



The University of Manchester

Digital Solutions Hub: User Research Report

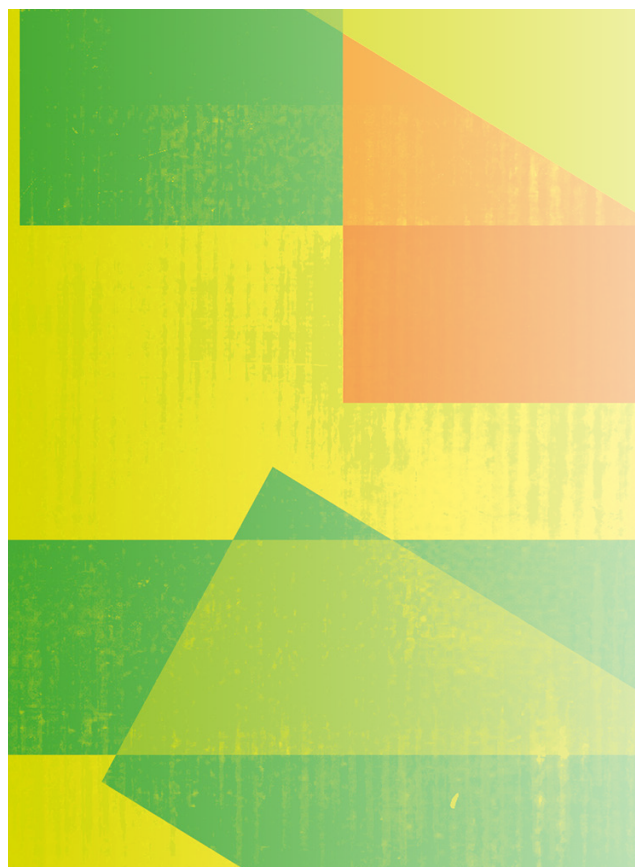


Open Data
Manchester

Contents

1. Executive Summary	01
2. Methodology	05
3. Glossary of Key Terms	11
4. Understanding the Data System	12
5. Data Checklist	13
6. User Stories	14
Summary of User Stories	15
Understand	16
Comprehend	17
Prioritise	18
Find	19
Access	22
Verify	24
Gather	25
Process	26
Test	27
Analyse	28
Present	30
Promote	31
Iterate	32
Manage	33
Supporting the Environmental Data Community	34
Keeping in Mind Specific Needs	34

7. User Archetypes	35
Summary of User Archetypes	36
Analysts Monitoring the Environment	37
Authors of Monitoring Frameworks	38
Analysts Answering Questions with Data	39
GIS Specialists	41
Data Support	42
Data Leaders	43
Data Stewards	45
8. Participants' Suggested Features	46
Appendix: Ways of Clustering Users	52



01. Executive Summary

Context

The Natural Environment Research Council (NERC) hosts a vast array of environmental and earth science data within its data centres, which is related to the work of thousands of scientists undertaking research both within the UK and internationally. Over 40 petabytes of data are combined with NERC's JASMIN supercomputer, which provides a globally-unique data analysis capability covering topics such as climate change, oceanography, air pollution, earthquake deformation, and analysis of wildlife populations.

The Digital Solutions Hub (DSH) is an £8 million investment by UK Research and Innovation (UKRI) awarded to the University of Manchester. Its purpose is to build a gateway to these data and resources, allowing anyone who wants to use environmental data in their work to access the data and leverage JASMIN's computational resources.

The success of the DSH programme relies on the tools, resources, and underlying data meeting the needs of a growing community of potential users. These users recognise that understanding and utilising environmental data can contribute to better policy-making, informed decisions, and increased awareness of how the environment and climate change shape our lives.

Open Data Manchester was commissioned to undertake a programme of user research that delved into the processes, workflows and workarounds of people as they tried to use environmental data, and sought to understand the concerns, challenges and barriers that they faced.

Across the UK, 97 individuals participated in 12, day-long, in-person workshops, and additional online workshops. They generously dedicated their time and provided valuable insights about how DSH might assist them in achieving their goals regarding environmental data usage.

Whilst this work was focussed on use of

environmental data and understanding user needs for the DSH, participants shared knowledge of many projects under way to improve access and use of data across many organisations and agencies. Though some sector or role-specific user requirements are likely, the research clearly points to many shared needs, challenges and hopes of those working with data in a wide variety of roles. This report represents the culmination of that research and the findings should be of value to anyone developing similar data services.

Identifying and addressing the needs of different users

There are a broad spectrum of users and potential users of environmental data working across public, private, academic and voluntary/civil society sectors, their roles range from policy through to analysing and translating insights so that others can understand environmental impacts. Analysis of the research cohort revealed a number of common behaviours and traits these were distilled into seven user archetypes which DSH should primarily be designed for.

These were:

- Analysts monitoring the environment
- Authors of monitoring frameworks
- Analysts answering questions with data
- GIS specialists
- Data support specialists
- Data leaders
- Data stewards

Three further archetypes were identified relating to voluntary/civil society, policy professionals without advanced data skills and IT specialists, but these were under-represented within the research cohort so most traits and behaviours have been integrated into the main seven archetypes.

Significant differences between these groups were:

- Needing to know what data is available to plan and prioritise what can be achieved with data versus seeking out datasets to answer specific questions. The former tended to be associated with more senior roles or people who were supporting others to explore data.
- The distinction between those supporting others to explore data versus answering questions themselves. The former have specific needs around visualising data and trying to move from reacting to requests to helping colleagues self-serve.
- Whilst data management is everyone's responsibility, some people are more directly focussed on this, and have needs associated with instituting good data management systems and cultures in their organisations or networks.
- For those working with spatial data being able to explore trends visually was particularly important. Those working on public health had distinct needs around navigating access to sensitive health data, and combining with other types of data.

Common journeys

The archetypes reveal a set of attributable stories that map out a journey to achieving their goals when using environmental data. The report goes into depth mapping out the user stories and journey, and reveals key findings in the next section of the report, but in summary it was found that user stories contained all or some of the following stages:

- **Define** - Understanding the problem, comprehending the challenge and the resources needed, and prioritising the actions needed to achieve a successful outcome.
- **Get** - Finding the data, assessing its accessibility, verifying it meets the needs of the task in hand, gathering new data if the data doesn't exist and processing so that it is usable.

- **Analyse** - Assessing the appropriate analytical technique to use, testing to see if it works and creating analysis using the chosen method and data.
- **Showcase** - Presenting the insights and outputs of the data analysis created and promoting it to the appropriate audience.
- **Iterate** - Soliciting feedback and integrating into future work.
- **Manage** - Ensuring that data and products are stored and managed in an appropriate way that enables others to discover, access and use them.

Shared challenges

The **Get, Analyse** and **Manage** stages were where most people experienced challenges with using environmental data.

At the Get stage users highlighted a lack of suitable data to satisfy the task in hand, with many datasets being inaccessible due to lack of discoverability, paywalls, the proliferation and confusing nature of different platforms that required bespoke access approaches, inchoate data formats and methods of retrieval and unclear provenance. This was coupled with some users working with locked down systems due to security or data protection requirements.

Within the **Analyse** step, two themes emerged:

- **Technical** - through restrictive IT work environments hindering the use of environmental data, and similarly to the Get stage, a lack of power to undertake computationally intensive processes was also a barrier.
- **Human** - through lack of access to topic or wider analytical expertise restricting the use of environmental data, many people also found it hard to keep up with latest developments.

The **Manage** stage highlighted poorly developed data culture being a barrier in some organisations. This manifested itself in a lack of a coherent and centralised data management infrastructure, prevalence of legacy datasets, difficulty in getting people to openly share data when they could and

the fear of inadvertent data disclosure or overzealous data protection regimes stymieing data reuse.

Summary requirements

There are a number of key areas which the DSH must address to meet the needs of its potential users. Some of these are technical requirements which should be designed into the build of the hub, whereas some of these are broader, addressing social or cultural aspects, or contributing to the wider data ecosystem the hub will exist within.

While these requirements are discussed in detail throughout the body of the report, here we summarise key elements the DSH team should incorporate into the development.

Reputation, awareness and coordination

To drive engagement and use the DSH must clearly articulate what it does, what data it contains and who it is for. Clear communication, marketing and SEO to ensure potential users know about the hub, can easily find it and make regular use of it, is critical. Ensuring the hub has a reputation for quality data, ease of use, high quality tools and good data management practices will be key to building trust and adoption.

A common concern was around the lack of coordination around how data is created and used and worries about duplication of effort so the DSH should work closely with other key stakeholders to coordinate efforts to ensure the DSH can play a crucial role in enabling greater access to and use of their data.

Promoting better data, better meta data and making more data available

As the DSH is not producing data it has limited ability to directly improve the quality of the data available, create more data or improve metadata, however it could use its influence to promote best practice in these areas and support the wider data ecosystem in improving the overall quality and quantity of data and associated metadata.

Creating and supporting communities of practice

The technical capabilities of the DSH are a key draw for users, but the development of a community of practice alongside the hub was seen as a significant benefit. Providing users with a way to connect with others working on similar challenges to share knowledge, advice, scripts and templates, as well as sharing 'analysis ready' datasets to minimise duplication of effort would add a lot of value to users and embed the DSH in their working practices.

Discoverability, exploration and weeding out

Finding what data is available in both a targeted and exploratory way was something many users currently struggle with, building in sophisticated search functionality along with the ability to filter based on a variety of parameters so they can weed out irrelevant data will help users to more easily find the data they need. Filters could include paid/free, topics, geographies, publishers, resolution, 'coming soon', verified etc. Users also talked about the usefulness of a 'smart suggestion' feature where the DSH might return results which others working on similar tasks had used (akin to the 'people who bought x also bought y' that many ecommerce sites use).

Working with the data

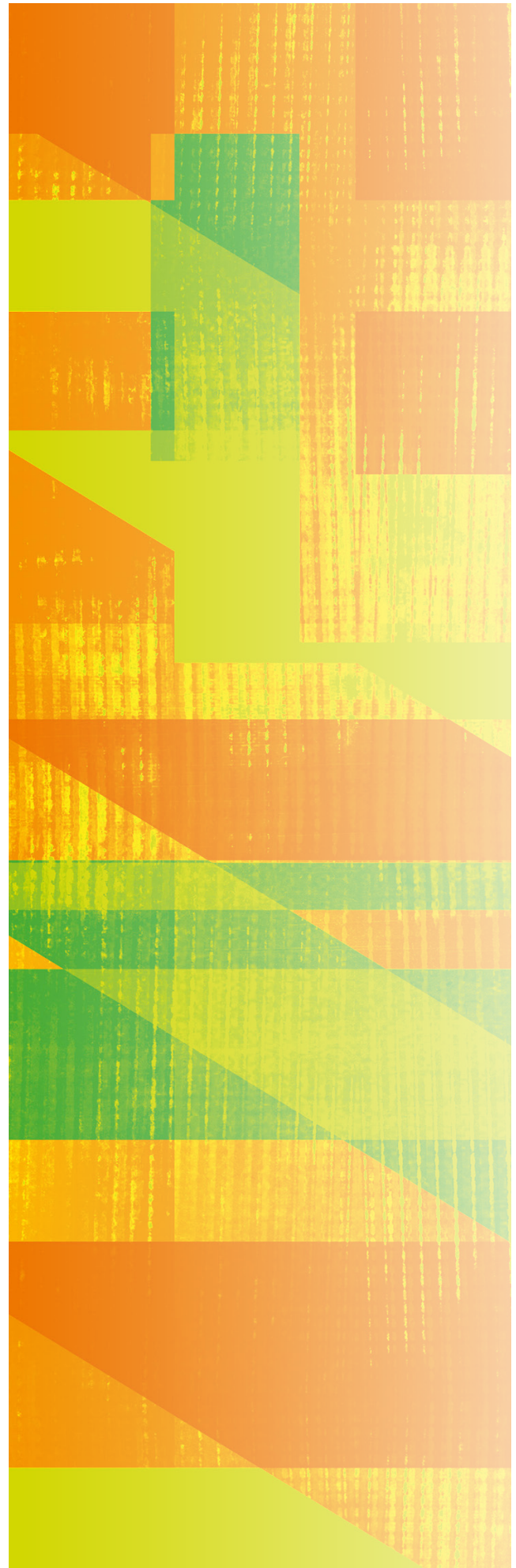
To be of use to the widest group of users possible the DSH should both enable users to download datasets to be processed and analysed locally, and also allow users to work on the data within the DSH. Some users want to be able to upload their own data to the Hub, some just to combine it with DSH data for analysis, some to publish it for others to use. Being able to keep track of what data they have used and what processing they have done is an important feature for many.

Easy to use, stay informed and understand the impact

Users need to be able to quickly view and assess if data on the hub is suitable for their needs, ideally users want to be able to see or use a sample of the data to check it meets their requirements before committing more time. Any barriers to this, like requiring logins or putting basic information behind paywalls will be a barrier to adoption. Logins for accessing full datasets, carrying out processing and publishing data is not an issue.

Once logged in the DSH should provide users with notifications about datasets they use (e.g. if they are updated, substantially changed, retired etc), if 'coming soon' data is now available, if they have been granted access to data that is not openly available or if another user is requesting access.

Users are also keen to get data and feedback on the ways in which data products they have published on the hub are used by others so they can evidence impact in their organisations.



Our aims

The development of the Digital Solutions Hub (DSH) offers the potential to harness the breadth of data held by NERC with the power of a suite of computational tools. This will support businesses, governments, and society to understand the complex environment we live in and make appropriate and valuable decisions that will benefit people, our planet, and prosperity.

To make this a success, it is important to start by understanding who the potential users of the DSH are and what their needs are. These were the overarching questions for this piece of user research.

For the DSH to work well, it cannot be just a stand-alone product, but must sit within a community of people who are aware of it, use it, and help shape its development. So our approach was designed to also help develop this community – raising awareness of the DSH and forming links with potential users who could shape its ongoing development.

To achieve both these aims, Open Data Manchester ran a series of 16 online and in-person workshops across the UK for

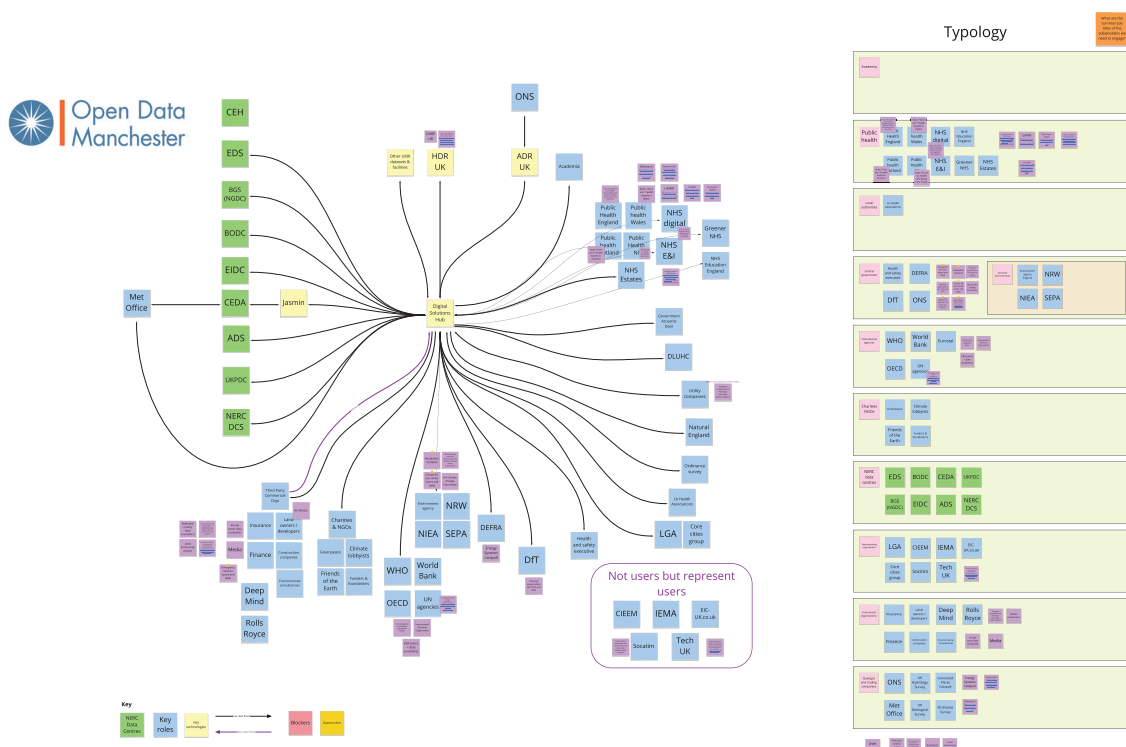
potential users of the hub. The workshops focused on participants' experiences of utilising environmental data and the challenges that come with its use. This was then used as a basis to discuss potential solutions and future developments.

Designing our approach

To understand the scope of the user research task a stakeholder mapping exercise was undertaken with the project team at the University of Manchester, and key organisations and networks that could potentially interact with the Digital Solutions Hub were identified. This exercise was necessary due to the project's complexity and the need to identify and understand the many different types of users who could benefit from it. The mapping exercise was conducted online using the Miro collaborative workspace to ensure effective collaboration and communication among the project team members.

After conducting the initial stakeholder

Figure 1 (below): Stakeholder mapping and typology exercise



What we know – evidence • What we think we know – assumptions • What we don't know - unknowns



Figure 2 Assumptions map

mapping exercise, the project team developed a typology of user types to identify the specific individuals and organizations that would benefit from the planned workshop programme. This typology also aided the team in monitoring engagement and allowed them to target under-represented user types during the workshop programme.

To ensure the success of the workshop programme, Open Data Manchester's user researchers collaborated with the DSH team to identify the assumptions they had about potential users. A mapping exercise was undertaken that identified what was known, what was assumed, and what was unknown. The results of this exercise were then used to define the form and narrative structure of the workshop programme.

The researchers also considered the outputs needed to answer the identified questions and designed an approach that would allow them to create:

- A set of archetypes or personas representing the diverse potential users and their different requirements.
- A set of user stories and an overarching epic, the approach aimed to convey

what users aimed to achieve and the requirements they needed to meet along the journey.

To create a workshop structure that would effectively address these needs, a series of workshop activities and a day-long workshop structure were developed. The team used four online prototyping workshops to test out their hypotheses and iterated the workshop model based on feedback received. Additionally, they used the first in-person workshop in Manchester to test how the activities worked in a physical environment. The feedback from this workshop was used to create a final iteration that was then implemented throughout the rest of the workshop programme.

Participants were recruited through a process of targeted recruitment that leveraged the networks that the University of Manchester and Open Data Manchester were involved with and open calls that were publicised through social media platforms such as LinkedIn and Twitter.

Workshop exercises

A series of activities was designed that would guide participants through a journey over the course of the day, encouraging them to share their own experiences with one another, and with the user research team. Effort was made to design a process that enabled participants to feel welcome and comfortable with a light-touch, quirky Data Explorer exercise that helped reveal who they were, their skills across 5 dimensions and an opportunity to draw a giraffe. The exercise acted as a warm-up and ice breaker.

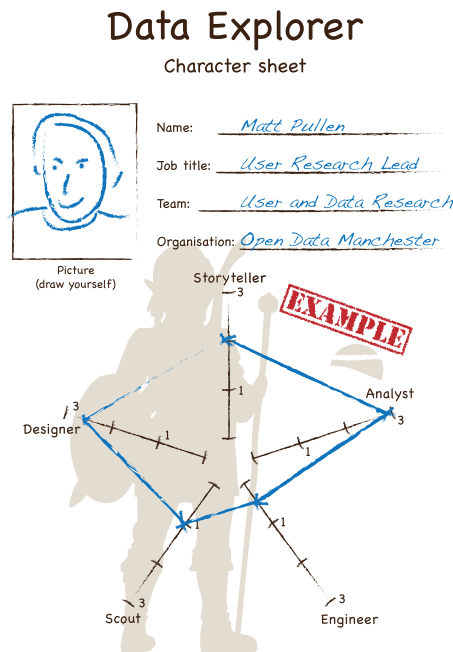


Figure 3 Example data explorer exercise.

Participants were then asked to describe their processes and the journey relating to environmental data use. This involved participants describing the steps they undertook and the processes, data and tools used, giving their thoughts at different stages of their workflow.



Figure 4 Example experience map front

This was followed by an exercise that asked them to identify the challenges faced when trying to use environmental data, their current workarounds and ideal solutions. Participants were also asked to identify the barriers to implementing their ideal solutions.

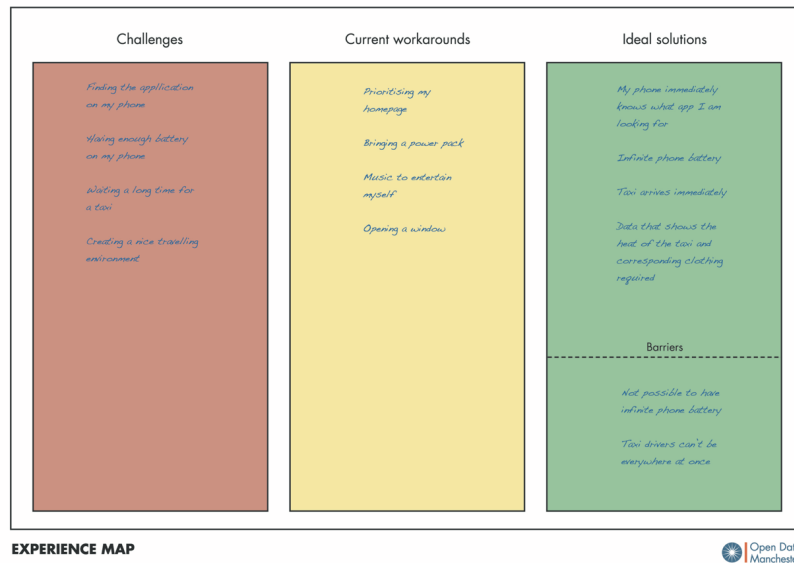


Figure 5 Example experience map back

Participants were then introduced to possible forms that the Digital Solutions Hub could take which led to a final reflection exercise that asked participants to think about ideal tools, data and solutions that the DSH could make available to help people use environmental data.

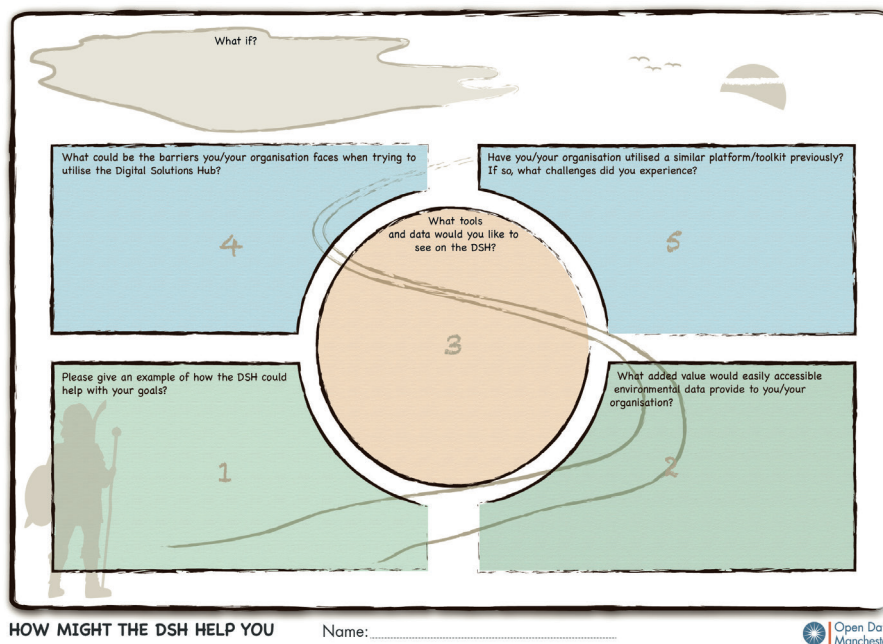


Figure 6 Example requirements map

Workshop programme

The programme sought to engage and listen to the voices of potential users from across the four nations through 12 day-long workshops undertaken across the cities of:

1. Manchester
2. Leeds
3. Newcastle
4. Plymouth
5. Bristol
6. Southampton
7. Norwich
8. London
9. Glasgow
10. Belfast
11. Cardiff
12. Birmingham

In total 97 people engaged with the in-person workshops representing, governmental departments, devolved administrations, local government, NHS, business, academia and civil society. The workshop programme also encouraged people to sign up and engage with the Digital Solutions Hub programme as it developed.

Who took part?

97 people participated in these workshops. A breakdown of types of organisation represented is shown below. 5 of these participants attended from the Natural Environment Research Council (NERC) itself; a small number of the academics listed also had some link to the project.

A vital part of making the Digital Solutions Hub a success will be working with key stakeholders in the environmental data community, such as those managing NERC data centres. Our focus here is on describing potential users of the hub rather than mapping out these stakeholders.

Type of organisation	Number of participants
Charity	6
Community sector	3
Private company	6
Environmental data centre	4
Data-focussed quango	9
Environment-focussed quango	16
Government department	14 ²
Local Authority	12
Natural Environment Research Council	5
Other Research Councils	2
Parliamentary support	1
Other data provider	1
Public health body	5
Academic	12

²One also worked for a quango but is recorded here.

While those who attended broadly matched the targets that were identified there was a slight skew towards those in more senior roles. It is also worth noting some groups who may be relevant but were not focussed on:

- Users from community groups, small charities, campaigning groups, or interested members of the public – although a small number of potential users in this group did take part in this research, this is not sufficient to properly assess the needs of a potentially very broad group. Doing so would require a broader piece of research that is directly focussed on this group.
- Policy professionals without advanced data skills – some of those we spoke with were supporting policy colleagues to explore data, and there was a suggestion that more might be done to enable this group to ‘self-serve’. Exploring needs around this with policy professionals could be an interesting area of further investigation.
- IT specialists – this research focussed on the needs of those who are working more directly with data, whose needs the hub will be designed to meet. However, working closely with IT specialists within the users’ organisations will then be an important part of meeting these needs effectively.

A set of archetypes is included in this report to describe the variety amongst users. A small number of participants are included in the findings but not represented by specific archetypes.

Developing the outputs

The learnings from the workshops were used to draw out key insights and develop the outputs shown here. Participants were first grouped by focussing on the information they’d given about their work, how they use environmental data, and their journeys of using environmental data.

From this information archetypes were created. **Archetypes** were chosen rather than personas to make observations about important differences amongst

types of users rather than generating detailed fictional characters that would encompass each and all potential users. These archetypes are based around the goals a type of user is trying to fulfil and the subsidiary needs associated with fulfilling this goal.

A pattern-matching process was used to analyse participant’s journeys – the steps they were going through, and the challenges they faced along the way. This revealed considerable overlap in the journeys of different users which led to the creation of an **assimilated user journey**. Although the assimilated journey was common to all archetypes, there were still unique processes or actions that needed to be described as **unique user journey elements**.

A single set of user stories that described the stages found within these journeys and the **user needs** at each of these stages was developed. Particular focus was on understanding and distilling the **challenges** faced and **requirements** for success.

This overarching user journey is combined with a set of archetypes that convey the diversity that exists amongst potential users of the hub. These describe in a more summarised way their different goals, and the steps they go through to achieving these, as well as giving further detail about the user types within these groupings.

It is intended that these outputs are a tool that reveals the thinking and motivations driving different potential users of the Digital Solutions Hub, which should be developed further over time. Identifying the goals of different user groups, alongside common steps that different users go through, the challenges they currently face, and some of their key requirements has been the primary goal of this research, rather than what users think they want from the DSH.

Glossary of Key Terms 03.

Some of the terms used in this report may not be familiar to all audiences, or have definitions that are contested or used differently by different communities. To avoid confusion, we've included a glossary of how we've used some key terms below. Links for further reading about these concepts have also been included.

Data product – we've used this term as a broad way of referring to both datasets, and the tools that are created based on these. This could include, for example:

- An individual dataset.
- A tool that helps user explore datasets such as a GIS product or a dashboard.
- A wider range of tools that enable users to get relevant answers or outcomes via the use of data.

In talking about the creation of data products we intend to encompass in a broad way:

- Gathering new data
- Combining existing data
- Translating complex data into something that is easier to grasp and use, for example showing a continuous variable as broad categories.
- Drawing out data of most relevance to users within large datasets.
- Analysing data to identify significant patterns and relationships, and creating models simulating what will happen under different conditions.
- Ways of presenting such information and insights, including creating tools that allow users to readily explore datasets, or that provide relevant answers or outcomes for users via the use of data.

Further reading:
<https://bit.ly/ODM-DSHL1>

Data management – creating and managing systems that ensure the creation, storage, sharing and use of data within an organisation or network works effectively.

Further reading:
<https://bit.ly/ODM-DSHL2>

Data governance – setting and enforcing standards and other policies for ensuring that the way data is created, updated, stored, and shared are appropriate.

Further reading:
<https://bit.ly/ODM-DSHL2>

Metadata – information that describes what is within a dataset, such as how data was collected, what resolution it is at, what format it is in, and how the dataset is structured.

Further reading:
<https://bit.ly/ODM-DSHL3>

04. Understanding the Data System

This research has not focussed on mapping out the data ecosystems that the Digital Solutions Hub must work with and become part of, but instead looked at its users and the systems in which they operate. This is not to ignore the vital importance of attending to the wider landscape and as the hub develops it will be critical for DSH team to continue to engage with stakeholders across the data ecosystem to maximise interoperability, collaboration and integration and to prevent the DSH becoming 'another' separate system or platform in isolation.

Here the focus is on the system within which users work, the process or data pipeline they navigate to achieve their goals. To help communicate how different parts of this system fit together we have mapped out a simple picture of the different processes

that go into making this system work. Although shown in a linear way these aren't always sequential steps. Users may only perform a subset of these steps, or may go through numerous iterative loops. Starting with this simple picture may make it easier to reflect on the more detailed information that follows in this report.

Throughout these steps, users are reviewing data and assessing its suitability for the task in hand. While the standard to be met varied from organisation to organisation or from task to task, there was broad consensus on the kinds of questions that needed to be answered in order to reach a decision on verifying a dataset's acceptability. These criteria tell us both what users require of datasets and what they need to be able to ascertain about datasets to tell if they will meet their needs.

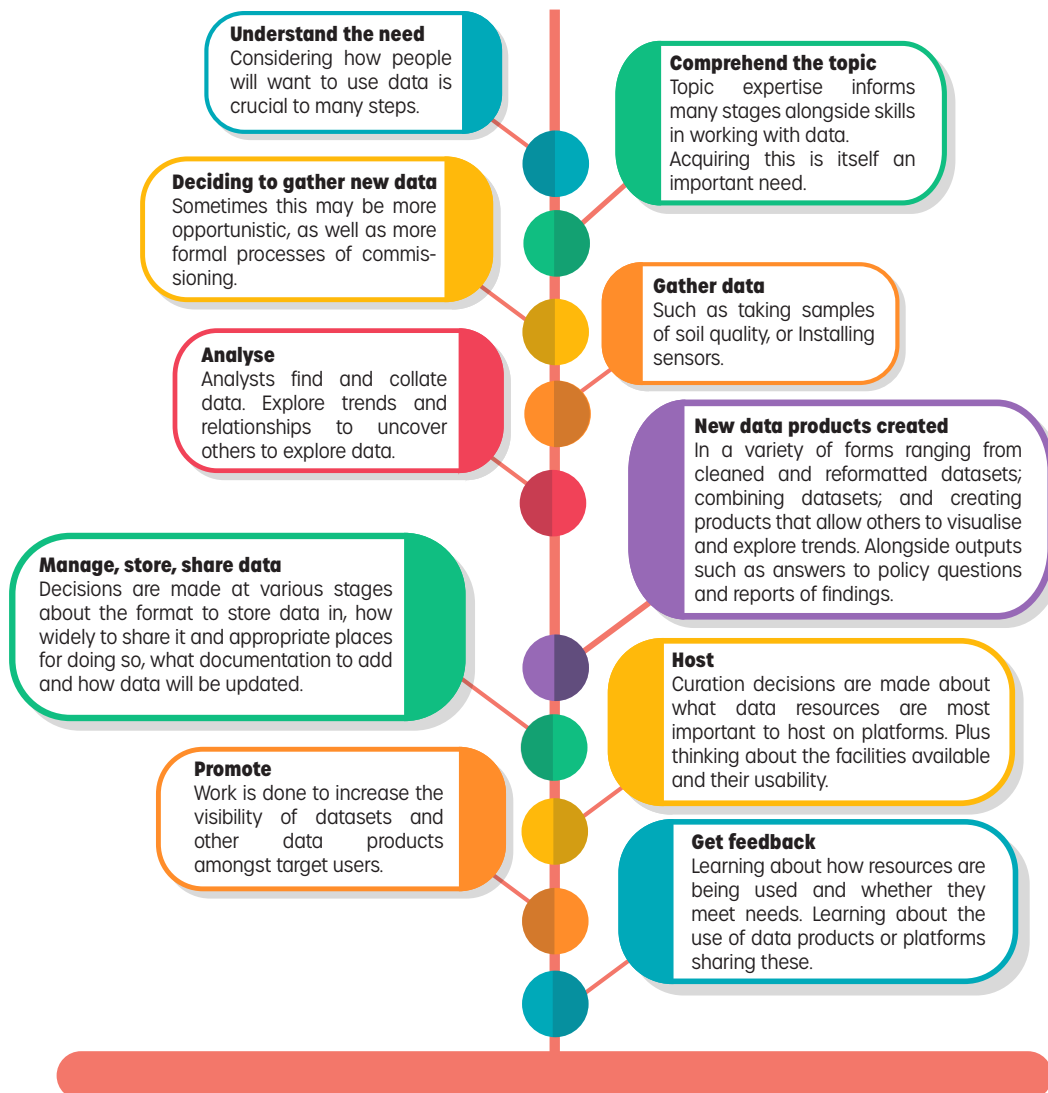


Figure 7 Data system or pipeline users operate within

Users want to be able to ascertain this information as readily as possible – such as knowing that data from particular sources reliably meets certain standards, or being able to find out key information through metadata, rather than taking more onerous steps such as downloading large datasets to quality assure them. Having contact details for someone responsible for the data was seen as a way of being able to fill any blanks in understanding the data, where necessary.

What the data covers

For example:

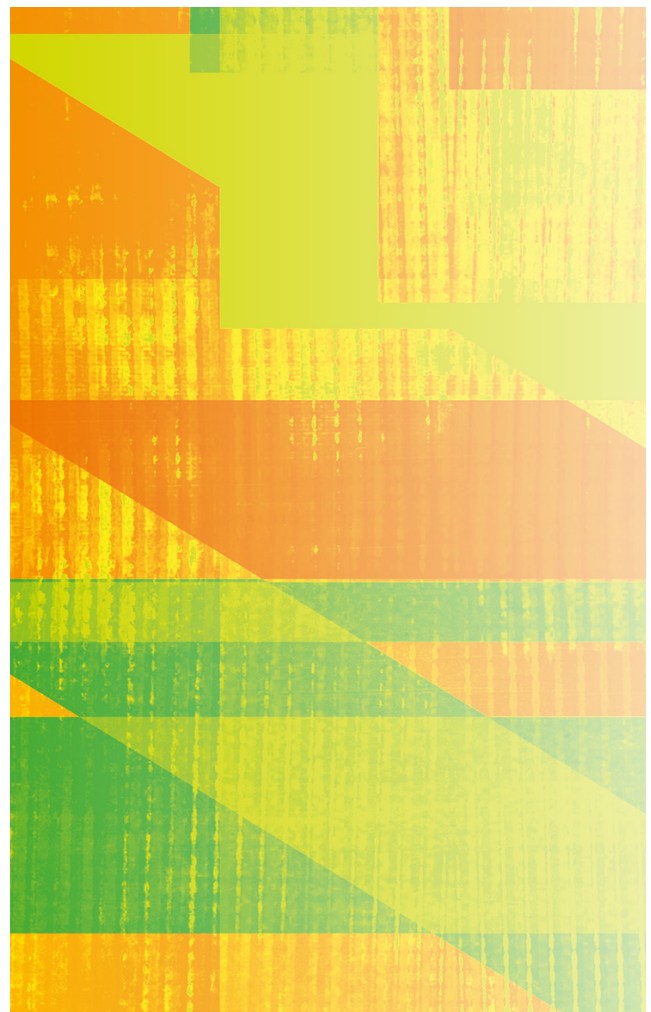
- At the required level of geographical resolution (including sufficiently consistent resolution across different datasets being brought together).
- Covering geographical areas of interest, which do not always map onto administrative boundaries.
- At a comparable resolution to other kinds, such as “species-level” data.
- Able to show trends over time.
- How up to date is it? Is there a later version? When will it be updated?

The quality and format of the data

- Understanding how data has been gathered, and whether this provides sufficient confidence when used in the way I intend. Being able to trace back to original sources is an important part of this.
- Can I be confident that the data meets high enough levels of accuracy and consistency, if not, is it feasible to quality assure and clean it?
- Is the data in an appropriate format for my needs, if not, is it feasible to transform it?
- What does a dataset contain? How is it organised? (Making this information readily available can speed up time getting to grips with a dataset and help ascertain how feasible it is to use).

How data can be accessed and used

- Can I, for example, create stable links to the data, or download it?
- Is it manageable to access and use the required parts of a dataset within local computing capacity, IT policies, and available staff time and skills?
- Will it become available within required timeframes (if not already available)?
- Is it free or affordable?
- Can I use it in the way I need, such as being publicly shareable (both legally and ethically), including risks of accidental disclosure?



06. User Stories

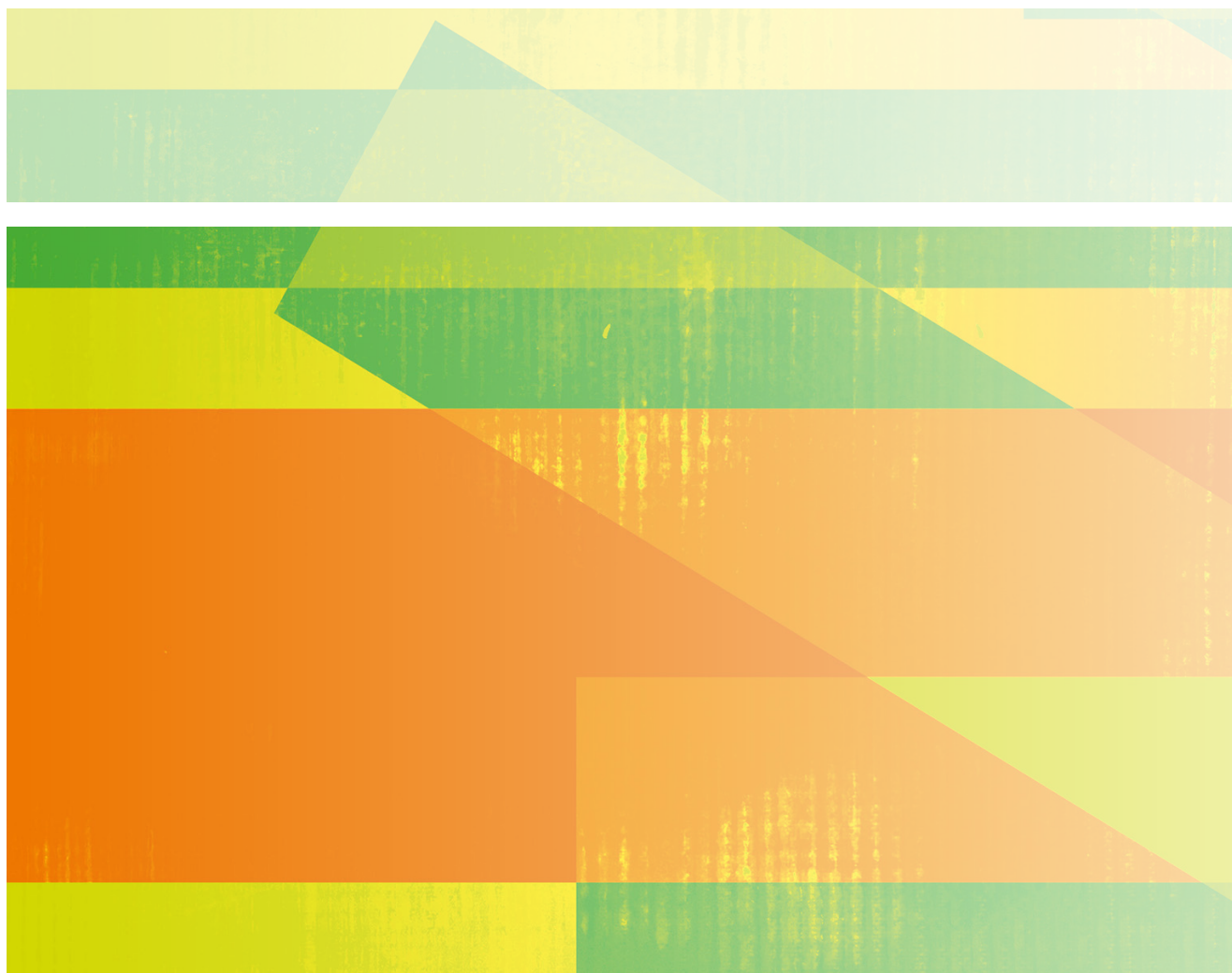
What is Shown Here

These user stories describe the steps that users of environmental data go through as part of their journey of working with environmental data. They describe a series of needs that have to be met in order to achieve their goals. They also show the challenges currently faced for achieving these, and a set of requirements for enabling these to be met through the Digital Solutions Hub.

Of course, there is variety across those we met with about the goals they are aiming for, and the needs they face along the path to achieving these. However, we found considerable commonality within the journeys that different users go through. For this reason we have presented user stories as a single set, that summarise what we found across all research participants. We have then broken down some of the differences between different users, including their overall goals, within the following set of user archetypes.

These user stories are loosely ordered into an overarching journey or 'epic', however the order they appear here does not always reflect reality. Sometimes they happen in a different order for some user groups, some stages may also overlap or are ongoing rather than straightforwardly being a sequence of steps. Not all users have all the needs shown here.

Some of these user stories are more relevant for the design of the Digital Solutions Hub than others, we have highlighted these key stages with an orange page background.



Summary of User Stories

FUNCTION	USER STORY	USER NEED
DEFINE	Understand	I must understand what need I am trying to meet in order to begin designing an appropriate data product.
	Comprehend	I need know enough about the field I'm working in to understand what data will meet my needs.
	Prioritise	I need to decide where to focus my efforts so that I can achieve the most impact within limited means.
GET	Find	There are some different starting points here, which come with slightly different needs: a) I have a specific goal, such as answering a specific policy question. I need to find sufficient data to meet this aim. b) I have a somewhat defined aim. I need to understand what data is (or will be) available, and where the gaps are. c) I want to explore existing data around a broad topic in an open-ended way to seek out interesting trends and correlations. I need to find data that will allow this.
	Access	I need sufficient access to the data for creating my final outputs.
	Verify	I need to establish as quickly as possible whether datasets I find are suitable for my needs.
	Gather	I need to gather or commission new data to fill shortcomings in existing data.
	Process	I need to check the data is accurate and consistent enough. I may need to clean the data to achieve these standards of consistency and accuracy. I may need to transform the data into a suitable format.
ANALYSE	Test	I need to test out how I'll analyse the data to check if it will work in the ways I require before committing to analysing all the data in this way. I need to mock-up outputs so I can get feedback from users and check whether these outputs will meet their needs, before developing these fully.
	Analyse	I need to analyse data to get the insight I require from the data, or to establish trends and relationships that it will be useful for end users to explore themselves.
	Present	I need to present the insights I have drawn from the data to my intended audience in a way that is compelling and sufficiently easy to comprehend. I need to develop ways of presenting data that allow others to readily draw out insights that are relevant to their needs.
	Promote	I need to make those who will benefit from my outputs aware of them. Doing this will allow me to achieve impact from my work (and may also allow me to evidence this to those I am accountable to).
ITERATE	Iterate	I need to receive feedback from the end users of my outputs so I can continue to develop these outputs to best meet their needs.
MANAGE	Manage	I need to ensure data products are stored and managed in a way that allows them to be readily discovered, accessed and used by myself and others.

Understand

User Need

I must understand what need I am trying to meet in order to begin designing an appropriate data product.

This is done when...

- I have an adequate understanding of what information is required, and for what purpose.
- I have an adequate understanding of how this information needs to be presented to meet these needs.
- I may need to balance the needs of different users and prioritise based on how I can have most impact or add most value.
- In some cases a requirement is achieving a sense of 'co-ownership' or 'partnership' with end users.

Steps Taken

Steps I may take to do this

- This varies considerably across different users, but may include:
 - Speaking with colleagues or partners in policy teams.
 - Speaking with colleagues or partners managing physical environments on the ground.
 - Engaging with a wider range of affected stakeholders.
 - User research or audience research.

Some of this work is reactive rather than proactive. Some users were trying to balance responding to requests alongside taking a more proactive approach.

This task is not always a one-off process but can be an ongoing part of a journey that involves some degree of partnership with end users.

Challenges

What is hard about doing this?

- Getting an adequate understanding of needs, particularly where users are a diverse group.
- A lack of understanding or consideration amongst end users of the challenges of using data. This creates work in translating their requirements into something realistic from the data available.
- Time spent responding to fragmented requests rather than being able to take a more proactive and prioritised approach that then allows end users to self-serve.

Requirements

Requirements for the Digital Solutions Hub

- This is a stage that would happen prior to engagement with the hub. Requirements here are around how users interact with their colleagues and other end users of data products.

Comprehend

User Need

I need know enough about the field I'm working in to understand what data will meet my needs.

This is done when...

- I have adequately understood what type of data will allow me to perform the analysis and presentation required to arrive at data products that meet the needs of end users. Including:
 - Understanding levels of confidence required, and how these relate to methods used in collecting the data.
 - Understanding topic-specific language sufficient to make sense of the data.
 - Understanding how to approach interpreting and analysing the data (see Verify and Analyse).

Typically, this is a stage that happens early on in journeys, but these aims are also supported by checking back in to get feedback on early tests.

Steps Taken

Steps I may take to do this

- Reaching out early on to topic experts within or outside my organisation. Often there were other reasons to reach out to this group at the same time.
- Conducting literature reviews of existing studies.

Challenges

What is hard about doing this?

- This requires bringing together both analytical skills and topic expertise. There are different tactics for this. Some emphasised developing both within one person, with analysts who have both statistical skills and topic expertise – this takes time and staff retention can make this a challenge. At other times analysts were reaching out to topic experts to combine their knowledge with the analysts data skills – challenges here were around knowing who to contact and being able to secure their time.

Requirements

Requirements for the Digital Solutions Hub

- Supporting users to connect with topic experts, and to other sources of guidance, about the topics they are working on.

Prioritise

User Need

I need to decide where to focus my efforts so that I can achieve the most impact within limited means.

This is done when...

- I have a decent understanding of why particular information is required.
- I have a decent understanding of what data already exists and where the gaps are.
- I have a decent understanding of what I can achieve with the resources available.

Steps Taken

Steps I may take to do this

- Investigating the requirements of end users (see Understand)
- Finding what data is available (see Find)
- Investigating what data will meet my needs (see Verify and Data checklist)
- Deciding how to approach analysing and presenting the required data and estimating the work needed for this.

Challenges

What is hard about doing this?

As for:

- Understand
- Find
- Verify

Requirements

Requirements for the Digital Solutions Hub

- Largely covered in other sections, especially Find and Verify.
- Enabling users to communicate more easily with others who are working with data to coordinate efforts, minimise duplication of work and maximise added value.

Find

User Need

There are different starting points which come with slightly different needs:

- a) I have a specific goal, such as answering a specific policy question. I need to find sufficient data to meet this aim.
- b) I have a somewhat defined aim. I need to understand what data is (or will be) available, and where the gaps are. This will allow me to decide what can feasibly be achieved in relation to this aim. With this knowledge I can prioritise where to focus my efforts, including knowing what cannot feasibly be achieved. I may be able to commission new data if I am confident that the data I require is not already available.
- c) I want to explore existing data around a broad topic in an open-ended way to seek out interesting trends and correlations. I need to find data that will allow this, as well as being able to readily explore trends and relationships across existing datasets.

To some extent this aspect of how users work with data is inseparable from that described below under the heading 'Verify'. However, we have separated them to focus on specific features of each.

This is done when...

These different needs, come with different endpoints:

- a) I have access to sufficient data for my needs. (Access may take different forms, such as downloading, or linking to, data).
- b) I am sufficiently confident that I understand what data is, and isn't available, to inform decisions about where to concentrate efforts or where to commission new data gathering.
- c) I have been able to find sufficient data to allow me to identify trends and relationships within existing datasets. (Meeting this need is broader than just finding data, but is a particular way of approaching this 'Find' stage).

Steps Taken

Steps I may take to do this

- Speak to colleagues, partners, or external experts to understand what already exists (or will soon exist) and where to look, including data that is not currently public.
- Search internal data stores.
- Search known reputable sources of data.
- Search the web.
- Review the data I do find to assess its suitability (see 'Verify' and 'Process')
- Carry out a literature review.

Find cont.

Challenges

What is hard about doing this

There is a lack of required data

Data is collected for lots of different reasons, resulting in a 'patchy' landscape in which it's hard to get the data you need.

It is particularly hard to:

- Get data at an appropriate resolution, particularly at a more local geographic scale. Matching similar resolution between datasets is a key challenge.
- Show change over time, this needs consistent data to continue to be available.
- Find data matching the areas you are interested in, particularly where these don't match common administrative boundaries.
- Find data for some geographic areas which are poorly served, this was a particular issue in Northern Ireland.

Paywalls are barriers to access and discovery

- Having to pay for data is a barrier to using it for some groups.
- Paywalls can also be a barrier to discoverability, respondents talked about data being 'tucked away' behind paywalls.

Data is held in lots of different places, and is not always easy to find and access

- There are lots of places data is held, rather than an easily searchable 'one-stop shop'.
- It's hard to keep up with the 'sheer range' of different platforms.
- It can be hard to work out the purpose of individual platforms and the variety of data they contain.
- Some platforms are 'clunky' to learn and use.
- Data can be held in formats and on systems that make it hard to search for (for example field notes, journal articles) or are only known amongst specific academic circles.
- It can be hard to uncover the provenance of datasets, which leads to a risk of using several sources that are all drawing from the same underlying data.
- Uncovering the data used in visualisations, and its provenance, is often hard.
- Information is currently held inside people's heads. This requires work to make connections with others. And this is problematic when staff within organisations move on and take knowledge with them.

Find cont.

Data coverage:

Requirements that are currently hard to meet are:

- Datasets with sufficiently granular resolution (particularly at a local geographic scale).
- Datasets showing change over time.
- Datasets covering the geographic areas users need.
- Datasets that can readily be combined with others (in terms of extent covered, and similar enough resolution).

More work will be needed to assess which specific datasets are required on an ongoing basis.

Navigating paywalls:

- Users need to readily know whether they must pay to work with specific datasets or not.
- Some users require data to be free and will want a way to easily ignore paid for data

Awareness of the DSH:

- Users need to know of the DSH as a trusted source of data.
- Relevant datasets on the DSH need to appear in web searches to aid discovery (what one respondent called 'google-ability').

Where the DSH fits in the landscape:

- It should be obvious what the purpose of the DSH is and the type of data it contains. Some themes from respondents about the suggested purpose and the 'unique selling point' (USP) of the platform are shown below.
- The DSH should not increase duplication but instead add value by coordinating effectively with other platforms.

From participants' comments some suggestions emerged around what the role of the DSH should be:

- A reliable source of 'analysis ready' data that meets high standards of quality, including clear metadata.
- Bringing together data from different fields in one place.
- Moving away from a 'fragmented landscape' towards a 'one stop shop' for data. It was suggested that this could include paid for data, as well as free data.
- Contributing to coordination with the data system.
- Sharing data deemed valuable by academics that is not currently 'on the government's radar as important'.
- A source of data that could support the development of a digital twin, or inform AI.

Find cont.

Requirements for the DSH

Discoverability of data on the DSH:

Discoverability of data on the hub will be a key area for further investigation. Important requirements within this are:

- Finding datasets relating to areas of interest, including via drawn shape. Doing this with minimal work, for example being able to upload shape files for this purpose.
- Finding datasets at appropriate levels of resolution, and matching similar levels of resolution across relevant datasets.
- Being able to readily identify the provenance of data, including that used in visualisations.
- In some cases, wanting to explore trends and relationships across datasets relating to a broad topic or geographic area in a more open-ended way.

Access

User Need

I need sufficient access to the data for creating my final outputs. This access will typically need to allow for:

- Quality assuring the data
- Transforming the data into a suitable format for analysis or presentation
- Combining it with other datasets
- Conducting analysis
- Creating, and maintaining, final outputs

This is done when...

This varies depending on the mode of access required:

- The dataset is downloaded onto a local system that is capable of performing the transformation and analysis required.
- Users are able to perform the processes outlined above, and keep sufficient records of their work, without copying the data locally.
- Users are confident in their ability to link to the data over extended periods of time.

Steps Taken

Steps I may take to do this

There are quite different forms this may take:

- Some may download the data to use locally.
- Others may be able to meet their needs without copying the data locally.
- In some cases users are linking to data over extended periods of time.
- Where data is not publicly available users may need to request access. For examples, it was a common requirement for those working on public health who are working with sensitive personal data.

Access cont.

Challenges

What is hard about doing this?

- Data is held in formats, and on systems, that prevent interoperability. There are not the facilities to access data via APIs or persistent links.
- Not being able to rely on links or APIs to continue to function stably, or that notifications will be given if any changes are made.
- Limited computing power locally, and restrictive organisational IT policies (these were particularly problematic in the public sector).
- The size, and complexity, of datasets makes these complicated and computationally demanding to work with.
- Some platforms are 'clunky' with multiple clicks, or arduous registration processes needed before accessing data.
- Access to some datasets needs to be controlled either due to sensitivity of the data, or to protect intellectual property.

Requirements

Getting at the data

- Working with the computing resources, skill and time available, I need to be able to either:
 - Download the data I require onto local systems.
 - Link to datasets over extended periods of time, knowing that APIs or links will not change without me knowing.
 - Carry out required processes with data online through the DSH: quality assuring, cleaning, transforming, combining, analysing, creating and exporting or sharing final outputs.
- Breaking up datasets into smaller or simpler parts may help me achieve this.
- I may need to rely on receiving updates if a dataset I have used is changed or updated.
- I need to keep track of the datasets I will be using.

Navigating restricted access

- I may need a way of easily requesting access to datasets that cannot be shared publicly.
- As well as personally having access, I may need to be able to straightforwardly share this access with other colleagues or partners.

Easy to use

- I am likely to use several platforms to get the data I need, so it must be quick and straightforward for me to work out how to perform these steps and carry them out. There are distinct stages of my journey that can be designed for: find; verify; quality assure, clean and transform; test the water; and analyse.

Verify

User Need

I need to establish as quickly as possible whether datasets I find are suitable for my needs.

This is done when...

- I have an adequate understanding of:
 - What the data describes.
 - Its geographic and temporal extent.
 - What resolution there is in the data.
 - The currency of data, and when later versions will be published.
 - What level of confidence there is in the data when used as I require. Including understanding enough about how the data has been produced.
 - What is the quality of the data, including the accuracy and consistency of the data contained.
 - What format the data is in.
 - How large or complex the data is.
 - Any restrictions on how this data can be used and shared.
- See also Data Checklist.

Steps Taken

Steps I may take to do this

- Reading metadata.
- Reaching out to someone who knows the data more intimately, such as the team who created it.
- Looking at the data itself.

Challenges

What is hard about doing this?

- Adequate metadata is often lacking.
- A lack of consistency in metadata is a barrier to quickly and easily assessing data.
- A lack of information about provenance and unclear information about data licensing were pronounced frustrations.
- It can be hard to interpret metadata when working on areas you are not a specialist in.
- Not knowing who to talk to about a dataset.
- It can be hard to find the original data that has been used by some data products.
- Having to go through several steps before being able to do this, such as several clicks online, registering on a platform, or downloading a dataset, only to realise the data does not suit your needs.

Requirements

Good quality, consistent metadata

- Metadata that conforms with good practice standards (more detail on required information is included in the full report).
- Sufficient consistency in metadata across the DSH, and similar platforms, to make this easy to read for humans, and for machines.
- Knowing that data on the DSH meets certain standards would remove some verification needs.
- The ability to record metadata in two languages is an important need for some Welsh users.

Gather

User Need

I need to gather or commission new data to fill shortcomings in existing data.

This is done when....

- I have access to data sufficient to my needs, including data that is of good enough quality and in a format suitable to my needs.
- This has been achieved within the resources available.
- Any legal or ethical restrictions on the use of this data can be appropriately met when the data is used as I require.

Steps Taken

Steps I may take to do this

- Those in more senior roles may be in a position to commission the collection of new data where required.
- In some cases this may involve seeking out new sources of funding, or developing new partnerships.
- In other cases the creation of new data is a main focus of their role, within set requirements.
- Some users were looking creatively for low-cost opportunities to gather new data, such as enabling users of data products to add data on their own projects into these resources.

Challenges

What is hard about doing this?

- Having the skills, mindset and organisational culture to realise creative opportunities for further data gathering.
- There were some concerns about a lack of coordination in the wider sector around what data gets created.
- The challenge of over-viewing what already exists and not wasting resources reinventing the wheel.
- Publishing and maintaining newly created datasets and governance around their use (see 'Manage').

Requirements

Requirements for the Digital Solutions Hub

- Largely beyond the scope of the DSH, though making it easier for users to assess what data already exists and its suitability may make the process of deciding to commission new data more efficient and effective as gaps would be more readily apparent.
- There was a suggestion that the DSH could help with coordination around what data is created, such as supporting users of the hub to identify important gaps in available data with funding available to help fill these.

Process

User Need

I need to ensure the data is of a good enough quality for my needs, that it is cleaned of inconsistencies and errors and is in a format which enables me to achieve the outputs I require.

This is done when...

- I am confident that the consistency and accuracy of the dataset is sufficient for my use of this data.
- The data is in a format suitable for analysis or presentation, including being combined with other datasets and ingested into required tools.
- I have recorded the work I've done on the dataset, where required.

Steps Taken

Steps I may take to do this

- Reviewing the data to identify errors or inconsistencies.
- Cleaning the data to remove these errors or inconsistencies.
- Contacting the source of the data to ask them to clean it or clarify inconsistencies.
- Transforming the data into required formats.
- Ingesting into required tools and combining with other datasets.
- Keeping records of work done on the data (e.g. cleaning), where required.

Challenges

What is hard about doing this?

- Many datasets are poor quality or inconsistent, requiring significant time and effort to be spent on quality assurance and cleaning.
- Datasets are held in a range of formats, and in formats that do not match users' needs, meaning significant time and effort is required to transform data so that it can be combined, analysed or presented. In some cases there is a concern that this work is not as well automated as it could be.
- There was some concern that this work is being duplicated, with multiple people cleaning and transforming the same data rather than being able to widely share cleaned and transformed versions of data after doing this work.

Requirements

Achieving required quality and format of data

- Ideally, users want to know that all datasets on the DSH meet accepted standards of quality and are in appropriate formats.
- If this is not the case, they need to be able to verify what standards of quality and format individual datasets meet, so they can select data that is feasible for them to use (See 'Verify').
- Failing this, they need to be able to quality assure, clean and transform data within the time, skill and computing resources available. Users need to be able to keep track of work they have done on datasets.

Avoiding duplicated effort

- To avoid duplication of effort users would like to be confident that where others (who they can rely on) have already done this work, this can be found and accessed. And to likewise share the work they have done on datasets with others in a way that can be relied upon.

Test

User Need

I need to test the scripts I will use to analyse the data to check if it will work in the ways I require before committing to analysing all the data in this way.

I need to mock-up outputs so I can get feedback from users and check whether they meet their needs, before developing these fully.

This is done when...

- I am sufficiently confident that my chosen approach will meet my or my end users' needs, so that I feel justified in investing more time in it.

Steps Taken

Steps I may take to do this

- Carrying out analysis and testing my script on a smaller sample of the data.
- Discussing the method and results with colleagues, partners or other experts to sense check the approach.
- Creating basic mock-ups of final outputs and sharing these with end users to obtain feedback and check these will meet their needs.

Challenges

What is hard about doing this?

- Similar barriers to accessing and analysing data in general (see Access and Analyse).

Requirements

Requirements for the Digital Solutions Hub

- Similar requirements as for 'Access' and 'Analyse'.
- Being able to test their process and intended outputs based on the data available in a straightforward manner, even if it's less rigorous, was an important need for some users, especially those with limited resources and time.

Analyse

User Need

I need to analyse data to get the insight I require from the data, or to establish trends and relationships that it will be useful for end users to explore themselves.

This is done when...

End points look different depending on different needs, and include:

- I am confident that I have reached an interpretation of the data that is good enough for my needs.
- I have answered the question I set out to answer with sufficient confidence for my needs.
- I have obtained insights that adequately meet my needs, or those of the end users of my findings.
- I have identified patterns and relationships within the data that are likely to be useful for end users to be able to explore, and which I can enable them to readily explore.
- I have a robust and repeatable process I can apply to updated data at required intervals.

Steps Taken

Steps I may take to do this

- Discussing proposed approaches to analysis, and interpretations of the data, with colleagues, partners or other topic experts.
- Some analysts valued exploring trends visually as an early stage of analysis, particularly those working with spatial data.
- Finding suitable scripts for performing analysis or developing and testing my own.
- Combining different datasets, including data found on platforms with data I already have access to.
- Responding to error codes and tweaking code to ensure processes operate as intended.
- Documenting my approach.
- Getting colleagues, partners or topic experts to review outputs.

Challenges

What is hard about doing this?

- Some groups, particularly those in the public sector, struggled with a lack of local computing power and restrictive organisational IT policies which prevented access to the data in the ways required, or prevented access to the latest versions of required software.
- Working out best approaches to interpretation and analysis is a challenging task, combining topic expertise with wider analytical skills (see also Comprehend).
- Analysis is a fast-moving field, and it can be hard to keep up with the latest developments.

Analyse cont.

Requirements

The role of visualisations

- Being able to explore trends and relationships in a visual way was an important early stage of analysis for some groups, particularly layering spatial data.

Testing the water and prototyping

- Given the requirements of analysing larger datasets in a rigorous way there was some interest in being able to more quickly and easily test out approaches and mock up outputs, even if done to a less high standard.

Combining data

- Analysing different datasets relating to a common geographical area was a common need. There was also interest in other ways of linking data, for example those working on public health were interested in being able to look at specific groups within wider populations.
- Often users were interested in combining their own data with data from external platforms as part of analysis.

Users need ways of organising the different resources they are working with as part of analysis.

Learning from others

- This is an important part of how analysts develop their skills, including getting feedback on their work. This was particularly valued when working in topic areas that were less familiar. There was some interest in the DSH supporting such communities of practice, for example through supporting forums in which people could share advice and resources.
- Users need to be able to find, adapt or develop suitable scripts for analysis and modelling. Being able to access and adapt existing scripts was seen as a potential benefit from such a community.

Working within computational power and IT policies

- Users need to be able to run demanding processes despite the barriers they face. Working with data in the cloud was seen as one possible response. There was also interest in ways of reducing these demands by breaking complex datasets down into more manageable parts.
- Users need to be able to keep a record of the work they have done.

Present

User Need

I need to present the insights I have drawn from the data to my intended audience in a way that is compelling and sufficiently easy to comprehend.

I need to develop ways of presenting data that allow others to readily draw out insights that are relevant to their needs.

This is done when...

- My findings are presented in a way that is sufficiently easy for my intended audience to understand. It may also be important to present these findings in a way that will stand out and tell a compelling story.
- Data is presented in a manner that allows end users to readily explore trends and relationships relevant to their needs and have confidence in these.
- I have clearly communicated what confidence it is appropriate to hold in the data when used as I have used it, or as I am enabling others to use it.

Steps Taken

Steps I may take to do this

These vary depending on the type of outputs that users are developing:

- Thinking about how to visualise important trends and relationships in the data:
 - If the key output is an answer to a question and an explanation of how this was reached then presenting this in a report is often the output of choice.
 - If the key output is to enable others to explore the data and discover trends and relationships that are significant for themselves then developing tools like maps, dashboards or pivot tables will be the focus.
- Present, explain and encourage use of the outputs amongst colleagues and partners.

Where users are enabling others to explore the data further themselves, key tasks included:

- Bringing together different datasets.
- Deciding which features of this data, what level of detail and the relationships within it, are most important to show. Presenting these in a compelling and easy to use format.

Challenges

What is hard about doing this?

- Presenting outputs in a way that is understandable for users who lack data literacy and/or topic specialist knowledge.
- Needing to simplify the outputs to such an extent that users worried about losing nuance and important details.
- The data not 'saying' what the end users wanted it to and navigating the politics of this.
- A need for visualisation skills in addition to other skillsets, these are sometimes provided by specialists.

Requirements

The role of visualisations

- There was a suggestion that being able to explore data visually may allow policy professionals without strong data skills to 'self serve'. This needs more investigation.
- Users are working on different types of outputs with different requirements. For some users creating, saving and sharing visualisations is an important requirement, though there are different ways this can be met.

Promote

User Need

I need to make those who will benefit from my outputs aware of them. Doing this will allow me to achieve impact from my work (and may also allow me to evidence this to those I am accountable to).

This is done when...

- The target audience for my outputs is familiar with these outputs, able to make use of them, and appreciates the potential benefits of doing so.

Steps Taken

Steps I may take to do this

For some users this is a larger portion of their work than for others and may include:

- Understanding the target audience (whether within their organisation, partner organisations or wider audiences).
- Planning and executing dissemination based upon understanding the need for the developed data product and employing different strategies and formats to reach out to end users in the best way.
- Using relationships built up at earlier stages of work to present outputs in person to policy teams.
- Using social media networks to raise awareness.
- Sharing outputs through data portals to raise visibility.
- Gathering data on uptake/use of my outputs.

Challenges

What is hard about doing this?

- Understanding, and prioritising, the needs of diverse groups of stakeholders.
- Changing culture to encourage use of new data products.

Requirements

Requirements for the Digital Solutions Hub

- Provide users with a place to increase the visibility of the end products they create and datasets developed to maximise their use and value.
- Enable users to access information on use of their data products, (e.g. number of uses, how they have been used, for what purpose, and what they have been able to achieve) so that they can use this as evidence of impact and be motivated to continue to engage and contribute.

Iterate

User Need

I need to receive feedback from the end users of my outputs so I can continue to develop these outputs to best meet their needs.

This is done when...

- I have sufficient feedback to justify either making changes or continuing with an output in its current form.
- In some cases this is an ongoing process of feedback and iteration over time.

Steps Taken

Steps I may take to do this

Different versions of this include:

- Demonstrating draft outputs to colleagues or partners in policy roles or working practically on environmental management, obtaining feedback, and making adjustments as necessary.
- Getting audience feedback about datasets intended for the public domain.
- Consulting with affected stakeholders about conclusions that affect policy.

Challenges

What is hard about doing this?

- Getting people's time for reviewing outputs

Requirements

Requirements for the Digital Solutions Hub

- This requirement is less relevant to the hub itself, being more focussed on how data users work with colleagues and other end users.
- Steps that help users to quickly create prototypes will support this.
- Being able to receive feedback on datasets shared on the hub may also be useful for some users.

Manage

User Need

I need to ensure data products are stored and managed in a way that allows them to be readily discovered, accessed and used by myself and others.

This is done when....

- Data is stored in a secure and stable format.
- Data is stored in a format that others can readily comprehend and use, including having sufficient metadata.
- Data is discoverable and usable for as wide an audience as appropriate, factoring in legal and ethical restrictions on this.
- Management systems are in place to enable myself and other colleagues to keep track of what data is where, including knowing what data is forthcoming, and ensuring that datasets are kept up to date.
- Those using a dataset will be updated about changes to the data, or new versions being released.

Steps Taken

Steps I may take to do this

- Cataloguing which data assets are held where within organisations or partnerships, what datasets are forthcoming, what are the restrictions on the use of datasets (including embargoes), when datasets will need updating (if applicable).
- Keeping track of the data I have compiled for my own work and making clear the data I have drawn from for my outputs.
- Storing datasets created by myself, colleagues or partners within organisational data vaults.
- Uploading datasets onto openly accessible data portals that are well-used by those who would benefit from this data.
- Checking that data is stored in an appropriate format, including appropriate metadata.
- Create a notification system that notifies people of when datasets are updated or depreciated.

Challenges

What is hard about doing this?

- A lack of appropriate central data stores within some organisations.
- The requirement to cope with legacy datasets is a barrier for developing appropriate solutions for being able to bring together different datasets in one place.
- Poorly developed data management culture and practices within some organisations.
- Difficulty motivating people to openly share datasets - a lack of incentives for doing so, or of appreciation of the benefits of doing so.
- Given this reluctance, some of those supporting data sharing are spending time chasing up data rather than being able to invest this time in other ways.
- Concerns about data protection and inadvertent disclosure are perceived as a barrier for sharing datasets. There was a suggestion that these are often perceived as more of a barrier than is accurately warranted.

Manage cont.

Requirements

Supporting good data management

- As someone with responsibility for data that I create, or help manage, I need to understand appropriate data management requirements for meeting the needs of those who will use this data.
- I also need a straightforward way of implementing data management within my own work, or within the organisation or network I work with.
- I need to share datasets on suitable platforms that will allow others to easily find, access and use them.
- It was suggested that the DSH might provide an opportunity to promote better data management, for example by sharing useful resources or using the incentive of sharing data on the platform to promote good practice.

Knowing how datasets are being used

- Being able to know how others have used datasets I have shared, and getting feedback from them, may allow me to improve my data management practices and evidence the impact of doing this work.

Supporting the Environmental Data Community

There was a common concern that there is a lack of coordination around how data is created and used. Respondents described a 'patchy' landscape of data collected for a wide range of different purposes and existing across a proliferation of different platforms.

Accompanying this was concern about a duplication of effort. Some analysts worried that others elsewhere may be doing the same work as them, for example both independently cleaning up public datasets rather than benefiting from each other's efforts.

At the same time, analysts tend to be time poor, having to work across many different platforms to try and find the data they need, often having to spend time cleaning this up and transforming it before it can be used. One estimated that 80% of their time is spent finding and cleaning up data rather than performing analysis on the data.

Currently relationships with others form an important part of how users find relevant data, how they develop their practice, such as getting feedback on proposed approaches and whether their interpretations are sound.

Many saw the DSH as an opportunity to respond to some of these realities. There was a lot of interest in how it might support communities of practice, such as through forums sharing advice about fields of study and useful datasets within these. And interest in how it might improve coordination and long-termism in the sector rather than adding to the proliferation of platforms and short-lived initiatives.

Keeping in Mind Specific Needs

This report is focussed on common themes across respondents, but it is important to remember specific needs within this, including those found in different parts of the UK. Two prominent examples of these were, the need to work in two languages for many in Wales, and challenges around being left out of some environmental data coverage in Northern Ireland.

What is Shown Here

User archetypes are a tool for conveying the detailed reality of those who may use a service in a simplified and manageable way. For such a wide-ranging service there is a particularly diverse body of people who may be affected. We have focussed on trying to overview some of that diversity. To achieve this, our archetypes have been built largely around the different aims of people working in this space, and the steps they take to achieve these, rather than focussing on which type of organisation, or the type of data, they work with. These other distinctions may well draw out further insights, such as the distinct needs of those working with spatial rather than other types of data or those dealing with sensitive health data.

While these archetypes generally reflect different people in this space, in some cases people may identify with more than one archetype within different aspects of their role.

It is important that these archetypes are treated as what they are: an early attempt to convey this diverse body of potential users. We hope that these will be used to dig further into some of these groups and to help identify whether there are potential users who are currently missing from this picture. We expect these archetypes to evolve over time.

For each archetype, we have stated the aim of this user group, an summary of the type of people found within this group, an overview of the steps they go through to meet these aims, and some of the important challenges that are particular to this group.

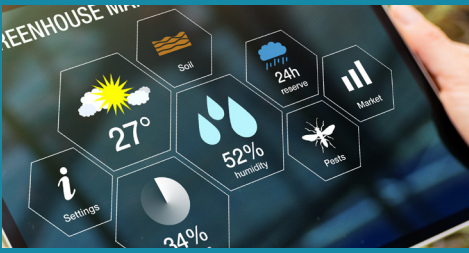
As already mentioned, a vital part of making the Digital Solutions Hub a success will be working with key stakeholders in the environmental data community, such as those managing NERC data centres. Our focus here is on describing potential users of the hub rather than mapping out these key stakeholders.

Again, it is worth noting that there are three other important groups we have not included here:

- Users from community groups, small charities, campaigning groups, or interested members of the public – although a small number of potential users in this group did take part in this research, this is not sufficient to properly assess the needs of a potentially very broad group. Doing so would require a further piece of research that is directly focussed on this group.
- Policy professionals without advanced data skills – some of those we spoke with were supporting policy colleagues to explore data, and there was a suggestion that more might be done to enable this group to 'self-serve'. Exploring needs around this with policy professionals could be an interesting area of further investigation.
- IT specialists – we have focussed for now on the needs of those who are working more directly with data, whose needs the hub will be designed to meet. However, working closely with IT specialists within the users' organisations will be an important part of meeting these needs effectively.



Summary of User Archetypes



Analysts Monitoring the Environment

I need to use data to demonstrate the condition of the environment in a consistent format that will allow the health of the environment, and the performance of environmental policy, to be scrutinised over time.



Authors of Monitoring Frameworks

need to identify measures of environmental health to use for scrutinising and evaluating policy, and informing future decisions.



Analysts Answering Questions with Data

I need to find and analyse data to identify correlations, report findings and make predictions about the likely impact of actions.



GIS Specialists

I need to create data products that allow my intended audience to readily comprehend and explore information that is relevant to their needs, using map-based formats.



Data Support

I need to support others to more readily explore data and obtain useful insights.

To do this I find and combine relevant data, analyse it to identify significant trends and relationships, and create data products that allow others to readily explore information relevant to their needs.



Data Leaders

I need to make sure my organisation is working effectively with data.



Data Stewards

I need to ensure that datasets I have responsibility for meet appropriate standards, and are appropriately stored, shared, and kept up to date. I do this so that they can be readily discovered and used by those who might need them.

Analysts Monitoring the Environment



What are my aims?

I need to use data to demonstrate the condition of the environment in a consistent format that will allow the health of the environment, and the performance of environmental policy, to be scrutinised over time.

Who am I?

I am likely to work for an organisation with set responsibilities for environmental monitoring, though some other groups are also doing similar work.

I am likely to have topic expertise and analytical skills, though am not required to look as deeply into the data as some other user groups. I generally have to stick to broad approaches, so have less flexibility than some other user groups.

How do I meet my aims?

- I must understand the data required, including talking with topic experts.
- I must gather together the data needed, generally this comes from a set of well-established sources.
- I must verify that this is the right data, such as the latest version.
- I must quality assure, clean and transform the data into an appropriate format for analysis.
- I must analyse the data to reach required outputs, such as a categorising the

environmental health of areas against set standards. Generally, this is done in quite standardised ways.

- I must present these findings in a clear format. This typically involves written reports, maps, and charts.
- Generally, I must ensure datasets I have created are stored in an appropriate format and uploaded to appropriate, well-used portals.

What challenges are faced in particular by this group

- There was a suggestion that more could be done to increase the value of the datasets created for these tasks by combining these with other relevant data, so that it wasn't 'single use'.
- There was a suggestion that more could be done to enable others to access the underlying data used to create final outputs.



Authors of Monitoring Frameworks



What are my aims?

I need to identify measures of environmental health to use for scrutinising and evaluating policy, and informing future decisions.

Who am I?

I work within a government department or for an environmental quango. I work in a managerial role, with responsibility for deciding on, and implementing, appropriate approaches for monitoring environmental health.

How do I meet my aims?

- I must understand what measures would provide useful insight, and how these might be generated – I will speak to a range of stakeholders to achieve this, as well as looking at what has been done elsewhere to learn about good practice and avoid duplicating work. Speaking to stakeholders and other topic experts is also used to understand what kind of data, analysis and outputs would generate the required insights.
- I must understand what data is available, verify whether datasets will meet my needs, and get hold of data sufficient to meet my needs. Relationships are an important part of this discovery and reaching out to the originators of datasets is often how I overcome shortcomings in the information available within published metadata. I may need to add adequate metadata to datasets when these are

placed in data stores. I often need to publicly share the data used, which can create a barrier to using some datasets.

- I must commission new data to fill gaps, ensuring standards of data quality, openness, and data management are established at source.
- I must ensure data is quality assured, cleaned, and transformed ready for use.
- I must ensure data is analysed and findings presented in a clear and compelling way. Typical outputs I create are reports, charts, and in some cases dashboards to allow further exploration.
- I must disseminate these outputs and consult with affected stakeholders to make ensure these are accurate and useful.
- I must ensure the underlying data is shared in an appropriate format. I must put in place data governance processes to make sure all data is stored appropriately, meets appropriate data standards, and is kept up to date.

What challenges are faced in particular by this group

- A lack of data, including data at a good enough standard.
- A lack of access to data, including in a timely manner and the existence of 'silos' within and between organisations were all seen as significant barriers in some cases. The need to publicly share datasets used was a barrier to the use of some existing data. Concerns about keeping data up to date were cited as barriers to data sharing by some.
- Inadequate metadata is a barrier to readily verifying whether data would meet requirements.
- The format of data often requires effortful transformation of it to make it useable.
- Deciding how to clearly communicate complex findings.
- Putting data governance processes in place to ensure datasets are created in a form that matches appropriate standards, and is stored, shared, and kept up to date appropriately.

Analysts Answering Questions with Data



What are my aims?

I need to find and analyse data to identify correlations, report findings and make predictions about the likely impact of actions.

Who am I?

I have expertise in the topic I work on and combine this with strong analytical skills. I look for statistically significant patterns and correlations in data and use this to create predictive models about what will happen if particular actions are taken.

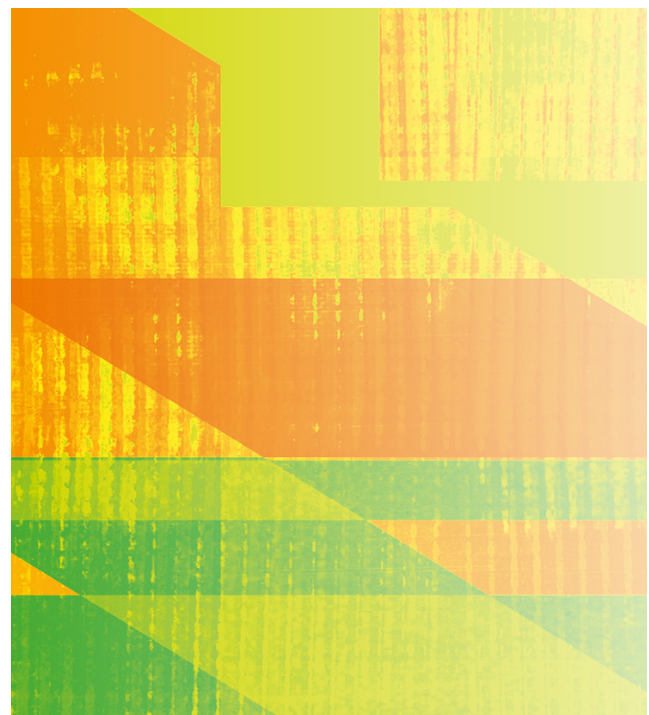
I often work within government departments and other large environmental bodies to inform practical decisions, such as how best to manage a peatland, or whether stronger regulation is needed to protect against a pest. I may also be found working in an academic context.

How do I meet my aims?

- I need to understand the need and understand what data might meet this. I may be working on relatively clear-cut questions. Or I may need to spend time understanding the requirements of those working on delivery, to know what insights would be most important to them. I frequently have to rely on goodwill to access topic expertise to help me understand the questions I am working on.
- I need to find data that will answer my questions, with sufficient confidence to

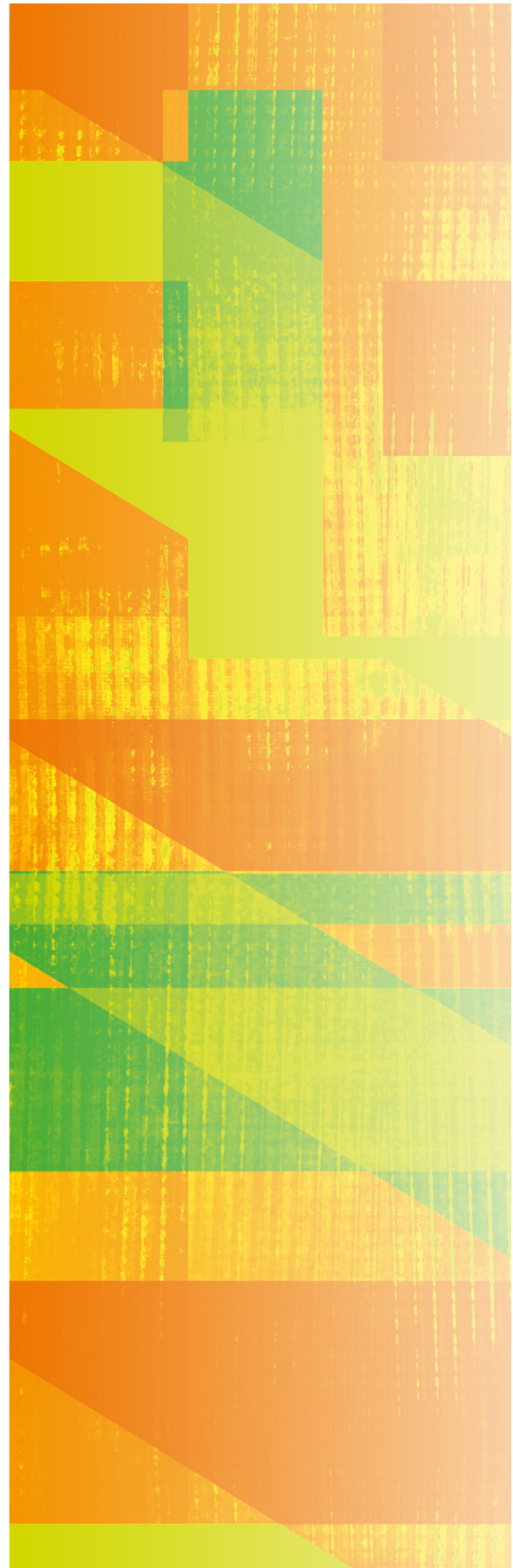
meet my needs. I need to verify that the data I do find will be suitable. I may need to extrapolate from existing data to plug gaps in what is more readily available.

- I need to quality assure, clean, and transform data into formats required for analysis, combining different datasets together.
- I need to analyse data to identify significant patterns and correlations, creating models to predict what will happen if particular actions are taken. I will look for existing scripts I can use or adapt, or develop my own using programming languages such as Python and R. I will often test out analysis early to check it will work, before working with a larger body of data. I'll talk with colleagues or other experts to discuss my approach and interpretation.
- I need to present my findings, making clear the conclusions I have reached, and how I have arrived at these. This tends to be done through a report, using charts to visualise and communicate findings in a way that it easier to understand.
- I need to keep track of the different data sources I will need during this work. I may make sure datasets I have created can be readily discovered and used by others, by uploading these onto relevant portals in appropriate formats with clear metadata.



What challenges are faced in particular by this group

- A lack of data, particularly at the right scale, particularly for local levels.
- Data stored in ways that are hard to access such as journals or reports. Site data and local authority data are often held in difficult formats. Data may also lack useful details such as administrative boundaries.
- Difficulty readily accessing datasets, such as multiple clicks or lengthy registration, can mean putting a lot of effort into accessing 'red herrings'.
- Datasets only known in specific circles, such as academia, rather than being open or easy to find, meaning if I don't know it exists I won't use it as I'm unlikely to find it.
- Thinking about how best to analyse data and needing to rely on access to domain experts to help interpret data correctly or sense checking the analysis.
- Having to work with multiple sources and formats and then unifying these. As with other groups the format and quality of data, and a lack of standardisation, was a challenge that required work to transform this prior to use. Working with 'wildly different' scales was noted as a particular challenge.
- Those working on public health faced stronger challenges around getting timely access to data and navigating data protection and risks of accidental disclosure.
- IT barriers were a major issue for many, including lacking admin rights on work laptops, only having access to old versions of R because it has to be tested by IT teams first, and coping with the heavy computational demands of large datasets.
- Keeping up with the pace of innovation was seen as another challenge.
- Finding the best way to promote findings to achieve impact, especially for academics in this group, was a challenge.



GIS Specialists



What are my aims?

To create data products that allow my intended audience to readily comprehend and explore information that is relevant to their needs, using map-based formats.

Who am I?

I am a data analyst whose skills are focussed on using Geographical Information Systems (GIS). I create map-based data visualisations to allow others to explore information that is relevant to them. For example, these may be created for policy colleagues, external clients, or members of the public.

How do I meet my aims?

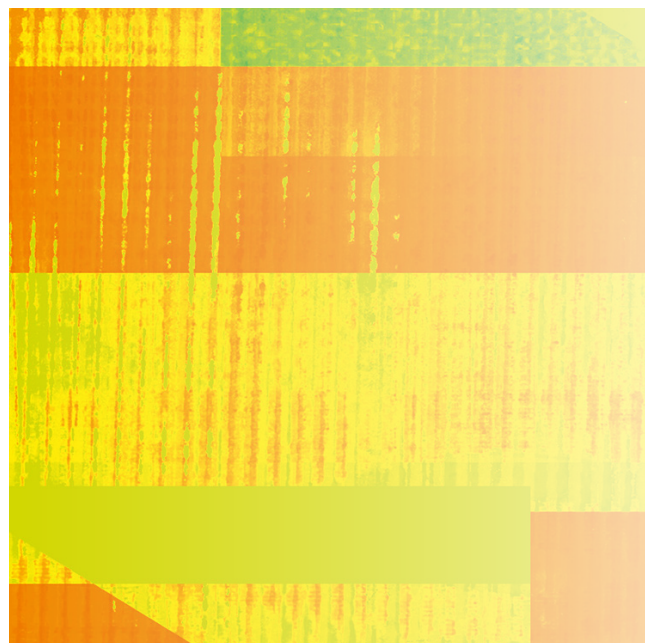
- I must understand the needs of my intended audience so I work closely with end users to do this, not just at the beginning of a task but checking back throughout to see whether the outputs being developed will meet their requirements.
- I must find data that I can use to meet these needs. This requires verifying that datasets I discover are appropriate for my purposes.
- I must quality assure, clean and transform these into appropriate formats. Combining different datasets is an important part of this. Bringing together datasets at different resolutions into a format that could work together is a significant part of this work.
- I must analyse the data to identify

significant features and decide how I can present these.

- I must present the data in a format that allows my intended audience to readily draw out the insights they need. Geographical Information Systems were used to clearly show spatial features and allow these to be explored.
- Testing out early prototypes and iterating based on user feedback is important for me to deliver products that closely match needs without wasting excess effort on unsuccessful approaches.

What challenges are faced in particular by this group

- Some described having to react to requests that don't always consider the difficulties of using data, rather than being able to prioritise and work more proactively.
- Data is held in lots of places.
- A lack of data, particularly at appropriate resolutions.
- A lack of consistent data standards.
- Datasets that are very large or complex rather than packaged into more manageable, frequently used chunks.
- The work needed to clean and transform data.
- Developing teams with sufficient skills.



Data Support



What are my aims?

I need to support others to more readily explore data and obtain useful insights.

To do this I find and combine relevant data, analyse it to identify significant trends and relationships, and create data products that allow others to readily explore information relevant to their needs.

Who am I?

I have skills in data analysis and communication. Rather than focussing heavily on specific topics and developing expertise in these I provide support with using data across a broader range of topics.

As such I am likely to work for an organisation with a less specialised remit, such as a local council or a private company. My role is about supporting policy work, at times including communicating information to a wider, public audience.

How do I meet my aims?

- I must 'understand the ask', getting to the bottom of what has been requested, why this matters, and the best way to meet these needs.
- I must find sufficient data to meet these needs, verifying that datasets are appropriate for my purposes. I may also need to seek help from topic experts to understand more about what data is appropriate.
- I must quality assure, clean and

transform these into appropriate formats for analysis, bringing together different datasets.

- I need to analyse patterns and relationships within the data to identify significant features, informing what information I will present.
- I must create products that allow my end users to readily explore information relevant to their needs, allowing them to 'self-serve', such as dashboards or pivot tables. I may also create reports and visualisations of data using charts. I may present findings in person and provide guidance and support in using these tools. Generally, these outputs are for internal use and are not widely shared, though sometimes I will be designing outputs that could be used by a public audience.
- I may promote my outputs to relevant users to encourage and support their use.
- I need to gather feedback from end users about these products to develop them further. I may share early prototypes as part of iteratively seeking feedback and adapting.
- I may also think about ways of advocating for more or better data to be created.

What challenges are faced in particular by this group

- Time spent 'understanding the ask'. A lack of consideration about how to use data within wider teams was seen as contributing to these difficulties.
- Working with 'patchy' data in a wide range of formats. Having to access different systems to find and access data. One respondent described their outputs as 'increasingly being pleas for better data'.
- Thinking about how to communicate complex information in an easy to understand way.
- Promoting outputs to ensure they are used. There were concerns that outputs are not being as widely shared and re-used as they could be, especially beyond the organisations that created them.

Data Leaders



What are my aims?

I need to make sure my organisation is working effectively with data.

Who am I?

I hold a management role that focusses on implementing an effective data strategy.

I have previously worked in data analysis, but now apply this knowledge to think about how data functions within my organisation.

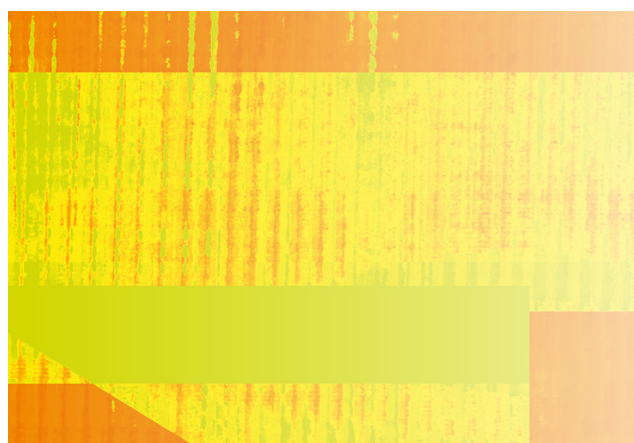
Organisations working with environmental data tend to think a lot about how their data serves a wider community, and so this is an important part of my role. In some larger, or data-focussed organisations, there may be several people working in roles like this. For many organisations, my role will involve thinking about how a wider network of partners works together.

How do I meet my aims?

I think about the whole journey, or system, of how data operates:

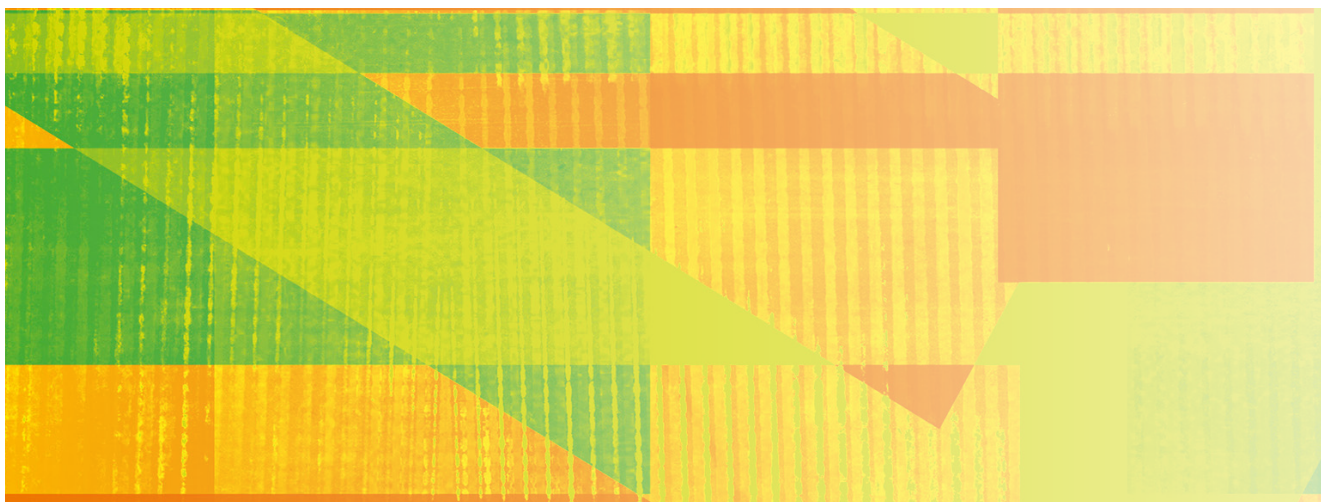
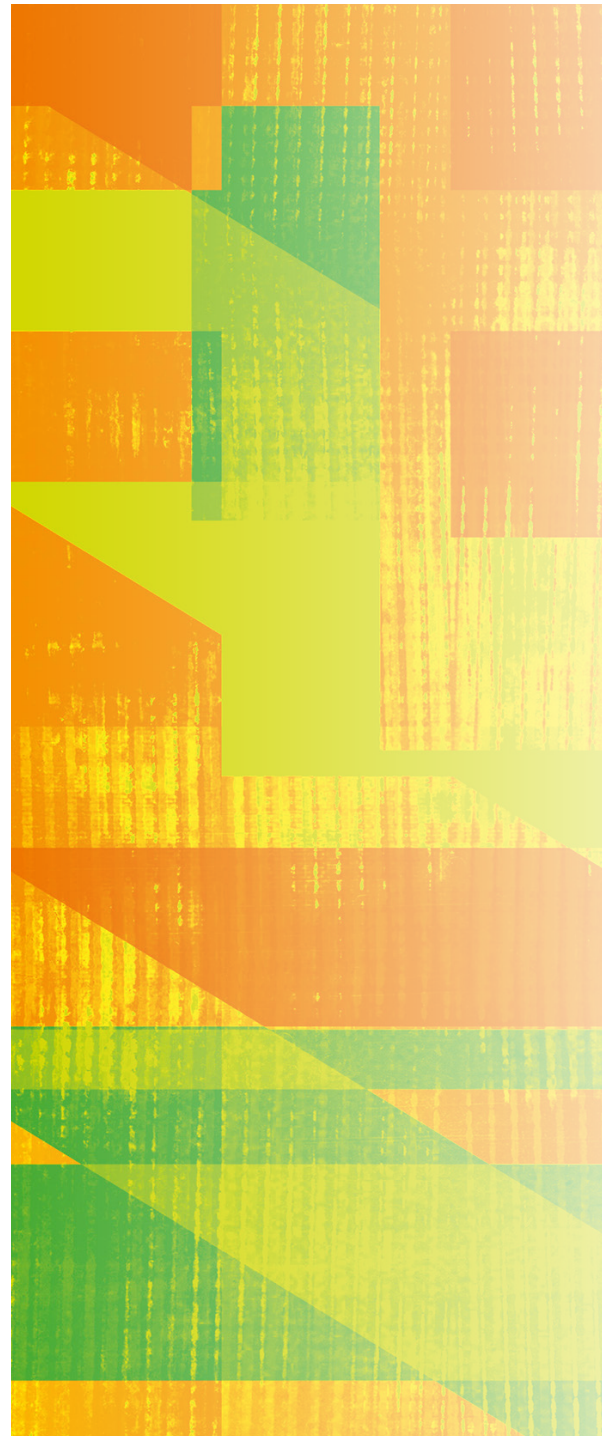
- I need to understand the needs of data users, such as policy colleagues or colleagues managing environmental sites on the ground. This does not just happen at the start of a journey but is often an ongoing process. I may look at encouraging 'co-ownership' of data products and processes with the end users of these products.

- I need to understand what data is available and where the gaps are. I need to verify the data I find to work out if it is suitable for my needs. Knowing what is out there helps me prioritise what can feasibly be achieved to have most impact within the resources available. As well as doing some 'detective work' to find data by looking in known platforms, and searching online, I will also reach out to others, both within my organisation and externally, to establish what is available where.
- I may need to commission the collection of new data to fill gaps. I may think quite creatively about ways of gathering data, such as encouraging users of data products to add-in data about their own projects.
- I will support strong analytical, presentation, and dissemination practices amongst my team.
- I need to ensure data assets are managed correctly - that they are kept in appropriate formats, meet standards of quality and accuracy, have appropriate metadata, are easily findable, and are updated where needed.
- I need to make sure final products meet user needs by getting feedback and iterating on it. I may also provide some support and guidance in using these outputs.
- I may collaborate with a wider data community to support an effective data system looking beyond organisational boundaries. In general, my role is likely to involve working with a range of partners who are connected with my organisation's work.



What challenges are faced in particular by this group

- Understanding user needs – this was a particular challenge for organisations serving a wider public audience, who had to prioritise between diverse needs.
- Getting an overview of what's out there – data is held in lots of places, it's hard to know what is available.
- A lack of data – the data that exists is 'patchy', it is particularly hard to get data at required levels of granularity.
- Pay walls and data protection barriers can prevent access to data that does exist. Given this reality, there was some mention of starting with what is available and thinking about what can be achieved with this, rather than always starting with the question to be answered.
- Verifying data – where datasets exist, it is not always easy to quickly identify key information such as what data is contained, what format it is in, and the method used to gather it.
- A lack of data standards – data is held in different formats, often not the most appropriate ones for storage and reuse. Systems are not as interoperable as they could be. Data quality is variable, and metadata is less standardised and clear than it could be.
- A lack of coordination in the sector and concerns that effort is being duplicated rather than more joined-up work to improve the availability of data. There was also interest in stronger communities of practice to support one another.



Data Stewards



What are my aims?

I need to ensure datasets I have responsibility for meet appropriate standards, and are appropriately stored, shared, and kept up to date. I do this so that they can be readily discovered and used by those who might need them.

Who am I?

I work for an organisation, or network, with large numbers of datasets. While good data management is everyone's responsibility, my role (or part of it) is to make sure appropriate data governance is in place for these datasets.

I understand good data management practice, and work with those creating and sharing data to make sure standards are met. My work is informed by an understanding of the needs of data users.

How do I meet my aims?

- I must understand enough about the needs of data users to ensure that datasets shared meet their needs. Those in more senior roles were taking a publishing mindset to what is shared, focussing on delivering outputs that combined and translated data into datasets that would be readily usable for key audiences.
- I work with those creating datasets to ensure appropriate standards are met, such as adequate accuracy and

consistency, appropriate metadata, and that data is shared in an appropriate format. I provide guidance, standards, and tools to assist them.

- I work with those creating datasets to identify when data will be available, and to understand limitations on sharing this data, including risks of disclosure, proprietary issues, and respecting any embargoes.
- I receive data from creators, but often have to 'chase up' to get this.
- I quality assure datasets before they are shared. I may have to clean and transform them into an appropriate standard.
- I must ensure limitations on data sharing are respected, and that systems are in place to keep data up to date where relevant.
- I upload datasets onto portals that are well-known and well-used amongst the relevant audience.
- I catalogue what data is held where.
- I may also need to document work I have done on datasets.

What challenges are faced in particular by this group

- An 'incomplete picture' of the needs and priorities of those using data.
- Getting data in a timely manner, with effort currently being spent chasing up by email to get hold of datasets.
- Getting those creating data to meet appropriate standards, including appropriate, standardised metadata.
- Having to work with lots of different systems and with data in different formats, including formats that are hard to work with.
- Working with large datasets that are hard to transfer.
- Working with outdated databases that can't cope with high volumes and diverse types of data, leading to the creation of bespoke solutions that don't 'talk to each other'.

08. Participants' Suggested Features

At the end of workshops the University of Manchester team provided some information about what the Digital Solutions Hub (DSH) could look like. Participants were then asked a set of questions about what features they'd like from the DSH, what barriers they'd experienced when accessing similar platforms, and what value they thought the DSH would add to their work. Common themes from these responses, outlining what participants felt were important design features for the platform, are summarised below.

What Data Exists

Filling data gaps

Participants were interested in a wide range of data but there were some themes around data gaps that the DSH might help plug. Respondents talked of the data landscape being 'patchy' and were interested in improving on this.

Prominent unmet needs were:

- Data at a local geographic level, or a finer degree of resolution more generally.
- Being able to link environmental and socio-economic or health data.
- Being able to show change over time.
- Suggested areas of focus included:
- Data relevant to current policy ambitions to support decision-making and scrutiny.
- Providing layers of different data across the UK.

Meeting consistent standards of data quality and format

A frequent complaint was the quality of available datasets, which required a lot of effort to quality assure, clean, and transform ready for analysis. In contrast, users emphasised the importance of having 'analysis ready' datasets that need minimal cleaning and transformation.

Key requirements included:

- Common, and appropriate, file formats.
- Well-structured datasets.
- Cleansed datasets.
- The compatibility of different datasets.
- Datasets being kept up to date (where applicable).
- Consistent, good quality, metadata (covered below).

It was felt that curation of the data on the DSH would be important to maintain these standards.

Getting at this Data

Being clear about the purpose of the DSH

There are lots of platforms out there. It can be hard to understand the purpose of platforms and the range of data available on each site. It must be clear to users what the DSH is for, what niche it fills in this landscape, and what range of data users can expect to find on the site.

From participants' comments some suggestions emerged around what the role or 'unique selling point' of the DSH should be:

- A reliable source of 'analysis ready' data that meets high standards of quality, including clear metadata.
- Bringing together data from different fields in one place.
- Moving away from a 'fragmented landscape' towards a 'one stop shop' for data. It was suggested that this could include paid for data, as well as free data.
- Contributing to coordination within the data system (as discussed further below).
- Sharing data deemed valuable by academics that is not currently 'on the government's radar as important'.
- A source of data that could support the development of a digital twin, or inform AI.

An easy to use platform

Data users are accessing several different platforms to find the data they need. Some of these platforms are not easy to use, meaning that users have to invest a lot of time in getting their hands on datasets, sometimes doing so only to find the data is not suitable for their needs.

The DSH needs to be straightforward

to use for someone in this situation. It should be clear what the platform is for, and how to use it. With minimal clicks, registration steps, or other barriers, users need to be able to a) work out if available datasets will meet their needs b) get sufficient access to these.

Appropriate metadata and documentation

A key requirement is consistent, good quality, metadata. The 'Data checklist' shared earlier outlines key information needed. A way of contacting someone responsible for the data was seen as a way of being able to fill gaps in the information provided via metadata.

Documentation that explained the rationale, context and potential uses of the data was also useful.

Being able to stably link to data

Users often wanted to be able to link to data or access it via an API. They needed to be able to rely on this being stable, including being able to trust that they will be contacted if any changes are made to these links.

Downloading data

Users often wanted to be able to download data to use locally, including combining it with their own data. They need it in an appropriate format to use at their end. Sometimes the size of datasets can make this a computationally demanding step. There was some

interest in being able to download only parts of a datasets that are relevant, including being able to sample datasets to tell if the data would be suitable for their needs before downloading larger sets. Chunking up larger datasets into commonly used parts was seen as another way to support this.

Strong discoverability and exploration

Data on the DSH must be readily discoverable. Most users talked about needing to find data that would answer specific questions. However, there was also some interest in being able to explore the data that is available on a broader theme in a more open-ended way.

As well as needing to find data suitable to their needs, some respondents also needed to work out what data does not already exist. This allows them to know whether new data gathering needs commissioning.

One suggestion for improving discoverability was to make it easy to browse provenance backwards and forwards, i.e. being able to click 'back' through to the original datasets that a particular dataset or other data product was created from, and being able to click 'forwards' through to datasets or data products that had made use of this dataset. Currently relationships form an important part of how users discover the data they need. Supporting these connections may be an important aspect of improving discoverability.

Tailoring to different audiences, and providing data visualisations

Towards the end of the workshop, some suggestions for what the DSH could look like were presented by the University of Manchester team. One suggestion was tailoring the platform to different users with different skills. Visualising data was part of this idea.

There was some debate around this idea. Some were sceptical about whether there was a sizeable public audience that would be likely to make use of such features, with a suggestion that data analysts looking for raw data may instead be the main audience and will require much less effort to serve than maintaining more publicly accessible content.

However, those working with spatial data were interested in being able to visualise, and layer, data. This was seen as an important early step in exploring trends and relationships. There was also talk of such tools allowing analysts to readily explore or 'play with' data.

There was also a suggestion that policy professionals who are not experts in working with data, may be an important audience that might benefit from such features. As most participants of the workshops were data professionals, it was not possible to ascertain the validity of this suggestion.

Getting the resolution and extent you need (particularly geographic)

The geographies that people are interested in don't always match up with those that datasets are built upon (which are often based on administrative boundaries). Finding sufficiently granular data was a particular difficulty. Being able to bring together datasets was a significant related challenge, with one respondent expressing the challenge of having to work with 'wildly different resolutions'.

As well as indicating the data that users require, these challenges also relate to being able to readily extract data for relevant areas. Suggestions to aid this included:

- Being able to snip data via drawn shapes that could be uploaded.
- Being able to readily extract data for well-defined regions such as devolved nations or local authorities.

Promotion

People need to know about the DSH, and how it develops over time, including new data that is added.

Online analysis and processing tools

Restrictive IT policies and local computational capacity were a barrier for many. There was interest in being able to process and analyse data online as a way around these, including saving such work.

As mentioned above, being able to visualise spatial data was seen as important part of beginning to analyse such data.

Creating statistical models was also seen as important. Being able to access and adapt some pre-built models was seen as a way to support this.

Adaptable features that help with managing work

There were a variety of feature requests that were about being able to use the DSH in a personalised way that would help users to organise or manage their work. For example:

- Being able to create personalised dashboards.
- Being able to add and annotate existing metadata.
- Tools that help data professionals with managing their own data, which integrate easily with the DSH, and could encourage them to publicly share their data in the process.



The Data Community

A community of practice and sharing resources

Analysis is a complicated enterprise. To tackle this, relationships with others were important for helping analysts get to grips with different topics, uncover new sources of evidence, and get advice about proposed approaches and interpretations.

Several users were interested in online forums for sharing tips and tricks on how to use the platform, including topic specific communities within this.

There were also concerns about duplication of effort amongst analysts, and whether more could be done to share the fruits of their labour. This was not just about sharing datasets and other products created, but also sharing analytical models, or cleaned versions of existing datasets. It is important to consider how the Digital Solutions Hub might support such communities of practice and sharing of resources. More work would be needed to explore which solutions would work best for this.

Contributing to coordination of the wider sector

There was a common concern about a lack of coordination around the data that is created. There were a lot of comments about the need to avoid reinventing the wheel, and instead think about how the DSH could add value and improve coordination in the existing landscape, exploring partnerships in the UK and beyond.

There were a number of suggestions here, including:

- Recording how data is used and alerting data providers to this.
- Supporting a community of DSH users to identify data gaps that are most important to this community and attract funding for plugging these gaps.

Appendix: Ways of Clustering Users

A

There are many ways of thinking about, and distinguishing between, different potential users of the Digital Solutions Hub. We have included some of these ways of reflecting on, and dividing these users as an appendix, if helpful for informing future work.

These are a set of questions to ask about potential users, and some bullet points illustrating the diversity found under each of these.

What topic?

Users combining health or social data with data on the physical environment have particularly distinct needs versus those working primarily on the physical environment.

What form is the data in?

- Spatial or not.
- Qualitative or quantitative.

Where is the data from?

Data collected within their own organisation or combining and analysing datasets that others have created.

What type of output are they working on?

For example, an answer to a policy question, a report, a map, a graph, a dataset, a dashboard.

Which skills do they have?

- Being plugged into the needs of data users (such as understanding how data can inform policy, or developing data products in partnership with users).
- Topic knowledge:
 - Knowing what data is out there.
 - Knowing where to look or who to speak to.
 - Knowing what it means and how it can be used.
- Statistical analysis (and the combination of this with topic knowledge).
- IT skills underpinning data systems.
- Data management and data sharing.
- Communicating findings.
- Visualisation.

- Outreach and promotion.
- Training others in using data better.
- Managing relationships and networks.
- Thinking about the big picture of how data gets created, managed, used, shared. Building change within organisations and networks.

What type of organisation do they work in?

For example:

- Academia.
- Local government.
- A quango.
- A research institute.
- Central government.
- A large charity.
- A community group, small charity or individual campaigner.

What is the culture or level of ‘maturity’ around data in their organisation, or the organisations they work with?

Data-driven, data-focussed organisations with highly skilled data teams, or organisations just beginning on their data journey with little or no inhouse data expertise.

What power or influence do they have within data systems

Funder, commissioner, supplier, customer? Big budget? No budget?





Open Data
Manchester

Open Data Manchester

Holyoake House, 8 Hanover Street
Manchester M60 0AS

hello@opendatamanchester.org.uk

+44 (0)161 885 3185



MANCHESTER
1824

The University of Manchester