DSH
Digital Solutions Hub

# Second User Research Report

## User Engagement for Requirements Elicitation

May 2024

MANCHESTER 1824
The University of Manchester

UKRI Natural Environment Research Council

**Funders**
This work was supported by the Natural Environment Research Council (NERC).
[grant number NE/W001985/1]

**Document Purpose**
This document is produced by Digital Solutions Programme team at the University of Manchester, as part of the user engagement strategy to design a user-centric Hub for NERC's environmental data.



# Contents

## A. Second Phase of DSH User Research 3

Why follow up? – Completing the Picture 3
Methodological Choice: Interviews 4
Interview Design 5
Participants Demographic 6
Limitations and considerations 6
From User Research to Requirements Elicitation 7

## B. Demystifying Terminologies Around Data Use 8

1. "Data Granularity" 8
2. "Data Quality" 9
3. "Data Maturity" 9
4. Trustworthy Data(sets): 9
5. "Legacy Data(sets)" 9

## C. Identified Data Use Scenarios 10

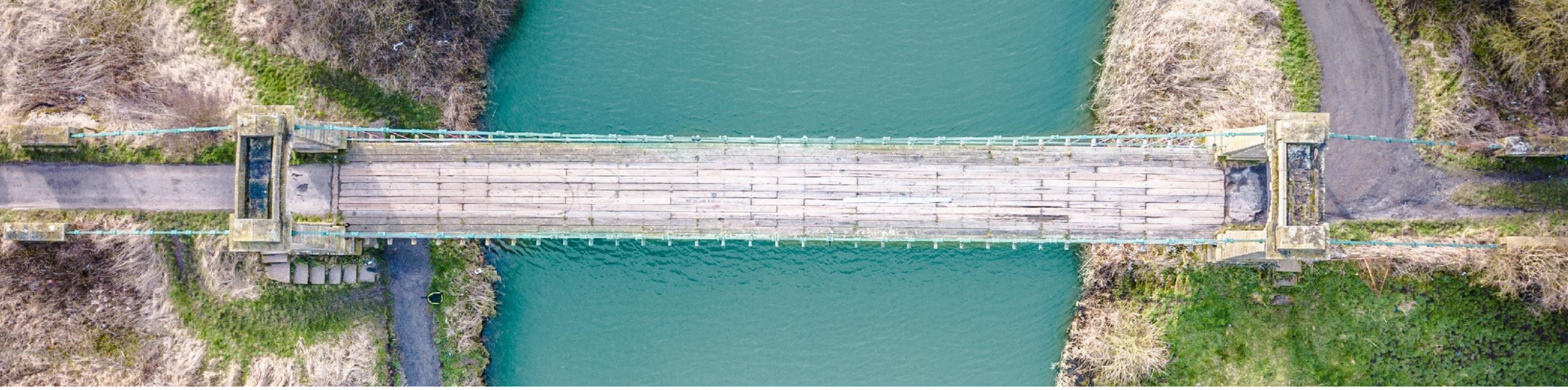
Users' Intended Outcomes and Output 10
Research Topics for Impact 10
Data and Expertise as a Desired Outcome 11
Data Use Scenarios Mapping 11

## D. Recommendations for NERC Digital Solutions Hub 15

Reflections on Requirements Catalogue 15
Recommendations for Metadata Quality 16
Recommendations for Data Quality 16
Recommended Tools 17
Recommended Data Formats 17
Recommendations for DSH Value Proposition 18

## Next Steps for DSH User Engagement & Requirements 19

Immediate next steps 19
Medium term steps (till end of 2024) 19
Long-term steps (2025 and beyond) 19



# A. Second Phase of DSH User Research

The Digital Solutions Hub (DSH) is a £7M gateway to a broad set of data and toolkits, funded by The Natural Environment Research Council (NERC). DSH targets the delivery of FAIR[1] access to data and decision support tools to a wide variety of users, with the needs of these users at the heart of the design and development of the hub.

Hence, user research on the Digital Solutions Hub has been a key component in driving the development of design requirements and features according to what users told us they need. This has been done through a multi-phased journey of user engagement activities:

1) Stakeholder scoping phase in early 2022,
2) Four online exploratory workshops in Summer 2022,
3) Twelve in-person user workshops in Winter 2022-23 in cities across the four nations,

and the phase presented in this report:

4) **Follow-up user interviews in Winter 2023**.

This report demonstrates the design, execution and insights derived from the user interviews in Winter 2023. A further set of interviews are currently underway and being led by the Connected Places Catapult to understand the business case for DSH.

[1] More on FAIR Principles.
[2] User Research Report published on DSH website here.

## Why follow up? – Completing the Picture

Following the in-person workshops in 2022-23, University of Manchester (DSH) and Open Data Manchester produced the first user research report[2], describing the abovementioned steps 1, 2 and 3. From there, the DSH team worked on synthesising these insights into a Requirements Catalogue[3] for the hub.

Upon synthesising this data, we identified three key topics that required further attention to support the DSH design. We designed interviews to address gaps in the analysis from phase 1 (described in the following methodology section). The findings analysis in this report addresses these three aspects in parts B, C and D.

The different user engagement stages are part of a long-term iterative agile software development process to develop the hub. Hence, building a continuous rapport with our users was also an inherent objective of the follow-up interviews.

[3] *A Requirements Catalogue* describes the characteristics that the software must have to satisfy the stakeholder needs (source: British Computer Society BCS).

## Methodological Choice: Interviews

DSH has adopted a multi-phased user-centric process (Figure 1), starting with scoping and context exploration, funnelling down in detail, through the different stages.

We used workshops in steps 1, 2 and 3 to explore the breadth of scenarios associated with multiple user groups in dealing with environmental data.

For step 4 (presented in this report), the objective was to zoom into the details of previously mapped scenarios and fill in gaps on user needs. Hence, interviews were selected to offer in-depth discussions with each stakeholder, rather than the less person-focused format of a workshop.

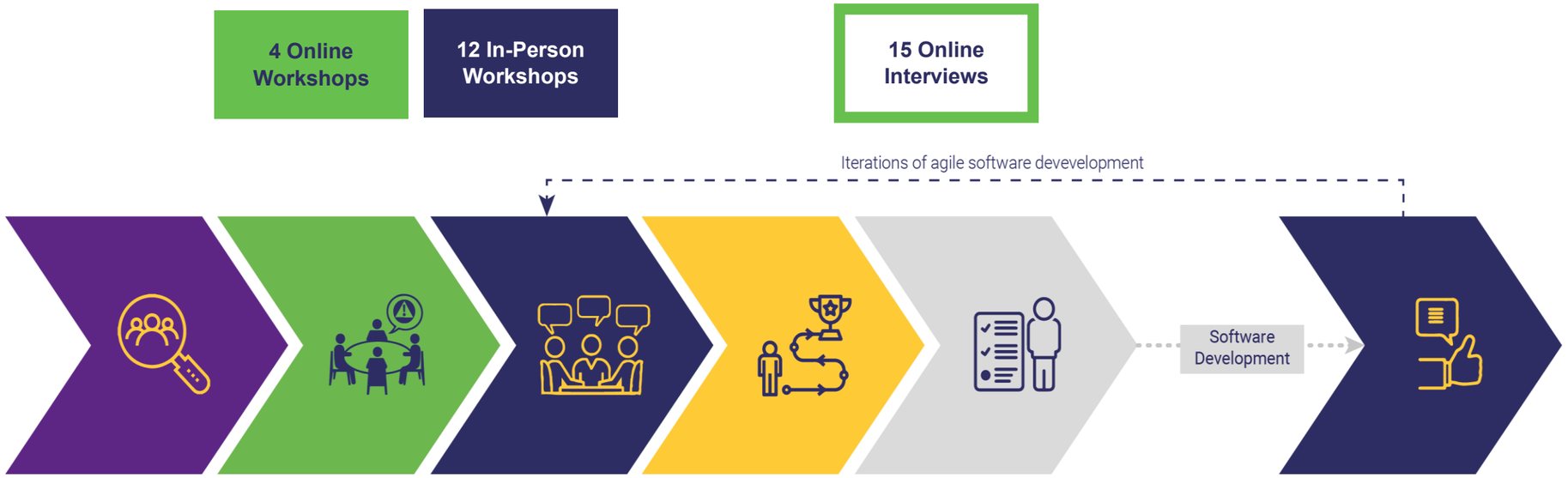



*Figure 1 - DSH Multi-phased User Research Strategy*

## Interview Design

Three main aspects were targeted in the design of the interviews and developed into **three interview models**. Each of these focused on a specific angle in the user's journey with environmental data, allowing an in-depth conversation on certain elements of this journey.

The **interview structure** was consistent across the three models. All three models walked the user through the different stages of their journey with data (identified in our previous user research report) (Figure 2). This helped users recall details and examples on each question and prompt, as questions followed the logical sequence they follow in their role/organisation. All interviews started with the user's desired outcome and output, going through the different stages along the user journey, and finally ending with a reflection on the overall journey and blue sky thinking of potential future developments they would like to see on DSH.

The interview models were:

- **Type 1 - Data model** focused on what users considered as trusted and high quality data, and which data they would use in their work.
- **Type 2 – Tools model** tackled the technologies and tools users already use to handle data, which would ideally integrate into DSH, and processes of searching for, cleaning, transforming and analysing data.
- **Type 3 – Collaboration model** focused on how teams operate in a collaborative environment to arrive at a common goal with data.

The interviews were designed to be **visually interactive** (**Error! Reference source not found.**), conducted using Miro. The journey was mapped on a "climb" metaphor, with journey stages mapped across going "up a mountain". The visual metaphor was well-received, and interviewees expressed they remained engaged even through a sometimes 90 minutes long interview.

The interview **medium** was online via Microsoft Teams. Interviews were audio recorded, with the Miro board on screen (cameras off). This allowed users to focus solely on the questions and visual, which was positively viewed in users' feedback.

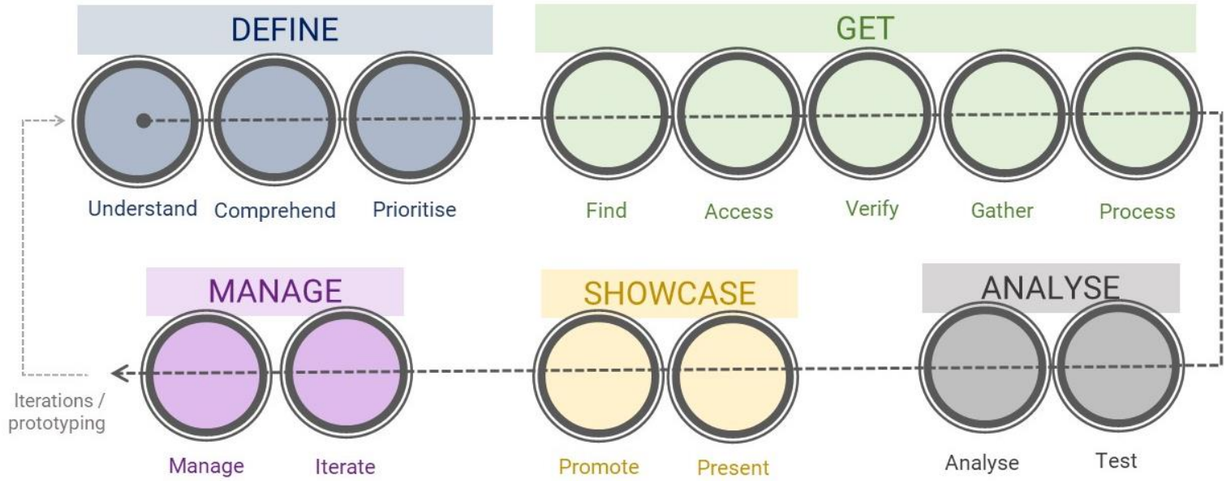



*Figure 2 - High Level User Journey (identified in phase 1 of DSH user research)*

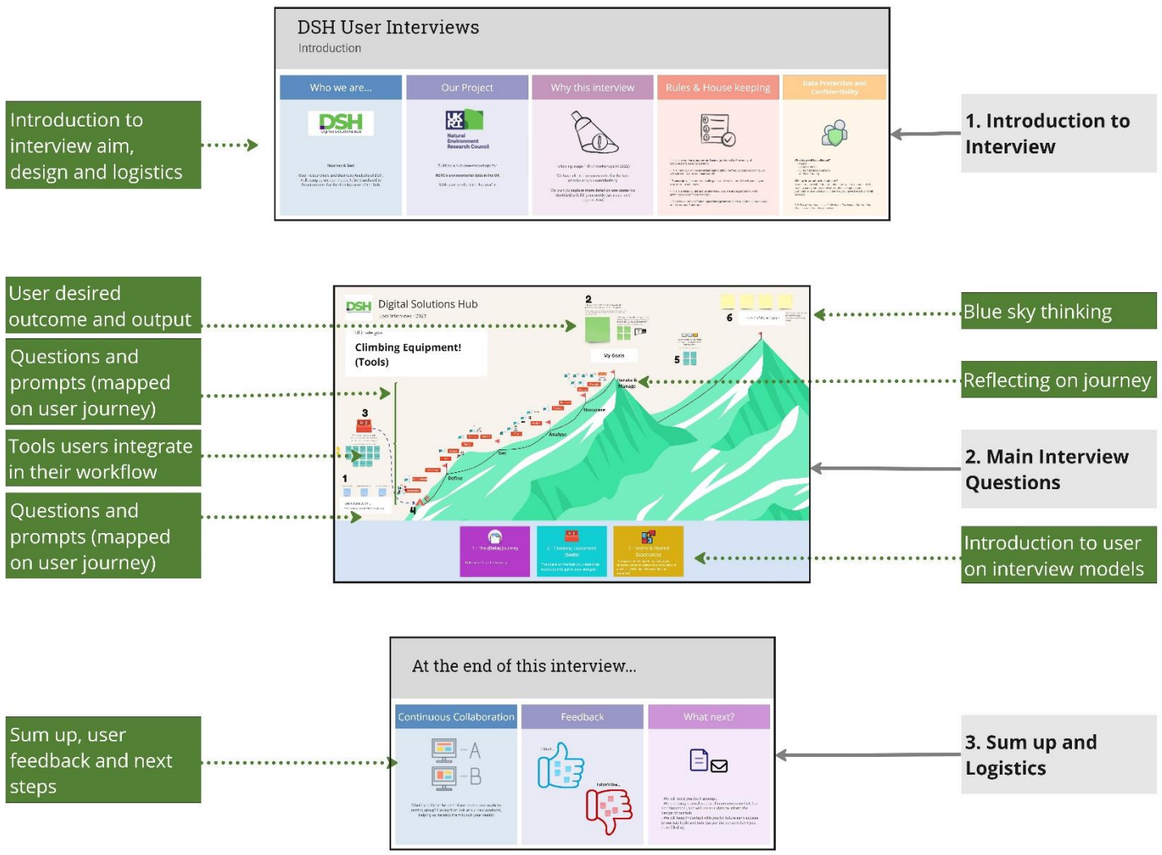



*Figure 3 – Visually interactive interview design on Miro*

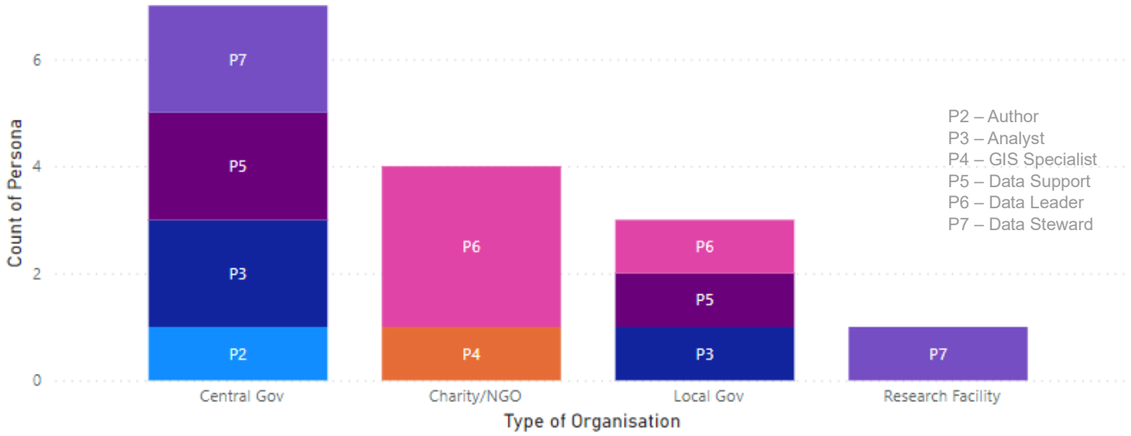



*Figure 4 - Interviews participants demographic – User Personas vs Types of Organisations*

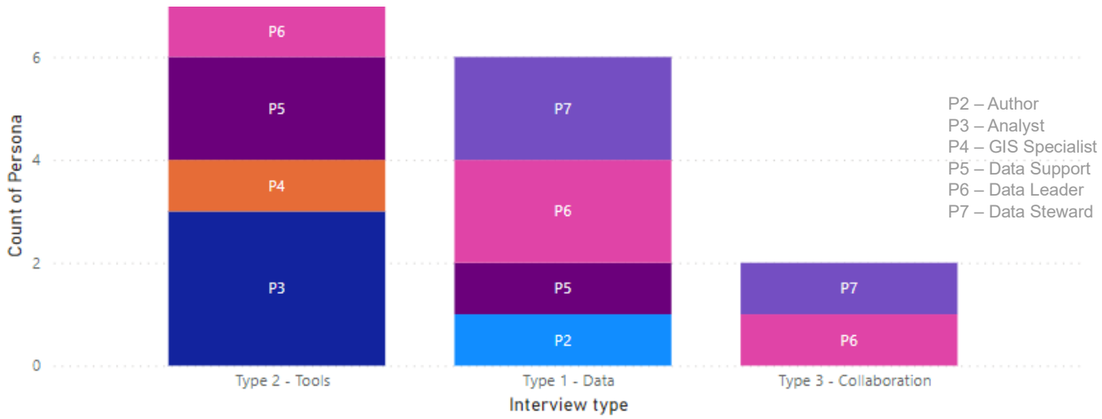



*Figure 5 - Interviews participants demographic – User Personas vs Interview Models*

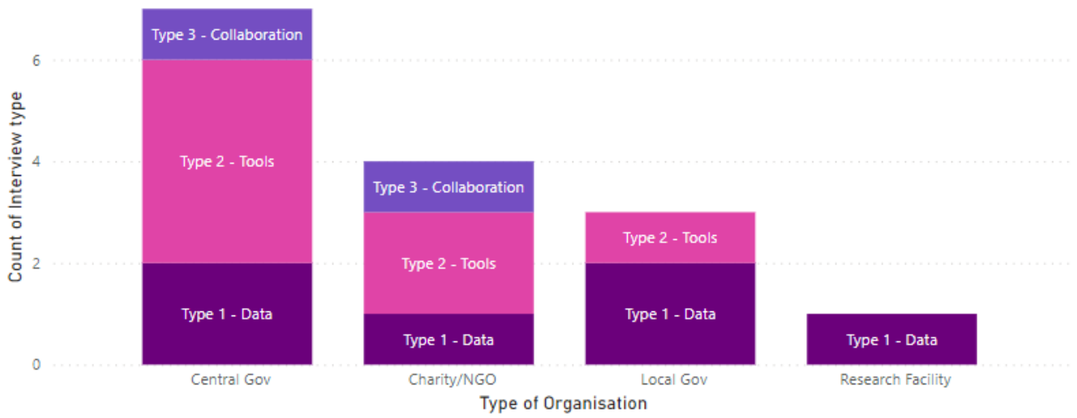



*Figure 6 - Interviews participants demographic – Interview Models vs Types of Organisations*

## Participants Demographic

For all user research phases, we worked with a range of participants in local and national government, health sector, urban and regional planning, environmental science and a whole range of local and national agencies. The main common factor between all our targeted stakeholders is the use of environmental data in their roles and workflows to arrive at informed outputs and decisions.

We held 15 interviews with participants from Central gov (7 users), Local gov (3 users), charities (4 users) and Research Facilities (1 user), representing multiple user personas[4] (Figure 4). All interviewees were from the original workshop attendees' sample (100 users from phase 1). This was important as it implied that the conversations were a continuation of user input from the workshops.

The choice of user personas per interview models (Figure 5 and Figure 6) was as follows:

- **Type 1 (Data)** Multiple personas to capture the perspectives of users from different sides of the data use spectrum.
- **Type 2 (Tools)** Technically oriented personas, who are more acquainted with technologies and tools in data use.
- **Type 3 (Collaboration)** Users more likely to work in a team within their roles.

More focus was placed on interviewing users from different types of organisations for interview models 1 and 2, as these represent the pillars of functionality expected in the first iteration of development within DSH (Figure 6).

## Limitations and considerations

As is the nature of qualitative data collection, an inherent level of generalisation is expected. Insights are based on the answers of the 15 participants, who represent their organisations and roles. With the large scale of some of our key central and local government organisations, it is important to acknowledge that each organisation has multiple user typologies and roles, as well as runs multiple lines of projects that require different expertise. It is also important to acknowledge that with different (or greater number of) participants, different angles might have come to light in this investigation.

[4] Further details on the personas in the User Research Report published on DSH website here.

## From User Research to Requirements Elicitation

The outputs of user research (phases 1 and 2) have been elicited into Software Requirements that constitute the DSH Requirements Catalogue[5]. The Catalogue consists of functional[6] and non-functional[7] requirements which define the capabilities and functionalities that users would like to see on DSH to address their needs.

To construct the Requirements Catalogue, the elicited requirements were grouped into high level capabilities (extracted from the report). Under each Main Requirement different granular functionalities were clustered and illustrated as per the users' input. Together, the Main and Granular requirements form the Requirements Catalogue were derived (Figure 7).

The catalogue informed identification of enablers (backend functionalities), capabilities (frontend functionalities) and users need (Figure 8). These are aligned with DSH Architectural Vision, Data Orchestration Model and identified Use Cases. This process feeds into the software development of DSH, while striking a balance between user needs and technological innovation. This stage is not covered in this report.

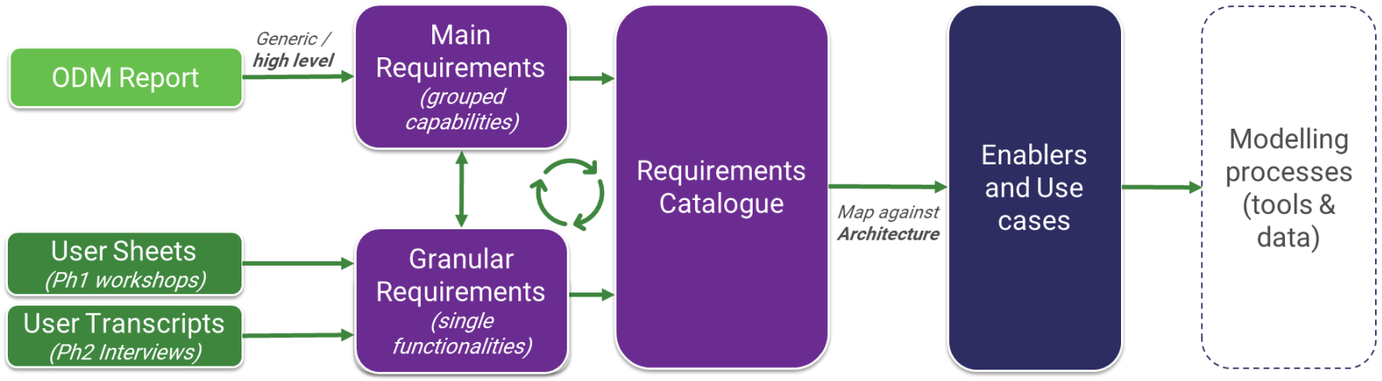



*Figure 7 - Requirements Elicitation from User Research to DSH Requirements Catalogue and Models*

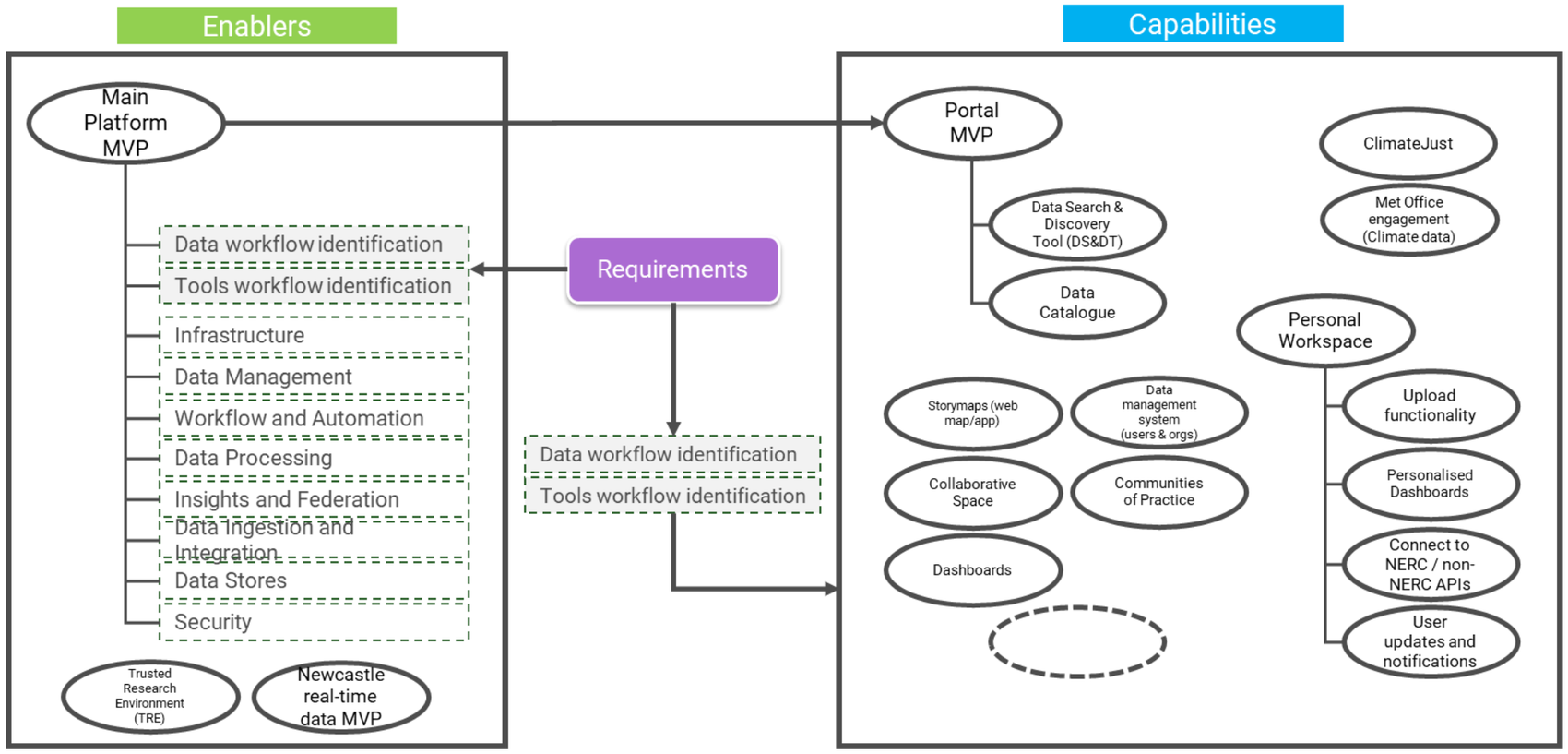



*Figure 8 – Translation of Requirements into enablers and capabilities for software development (based on phase 1 data)*

[5] *A Requirements Catalogue* describes the characteristics that the software must have to satisfy the stakeholder needs (source: British Computer Society BCS).
[6] *Functional Requirements* focus on functions and features of a solution (source: BCS).
[7] *Non-functional requirements* focus on how the behaviour of a solution is expressed or constrained, e.g. performance, security, usability, or portability (source: BCS).

# B. Demystifying Terminologies Around Data Use

From phase 1 research, we concluded that different user groups view certain terminologies differently according to their field of interest, technical background, or use within certain organisations. This usually causes confusion in collectively understanding what users need in the data space. For this reason, the interview design focused on exploring what users refer to when mentioning key terminologies that relate to their data use. Terminologies explored here are:

- Data Granularity
- Data Quality
- Data Maturity,
- Trustworthy Data(sets) and;
- Legacy Data(sets).

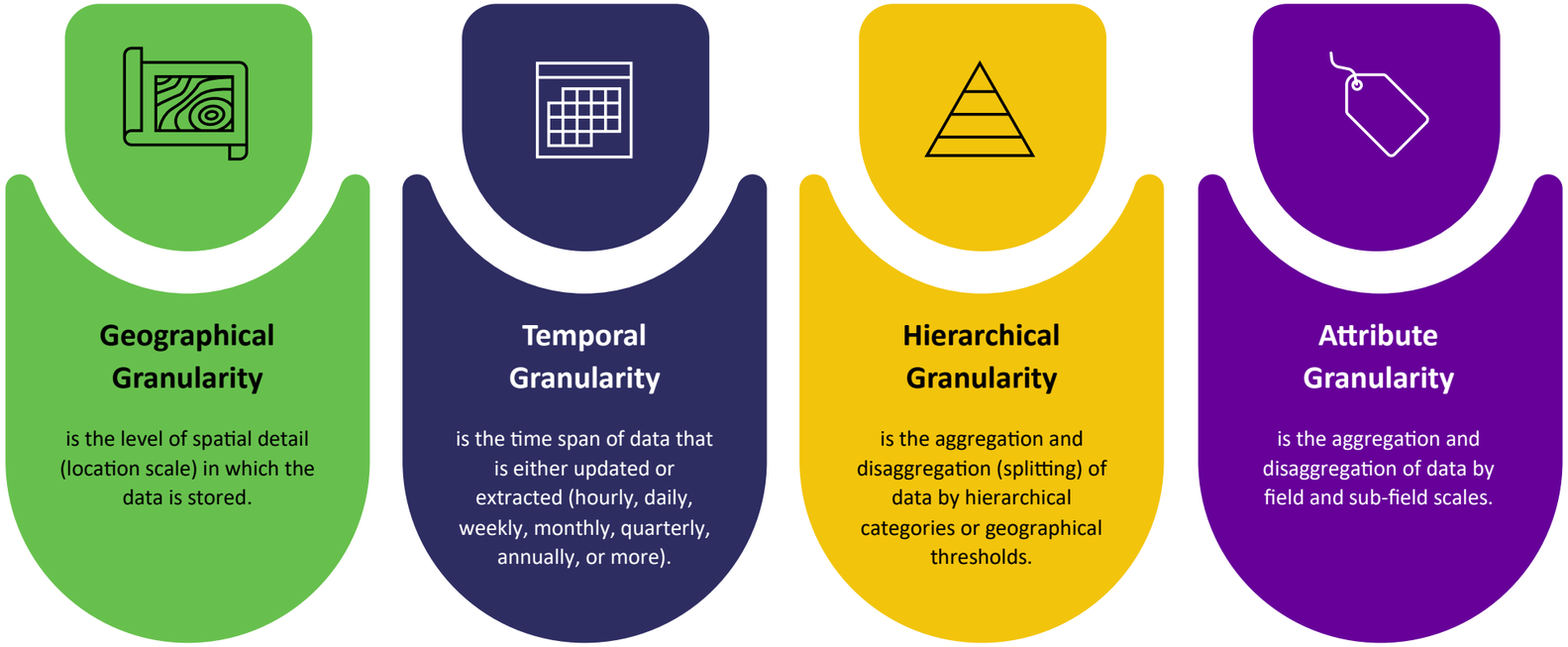
## 1. “Data Granularity”

In investigating types of data granularities, we realised that users unintentionally group all types together, leading to confusion about terminologies. We attempted to decouple these, to be able to design use cases and data processes for each type as needed from a data orchestration perspective.

Digital Solutions Hub

DSH
Digital Solutions Hub

## TYPES OF “DATA GRANULARITIES”




| Sector | Mostly interested in |
|---|---|
| Local Gov | local geographies at ward, MSOA, LSOA, postcodes or individual land plots. |
| Central Gov & Third Sector | Spans from individual sensor data to aggregated regional or national datasets. Interested in local scale when relevant to studies of a unique environment. |
| Scientific User Groups | Spatial detail very much depended on the projects’ scale and collaborations. |

| Type of Data | Relevant time spans |
|---|---|
| Health | Updated monthly. |
| Long-term Health Conditions | Do not change as frequently over time. |
| Census | Every 10 years. |
| Deprivation | Every 4 years. |
| Air Quality | Do not change as frequently. Changes for different times a year |
| Climate Change | Do not change as frequently over time. |
| Biodiversity and Plant-related | Mapped by seasons rather than temporal units (quarterly data flow needed). |
| Soil carbon concentration | Spans across 5 years |
| Environmental | Spans across 5-10 years |
| Water | Changes daily or monthly |
| Climate-based Observation | More longer term. Typically hourly averaged. |

| Example desired outcome | Granularity processes |
|---|---|
| Investigate data relevant only to river catchment scales | Data broken down to go just below catchment scale. |
| Interested in sub-field data for certain scales | Data broken down to sub-fields according to scales. |

| Example desired outcome | Granularity processes |
|---|---|
| Interest in analysing a given field in detail | Data broken down to sub-fields. |

**Note:** There was a common use among users to refer to Hierarchical and Attribute Granularities interchangeably.

**Granularity Type Significance:** To allow for meaningful and pragmatic discussions of data across different scales.

**Most used for:**
- Catchment data
- Data on green spaces

## 2. “Data Quality”

When asked “what does Data Quality mean for them”, users told us numerous variables they need to see in a dataset to consider it of high quality. These are:

- Data that **provides the needed coverage** - it aligns with the identified granularity preferences.
- Data that is **complete, up to date and timely**.
- Data that is **correct, clean**, with relevant and “**tidy**” content.
- Data that displays a **high level of confidence** and the user identifies is **from a trusted source**.
- Data that has **reliable metadata** which is captured and communicated.

## 3. “Data Maturity”

Users have correlated “maturity” with the term **“well-established” data**. This can be the maturity of the data, the provider or the end user.

**Mature Datasets:**

- Data published periodically and consistently (over a long period of time).
- Data from a trusted organisation.
- Data collected using well-known and widely used methods in the respective field.

**Mature Skillset:**

- Users who have the needed skillset that enables them to make most use of the data add to the maturity data use process.

## 4. Trustworthy Data(sets):

To trust data, users told us they need to also trust the data providers.

- Data **collected long-term** by established programmes (“old data”).
- Data of thorough **provenance records** (source? expertise and experience of the provider?)
- Data that meets **ethical standards** for data collection and sharing.
- Data from research groups with a **consistent track record** of producing data on a particular area.
- Data that has **testimonials** from those who have used it.

## 5. “Legacy Data(sets)”

Legacy datasets was referred to as being:

- **Historical** data.
- Data that was measured in the past and **has gone out of date**.
- Legacy data is potentially **not accurate**.
- Datasets where the **data collection methodology would have changed**.
- Data **stored** in a way that cannot be easily found.
- Data stored potentially in **formats** that have changed / gone out of date.
- Data that’s **accessibility** has changed/gone out of date.
- Data that provides **value** through comparisons and analysing **trends over time**.

# C. Identified Data Use Scenarios

After understanding different terminologies with data use, we started mapping user scenarios relevant to searching, accessing, processing and analysing data towards achieving outcomes (as mapped from the phase 1 main user scenario in Figure 2).

## Users' Intended Outcomes and Output

As part of the interview questions and through a pre-interview survey, we asked users about their intended outcome[8], how that translates into an output[9], and how this outcome supports an impact[10] using data.

## Research Topics for Impact

Three main themes prevailed amongst interviewees, which were 1. Environmental Management, 2. Urban Planning, and 3. Public Health. User's desired impact ranged from monitoring environmental phenomena, scaling outcomes on different scales (aggregating or disaggregating results), or informing design of space/land use, services for communities or tools supporting communities.

[8] **Outcome** here is the expected direct result or goal the user aims to achieve from the process dealing with data.



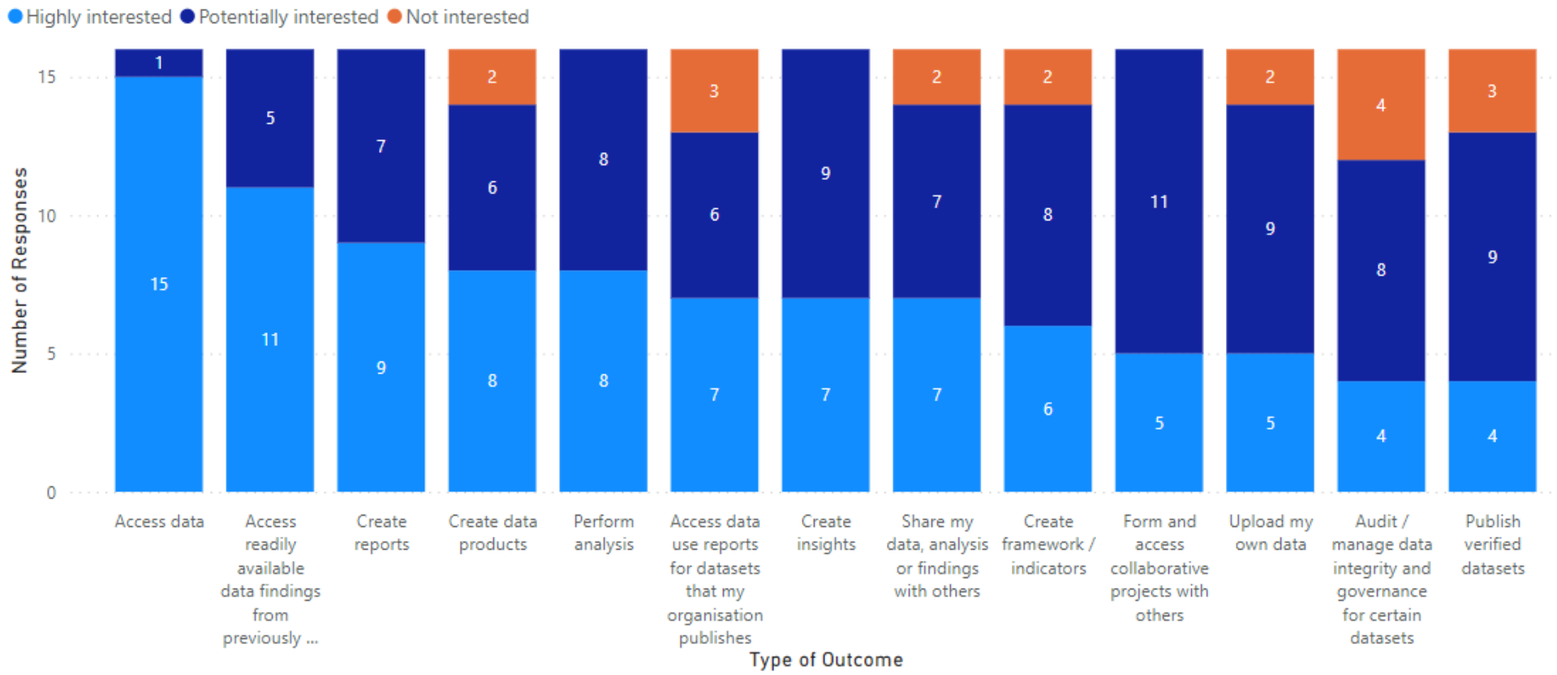



*Figure 9 - User preferences on Outcomes - Source: Users Responses to Pre-interview survey.*

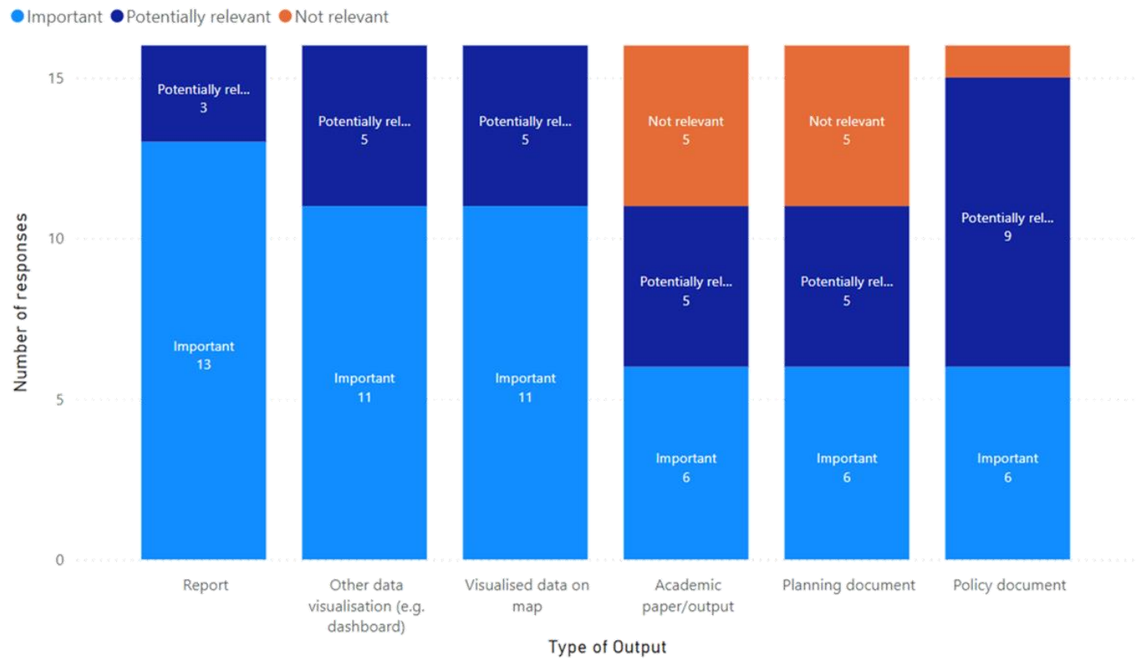



*Figure 10 - User preferences on Outputs - Source: Users Responses to Pre-interview survey.*

[9] **Output** here is the medium in which the outcome is conveyed, which can ideally be shared with others (e.g. a report or digital media).
[10] **Impact** here is the broader influence of the process of data use on the target audiences.

## Data and Expertise as a Desired Outcome

Some users were interested in an outcome being a different version (possibly via data transformations or content analysis). Users in Central Gov were particularly interested in DSH supporting them with finding more about how others use their data, and the gaps in them. National data providers were interested in aggregations of data on a national scale, while users from Research Facilities were interested in matching user network expertise with the relevant data.

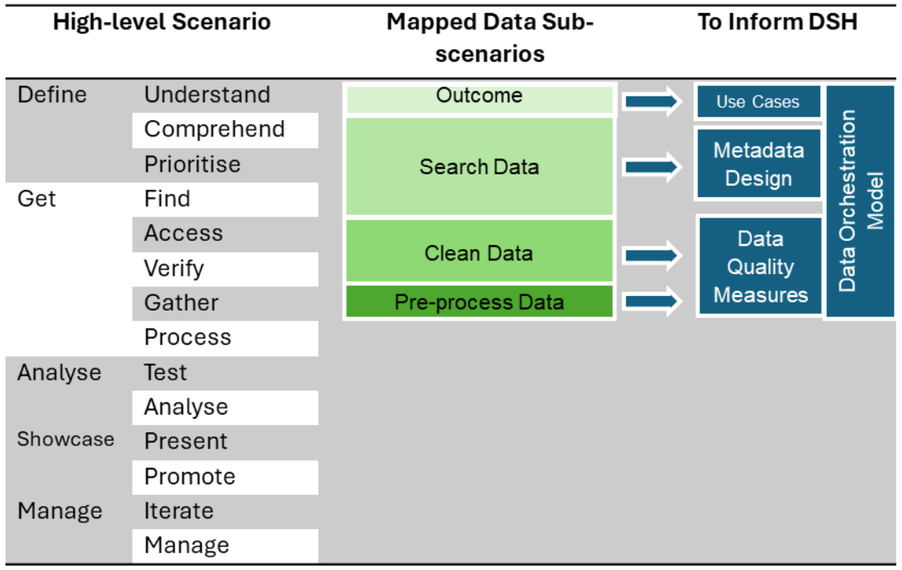
*Table 1 – User mentioned desired outcomes relevant to data and expertise.*

| User Group | Data Input | Desired Data Output |
|---|---|---|
| **National scale data providers** | **Local data** (possibly disconnected datasets) on certain environmental measures | Developed **national scale datasets** for these measures |
| **Data providers** | Organisation's **shared datasets** | Information on:<br>- **Usage** of these datasets<br>- Which **data** was **combined** with them |
| **Research Facilities** | A complex question that **requires multiple stakeholder expertise** to tackle | - **Matchmaking** and identifying **key expertise** to collaborate to identify "unknown unknowns"<br>- Ensuring the data is fully **developed for impact**. |

## Data Use Scenarios Mapping

We extracted different sub-scenarios as alternatives to achieving the user outcome at each stage of handling data. Our first finding was that although the high-level scenario (shown in Figure 2) corresponds to the experience of all users we spoke to, the sub-scenarios hugely vary and sometimes happen in-parallel or in-series according to the context of data use.

*Table 2 - User Sub-scenarios mapped against initial Scenario and expected input into DSH design*



## *1. Searching Data*

We mapped several techniques that users told us they use to find the data they seek. All which map to one holistic process of identifying the aim, initial search, focused search and finally assessment of suitability of the result (Figure 11).

| Process Steps / Type of Search | Identify Aim/target | Identify Source | Perform Initial Search | | Perform Focused /Refined Search | Assess Results Suitability | |
|---|---|---|---|---|---|---|---|
| **Pre-search** | Identify search aim / target | Experiment with search terms to check capabilities of the search functionality | | | | | |
| **Keywords search terms**<br>**OR; Scientific Literature Search Equivalent** | Identify search aim / target | | Perform Initial search | See relevant synonyms | Refine search | Explore results | |
| **Search by institution / data provider** | Identify search aim / target | Identify data provider of interest | Perform Initial search | | Check newest data by this provider | Explore results | |
| **Explore broad topic (using a visualiser)** | Identify search aim / target | | Perform Initial search | | Access visual (graph or other) which shows connections / relations between datasets | Check if there is an existing visual of the data | Check visual to assess data is fit for intentded use |
| **Map-based search: Searching data available for certain locations** | Identify search aim / target | | Perform Initial search | | | Check wider geography → Data not found<br>Explore datasets in set location → Data found | |
| **Explore data from academic papers** | Identify search aim / target | | Perform Initial search<br>*Connection to Research Portals (academic outputs)* | | Check relevant papers with datasets | Follow citation trees to check other works using the data | |
| **Prompt / challenge statement based search**<br>**(AI approach)** | Identify search aim / target | | Write prompt / challenge statment / open ended question | | Further prompts to the system | Explore results | |
| **Search data via API connection** | Identify search aim / target | | Find API packages that connect to the data needed | | | Explore results | |

*Figure 11 – Mapped processes for "Searching Data"*

## *2. Cleaning data*

The focus of the conversations were spatial data, and particularly environmental datasets. Hence, the most prominent type of data users told us they clean was **tabular data**. We mapped all stages users would go through with their datasets to arrive at a "clean" usable data (Figure 12).

| Process Steps / Type of Data | *Access source dataset* | *Filter source data* | *Get source data* | *Overall Transformations* | | *Field cleanup* | | | *Entries cleanup* | | | *Data Trimming* | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Tabular Data** | Find dataset | Choose / filter needed portion of data | Download dataset | Ensure there is one table per sheet | Ensure each sheet has headers | Ensure consistent field types | Compare datasets of same variables | Ensure no fields are missed<br>Ensure all fields are in the right place in the dataset<br>Ensure all fields are labelled correctly * | Check for data entries consistency | Check all entries have same number of fields<br>Check if entries have extra fields<br>Check if any entry has missing/incorrect data *<br>Check if missing values can be added manually<br>Explore why data might be missing | | Trim down data into only user needed fields/entries | extract trimmed down data into users system |
| ***Data from API*** | Find data through API | Choose / filter needed portion of data through API (before pulling it) | See relevant synonyms | Done on the user's system (not DSH) | | | | | | | | | |
| ***Temporal Data*** | | | | | | | | | Check if time dataset has missing values | Block time slot that has missing data (use data without it) | | | |
| ***Spatial Data*** | | | | Run CRS transformation<br>Transform the coordinate reference system (re-project data) | | | | | | | | | |

*Figure 12 – Mapped processes for "Cleaning Data"*

*** Criteria for correct field labels (same for checking missing data):**

- Name matching.
- Duplication searches.
- Cleaning names.
- Cleaning geographical locations.
- Letters are correctly cased (uppercase/lowercase).
- All fields similarly set up (Kebab case or Snake case).
- All numbers are correct.
- All dates are accurate and adhere to ISO-8601 (year-month technique).

## *3. Transforming Data*

Here we expanded on stage 4 on the data cleaning process (overall dataset transformations). We mapped all transformations users go through to "change" datasets in terms of format, aggregation, scale, etc (Figure 13).

| Process Steps / Pre-processing Task | Identify Desired Outcome | Methodology | Process | | | |
|---|---|---|---|---|---|---|
| **Changing data format** | Getting the data in a format that would be useful | Change format | Check if data format is suitable | Check options of available formats | Change format to preferred option | |
| **Changing data projection** | Getting the data in British National Grid format | Re-project | Check data format and projection | Change other projections to a British National Grid projection | | |
| **Change granularity** | Change data granularity to have a different perspective in the analysis | aggregate / disaggregate (e.g. Hexpin data) | Check if different perspective is needed on data | Aggregate data<br>OR: Disaggregate data | Choose options of data layouts / functions | |
| **Changing units of fields** | Get same information in different units | Calculation-based new field | Check if data units are suitable | Find needed calculation to transform data units | Create new field with new data according to transformation | |
| **Data scaling** | Process data to be ready for modelling (scaling in terms of extrapolation or intrapolation) | Extrapolate / intrapolate data | Align model outputs that are working at different scales | Check if coverage/scale is suitable/sufficient to model | Interpolate data<br>OR: Extrapolate data | Add data to model |
| **Create indices or scores** | Process data to be ready for modelling (create indices to organise messy data) | Create indices or scores | Align model outputs that are working at different scales | Create Indices or scores | Model data | |
| **Trim data** | Select areas of interest and trim the data to fit these areas only | Scripts or map-based tool | Decide on area of interest | Choose tool to use to trim the data | Trim data | |

*Figure 13 – Mapped processes for "Transforming / Pre-processing Data"*

# D. Recommendations for NERC Digital Solutions Hub

The last area of focus within the interviews analysis was to reflect on the Requirements Catalogue (enablers and capabilities initially presented in Figure 8 above) and other data orchestration recommendations considering the new insights from users.

## Reflections on Requirements Catalogue

With more in-depth conversations in the interviews (as compared to the workshops), users told us more about the functionalities they would like to access on DSH.

New capabilities mentioned by users were (updated in green in Figure 14):

- Access to an API (Application Programming Interface).
- Access to AI-powered search.
- Access to HPC (High Performance Computing) capabilities, potentially via VMs (Virtual Machines).

We also mapped additional requirements on existing capabilities directly into our Requirements Catalogue (not covered in this report).

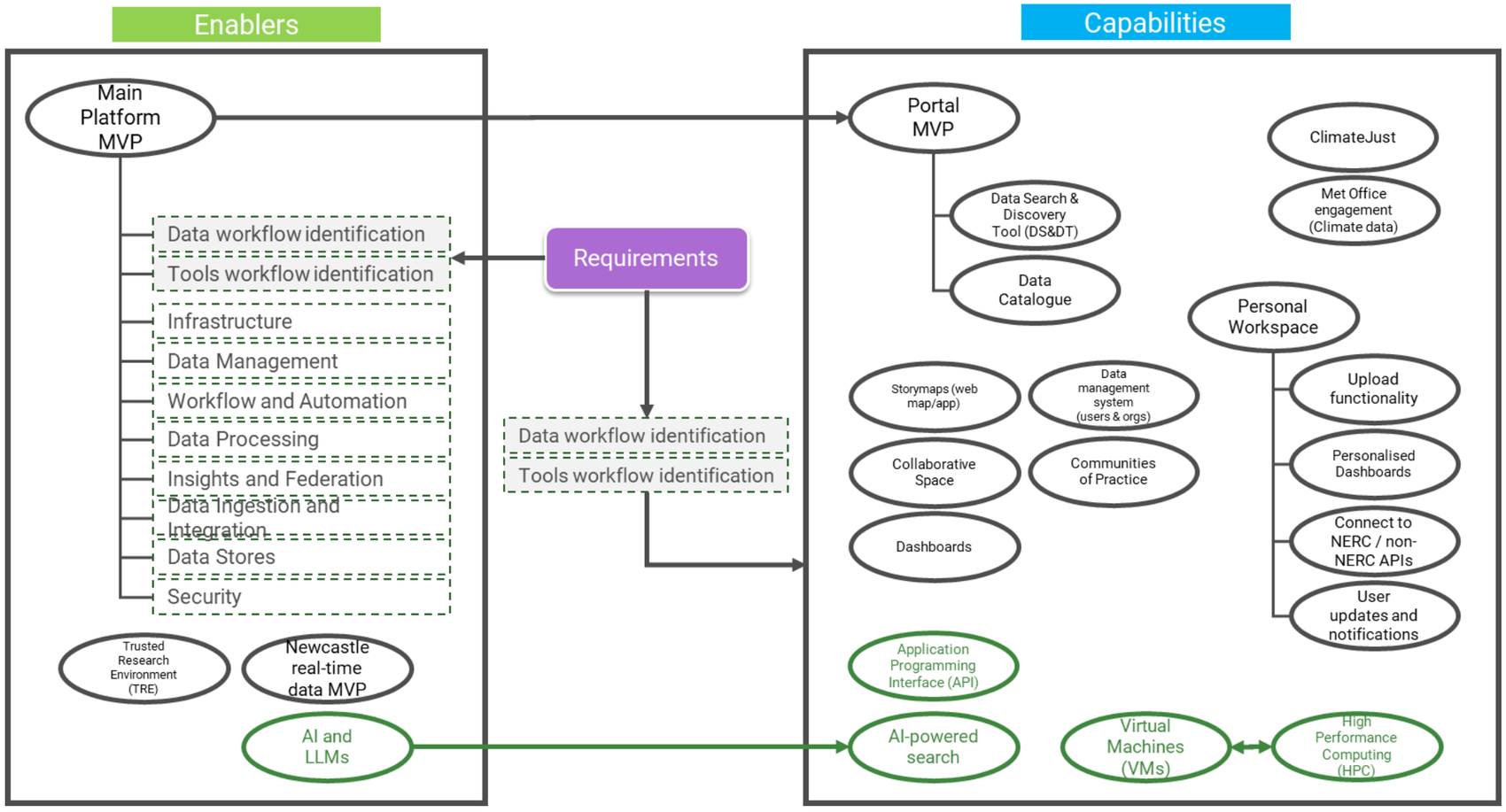



*Figure 14 – Updated DSH Enablers and Capabilities considering phase 2 input from users*

## Recommendations for Metadata Quality

These are recommendations based on what users told us they would like to know about datasets before downloading/using them. These could be addressed as additional metadata descriptors or could be added elsewhere attached to the datasets.

**Consistent, good quality metadata should include:**

- Date of latest dataset update.
- Data use and purpose of datasets.
- Data Coverage.
- Considerations or aspects to address before handling this data.
- Data Provenance.
- Data Lineage Record.
- Metadata tags.
- Whether data is machine readable.
- Data IP and licensing information.
- Account of data collection and sampling methodologies.
- Protocols (and processes) that were deployed to both collect and validate the data.
- Ethical considerations for data collection, processing and sharing.
- A measure of consistency of methodology of data collection over different versions of the dataset.
- Expected confidence intervals in the data and sensitivity analysis of expected impact on output.
- Expected risks, limitations, and biases relevant to the dataset (e.g. limitations in the data collection, or reason for higher margin of error in the data).
- Data Impact: Dataset-derived publications or peer reviewed academic content.
- Protocols (and processes) that were deployed to both collect and validate the data.

## Recommendations for Data Quality

These are recommendations based on challenges users have identified regarding the data, which would inform the data orchestration processes and pipelines within DSH.

**Consistent, good quality data should be:**

- Data that provides the needed coverage (aligning with identified granularity preferences).
- Data that is correct.
- Data that is clean, with relevant and “tidy” content.
- Data that is up to date and timely.
- Data is complete (as much as possible).
- Data that is available (as open as possible).
- Data representation (of UK, particular to central gov organisations).
- Data that can be interpreted correctly.
- Displays high level of confidence.
- Data the user identifies as trusted / reliable.
- Data the user identifies as mature.
- About reliable Data Collection.
- Data in convenient formats.
- Data that has the proper IP and licensing information attached.
- Data with reliable metadata captured and communicated.

## Recommended Tools

We asked users about tools they integrate in their workflow that they would like to see on DSH. We also asked users about tools they do not currently use but would like to have access to as part of the services by DSH. These were:

*Table 3 – Tools mentioned as preferred by users in their workflows.*

| Tool purpose / Tool | Visualisation and sharing | Spatial Analysis | Machine Learning and AI | Interoperability | Development & Working in Teams | Analytics and Statistics |
|---|---|---|---|---|---|---|
| Analytical models hosted online | | | | | | |
| Aristotle System | ■ | ■ | | | | ■ |
| Atlassian tools (Jira, Confluence and other) | | | | | ■ | |
| Business analytics tools | | | | | | ■ |
| CEH E-Planner | | | | | | ■ |
| Digital Twins | | ■ | ■ | | | ■ |
| ESRI / ArcGIS (online and enterprise) | ■ | ■ | | | | ■ |
| Genstat | | | | | | ■ |
| Github | | | | | ■ | ■ |
| Google analytics | | | | | | ■ |
| Jupyter Notebooks | | | | | ■ | ■ |
| MS Excel | ■ | | | | | |
| MS Office | ■ | | | | | |
| MS Teams | | | | | ■ | |
| Other GIS tools (e.g. QGIS) | ■ | ■ | | | | ■ |
| Power BI | ■ | | | | ■ | ■ |
| Python | ■ | | | ■ | ■ | ■ |
| R (and R studio) | ■ | | | ■ | | ■ |
| SAS | | | | | ■ | |
| Slack | | | | | ■ | |
| Social Media / Content Streaming | ■ | | | | | |
| SPSS | | | | | | ■ |
| SQL | | | | | | ■ |

■ *User-identified purpose for using each tool*

## Recommended Data Formats

Preference for non-technical users: PDF and JEPG (image formats).
Preference for technical users: API endpoints for finding data of different formats.

Users told us they would like access to “standard data formats”, mentioning the following:

1) Tabular formats: Machine readable tables such as CSV/Excel.
2) Downloading GIS information and types of apps that can be mapped to coordinates.
3) Useable documented API format.
4) GIS formats: shapefiles and geopackage.
5) PDF and JPEGS (image data).
6) NET CDF.
7) Geo-Json.
8) LIDAR file format (radar data)
9) SQL build formats.

# Recommendations for DSH Value Proposition

In discussions on scenarios and impacts users would like DSH to assist them with, we mapped recommendations on areas where DSH can provide value to users (Value Proposition). Users stated the following propositions for DSH to potentially bring value to its end users:

**Products and Services**

1. DSH to be a one stop shop for the data that would allow datasets to be joined and then analysed much easier and have quicker response times for the different scientific areas to inform policy and write scientific journals. It's to generate interest for external funding as well.
2. DSH as a facilitator for developing new datasets and increasing the use of data for national work.
3. DSH as a platform to develop datasets out into the much wider areas or allow users to take data that has been collected for one purpose and apply it to other purposes or to a more national scale.
4. DSH to have the power of Jasmin.
5. DSH to allow users to use their own data, upload it to DSH and modify it, and have it be open and have access to tools and analysis to derive insights from the data. That can support: 1) A citizen science approach for interested members of the public being able to explore it and find out meaningful insights to them, and 2) Academic and policymakers' inquiries, with questions to find answers to using the data as guidance and supporting evidence to decision making.
6. DSH to support API functionality, which users told us they are already pushing for within their organisations. This might be an interesting feature to market to such organisations.
7. To build a data hub and analytics platform that links to other data hubs, e.g. Agri-Food and Biosciences Institute (AFBI) and Environment Agency (EA) data hubs.
8. DSH to create access to some new embedded tools that users can explore how it could be used in their work.
9. DSH to enable users to search and find data and then download individual sets of data and work on it.
10. DSH to build data flows connecting through an API.

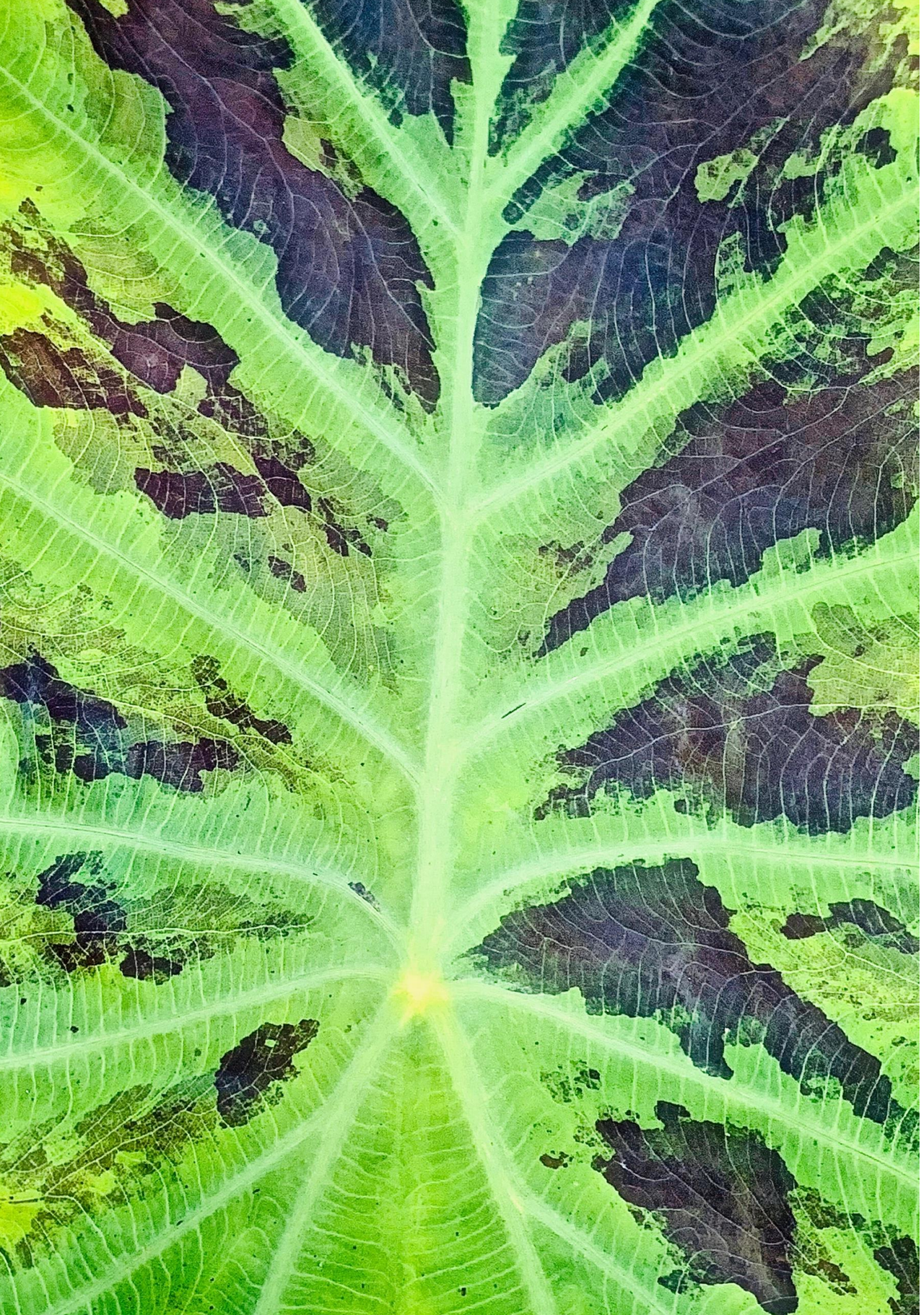
**Pain Relievers**

1. The DSH would be useful in pointing out certain data that is being produced that is of interest to Central Government Organisations (or other types of organisations).

**Gain Creators**

1. Sharing data directly with different interested groups and potentially hosting this data. Having a more open and accessible collaboration platform that is useful to open more ways of directly working with different academic institutions.
2. DSH as a platform where communities of practice and others can find spaces to work creatively together but with urgency and brings about ways of helping non-technical users be able to manipulate and explore evidence in a range of different ways.
3. DSH to demonstrate/show to potential users 'how' the hub works by demonstrating what can be done via the hub (leading users with examples). The hub should have a section where users can see what other people have done on the hub and how they have made benefit of it (demonstrating impact).
4. Making known the benefits of using the DSH and how it can be useful to their work.

Christabel Pankhurst Institute, Manchester.
(Digital Solutions Programme offices)





# Next Steps for DSH User Engagement & Requirements

At DSH, we are deploying a long-term engagement approach with our users to ensure all our tools are built to suite users' needs. The following are our immediate, medium term and long-term engagement plan to this end:

## Immediate next steps

- Update DSH Requirements Catalogue based on this data.
- Inform the prioritisation and scope of DSH use cases (in coordination with tech team).
- Modelling business processes for DSH use cases based on requirements.

## Medium term steps (till end of 2024)

- Work with partners to map sustainable business case alternatives (Connected Places Catapult).
- Further user engagement activities with key stakeholders to map requirements for upcoming use cases.
- Further user engagement via usability testing for developed use cases (iterative prototyping and testing).
- Iterative management of Requirements Catalogue.

## Long-term steps (2025 and beyond)

- Streamline DSH pipelines for ingesting user needs data and transforming inputs into workable products on the hub.
- Knowledge sharing and exchange with NERC projects and services ecosystem.
- Expanding user base of NERC's environmental data via engagement with DSH data products.

The University of Manchester
Oxford Road Manchester
M13 9PL
www.manchester.ac.uk

Created by Digital Solution Programme
2024