There were questions about what and who is driving this activity, and the 'others' category included questions about 'another one stop shop', how the service's success would be measured and how it would support research; also included were cyber security, licensing issues and 'educating people'.

These data are not intended to show whether stakeholders would be for or against such an endeavour; rather they indicate that participants are aware of possible benefits of an integrated service, and show the areas that may need to be addressed in any such plan. The overall picture reiterates the perceived importance of the specialism / expertise of the data centres and their close connections with users, while recognising that there *might* be scope for enhanced efficiencies at the operational (data management) and more infrastructural levels, and opportunities for international, commercial and wider innovation. Of course, there are also questions about managing the costs, complexities and technicalities of bringing about such a service.

Session 3: Trends and the future of data

To obtain the summary statements in table 7 (overleaf), sub-groups were asked to reflect on how patterns of data use, and the influences operating in their environment, had changed over the preceding 10 years. Based on this, the groups were asked to cast their ideas forward to describe the environment in which data services will be operating, five years from now.

Five years from now, we will be operating in an environment that is...

Groups were consistent in suggesting greater internationalisation and openness, along with technological potential, but also emphasised these being constrained by restricted funding and resources. The importance of impact and funding opportunities was also stated, along with questions over the soundness of data.

The largest forces operating on data services will be...

Groups asserted that value for money and impact will be dominant forces as will availability of expert staff, infrastructure and data practices (interoperability, traceability etc).

People will be using...

Two groups who answered this section, suggested diverse data (in terms of discipline and sources) will be used more, along with technologies (online analytics, Google and Internet of things) that may be less rigorous.

...data for...

Policy uses for data were suggested as were research and the application of data to more everyday uses. Two groups also emphasised that unexpected uses will arise.

Overall participants felt the future would have:

- A greater international emphasis
- Enhanced technological potential to deliver societal / policy / scientific impact, but
- This will be constrained by funding and resources (both technological and human).

Table 7 – Summary of outputs from the activity 'Trends and the future of data'

Sub-group	A	B	C
Five years from now, we will be operating in an environment that is...	<p>International in focus (science is more international; impact is more local)</p> <p>Further away from the desktop, exploiting the cloud, remote apps etc. (but £ cloud persistence is an issue)</p> <p>Strapped for cash - even more focused on impact (bigger consortia)</p>	<p>Increasing internationalisation (despite Brexit)</p> <p>Widening gap between aspirations for data use and tech and resources/tech</p> <p>Backlash - an "expert-free" world</p>	<p>More directed towards funding opportunities (Impact agenda UK PLC)</p> <p>More data rich, more technologically driven (potentially more funding-poor)</p> <p>Has increased tension between openness/open data and expertise/collaboration/provenance/reproducibility/verifiability</p>
The largest forces operating on data services will be...	<p>Perceived VFM</p> <p>Upskilling and retention</p> <p>Exploiting cloud</p>	<p>Demand for data in/out (interoperability)</p> <p>Impact – justification</p> <p>Resources (staff, funding, infrastructure)</p>	<p>(QA/QC) /usability (access)</p> <p>Infrastructure (H/w s/w platforms) (Resilience)</p> <p>Training and skills (Diversity and domain expertise)</p>
People will be using...	<p>Google to access data (why bother with data centres)</p> <p>Combination of diverse data</p> <p>Data services (won't know where it comes from)</p>	<p><i>(Did not answer this specific component)</i></p>	<p>(Who? - cultural issues)</p> <p>Internet of things (issues: ethics, quality, security)</p> <p>Cross discipline data (blurs margins of NERC)</p> <p>On-line analytics (quality)</p>
Data for...	<p>We don't know!</p> <p>Unexpected usage but will need to advance knowledge (e.g. Local environment / global - climate impact)</p>	<p>Excellent research (international), ref NERC core mission</p> <p>Policy - at all levels</p> <p>Day to day information for wider consumer / stakeholder groups / unanticipated users</p> <p>Big data applications</p>	<p>Government / individuals - evidence-based strategy and policy (monitoring e.g. SDGs)</p> <p>Science inform knowledge (PIDS, DOIs etc. workflow, metaforward)</p> <p>Information-based lifestyle choices (individuals / businesses)</p>

Session 4: The ideal data service in five years

To produce the summary statements in table 8, subgroups were asked to individually reflect on what the ideal data service would be like. Individual responses were clustered by similarity, then each cluster was named, prioritised and assembled into the statements given in table 8.

To operate successfully in that environment, data services will need to be...

The groups described the ideal service as international in scope, robust (quality assured and secure), reliable (long term) and easy to use. That ease of use should include skilled experts, and the robustness should include responsiveness to changes in user needs and technology.

They will need to deliver...

The need for high quality and standards of data was stated, and the role of experts and specialists emphasised with two groups, stating training in the centres as important. The ease of data discovery and access was also stated and requires suitable tools, interoperability, sound data description and linked, scale-able data.

With this in mind, we suggest start...stop...continue...(actions to move towards the ideal service)

The groups reiterated the need for sustainable funding, skills and shared practice. They also stated the need for ongoing stakeholder engagement, clear governance and consideration of a wider perspective than just NERC. Interoperability was also suggested, as was a request to preserve the appropriately diverse ways of treating data.

Overall, the ideal data service was imagined as one that includes:

- A wide, international scope
- High quality and standards of data
- Sound tools and methods to allow access, use and interchange
- Sustainable resources and funding
- Ongoing engagement with users and their needs
- Ongoing access to specific expertise and training

Table 8 - Summary of outputs from the activity 'The ideal data service in five years'

Sub-group	A	B	C
To operate successfully in that environment, data services will need to be:	Robust long-term easy to use open service relevant beyond NERC interoperating globally Responsive to changing user and stakeholder needs Changing technology and changing user communities Retain skills to maintain and build upon quality of data services	Easy to use robust reliable open free(mium)	Quality assured (robust, and hence secure) staffed by technical and domain experts bigger than NERC (i.e. international in outlook)

<p>They will need to deliver:</p>	<p>High quality and well described data and meta data</p> <p>Which are easily discovered, accessed and understandable</p> <p>Supported by Domain expertise, analysis tools and environment</p> <p>Support services and training within the Data Centres</p>	<p>Linked, scale-able datasets maintained by appropriately skilled experts (human) for the respective, targeted users</p> <p>Access to domain and technical expertise, including training</p> <p>A standards-based approach to ensure interoperability</p> <p>Access in a multi-faceted way, which needs tools e.g. visualisation, discovery and fusion</p>	<p>Quality services, that are backed by experts and that are easy to use and sustainable</p>
<p>With this in mind, we suggest:</p> <p>Start</p> <p>Stop</p> <p>Continue</p>	<p>Ensure adequate staff and resources</p> <p>Provide clear governance</p> <p>Think beyond NERC (globally and other solutions)</p> <p>Identify your stakeholders</p> <p>Recognise data is a strategic asset to use, preserve and protect</p> <p>Trying to inappropriately homogenise treatment of all data</p> <p>To provide an excellent service (where you are)</p> <p>Consulting your stakeholders and user communities</p>	<p>You consult with the user community on an ongoing basis, e.g. at point of use, to ensure the end-product is used</p> <p>Ensure long-term, adequate and secure funding is in place to deliver</p>	<p>Start training next generation of data scientists</p> <p>Share best practice</p> <p>Internal/external interoperability</p>

Comments after delivery of these statements

After the statements were introduced, comments were invited in response, and are summarised here:

There is a need to be clear about whether a single data service or a single data centre is proposed, as any mis-reading of such a proposal could have far-reaching consequences. Also be clear about the resulting governance and structure.

Consider:

- How NERC brand may grow, individual identities diminish
- How to treat analogue data
- Links with Google
- NERC to become the go-to place for environmental data in the UK including access to other data sets, expertise, tech etc
- How to prioritise from here
- If it isn't broken, don't fix it
- Viability, predictability, trust that data centre will be there in the future

Maintain: link to expertise - curators, scientists etc

Appendix 1 – Participants and their organisations

Title	First Name	Surname	Organisation
Dr	Jen	Allanson	Facilitator, JMCD
Mr	Kevin	Ashley	University of Edinburgh
Dr	Nichola	Badcock	NERC
Mr	Garry	Baker	British Geological Survey
Mr	Edward	Baker	Natural History Museum
Dr	Victoria	Bennett	STFC
Mr	Adam	Branson	Card Geotechnics Limited
Mr	Mike	Brown	CEH
Dr	Sarah	Callaghan	Centre for Environmental Data Analysis
Mr	Jose Angel	Cañizares Fernandez	Oasis Hub
Dr	Gareth	Chisham	British Antarctic Survey
Mr	James	Cooke	United Kingdom Hydrographic Office
Prof	Martin	Culshaw	University of Birmingham
Dr	Paul	Duller	Tribal Group plc
Mr	Peter	Edmonds	The Crown Estate
Dr	Dave	Filipović-Carter	Facilitator, JMCD
Dr	Simon	Gardner	NERC
Dr	Jose	Gomez-Dans	National Centre for Earth Observation and University College London
Dr	George	Graham	Sir Alister Hardy Foundation for Ocean Science
Prof	Sandy	Harrison	University of Reading and World Data System
Mr	Matt	Harrison	British Geological Survey
Prof	Dominic	Hodgson	British Antarctic Survey
Dr	Elizabeth	Kent	National Oceanography Centre
Miss	Jessica	Ladd	NERC
Dr	Barry	Latter	STFC
Dr	Bryan	Lawrence	NCAS
Prof	Peter	Liss	Marine Data and Information Network (MEDIN)
Ms	Vicky	Lucas	The IEA
Ms	Miriam	Maina	Brunel University
Mr	Gerard	McArdle	TSP Projects
Dr	Robin	McCandliss	NOC
Dr	Jamie	McDonald	Facilitator, JMCD
Prof	Stephen	Mobbs	NCAS
Dr	Timothy	Moffat	British Antarctic Survey
Mrs	Niki	Newton	Joint Nature Conservation Committee
Mr	Jon	Parr	Marine Biological Association
	Helen	Peat	Polar Data Centre, BAS
Dr	Sam	Pepler	CEDA

Mr	Michael	Pinnock	British Antarctic Survey/NERC
Prof	John	Remedios	NERC – National Centre for Earth Observation
Dr	Beatrix	Schlarb-Ridley	BAS
Prof	Len	Shaffrey	NCAS/University of Reading
	Andrea	Sharpe	NERC
Mr	David	Shilston	Atkins Limited
Dr	Tim	Smyth	Plymouth Marine Laboratory
Dr	Helen	Snaith	NOC
Dr	Tracey	Stead	Facilitator, JMCD
Mr	Darren	Stevens	SAHFOS
Mr	Alex	Tate	British Antarctic Survey
	Ben	Taylor	NEODAAS
	Andy	Thompson	Oil and Gas Authority
	Mark	Thorley	NERC
Ms	Kay	Thorne	BODC
	Remi	Vergnon	Sheffield University
Dr	John	Watkins	CEH
Dr	Duncan	Watson-Parris	University of Oxford
Prof	Tim	Wheeler	NERC
Mrs	Joanna	Whittle	Cefas
Prof	Martin	Wooster	NERC NCEO

Appendix 2 – Outline of the event

Before attending, participants were invited to:

- Identify their reasons for coming
- Outline their work and interests regarding the data centres and services
- Reflect on the feedback they would offer on current data centres and services
- Identify how their use of the data has developed through time, the pressures driving that, and what their requirements may be in the future

09:00	Registration (and informal networking)
09:30	Plenary Welcome – <i>Tim Wheeler (NERC)</i> Introductions and Orientation to the day – <i>Jamie McDonald (JMCD)</i> NERC Data Policy – <i>Mark Thorley (NERC)</i>
10:00	Initial activity - <i>in an interactive process, participants introduce themselves to each other, and capture information on current use of the data centres.</i>
10:30	Data Centres & Recommissioning a Data Service – <i>Stephen Mobbs (NCAS)</i> Available Technology – <i>Bryan Lawrence (NCAS/CEDA)</i>
11:00	Break
11:15	Exploiting NERC Data - A Data Centre Perspective – <i>Matt Harrison (BGS)</i> Data Innovation – <i>Simon Gardener / Robyn Thomas (NERC)</i>
11:45	Trends and the future of data – <i>Trends in data use, as well as the pressures that drive data use and management are explored and possible future forces suggested.</i>
13:00	Lunch
14:00	Exploring the possibility of ‘one data service’ – <i>The implications of an integrated data service are discussed and explored.</i>
14.30	The ideal data service in 5 years – <i>Building on all of the above, the features of an ideal data service are examined and summarised.</i>
15:30	Break
15:45	Feedback & Findings - <i>Spokespeople present their findings from the preceding sessions</i>
16:15	Closing comments - <i>from NERC and course team</i>
16:30	Participant departure (after leaving brief feedback on event)

Please note there were some deviations from the programme above:

- Time for questions/reflection was given after each talk and participants were asked to note their 'positive', 'negative' and 'interesting' reflections after each talk (to set up the afternoon session).
- The precise format for deriving the summary statements in the morning and afternoon sessions varied between sub-groups as the facilitators responded to the needs and nature of their group. This did not affect the outputs, however.
- Time was given, after lunch, to report back summary statements from the morning, and to seek initial responses. This also enabled a clear context to be set for the afternoon sessions.
- The session on the proposal of a single data service was run from the front of the room, rather than in sub-groups, and gave a good energy boost to participants after lunch, and some of the outputs from that were fed back immediately afterwards, rather than later in the afternoon.