UK Centre for Ecology & Hydrology
Environmental Information Data Centre
#CC03 – 28th October 2025
10.30am – 12.00am
26 participants

Community Conversation

Co-Designing Environmental Information Data Centre Services

Report on #CC03: Enhancing data access and data deposit

Kathryn Harrison (Lead author: kath@ceh.ac.uk), Hannah Dean, Katherine Wright, Sharon Hayton, and Mary Preston. UK Centre for Ecology & Hydrology. 2025.

What did we do?

As the third in a series of online **Community Conversations,** this event focussed on Enhancing Data Access and Data Deposit, as part of the Environmental Information Data Centre's (EIDC) co-design development plan. The EIDC is part of the Natural Environment Research Council's (NERC) Environmental Data Service and is hosted by the UK Centre for Ecology & Hydrology (UKCEH). We manage nationally-important datasets concerned with the terrestrial and freshwater sciences. The EIDC supports researchers by ensuring that the data they produce are curated appropriately for the long-term, guaranteeing their security, accessibility and re-usability for the user community. This event builds on the first Community Conversation event, which focussed on enhancing discoverability and access to data, which highlighted the importance of rich metadata to describe data and AI driven searches to ensure data is findable and reusable. Dr. Kathryn Harrison, as the EIDC Operations Manager, opened and chaired the session, which involved two **presentations**, each accompanied by a **facilitated breakout session**. Prior to the presentation and breakout sessions, we conducted a series of **polls**, on participants' experience of depositing data with the EIDC, or elsewhere, and how they currently find data. Afterwards, we invited participants to complete an opinion **survey** on how the event was run.

Presentations:

Philip Trembath Senior Semantic Knowledge Specialist (UKCEH)	**Hannah Dean** Senior Data Steward (UKCEH)
An introduction to the EIDC and how to access the data it holds using the EIDC catalogue	An overview of the current data deposit process at the EIDC.

Facilitators: Matt Nichols, Katherine Wright, Philip Trembath, and Hannah Dean
Notetakers: Sharon Hayton, Dan Wright, Helen Rawsthorne, and Els Dhiedt

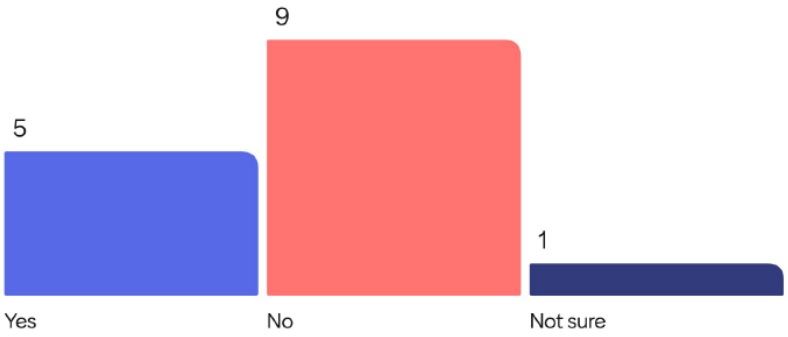
What did we find?

The core aim of the event was to understand more about how users find, access and reuse data, as well as gather their experiences of depositing data with the EIDC so that we can develop tools and services to meet their needs.

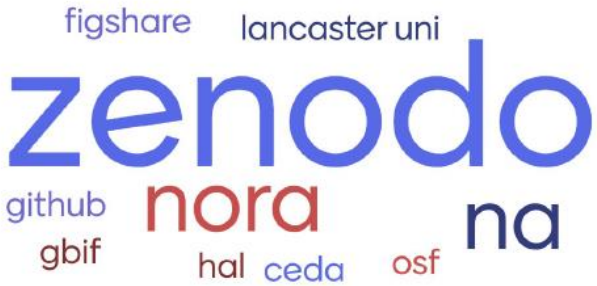
Some participants had prior experience in either depositing data or discovering data from multiple sources.

At the start of the event, we asked participants questions to gauge their previous experience of using EIDC services. The questions we posed were answered anonymously using Mentimeter and the results are shown below. Most participants **had not previously deposited data with the EIDC**. **Zenodo** was the repository most participants had deposited with. When looking for data, the three main routes participants reported using were **Google**, word of mouth and speaking to **colleagues**.

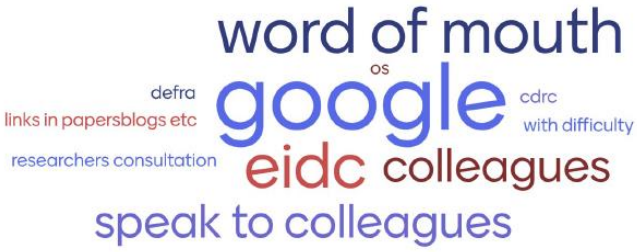
1) Have you deposited with the EIDC before?



2) If you have deposited elsewhere, where was that?



3) How do you currently find data (e.g. Google, EIDC catalogue, NERC data catalogue)?



Data needs to be better described so it is easier to find exactly what users want.

Participants reported that it can be time consuming to find data and often rely on colleagues to point them to datasets or use search engines such as Google. When they do search for data, they are often presented with a long list of datasets without knowing which are the most relevant for their needs. Participants also highlighted the need for richer, standardised keywords in metadata (e.g. scale, methods, units, biodiversity index used) to help them identify the data they require. They reported that they would value **AI-generated keywords or metadata extraction** to improve discoverability, as well as **support from thematic specialists to define standard keywords**. **Suggested additional or alternative datasets** would also save them time, for example, suggesting datasets that work well together, or more datasets in a series/collection.

On the challenge of finding data, one participant said: "I don't have time to go through all the datasets. I search with keywords but then have to explore each dataset individually to find out if it suits my needs."

Participants would like to preview datasets to avoid downloading data that doesn't meet their needs.

Participants reported that they often waste time downloading data which are not suitable for their purpose. Several participants said that a **data preview or data visualisation tool** (especially for data with associated charges) would allow them to check the data were suitable for their needs prior to download. Participants also said they would like information in the metadata to let them know **if the data are machine readable**.

Participants wanted faster, frictionless access to data

Participants wanted **quick and easy** ways of accessing data, **without forms and delays,** and **without needing to log in**.

Participants expressed interest in features that make data easy to discover, access, and reuse.

Participants expressed an interest in tools for **spatial/temporal clipping** of data, so they only download the data they need. There was a desire for **tools that support access and reuse**, especially for combining datasets with different units, scales or formats. Participants would like a better understanding of how to use Web Map Services (WMSs) and Application Programming Interfaces (APIs), as well as guidance on **assessing the reusability of datasets**. **How-to guides** for code and command line tools (e.g. Python, wget) with example scripts showing how to get data into applications for analysis.

An opt in service notifying users when a new dataset they are interested in is published was also noted as a useful development by participants.

Generally, however, they reported that **downloading from the EIDC was easy** once they had found what they wanted.

Depositing data should be quick and effortless, but the process is currently too manual

Participants who had previously deposited data with the EIDC found the process was **too long and felt too manual** with excessive back-and-forth communications with EIDC staff. They reported that they had experienced **inconsistent guidance** on data formatting (e.g. handling null values, text in numeric columns). Participants highlighted that due **to funding cycles and research culture norms**, researchers often move on to new projects/organisations before they have deposited data from a previous piece of work. They also have **difficulty in finding time to create the supporting documentation** required. Other suggestions that could save them time and simplify the deposit process were: the ability to **link the** Data Stewardship Wizard (an online automated Data Management Plan creation tool) to avoid re-entering information, for colleagues to **work collaboratively when creating** Service Agreements (an agreement between the EIDC and the depositor which details information about the data and the services required), and **clearer information on the FAIR principles**, including how the EIDC's standards align. Participants also reported that they wanted a **named contact person** who could help them with the process.

Participants want practical, example-based guidance on how to deposit data

Participants reported a strong need for practical, example-based guidance on the deposit process including:

- **Video tutorials** on how to upload data.
- **Case studies** demonstrating how EIDC datasets can be combined for integration and analysis.
- Training on **writing supporting documentation for data deposits** with published good examples and access to templates to help structure documents and ensure they meet EIDC standards.
- Training and guidance targeted at **Early Career Researchers/PhD students** who may be unfamiliar with the process.
- Clearer **guidance on FAIR principles**, FAIR scoring of datasets, and how depositors can improve the FAIR score of their data, e.g. how to structure and format data to improve reusability.
- **Checklists** to help depositors confirm that their data meet submission requirements, offering them clarity and reassurance during the process.

When discussing the current data deposit process guidance, one participant said: "The forms are 'a bit scary' at first but easier on subsequent sessions, when you're familiar with them."

Participants want to know what guidance and tools are available to them

Participants felt they needed to be more aware of **what EIDC guidance was available**, as well as the existence of tools, such as Google dataset search. They also felt more guidance was required on what data falls within the **EIDC's remit**. For example, what is 'data of long-term value', how do they decide which NERC data centre to offer their data to and what data requires depositing?

They also reported that they need specific guidance on:

- **When** to deposit and how **embargo periods** work.
- How to handle **large datasets** (approaching or exceeding ~1 TB), including file size limits and constraints.
- What is their role and responsibility in completing **Data Management Plans (DMPs)**.
- How to handle data that is only **partially NERC-funded**.

What would make a difference?

Data held by the EIDC should be Findable, Accessible, Interoperable and Reusable (FAIR). However, a tension exists between individuals who want to reuse data (and therefore need it to be well described, well structured, using standards and controlled vocabularies), and depositors, who pressed for time, may not have the skills, time or support available to publish data of a sufficient quality to enable reuse. In this section we outline three key community needs and the areas of work requiring development to meet these needs.

Improve dataset search and discovery

An **AI-powered search** tool and **automated keyword extraction** to enable users to pinpoint the exact data they need for their purposes and to improve the richness of discovery metadata, would make it easier to search using more specific keywords. Expanding metadata standards, with richer, structured fields, and expanding the use of **controlled vocabularies** for keywords, would greater enhance discoverability of the data we hold. By providing **case-study examples**, where datasets have been integrated for analysis, users would gain insight into potential reuses for the data or link to other, similar datasets they may find useful for their needs.

Enable pre-download assessment of data

By introducing dataset **previews/visualisations**, users would be able to assess the suitability of the data before committing to downloading large numbers of files. This is of particular importance where data access comes with an associated charge which is based on the amount of data requested. Similarly, by allowing spatial/temporal **clipping** before download, users would be able to access the specific data they required, rather than having to download data they have no use for. Including **machine-readability indicators** (e.g. information displaying the use of standardised formats or controlled vocabularies) in discovery metadata would allow users would be able to discover if the data were appropriate for access by machines, making integration easier.

Streamline access to data

Ensuring barriers, e.g. mandatory log in or form filling required to access data, are removed will streamline the process, making sure data are available for immediate download. By enabling tailored, opt-in **dataset release notifications,** users will find out immediately when data they are interested in working with are available to access.

Strengthen support for enhancing data reuse

Providing more **example scripts** (Python, wget) for data access with clear guidance on how to use them, will support users in accessing and working with data in new ways, potentially increasing reusability of the data we hold. Similarly, offering advice on how to integrate datasets with different scales and units and increasing the visibility **of tools** such as APIs, and when/how to use them, would encourage reuse and demonstrate the potential of the data.

By producing clearer **FAIR-aligned guidance** and **scoring indicators**, depositors would be able to identify specific areas of their data and supporting documentation that they could improve on to make their data 'FAIR-er' and therefore of a higher reuse potential. FAIR scores would also help users looking for data as higher scores would indicate more easily accessible and reusable resources.

Overhaul the data deposit guidance, ensuring consistency of information provided

Ensuring **consistent, centralised guidance** for all Data Centre staff and delivering **regular training** updates would ensure users were receiving correct, up-to-date guidance irrespective of which member of staff they are communicating with.

Enhancing the **EIDC guidance and support offering** by making checklists, templates and exemplars publicly available, together with video recorded tutorials, would ensure users were able to access support in an accessible format and more clearly understand the standards required to deposit with the EIDC. **Clarifying rules** on, for example, what data is within NERC's remit, how to deal with datasets under partial NERC funding, and how to deposit large datasets or datasets with embargoes would support users in the initial stages of the deposit process, reducing time wasted and facilitate a smoother experience overall.

If the **Data Stewardship Wizard** were integrated into the deposit process, there would be a reduction in the amount of information users have to enter, as previously provided details could be accessed and reused, thus saving time and effort. Similarly, if depositors were able to collaborate with co-authors when completing the EIDC Service Agreement, to document details about the data, time and effort would be saved in email communication to and from co-authors.

Reduce research cultural and process barriers

By demonstrating, supporting and championing the benefits of good project planning, highlighting the importance of allocating sufficient **time and resources** within a project **for data publication** we will grow a culture where this practice becomes the 'norm'. Providing **tailored support** for Early Career Researchers and new users ensures the culture is bedded into research practice

Simplifying the deposit process by reducing repetitive back-and-forth communication and **simplifying forms,** where possible, would encourage users to deposit data as it would take less time and effort.

What did participants think about the event?

We asked participants to feedback via a Microsoft Form on how they felt about the EIDC Community Conversation event. Only two participants responded, however, they both gave the event 5/5 stars, stating that: they found the event useful, that we provided enough contextual information to allow them to form an opinion, that they were able to voice their opinions easily, that they would recommend the event to colleagues, and that they found the Community Conversation just the right length.

What's next?

The participants' feedback and ideas summarised above will directly guide future key developments that will make a real difference to users discovering or depositing datasets with the EIDC. Developments will be prioritised by ease of implementation and user value.

Easier to implement tasks include:

- Reviewing and improving the guidance and support provided by the EIDC at all stages of the data deposit process, e.g. delivering targeted training to users, continued development of EIDC staff and providing guidance in different formats to increase accessibility.
- Introduction of a FAIR metric for datasets so that depositors are clear on how FAIR their data is and how they can improve their score.
- Published good practice examples of supporting documentation that meet EIDC standards, and case studies illustrating where data have been integrated or fed into novel workflows for subsequent analysis.

Longer term developments include:

- Streamlining and automating the deposit process to save depositor's time and effort.
- Developing a more intelligent, AI supported search tool to enable users to more easily find the data they require.
- Producing AI-generated keywords or metadata extraction to improve discoverability.
- Implement an opt-in dataset release notification service to make users aware of new releases of data they are interested in working with.
- Create a user forum and investigate more diverse ways of obtaining feedback on our tools and services.

Many of the longer-term developments are already in train. For example, a prototype search tool leveraging AI techniques was demonstrated at a previous Community Conversation event and is undergoing further testing and development. We are also trialling a new tool that we have developed which uses Large Language

Models to extract keywords from supporting documentation that can be used to enhance discovery metadata, making data more findable. Work on streamlining and automating our deposit process has just begun (Jan 2026), and we are planning to build an EIDC user forum over the coming months.

We will keep the community updated on developments, including the release of new tools and services, via our mailing list. If you would like to suggest topics for future EIDC-focussed Community Conversation events or other engagement formats, please email us at info@eidc.ac.uk.

You can also provide **direct feedback** on our tools and services via the **feedback button** located to the right of our catalogue and website pages.

Please sign up to the EIDC mailing list via our website for the latest updates on EIDC news and events, as well as future opportunities to engage in data centre events.

Special thanks to the **organisers** (Hannah Dean, Kathryn Harrison, Sharon Hayton, Katherine Wright, Mary Preston and Sophie Moran), **speakers** (Hannah Dean and Phil Trembath) and the **facilitators and notetakers** (Sharon Hayton, David Leaver, Pauline Shepherd, Michael Tso, Jennie Roebuck, Sophie Moran, Dan Wright, Helen Rawsthorne, Matt Nichols, Katherine Wright, Philip Trembath, Hannah Dean, Matt Dalle Piagge and Els Dhiedt), and most of all the **participants** for making this a rich and engaging conversation, and the valuable contributions that will help drive and inform our plans for the EIDC.