



FDRI Digital Workshop

Enabling reproducibility in hydrological research

Ali Rudd (UKCEH), Matt Fry (UKCEH), Gemma Coxon (University of Bristol), Kit Macleod (UKCEH)

Ref: WP2 WS1

Version 1
21.08.2024

Contents

1.	Executive summary.....	3
1.1	What was the aim of the workshop?.....	3
1.2	What did we do?.....	3
1.3	What will we do next?.....	3
2.	Workshop objectives and scope.....	5
3.	Our approach.....	5
4.	Workshop outcomes.....	6
4.1	Workshop discussion themes.....	9
	FAIR – Findable.....	9
	FAIR – Accessible.....	10
	FAIR – Interoperable.....	10
	FAIR – Reusable.....	10
5.	Next steps.....	12

1. Executive summary

In this report, we summarise the key findings and outcomes from the digital workshop on 'Enabling reproducibility in hydrological research' held in January 2024 as part of the UK Floods and Droughts Research Infrastructure (FDRI).

1.1 What was the aim of the workshop?

The FDRI digital infrastructure will bring a step-change in the UK's capacity for data discovery, data access, and data integration. This will enable more efficient analysis and exciting scientific exploration of a range of environmental data related to understanding and responding to hydrological extremes in the UK. The aim of the workshop was to, as a hydrological research community, think about how FDRI digital infrastructures could help transform our hydrological research to provide new and reproducible process understanding. To achieve this, we framed the workshops around the concept of reproducibility. Reproducibility is a critical component of scientific analyses but most hydrological science is not reproducible.

1.2 What did we do?

We organised a webinar with an international guest speaker, Dr Irene Garousi-Nejad from CUAHSI (The Consortium of Universities for the Advancement of Hydrologic Science), to hear from a long established (since 2001) world leading project on how digital research infrastructures can support data management and reproducible hydrological research. We then hosted two workshops with 47 attendees to discuss how we enable research involving hydrological data management, analysis and modelling to become more reproducible, and to deliver the FDRI objective of improved understanding of hydrological extremes in the UK.

1.3 What will we do next?

The community supports making research more reproducible, and sharing knowledge and skills. The workshop provided key insights into how FDRI can help facilitate this.

The FDRI team will:

1. In the short term (by the end of 2024) continue to grow the FDRI digital community though providing regular updates and sharing data, code, and related workflows openly online - so that people can benefit from shared learning. This will be delivered through a central online FDRI website and engagement with the wider UK research community at events and through focused engagements (Themes 1, 8, 9, and 10, below).
2. During the second half of 2024 we will work with the UK research community to soft launch an FDRI dataset showcase to increase the visibility of datasets and associated metadata across the hydrological community (Themes 1 and 9).
3. During 2025 we will provide near real-time open datasets on a cloud data platform and examples of their use. These will be accompanied by guidance on metadata and licensing information (Themes 1, 2, and 4).
4. During 2025 we will provide an example of a web-based collaborative platform that will support the analysis of FDRI and other data using research cloud computing, and the development and sharing of hydrological analysis toolkits (Themes 3, 5, and 8).
5. By the end of 2025 the FDRI digital team will have worked with the FDRI field team and wider community to review current field monitoring standards and data standards for monitoring data, and provide guidance to help increase the use of uncertainty data within hydrological research (Themes 6 and 7).
6. By the end of 2026 the FDRI digital team, will have worked with the FDRI field team and wider community to co-develop and share online data processing workflows to enable the standardisation of quality assured data and related code (Themes 3 and 5).

2. Workshop objectives and scope

The UK Floods and Droughts Research Infrastructure (FDRI) project is a capital investment by UK Research and Innovation (UKRI) enabling essential science and innovation to improve the country's resilience to hydrological extremes. The FDRI digital infrastructure will bring a step-change in data discovery, data access and data integration, enabling more efficient analysis and exciting scientific exploration of a range of environmental data.

The aim of the workshop was to, as a hydrological research community, think about how digital infrastructures could transform our research. We framed the workshops around the concept of reproducibility. Reproducibility is an essential requirement for iterative scientific research. There are many limiting factors in hydrological science which limit the reproducibility of most hydrological research, including the lack of openly published data, code and models. These limitations on reproducibility hinder the understanding, verification and further development of research outputs (Hutton et al, 2016¹).

In FDRI we want to support and adopt open and reproducible science, both in terms of technical solutions (i.e. linking open DOI'd code to open DOI'd datasets and platforms to run / re-run / adjust / republish analyses; where DOI stands for Digital Object Identifier) and capacity building to enable researchers to adopt best practices in this area. To ensure we maximise benefits for the community, we organised the workshop so we could gain input from the research community on where we should focus and prioritise efforts on digital infrastructure to meet their needs.

3. Our approach

To explore reproducibility in hydrological science, we first hosted an FDRI webinar to learn from a leading international initiative that, for over 20 years, has facilitated the interdisciplinary advancement of water science by making it easier for the water science community to work collaboratively and openly. Dr Irene Garousi-Nejad from CUAHSI (The Consortium of Universities for the Advancement of Hydrologic Science) presented

¹ Hutton, C., Wagener, T., Freer, J., Han, D., Duffy, C., & Arheimer, B. (2016). Most computational hydrology is not reproducible, so is it really science? *Water Resources Research*, 52(10), 7548–7555.

<https://doi.org/10.1002/2016WR019285>

on [How can digital research infrastructures support data management and reproducible research?](#)

Two collaborative workshops were then hosted and attended by hydrological scientists from universities, regulators, research institutes and industry to ensure that different perspectives could be heard. The workshops included introductory presentations on the plans for the [FDRI digital infrastructure](#) (Matt Fry), [reproducible science in hydrology](#) (Dr Francesca Pianosi), [how Digital Research Infrastructures can support FAIR \(findable, accessible, interoperable and reusable\) research](#) (Dr Jennie Roebuck) and [shared experiences and challenges on making research FAIR](#) from an early career researcher perspective (Saskia Salwey).

The presentations were followed by facilitated group breakout sessions where we sought feedback on *'How do we enable research involving hydrological data management, analysis and modelling to become more reproducible?'* Reproducibility in hydrology was considered using three scenarios of hydrologists trying to make their field monitoring, data analyses and modelling reproducible:

- Xavier is a researcher undertaking hydrological field monitoring and wants to make sure their data can be published to allow researchers to include it within further analysis and supports near real time understanding of hydrological conditions.
- Yvonne wants to analyse how flood metrics have changed over the last 20 years across the UK, potentially using 15-minute flow and rainfall data, and feels the metrics and methods might be useful to others.
- Zoe is involved in a large NERC project producing an ensemble of future hydrological drought projections to analyse changes in UK water supply restrictions and feels the methods and model outputs might be useful to others.

These discussions highlighted the challenges, benefits and digital needs to make hydrological data collection and research more reproducible.

4. Workshop outcomes

In total, 50 people attended the webinar and 47 people participated in the workshops. We had a diverse spread of attendees from across universities, regulators, research institutes and industry.

A Mural board was used to capture the workshop discussions, some of the discussion points are shown in the figures below.

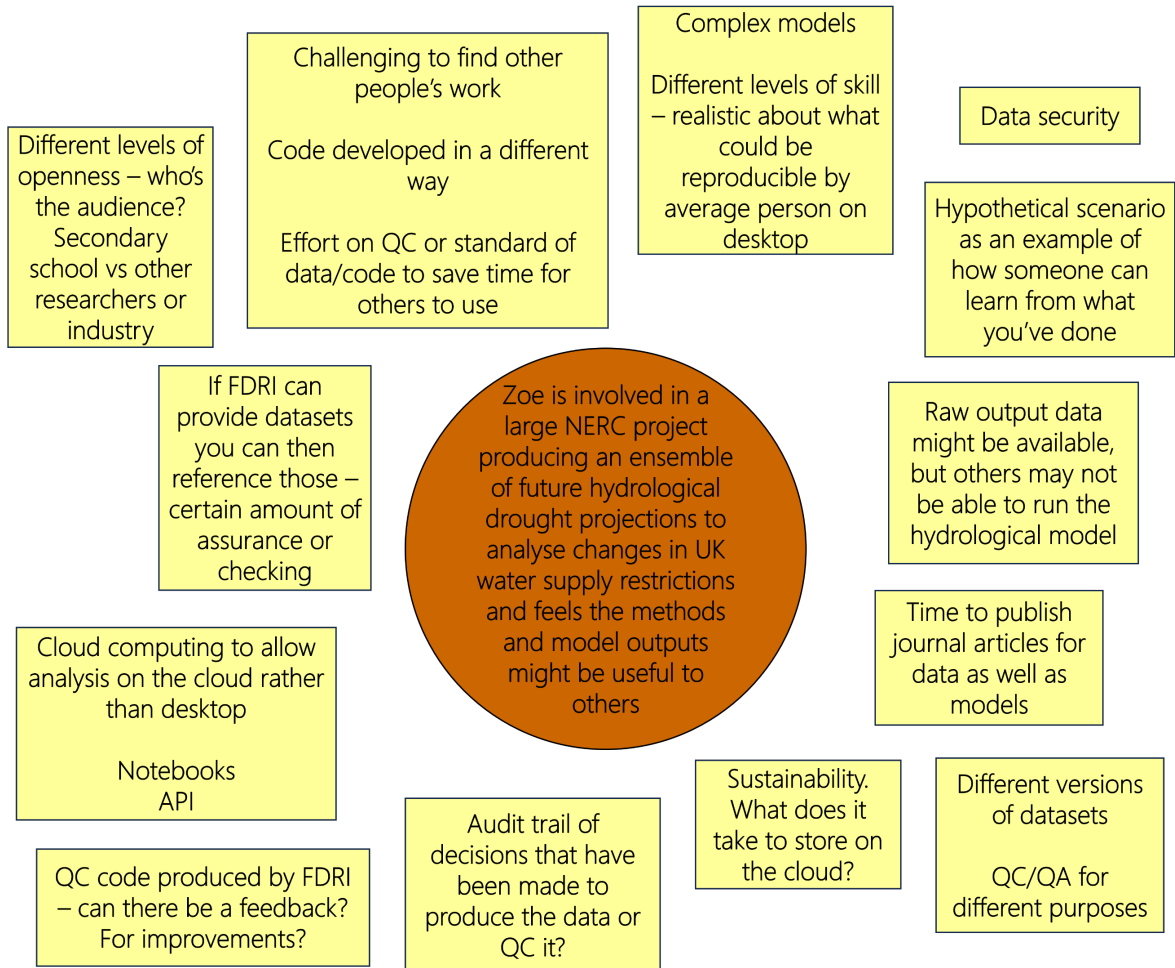
Figure 1: Example workshop notes from the breakouts
(indicative image of workshop Mural board)



Things to consider (computing resources, tools, policies and processes) required for:

- Referencing data sources?
- QA / QC and general audit trail / provenance
- Running analytical code / models?
- Sharing code / data during a project?
- How would data outputs be published?
- How would code outputs be published?
- Route to journal publication
- How would you share your outputs with academics/industry?
- How would the outputs be reused in future?

Figure 2: Example workshop notes from the breakouts
(indicative image of workshop Mural board)



During the workshops there were discussions around access and use of FDRI and 3rd party monitoring data, including data quality, licensing, and metadata concerns and opportunities. We also discussed analysis and modelling, and publishing of data outputs and code.

4.1 Workshop discussion themes

In general, the community supported making research more reproducible, and to share knowledge and skills. The workshop provided key insights into what the community might need to be able to achieve that. The key themes that came out of the workshop are summarised below according to the FAIR principles along with a short description of how FDRI plans to respond to each of the themes.

FAIR – Findable

1. Communicating what FDRI is doing – learning from other initiatives was discussed in the breakouts. It needs to be clear what the FDRI platform will offer and then ensure that the message is shared as widely as possible.

FDRI will continue to learn from international projects and initiatives. We will increase our online and face to face presence to enable engagement and uptake by the UK research community.

2. Guidance on metadata – the community would like FDRI to provide guidance on metadata, how to write it, what to include. For example, consider storing additional information when instruments were put in the field (e.g. photos, spot gauging). This could include providing guidance to enable people to produce or upload data and code that can be taken up by others.

FDRI will provide guidance on metadata and will work with the community to make sure it is fit for purpose.

3. Cloud data availability and processing – the community would like to know which data has/has not been published. There's an appetite for information on any data processing that has been applied to the datasets, sharing the processing steps with the community is needed for open science and it is hoped this will lead to re-use, shared learning and reduction of unnecessary repetition. The move towards doing the analysis where the data is rather than downloading files and doing it locally will require guidance, it will need to be easy to do. It may be beneficial to have simple notebooks on the platform that help people to browse a dataset and understand the quality of the data.

FDRI will make it clear what data has been published and how it has been processed. FDRI will provide online notebooks to enable the community to explore the datasets. FDRI will explore enabling the community to carry out their own analysis.

FAIR – Accessible

4. Licensing – we had a discussion around data licensing and making it clear what data can and cannot be used for.

FDRI datasets are going to be licensed to enable them to be openly accessible. We will engage with other data providers e.g. water companies for them to share data with open licenses wherever possible.

5. Diversity in user data processing needs – we discussed whether to have different levels of openness depending on the data, code and potential users. Technical skills could limit what someone can do in terms of re-running a model. It might not be feasible, or necessary, to allow everyone to re-run a huge analysis but a walk through of a model with an example dataset to understand how it might work could be considered.

We will explore how we could provide these different levels of data processing.

FAIR – Interoperable

6. Standards and guidelines – there was appetite for guidelines on how to do field monitoring, standards to follow to provide confidence in the data. If standards cannot be followed, information on what was done instead would be beneficial. The community would like FDRI to identify gaps in standards and address those and they would like guidelines for software generation, including examples to illustrate best practice. There was discussion around FDRI championing standard data formats and automation of metadata, promoting open documented formats.

FDRI will review current standards and provide guidance on which should be followed, it will also provide a means of capturing when standards have not been followed, for whatever reason. Where there are gaps in standards FDRI will look to address them or make recommendations.

FAIR – Reusable

7. Data quality and uncertainty metadata – the hydrological community would like standard data flags that can be used to highlight data quality issues. Users would like confidence in the data contained in the FDRI platform. It was discussed that with buy-in from the instrument manufacturers there could be some quality flags assigned to data as well as reasons for missing data (i.e. a range of standard data flags for missing data). Data users would like FDRI to capture instrument uncertainty in a standard way. Consider

capturing ranges, for example flow uncertainty ranges when converting from levels to flow using rating curves. Can FDRI encourage instrument manufacturers to provide uncertainty information in a machine-readable format?

As part of the data management work in FDRI, the team will suggest a set of standard data flags including uncertainty information and bring that back to the community for consultation. FDRI plans to investigate different sources of uncertainty data and how best to store and provide access for the community. In the longer term FDRI will discuss data quality flags with manufacturers.

8. Guidance on publishing and reusing code – there was appetite for guidance on how to produce software tools as well as helping researchers to turn their research code into publishable software packages. It will be important to lower the barriers to re-use of code.

FDRI will provide guidance for researchers to publish and share their code.

9. Publishing data – there was discussion around data journals as a route to publishing datasets, alongside data centres, something that as a community we should encourage each other to do. There will be different considerations for real-time data – how do you publish that? It's also important to acknowledge that as datasets are quality controlled or reviewed it may not be possible to completely reproduce a result e.g. flood and drought stats from real-time datasets.

Where policies already exist about publishing data, e.g. from the funder, these policies should be followed. FDRI will continue to make it easier to publish data in the NERC data centres, especially the EIDC.

Across the FDRI digital platform we also need to consider sustainability issues.

10. Sustainability – consideration needs to be given to the environmental impact of the FDRI, especially the digital elements. Do larger High Performance Computers (HPCs) run more efficiently than local HPCs? Are some formats more sustainable?

FDRI will discuss sustainability and engage with the community.

5. Next steps

FDRI aims to support a digital community where we make data, code and workflows open so that people can benefit from shared learning.

The FDRI team will:

1. In the short term (by the end of 2024) continue to grow the FDRI digital community though providing regular updates and sharing data, code, and related workflows openly online - so that people can benefit from shared learning. This will be delivered through a central online FDRI website and engagement with the wider UK research community at events and through focused engagements (Themes 1, 8, 9, and 10).
2. During the second half of 2024 we will work with the UK research community to soft launch an FDRI dataset showcase to increase the visibility of datasets and associated metadata across the hydrological community (Themes 1 and 9).
3. During 2025 we will provide near real-time open datasets on a cloud data platform and examples of their use. These will be accompanied by guidance on metadata and licensing information (Themes 1, 2, and 4).
4. During 2025 we will provide an example of a web-based collaborative platform that will support the analysis of FDRI and other data using research cloud computing, and the development and sharing of hydrological analysis toolkits (Themes 3, 5, and 8).
5. By the end of 2025 the FDRI digital team will have worked with the FDRI field team and wider community to review current field monitoring standards, and data standards for monitoring data, and provide guidance to help increase the use of uncertainty data within hydrological research (Themes 6 and 7).
6. By the end of 2026 the FDRI digital team, will have worked with the FDRI field team and wider community to co-develop and share online data processing workflows to enable the standardisation of quality assured data and related code (Themes 3 and 5).

We will host a series of digital community workshops to foster community engagement and continue the dialogue of reproducibility in hydrological research.

Acknowledgements

The authors would like to thank all the participants of the webinar and both workshops and also the following people for speaking, chairing, facilitating or helping organise the webinar and workshops and publishing the recordings and this report: Dr Irene Garousi-Nejad (CUAHSI), Dr Francesca Pianosi (University of Bristol), Dr Jennie Roebuck (UKCEH), Saskia Salwey (University of Bristol), Wiktoria Lawniczak (UKCEH), Faiza Samreen (UKCEH), Amulya Chevuturi (UKCEH), Sunita Sarkar (UKCEH), Marcia Spencer (UKCEH), Emma Bennett (UKCEH), Yvonne Keen (UKCEH), Faye Stokes (UKCEH), Paulette Burns (UKCEH), Alice Hope (UKCEH), Andy Sier (UKCEH).

Attendees and participants of the webinar and/or workshop were from the following organisations; Affinity Water, AtkinsRéalis, Arup, Biomathematics and Statistics Scotland (BioSS), British Geological Survey (BGS), Campbell Scientific, Cardiff University, Cranfield University, Consortium of Universities for the Advancement of Hydrologic Science (CUAHSI), King's College London, Environment Agency (EA), University of Oxford, Innovate UK Knowledge Transfer Network (KTN), JBA Consulting, JBA Trust, Lancaster Environment Centre, Labcell Ltd, Natural Resources Wales (NRW), Natural Environment Research Council (NERC-UKRI), Newcastle University, Nottingham Trent University, Queen Mary University of London, University of Edinburgh, Scottish Environment Protection Agency (SEPA), The James Hutton Institute (JHI), UK Centre for Ecology & Hydrology (UKCEH), University of Aberdeen, University of Birmingham, University of Bristol, University of Edinburgh, University of Leeds, University of Manchester, University of Southampton, Wallingford HydroSolutions.

Contact

We welcome opportunities for you to get involved with FDRI. We are looking for early adopters and beta-testers of the FDRI digital infrastructure, ideas for future workshops and feedback on your FDRI digital needs. Please get in touch with us by emailing FDRI@ceh.ac.uk to discuss opportunities.

Partners

Funder:

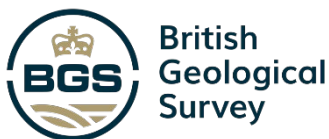
- UKRI-NERC (£38 million)

Lead partner:

- UK Centre for Ecology & Hydrology

Other partners:

- British Geological Survey
- Imperial College
- University of Bristol



IMPERIAL

