

Digital Research Infrastructure (DRI) at the National Environmental Monitoring Conference (NEMC) 2025

Dr Kelly Widdicks (Lead for Digital Innovation) and Prof Gordon Blair (Head of Environmental Digital Strategy) at the UK Centre for Ecology & Hydrology (UKCEH)

Environmental monitoring is becoming increasingly reliant on digital technologies from the point of data capture to its ingestion – enabling more automated and faster monitoring – but also for ‘smarter’ data capture and analysis e.g., through data science and AI techniques that help inform future monitoring. Innovations are also being made to enhance how environmental monitoring data is found, used, integrated, analysed and shared. These technologies and the digital objects they support (i.e., data, methods, models etc.) are all part of growing ‘Digital Research Infrastructure’ (DRI) in environmental science, enabling a necessary shift for more integrative scientific insights.

In the ‘Vision to Value’ session theme at NEMC 2025, we presented on-going DRI research and innovation at UKCEH for environmental monitoring – covering insights from key National Capability programmes including [NC-UK \(National Capability for UK Challenges\)](#), [FDRI \(Floods and Droughts Research Infrastructure\)](#) and NC-Data (funding the [Environmental Information Data Centre \(EIDC\)](#) as part of the [NERC Environmental Data Service](#)). These were formed under three topics: 1) Streaming Environmental Monitoring Data; 2) Data Science and Artificial Intelligence (AI) Techniques for Environmental Monitoring; and 3) Environmental Monitoring Data Discoverability and Accessibility. MentiMeter was used through the presentation to gather insights and input from NEMC attendees around the three topics.

1) Streaming Environmental Monitoring Data

Recognising that some monitoring activities can be automated – and given the varying stages in which data is shared (e.g., raw and published formats) – there is an opportunity for scalable DRI to gather and deal with diverse and near real-time data. We shared the automated ingestion process for [streaming time series data from FDRI monitoring infrastructure](#), visualised through a generic ‘Time Series Explorer’ that can be adopted for other time series data beyond floods and droughts. From this, we asked NEMC attendees: ‘what challenges or barriers do you foresee in tools that stream environmental monitoring data?’ where we received 69 responses.

A key challenge foreseen by attendees surrounded the data itself – covering topics such as **data quality, data standardisation, monitoring coverage and volume**. These raise important requirements such as the need for rapid and potentially automated quality control approaches, metadata completeness and standards for all data types, better connectivity for capturing data automatically nationally, and the provision of necessary storage and compute for the volume of such data.

Another theme emerged around how monitoring data is captured and shared. These included issues around the **costs** of such systems, the **maintenance** of these once funded projects end, the ability to respond to changes and **innovations** in digital technology over time, and the lack of joined-up consensus and **integration** for DRI in monitoring currently. There were also many responses that shared concerns about the **adoption** of such technologies, especially given the **varying skillsets** and literacies in digital technology, and emphasised the importance of ensuring systems are easy to use through intuitive and accessible user interfaces.

Attendees also considered challenges regarding the wider **responsible innovation and use** of such systems, covering issues such as the security and privacy of data and systems, the environmental impacts of digital technologies, the ethics of working with different technology organisations, and communicating uncertainties.

2) Data Science and AI Techniques for Environmental Monitoring

In this second section of the presentation, we shared how data science methods and AI techniques are currently distributed across environmental science, with varying method types from cleaning monitoring data to analysing it. Recognising that DRI can support the sharing of methods, we presented the [‘Environmental Data Science Toolbox’ being developed under NC-UK](#). This aims to provide a repository for sharing data science methods and AI techniques from across environmental science, ensuring these are discoverable across domains of science and interweaved with detailed narrative and examples for method uptake and usability. Following our short demonstration of the toolbox, we asked NEMC attendees: ‘what types of methods may you need to support environmental monitoring or analysis of monitoring data?’, receiving 39 responses.

The types of methods shared varied extensively, covering techniques for **different data media** such as imagery (e.g., ‘Computer vision methods’, ‘Image labelling methods’, ‘Automatic ID of images/sounds/species’), **bias correction** methods, methods to **detect anomalies or outliers**, **trends analyses** (e.g., ‘Trends over various timescales and reference periods or habitats’), **modelling** (e.g., ‘Geospatial modelling’), and other **specific types of analysis** (e.g., ‘Timeseries analysis’, ‘EO [Earth Observation] spectral signature analysis’, ‘Fourier analysis’).

Attendees also shared that they may need methods for more **basic approaches** (e.g., ‘We should teach learn basic methods not be reliant on software/AI’), including **data preparation techniques** covering standardised approaches to cleaning, quality control, and responding data errors or monitoring gaps (e.g. ‘Methods to deal with gaps in time series’). They also shared the importance of data **integration methods** (e.g., ‘Multivariate data integration’, ‘Integration methods for combining new tech and traditional approaches’), as well as **training and funding** for methods in environmental monitoring.

3) Environmental Monitoring Data Discoverability and Accessibility

The final topic of DRI explored in the presentation covered the discoverability and accessibility of data through searching and exploring data. For searching, we discussed how current searches rely on metadata, but that AI can improve data discoverability through utilising ‘Large Language Models (LLMs)’. We showcased a new AI-enhanced search tool being developed under NC-UK and NC-Data for the EIDC, which offers more relevant results to standard search approaches, enables users to ask questions of data rather than typing specific search terms, and generates summaries to help answer user queries more intuitively. This tool is currently in development but will be shared externally in 2026.

For exploring, we emphasised that monitoring data currently comes in various formats across different repositories, and that more interactive viewers could be developed to support data accessibility regardless of users’ skillsets. We shared the [‘Spatial Explorer’ being developed under NC-UK and NC-Data](#); this is currently available in its Beta version for the [Land Cover Map](#)

[2024 dataset](#) but will be expanded to support place-based data search and interaction across other monitoring schemes and datasets.

Following these demonstrations, we asked NEMC attendees: ‘if environmental monitoring data were to be more discoverable and accessible, what would this look like to you?’ and received 28 responses. These covered **mapping platforms** and **searching** as aligned with our showcased tools but also emphasised the **role of catalogues** and **the centralisation and integration of repositories** that connect systems and tools. Attendees also shared that discoverability and accessibility to them means **free, open and licenced data**; interactive, easy, uniform and **improved user interfaces**, including **narratives** that help users understand the data; Application Programming Interfaces (**APIs**); and access to **data at all stages** from raw data to analysed insights.

Call for Collaboration

We presented the importance of a joined-up approach to DRI for environmental science – ensuring we avoid duplication of effort across the community and draw on the skillsets and techniques of each other. We therefore used the presentation as a call for collaboration to 1) help shape the on-going DRI developments shared in the presentation and funded under the National Capability programmes above; and 2) work together to integrate the wealth of digital systems used in environmental monitoring and by the NEMC community.

The MentiMeter feedback above will feed into our DRI developments and National Capability programmes, so we thank all those at NEMC 2025 for participating in this activity. We also would like to thank the organisers of NEMC for giving us the opportunity to present and for all their help throughout the conference.

With our digital-focused talk at NEMC 2025 (slides 197-207 in the [NEMC 2025 Conference Slides](#)), we hope that this will spark more discussion about DRI and the sharing of digital and data integration challenges and opportunities at the conference moving forward.

If you would like to collaborate or are interested in any of the content shared in this summary, please do reach out to us at NC-UK@ceh.ac.uk.

Question and Answer Responses

1. *Whose responsibility is it to fund environmental monitoring data storage? Particularly for big data, e.g., acoustic data.*

Thank you to the attendee who asked this question, and we recognise there is no easy answer to this. The [NERC Environmental Data Service](#) as an existing data storage infrastructure and service is important here, ensuring they have the resources necessary to store data across NERC that is of long-term value to the community. However, this is tricky when it comes to raw streaming data and the vast amount of data collected in acoustics or other more data-intensive monitoring mechanisms. This will likely will come down to three key requirements and considerations: 1) ensuring the project, monitoring programme or organisation costs in the funding required to support the data collection for its lifetime *prior* to the activity being agreed, recognising that this is difficult given the changing costs of storage; 2) embedding ‘sufficiency’ into science i.e., only collecting and storing what is needed, and thus in turn

requiring data science methods and social understandings about what is required; and 3) utilising cheaper 'offline' storage for data which isn't as in demand now, but may be required in the future (e.g., for new scientific insights and developments) – ensuring that metadata online is shared to make others aware of the existence of this data offline.



Dr Kelly Widdicks and Prof Gordon Blair from UKCEH presenting on DRI at NEMC 2025.