UK Centre for Ecology & Hydrology
National Capability for UK Challenges
#CC04 – 3rd December 2025
10.00am – 12.00pm
40 participants

Community Conversation

Co-Designing Digital Infrastructure for Environmental Science

Report on #CC04: Enhancing Discoverability & Access to Environmental Exposure Data and Methods

Eunice Agyei (Lead author: eunagy@ceh.ac.uk), Jacky Chaplow, Michael Tso, Stephen Lofts, Michael Hollaway, Gemma Nash, Ezra Kitson, Carolynne Lord, David P. Green, Kelly Widdicks, and Gordon S. Blair. UK Centre for Ecology & Hydrology. 2026.

What did we do?

Our fourth in the series of online **Community Conversations** focused on Enhancing Discoverability and Access to Environmental Exposure Data and Methods and delivering integrated monitoring, modelling and data for the UK environment. We presented work to date on NC-UK's 'UK Environmental Exposure (UK-EEx) Hub' project, aiming to develop a centralised knowledge Hub for aquatic, atmospheric and terrestrial pollution data, and offer insights into the chemical state of the UK environment. The aim of the online workshop was to highlight datasets, tools and integration efforts in the project to guide next steps. The session was opened by Dr. Kelly Widdicks, the NC-UK Digital and Data Integration Co-Lead. Mrs. Jacky Chaplow, a contaminant data scientist and the UK-EEx project lead, chaired the session. There were five presentations, followed by an interactive **Q&A session**, and small group **breakout discussions**. We conducted a series of **polls** throughout on discovering and accessing exposure data and methods. Afterwards, we invited participants to complete an opinion **survey**.

Presentations:





Dr. Michael Tso	**Dr. Stephen Lofts**	**Dr. Michael Hollaway**	**Mrs. Gemma Nash**	**Dr. Ezra Kitson**
Data Scientist (UKCEH)	Soil and Aquatic Chemist (UKCEH)	Senior Data Scientist (UKCEH)	Web Developer (UKCEH)	Environmental Data Scientist (UKCEH)

Data and Model Integration		**Data Interaction & Exploration**		**Data Analysis**
Methods for data integration across compartments (air, water, soil, biota)	Model integration of WHAM-F_{TOX}, a model of the impacts of mixtures of acidity and metals on biota.	Methods for exploring and interacting with place-based chemical exposure (E.g. Exploration of chemical pollution	Examples of live data portals used in other areas of environmental science	Methods for Visualising Non-Target Analysis (NTA) Data

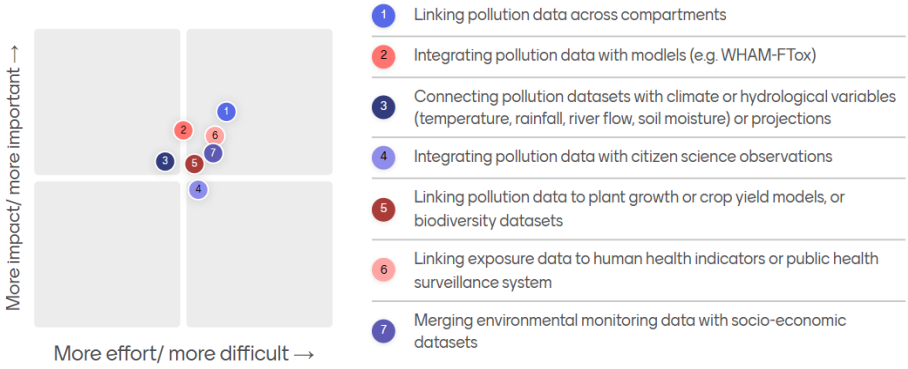
		from a place based perspective.)		

What did we find?

The event aimed to bring diverse stakeholders together in a collaborative setting to explore practical ways to enhance the discoverability and accessibility of environmental exposure data and methods across research domains. Our findings, presented below will inform the design and development of the UK-EEx Hub to strengthen our collective understanding of pollution exposure through an open, collaborative space for sharing knowledge, data, and methods.

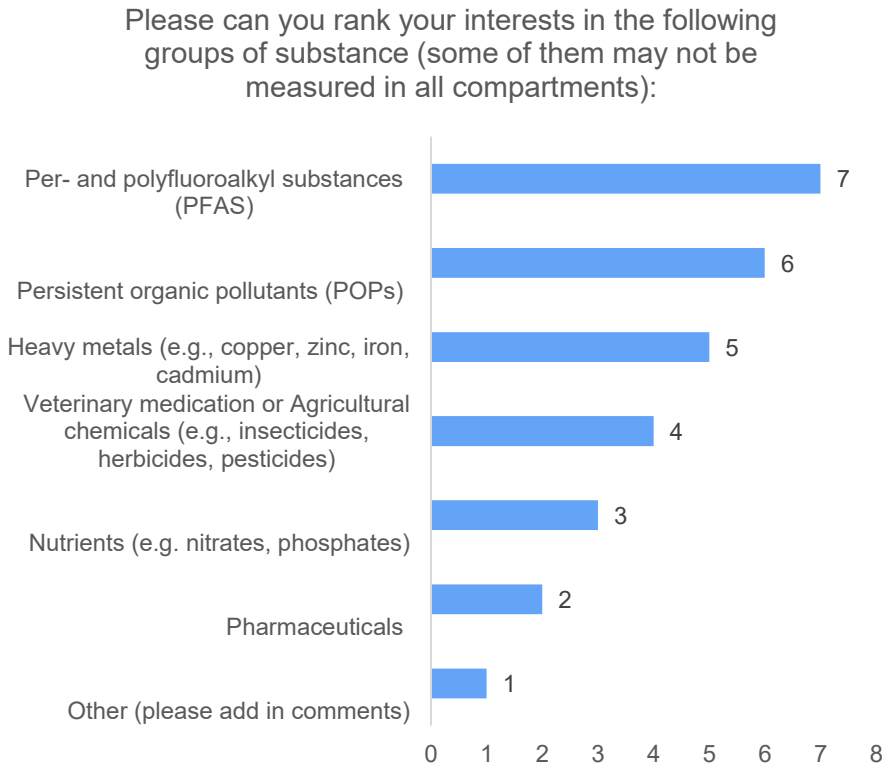
Integration activities are needed across compartments to build a coherent picture of exposure.

Integration activities involve harmonising data from different exposure compartments such as air, water, soil, food, and biota to ensure that datasets can be combined for robust exposure assessment. We shared examples of ways to integrate exposure data and asked participants to "**Rate each of these integration activities"** in terms of perceived effort in integrating, vs impact if integrated. Many integration activities were deemed as high-effort but high-impact, so success of the UK-EEx will depend on careful planning and strategic prioritisation to ensure the efficient use of resources to maximise long-term benefits.



High priority substances (e.g. PFAS and POPs) should guide data integration and modelling efforts across compartments.

To ensure the UK-EEx Hub delivers relevant pollution insights, we asked participants to prioritise a range of chemicals of current concern. This ranking exercise revealed high interest in Per- and Polyfluoroalkyl Substances (PFAS) and Persistent organic pollutants (POPs) and helped in the identification of priority substances for discovery, access, curation, integration and visualisation to ensure the UK-EEx Hub meets the needs of the user community.



Systems thinking ensures that integration accounts for the full complexity rather than treating datasets and other resources in isolation.

Participants were asked in breakout rooms **"What data or other resources (e.g., models and outputs) would be useful to connect together to understand exposure?** e.g., to examine air, soil and water contamination, or human and environmental health at the same time, or to connect environmental, social and medical science data." The responses collectively point to the inclusion of the following resources:

- Data about events such as wildfires together with earth observation data, utilising Artificial Intelligence (AI) techniques to predict future patterns and track historical patterns in land and waste management practices.
- Exposure data on animals including information about their prey, which is an important source of exposure.
- Biotic components (such as freshwater pollution) and abiotic components that could help connect parts of the food chain,
- Veterinary medicines which can include chemicals and thus contribute to exposure.
- Other known and relevant data sources e.g., NORMAN Ecotoxicology Database,
- Future scenarios e.g. Shared Socioeconomic Pathways (SSPs).
- Policy and regulatory updates such as dates for chemical bans, restrictions or new standards to help assess and understand changes in exposure trends.

Participants highlighted the need to understand relationships that may exist between data and other resources to ensure robust data integration and generate meaningful insights: *"linking everything to everything will be overwhelming"* and hence the need for *"systems*

thinking… approaches to examine relationships that might exist?". Also, they suggested was the need for opportunities to guide analysis and integration processes, using example questions such as: *"how are the poor affected vs the wealthy e.g. in relation to where they live?"* and *"is drinking water a danger to human health?".*

Data availability, accessibility, and quality matters for effective integration.

Across breakout discussions, three themes emerged from synthesis of recurring ideas in participant contributions: (1) data availability, (2) accessibility, and (3) quality reflecting user issues with discovering, obtaining and trusting data.

1. **Data availability:** Participants highlighted the need for an inventory of available data. It is often unclear what data an organisation holds, including government bodies, and/or the level of sharing permitted internally or externally. To support sharing where privacy is a concern, solutions could include changes in spatial resolution (where applicable) or anonymisation.
2. **Data accessibility:** Participants emphasised the need for accessible formats for data integration. In particular, it was suggested environmental data should be made available at a range of spatial resolutions to enable various explorations that each reveal a different pattern.
3. **Data quality:** Issues raised included the need for improved metadata to understand limitations in the data, including detailed sample processing, the rationale for monitoring frequency, Limits of Detection (LOD), measurement approaches, changes in instrumentation or techniques over time. To address issues with data quality, participants suggested the use of CREED (Criteria for Reporting and Evaluating Exposure Datasets), a framework developed to ensure that data has the minimum required information for reliable evaluation, and to establish a standard for ensuring metadata completeness and reliability. Also mentioned was the importance of a reliability threshold and the adoption of FAIR (Findable, Accessible, Interoperable, and Reusable) data principles.

For effective data integration activities, participants noted the need for "systems thinking [to examine …] relationships" and an "inventory of data" among other factors.

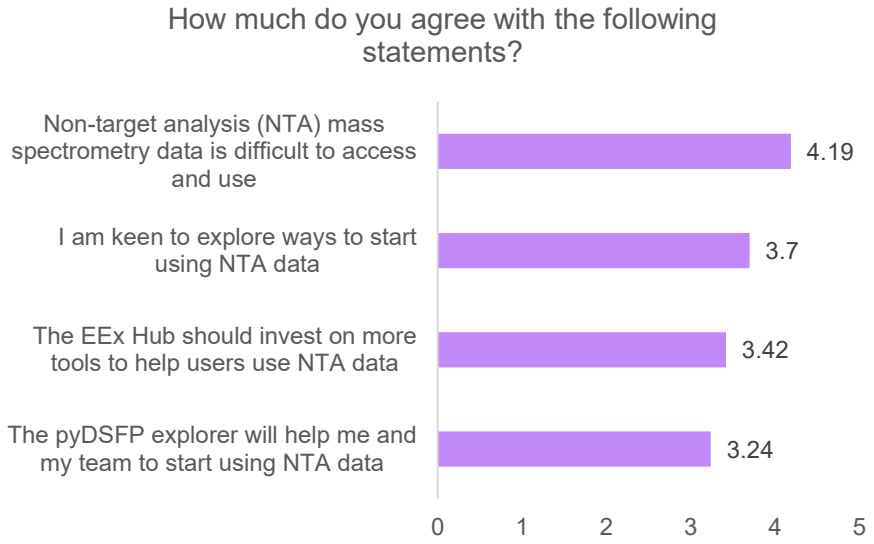
Providing a method to access NTA data from the UK-EEx Hub will enable exploration, integration and analysis of

Non-Target Analysis (NTA) is a powerful, discovery-based analytical approach used to identify unknown chemical, environmental, or biological compounds in complex samples without prior knowledge of their presence to identify known, unknown and emerging chemicals. NTA relies on Liquid Chromatography-High Resolution Mass Spectrometry (LC-HRMS) or Gas Chromatography-Mass Spectrometry (GC-MS) to generate large, information-rich datasets.

To enable exploration of this data, for example in integrated exposure assessments, we presented an open access method (named 'DSFP-

known and unknown chemicals.

PyExplorer', available in the Environmental Data Science Toolbox) to connect with the NORMAN Digital Sample Freezing Platform (DSFP). Participants were asked to rate statements on NTA data using a Likert scale. Respondents agreed that accessing and using NTA data is difficult and there was interest in exploring the published method. Participants indicated that tools like the DSFP-PyExplorer method will help users to start exploring NTA data, but additional tools and support may be needed.



Overcoming barriers to data integration opens opportunities to explore impactful questions and generate actionable insights.

Breakout groups were asked "**What environmental exposure and impact questions could you answer if resource integration (i.e. data, model outputs, models, scenarios, etc.) were no longer a limitation?**"

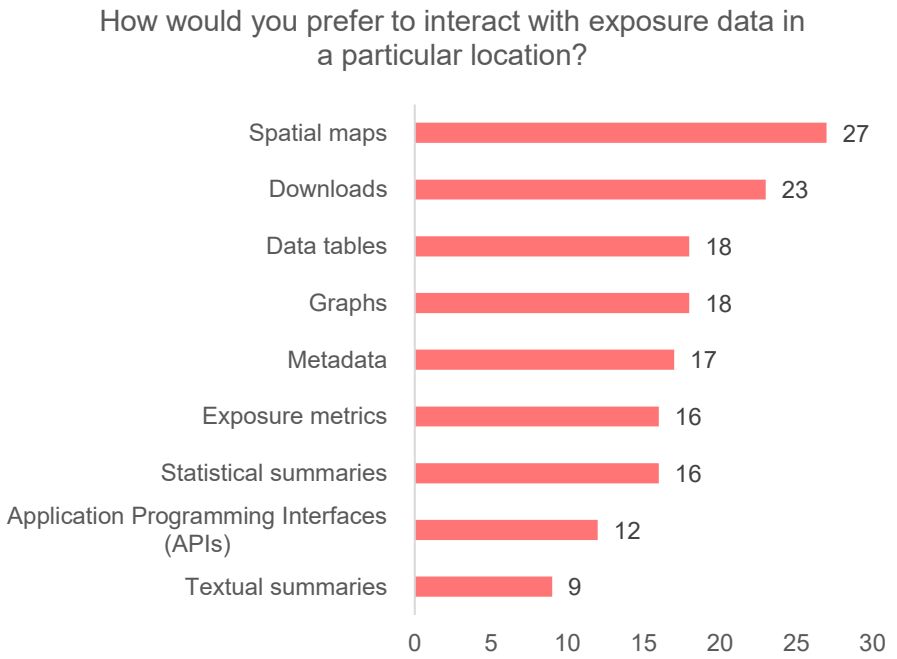
Participants shared several questions focussed on:

- The impact of historical chemical use, trends and environmental exposures to pollutants on human systems (e.g., public health, food systems, infrastructure), environmental components (e.g., air, water, soil, vegetation), diverse ecosystems (e.g., freshwater, terrestrial, coastal, urban), and broader effects including climate-related changes.
- Determining exposure baselines using multiple datasets across time to identify baseline conditions and meaningful deviations. This may help with situational awareness, recognition of emerging risks and appropriate response planning for exposure events.

To yield meaningful insights and therefore answer impactful questions, a participant stated the need to understand the "importance of variables to test hypothesis".

Spatial maps are the preferred way to interact with exposure data in a particular location.

Participants were asked "**How would you prefer to interact with exposure data in a particular location?".** The most popular response was via maps of spatial extent (n = 27; total = 30), then by download (indicating existing ability to independently interact with exposure data), then data presented in tables and as graphs. This was followed by metadata, exposure metrics, statistical summaries, Application Programming Interfaces (APIs), and textual summaries.



Further breakout discussions confirmed the need for maps that allowed for data interaction to overlay data to identify spatial hotspots, use a slider to change transparency drawing on best practices from e.g., the Defra MAGIC website, draw an area on the map to investigate and download data for a selected area, upload a point or shapefile to extract data for a specific point/location of interest, and use derived data for gap filling (noting uncertainty) e.g., for essential biodiversity or climate variables.

Participants shared their favourite websites for exploring spatial exposure data: Copernicus, Naiades; EU IPCHEM, Environment Agency Water Quality Sampling Sites (Open WIMS) data, Environmental Information Data Centre EIDC, UKradon, UK Air Pollution Information System (APIS), Forever pollution map, and the OHAT (OSPAR Hazardous Substances Assessment Tool) online platform hosted by the International Council for the Exploration of the Sea (ICES).

Given the wide range of interactive features identified for possible inclusion in the UK-EEx Hub, a participant reminded us that user preference will be "context dependent" and based on a specific outcome.

What would make a difference?

Enabling easier **discovery and access to high quality** environmental exposure **data for a** diverse user base including environmental scientists, policy makers, and the general public **is essential.**

Drawing on insights from the polls, Q&A, and breakout discussions, this section outlines four key community needs and priority areas of work to strengthen the UK-EEx Hub and its contents. These will help reduce current barriers and support better discovery, access, and use of exposure methods.

Enable easy discovery and access to key data and resources on chemical pollutants.

The UK-EEx Hub should enable discovery of relevant, high-priority data, resources, and methods for chemical pollutants, and identification of chemicals most studied, most concerning, or most uncertain with clear descriptions, rich metadata, and intuitive search and filtering tools. This will reduce effort and enable delivery of more impactful outputs and improve the quality of exposure analysis. It will assist users at different skill levels to navigate complex environmental datasets that can be challenging to access e.g. NTA 'big' data, or data collected at differing spatial and temporal scales.

Provide seamless access to complex and emerging data types.

The UK-EEx Hub should aim to simplify access to challenging data types, such as NTA mass spectrometry outputs. Clear guidance, standard data processing workflows, and user-friendly tools should be provided to enable users explore and analyse data that would otherwise be difficult to find or interpret.

Enable users to develop comprehensive understanding of environmental exposure.

The UK-EEx Hub should enable integration of chemical data across compartments from air, water, soil, and biota, alongside other datasets and resources such as the examples provided by participants above. Integrating these resources would empower users to ask bigger questions, compare patterns across domains, and gain insights that are not possible when datasets are isolated. This would give users a comprehensive understanding of environmental exposure.

Ensure users can know and assess data quality and reliability.

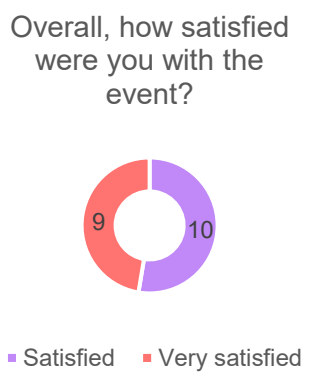
The UK-EEx Hub should provide tools and metadata that allow users to evaluate the quality, provenance, and reliability of data. Adoption of FAIR and CREED principles, with robust data anonymisation, may strengthen data usability by enabling users to make informed decisions and ensure any conclusions drawn from integrated data are trustworthy, and facilitate reproducible analyses. The Hub could also provide quality assurance and control measures to prevent and correct errors respectively.

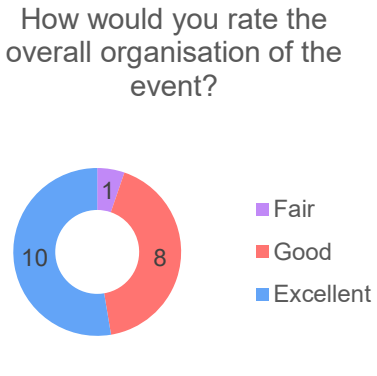
Support easy exploration and visualisation of exposure data.

The UK-EEx Hub should provide intuitive, interactive tools such as maps to allow users to explore and visualise pollution and exposure data across compartments (air, water, soil, biota), chemicals, locations, and time. Flexible interfaces and clear visualisations would help users investigate patterns, test ideas, and communicate findings effectively.

What did participants think about the event?

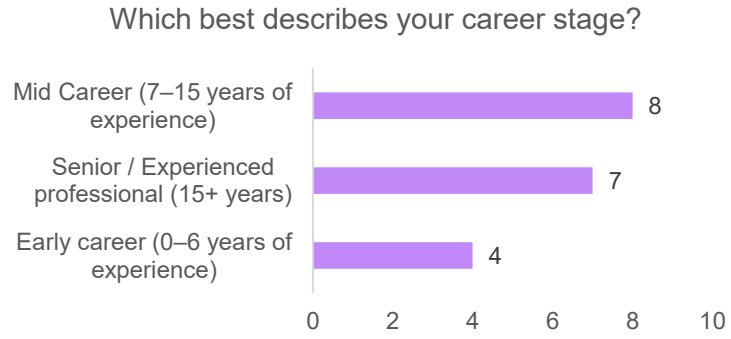
We invited participants to share their feedback on their experience of the event, along with their demographic information. Almost half (n = 19; total = 40) of the participants responded. These participants were either satisfied or very satisfied with the event and rated the event as excellent, good or fair.



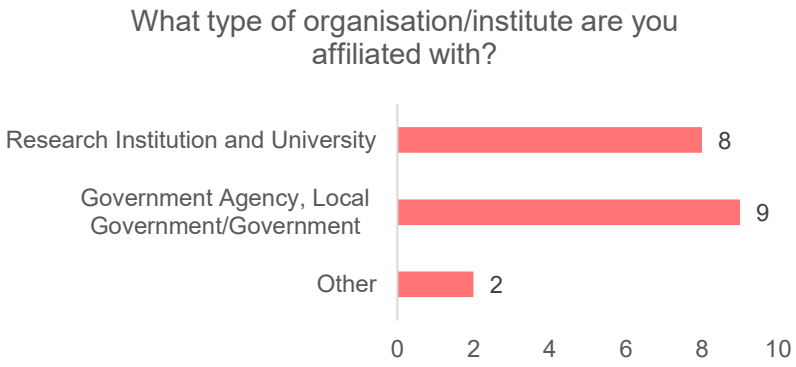


Who participated in event?

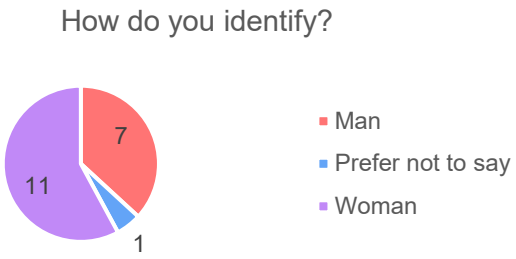
Participants represented a wide range of career stages, with mid-career professionals forming the largest group (n = 8; total = 19), followed by senior experienced professionals (n = 7; total = 19), and early-career researchers (n = 4; total = 10).



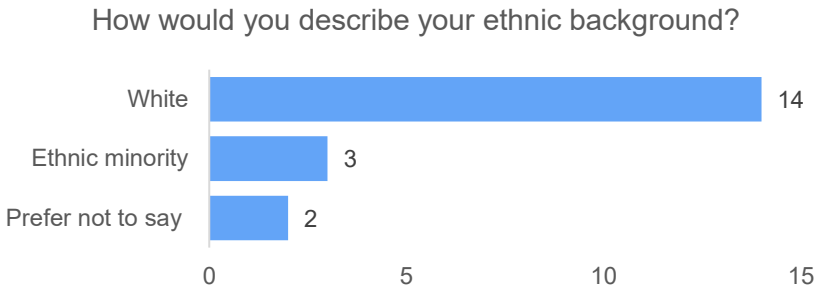
These participants came from mainly a mix of research institutions and universities, and government bodies.



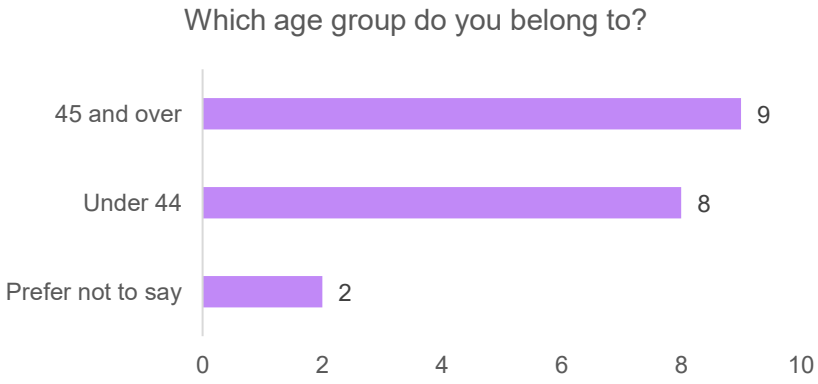
Gender representation was slightly skewed towards those identifying as women (n = 11; total = 19), with others identifying as men or choosing not to disclose.



Participants came from a range of ethnic backgrounds which were combined into two broad categories to protect anonymity along with participants who did not disclose their ethnic group.



A broad range of ages were represented at the event. To ensure anonymity, age ranges were combined into aggregated categories (under 44; 45 and over), with non-disclosed ages shown separately.



What's next?

To ensure the UK-EEx Hub becomes an established community resource, it will unify exposure knowledge to accelerate research and policy impact. Our next steps include:

- disseminating this report that outlines the challenges, opportunities and needs for the UK-EEx Hub.
- further developing data integration and visualisation methodologies for a suite of exposure resources, focusing on the high priority chemicals identified by participants.
- examining metadata needs to enable thematic links and integration of resources across compartments (air, biota, water, soil) in the UK-EEx Hub (and other portals) and enables a change in data archiving.
- developing user interfaces that surface these metadata connections and allow for users to explore related exposure resources, including building on the examples we shared in the Community Conversation and examining user journeys on how different users utilise the UK-EEx Hub.
- finding new collaborative opportunities to build on the UK-EEx Hub developments thus far with the environmental science community.

Please sign up to NC-UK mailing list via NC-UK Website for the latest updates on UK-EEx Hub, as well as for wider updates on the NC-UK programme and opportunities to engage (e.g., through future Community Conversations).

Special thanks to the **participants** of this workshop for making this a rich and engaging conversation, and their valuable contributions to in shaping the UK-EEx Hub. Thanks also to the **organisers, speakers, facilitators, and notetakers** (Eunice Agyei, Mary Preston, Jacky Chaplow, Michael Tso, Stephen Lofts, Michael Hollaway, Gemma Nash, Ezra Kitson, and Kelly Widdicks).

Appendix Responses to Questions & Answers

1. How do you plan to show uncertainty, variance, assumptions, consistency in instrumentation etc. in each of these tools? These parameters are very important when bringing together such disparate datasets.

Fundamental to connecting and integrating data through the UK-EEx Hub is ensuring users understand the data sources through quality metadata. To avoid duplication and misinterpretation of data, we aim to connect to datasets where they are currently stored together with their metadata. A lot of exposure information comes from analytical chemistry, for example, and it is important users can access metadata on how the data was generated – preserving the provenance of data as it is integrated.

For example, three methods showcased in the Community Conversation documented in the Environmental Data Science Toolbox include links to datasets at their source. We are also exploring other ways to discover data, methods and models across repositories (including via AI-enhanced searches), making connections to the source data they utilise.

	Regarding the uncertainties, variance, etc., we will look into that more through our development of the UK-EEx Hub user interfaces and future engagement with users. Sometimes the challenge of using one map is that it is very difficult to overlay the uncertainty on top. We are looking for ways to provide that information so that users can enable or disable uncertainty information behind the integrated estimation.
2. I'd love to link some of my biodiversity data to pollution data, but I have no idea about the persistence of the pollutants to make good decisions about the temporal matching needed between the pollution data and my biodiversity data. Will you be able to offer any guidance on this?	Our Community Conversation discussions highlighted the need for a more joined up, systems approach – moving away from working in silos. We have had discussions prior to the workshop about how to bring biodiversity data into the UK-EEx Hub, but this is another broader discussion that is needed to ensure that what we are connecting makes sense, requiring an understanding of other science domains such as biodiversity. The integration examples shared focused on the spatial aspects, though we recognise this is trickier when temporal aspects are included too – especially as substances such as PFAS are especially persistent. Considering this is complex but a valuable discussion for the UK-EEx Hub. We call for, and encourage, contributions to these discussions, aligning with our next steps of finding new collaborative opportunities to build on the UK-EEx Hub developments thus far.
3. I was thinking along the same lines, it would be great if the portal could be expanded to also include demographic and socio-economic data?	We believe that an important goal for the UK-EEx Hub is integrating and joining up different types of data. We gathered lots of different data sources from participants in the Community Conversation (including suggestions such as data from the Clean Air Programme). If available, including demographic and socioeconomic data could reveal unequal exposure patterns, identify vulnerable groups, and provide evidence for effective environmental decisions. We welcome suggestions for additional data that the UK-EEx Hub should connect to and are keen to explore opportunities to develop them together.
4. I would like to know whether and how importantly data including values below limit of detections impact analyses/ modelling.	Where available, we will aim to include data values below limits of detection. For modelling and analysis, techniques such as censored data analysis will be used and clearly documented.
5. In relation to [Model integration of WHAM-FTOX, a	We will be exploring creating various workflow and output access options for users. If you have suggestions or use cases, please contact us.

model of the impacts of mixtures of acidity and metals on biota ...] specifically the workflow; I think a key stage is the user options stage; would this be the point where data quality could be selected?	
6. How much coding skill is required to utilise the DSFP-PyExplorer] tool?	The DSFP- PyExplorer tool is available in the Environmental Data Science Toolbox and it is designed with narrative and visualisations alongside the code to help users understand what the code is doing. However, users would require some level of programming experience (e.g., installing Python, running the tool via command line) to run or edit the tool's code. We hope to offer a variety of ways in which users can engage with data, methods and models connected through the UK-EEx Hub – e.g., through interactive maps for those who do not have coding expertise, as well as methods like the DSFP-PyExplorer for those who wish to use or build their data science expertise. For the latter, there is an opportunity to further enhance information in the Toolbox for how to install Python to help beginners, and train users to use such methods. We had a Community Conversation on the Toolbox in July 2025, and key discussions and next steps can be found in this report....
7. Where can I access this dataset for place-based exposure?	The three data science methods: integrating chemical substance data across compartment, place-based analysis of multiple chemicals (currently developed for air quality and veterinary medications examples over UK regions), and non-target analysis of environmental mass spectrometry data (i.e., the DSFP-PyExplorer). These are all published and available in the Environmental Data Science Toolbox.
8. Will the integrated metals data include coastal areas (e.g. concentrations in water and sediment?	Yes, it could be extended to include coastal water and sediment data. Please contact us to suggest relevant data sources.
9. Will the [air quality] AQ data include that collected by local authorities	Defra have extensive holdings of UK air quality data including data from both the Local Air Quality Management (LAQM) and Automatic Urban and Rural Network (AURN) which are available through the UK-AIR Data Selector and through an API using the openair R

	through LAQM? It (generally) has a more extensive network than the AURN. I assume Defra hold all this somewhere as it's submitted for the Annual Status Report process	package. These data are freely available under an open government licence so therefore could be included if useful and relevant for place-based analysis. However, the UK-EEx Hub would only be able to carry a snapshot of the data at a given time and would require periodic updating to ensure the latest data was being made available. Future iterations of the UK-EEx Hub could include running a method or script periodically to ensure this data is up to date – balancing this with the need for provenance.
10.	How do we as a community ensure we support time for FAIR? How do you deal with data ownership and knowing who to acknowledge for any data that is acquired through the tool?	Thorough data licensing is key here for acknowledging the teams involved in collecting and analysing data. As the UK-EEx Hub aims to connect to data at its source, this will serve as another point of data entry rather than involving transfer of data ownership of or licence changes. There are also ways to provide the licence and attribution information with the data, so users know provenance of the data itself. For example, this can either be through the methods in the toolbox or supplied as metadata packaged with the data. Regarding time for FAIR, this is similar to a question we received for the Toolbox Community Conversation. Change is required at the community, project and funding levels to ensure there is appropriate time and incentives to support FAIR. Technologies and guidance can also support making FAIR compliance for data, methods, models and other digital objects quicker and easier. Please see Appendix question 9 on page 9 in our Toolbox Community Conversation report for our detailed response to this.
11.	Will the information needed to perform CREED assessments be considered for inclusion - in particular for reliability.	If there is enough demand by the community for CREED assessments, then yes, the UK-EEx Hub could look to including this. CREED is an approach for assessing the reliability and relevance of environmental exposure datasets. The issue around reliability is largely metadata on the analytical method used, quality control approaches, detection limits for chemical concentrations among other things.