

Creating FAIR Outputs

Data Stewardship Team, UK Centre
of Ecology and Hydrology



Introductions



Els Dhiedt: Biogeochemist Data
Scientist at UK Centre for Ecology and
Hydrology



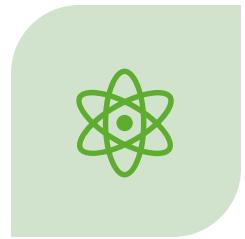
Jasmine Hunter: Data Steward at UK
Centre for Ecology and Hydrology



UK Centre for
Ecology & Hydrology

ceh.ac.uk

Agenda



1. WHAT IS FAIR DATA?



2. PREPARING FAIR DATA



3. PUBLISHING FAIR DATA



4. CREATING FAIR METADATA



5. TOOLS,
RESOURCES,
QUESTIONS



UK Centre for
Ecology & Hydrology

ceh.ac.uk

Section 1: What is FAIR data

1.1 What are the FAIR principles



UK Centre for
Ecology & Hydrology

ceh.ac.uk



A Short history of FAIR

A set of guiding principles for open and transparent data sharing

Developed in 2016 by representatives from the scientific community

SCIENTIFIC DATA

Amended: Addendum

OPEN

SUBJECT CATEGORIES

- » Research data
- » Publication characteristics

Received: 10 December 2015

Accepted: 12 February 2016

Published: 15 March 2016

Comment: The FAIR Guiding Principles for scientific data management and stewardship

Mark D. Wilkinson *et al.*^{*}

There is an urgent need to improve the infrastructure supporting the reuse of scholarly data. A diverse set of stakeholders—representing academia, industry, funding agencies, and scholarly publishers—have come together to design and jointly endorse a concise and measurable set of principles that we refer to as the FAIR Data Principles. The intent is that these may act as a guideline for those wishing to enhance the reusability of their data holdings. Distinct from peer initiatives that focus on the human scholar, the FAIR Principles put specific emphasis on enhancing the ability of machines to automatically find and use the data, in addition to supporting its reuse by individuals. This Comment is the first formal publication of the FAIR Principles, and includes the rationale behind them, and some exemplar implementations in the community.

Supporting discovery through good data management

Good data management is not a goal in itself, but rather is the key conduit leading to knowledge discovery and innovation, and to subsequent data and knowledge integration and reuse by the community after the data publication process. Unfortunately, the existing digital ecosystem surrounding scholarly data publication prevents us from extracting maximum benefit from our research investments (e.g., ref. 1). Partially in response to this, science funders, publishers and governmental agencies are beginning to require data management and stewardship plans for data generated in publicly funded experiments. Beyond proper collection, annotation, and archival, data stewardship includes the notion of ‘long-term care’ of valuable digital assets, with the goal that they should be discovered and re-used for downstream investigations, either alone, or in combination with newly generated data. The outcomes from good data management and stewardship, therefore, are high quality digital publications that facilitate and simplify this ongoing process of discovery, evaluation, and reuse in downstream studies. What constitutes ‘good data management’ is, however, largely undefined, and is generally left as a decision for the data or repository owner. Therefore, bringing some clarity around the goals and desiderata of good data management and stewardship, and defining simple guideposts to inform those who publish and/or preserve scholarly data, would be of great utility.

What is...?

F

- Findable**
- Feasible**
- Functional**
- Friendly**

A

- Automated**
- Accessible**
- Accurate**
- Achievable**

I

- Integrated**
- Interoperable**
- Intelligent**
- Innovative**

R

- Recyclable**
- Reproducible**
- Reliable**
- Reusable**



What are the FAIR principles?



Findable: Can the data be found easily by humans and machines?



Accessible: Once it is found, is the data easy to access and download?



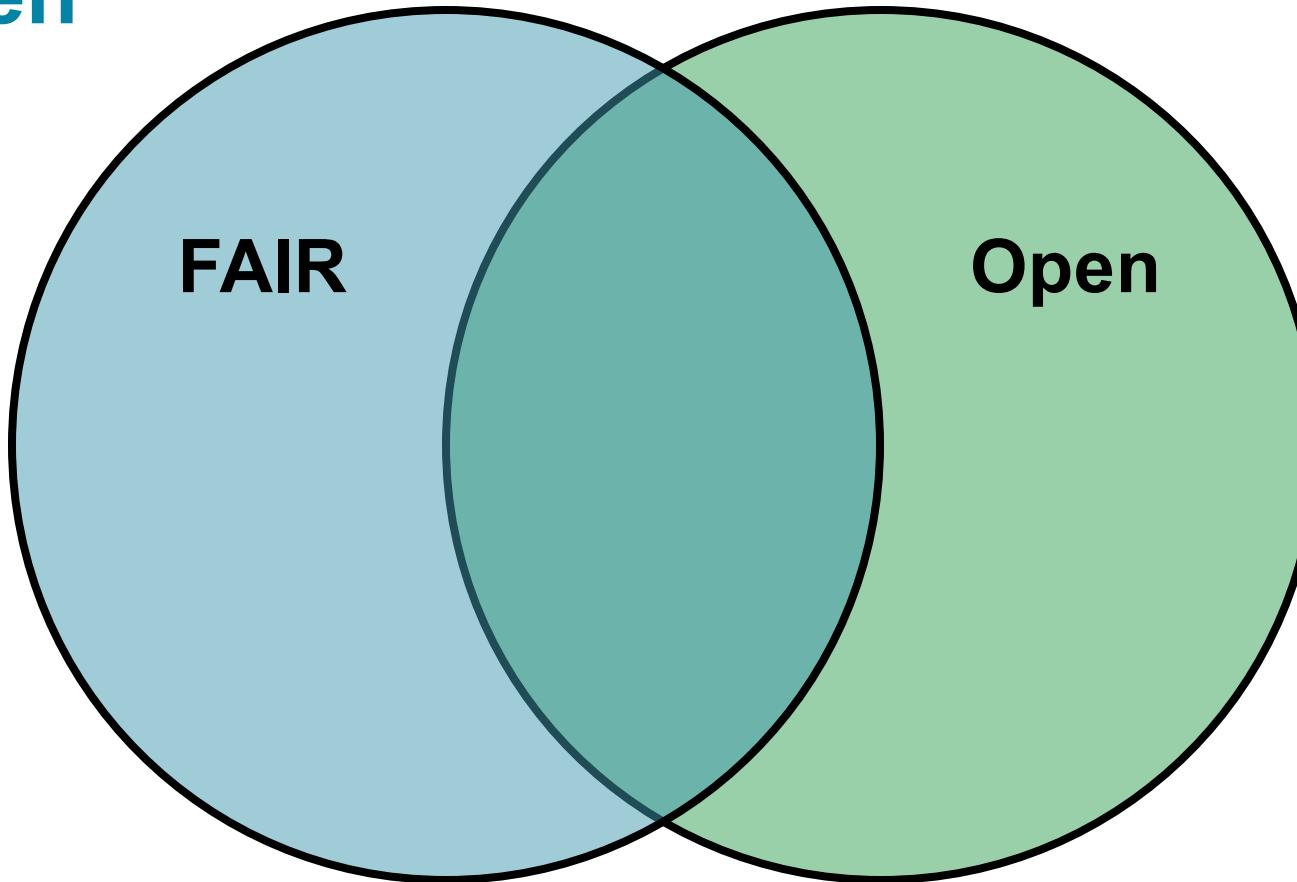
Interoperable: Is the data easy to integrate with other data, workflows, and applications?



Reusable: Is it easy for the data to be recreated and understood?



FAIR vs open



‘Data should be as open as possible, as closed as necessary’
ARDC 2020

1. What is FAIR Data

1.2 Why is FAIR important?



UK Centre for
Ecology & Hydrology

Obligations



Obligations



Funders

- Funder's data policy
- Accessibility, embargo periods, data publication



Obligations



 Publishers

- Requirement of published data



Obligations

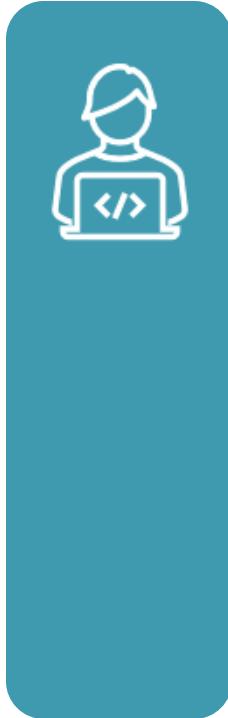


Legislation

- Personal data (e.g. GDPR)
- INSPIRE (Eu)
- Freedom of Information



Benefits



UK Centre for
Ecology & Hydrology

ceh.ac.uk

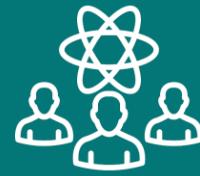
14

Benefits

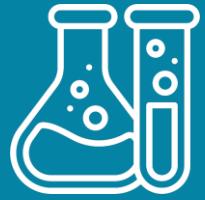


Data producer

- Credit and promotion of work
- Data authors
- More trust in your research
- Build your network and create new collaboration opportunities

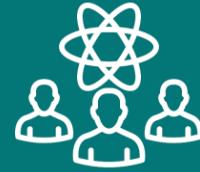


Benefits



Data user

- No need to reproduce data
- Easier to access and reuse other data
- Opportunity for collaboration



Benefits

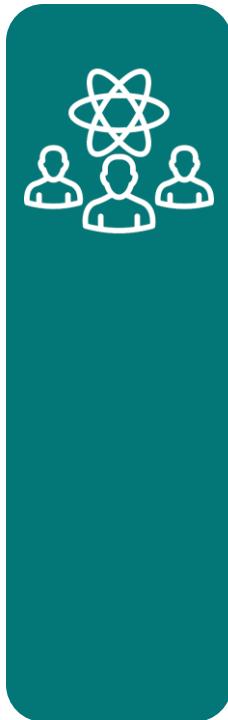


Scientific community

- Building on existing science
- Inspiration



Benefits



Society

- Progress innovation
- Access to data
- Transparent science



2. Preparing FAIR data

2.1 How to achieve FAIR



UK Centre for
Ecology & Hydrology



How to achieve FAIR? Exercise



How to achieve FAIR?



Findable

- Discoverable
- Web-accessible record
- Well-managed catalogue
- Unique identifier – e.g. DOI
- Discipline-specific keywords & description



How to achieve FAIR?



Accessible

- Persistently working link
- No access constraints or minimal and clear constraints via licencing
- APIs and user interfaces



How to achieve FAIR?



Interoperable

- Accepted languages, format, vocabularies
- Non-proprietary formats
- Machine readability



How to achieve FAIR?



Reusable

- Discipline-specific vocabularies
- Appropriate metadata standards
- Supporting documentation
- Clear licence describing terms of use



2. Preparing FAIR data

2.2 Key considerations



UK Centre for
Ecology & Hydrology

ceh.ac.uk

Preparing FAIR data – Community standards

What are the community standards for data in your area?

Techniques and methodologies	Control measurements	Uncertainties and error reporting
<ul style="list-style-type: none">• Clear descriptions of sampling methods	<ul style="list-style-type: none">• Geochemical reference material and lab standards included in metadata	<ul style="list-style-type: none">• Compulsory age model uncertainty ranges• Spatial resolution included

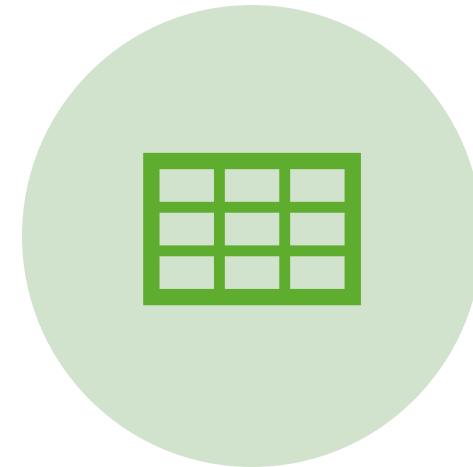


Community standards

Use the following to search for standards that may apply to your field of research:



FAIRsharing repository of standards: ideal for a detailed search. You can apply filters such as 'subject' (e.g. Ecology).



DCC list of metadata standards: general list of most common standards.



Preparing FAIR data – Data formats



Preparing FAIR data - Controlled vocabularies

Controlled vocabularies are lists of standardised terms. They are used by data creators to enhance searching through harmonizing terminology across research fields.

 **The NERC Vocabulary Server (NVS)**

NVS Home | Vocabularies | Thesauri | Search NVS | SPARQL | Other Tools | About NVS

Search for a term in a vocabulary collection

milligrams	P06	Search		
<input checked="" type="checkbox"/> Identifier <input checked="" type="checkbox"/> Preferred label <input checked="" type="checkbox"/> Alternative label <input type="checkbox"/> Definition <input type="checkbox"/> Exact match <input type="checkbox"/> Case sensitive toggle advanced options				
Found 15 records Show 100 ▾ 1				
Identifier ↑	Preferred label ↑	Alternative label ↑	Definition ↑	Date ↑
MGPG	Milligrams per gram	mg/g	One part per 10^3 (thousand) by weight of the measurand in the matrix.	2022-12-02
UMKG	Milligrams per kilogram	mg/kg	One part per 10^3 (thousand) by weight of the measurand in the matrix.	2022-12-02
MGSS	Milligrams per square metre per second	mg/m ² /s	10^{-6} of the SI unit of mass (kilogram) per SI unit of area per SI unit of time.	2021-04-11
MGSQ	Milligrams per square metre per hour	mg/m ² /h	Unavailable	2005-05-05
UCLC	Milligrams of carbon per (microEinstein per square metre per second) per milligram of chlorophyll per hour	mgC/(uE/m ² /s)/mgChl/h	Unavailable	2005-03-01

GENERAL INFORMATION

 R

Climate and Forecast Standard Names Parameter Vocabulary (CF Standard Names)

doi [10.25504/FAIRsharing.f8cd83](https://doi.org/10.25504/FAIRsharing.f8cd83)



Type Terminology artefact

Registry Standard

Description The Climate and Forecast Standard Names Parameter Vocabulary (CF Standard Names) is intended for use with climate and forecast data in the atmosphere, surface and ocean domains. The CF Standard Names are the list of standard names used by the CF conventions. The list includes the units recommended for each standard name (most common prefixes can be used with the units, e.g. kilo (k), hecto (h), Mega (M), etc). If a standard_name metadata attribute is associated with a data variable, its value must be chosen from the list published in the standard name table. It is not compulsory within the CF conventions to assign a standard name to a data variable, but including one helps data users to understand the contents of a netCDF file. A CF standard name is not a netCDF variable name, but instead a value for the standard_name variable attribute. A standard_name attribute identifies the physical quantity of a variable.

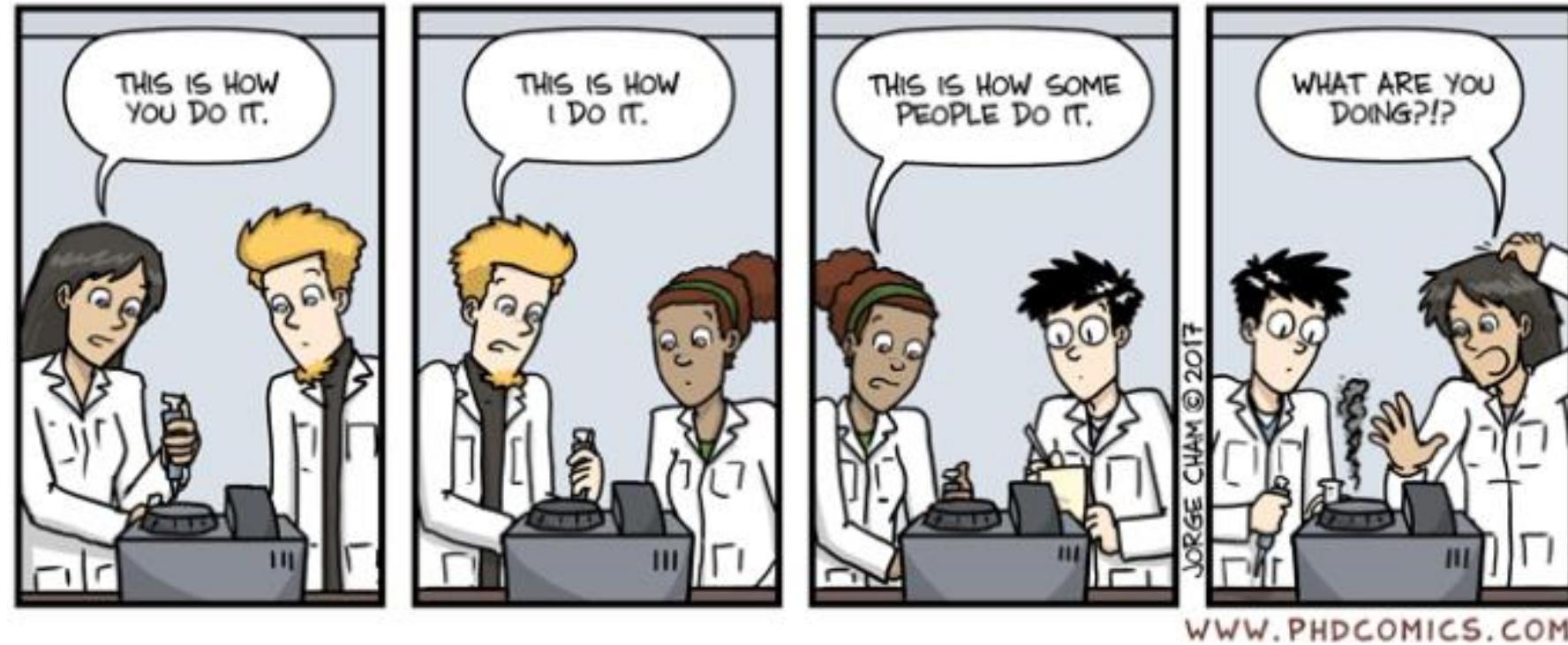
Homepage <http://cfconventions.org/Data/cf-standard-names/current/build/cf-standard-name-table.html>

Year of Creation 2018

Reproducible data transformation and methods

Take notes as you go!
This will help you create supporting information and ensures the quality of your dataset

This includes adding comments to code to make it reproducible



Any questions?



UK Centre for
Ecology & Hydrology



3. Publishing FAIR data

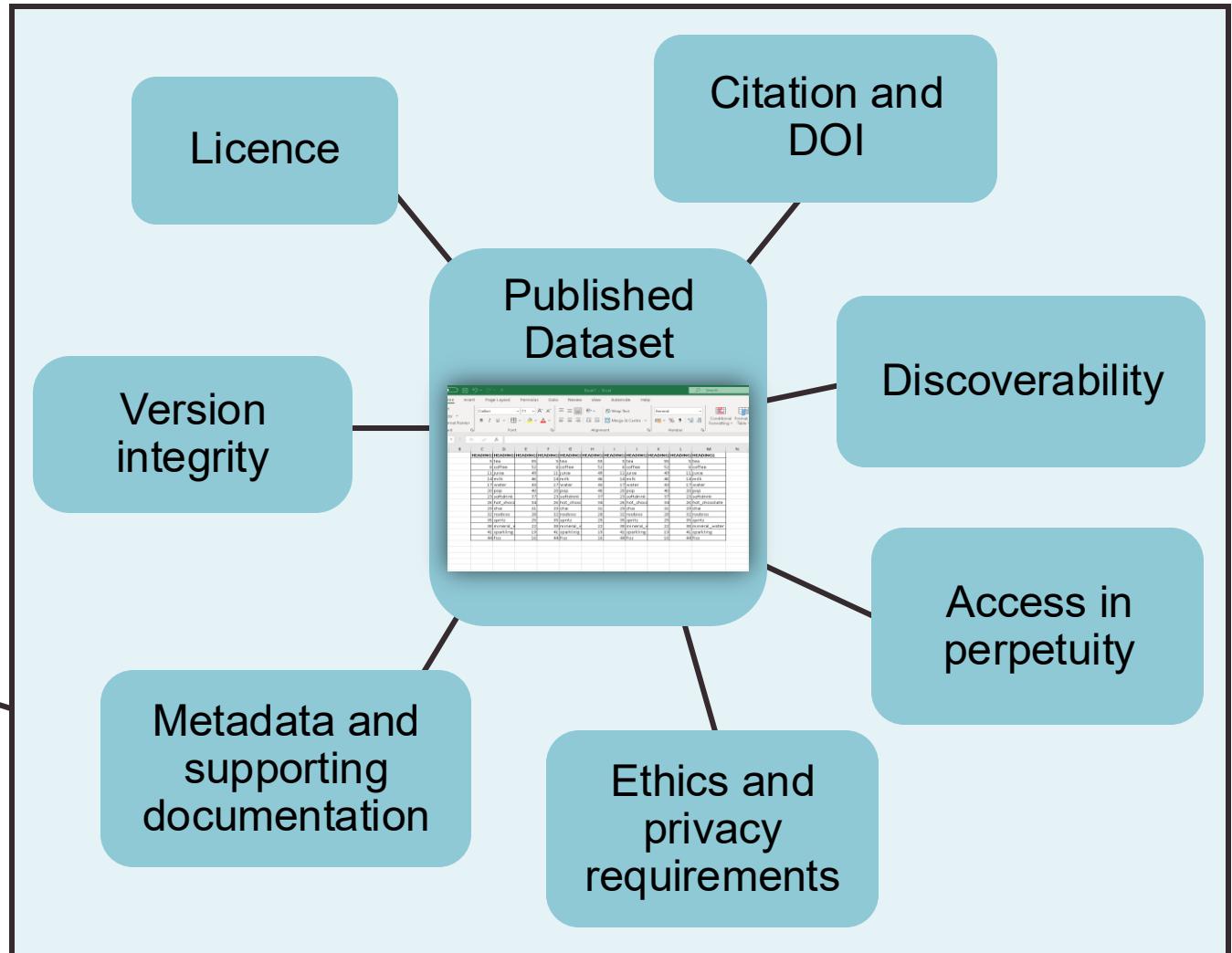
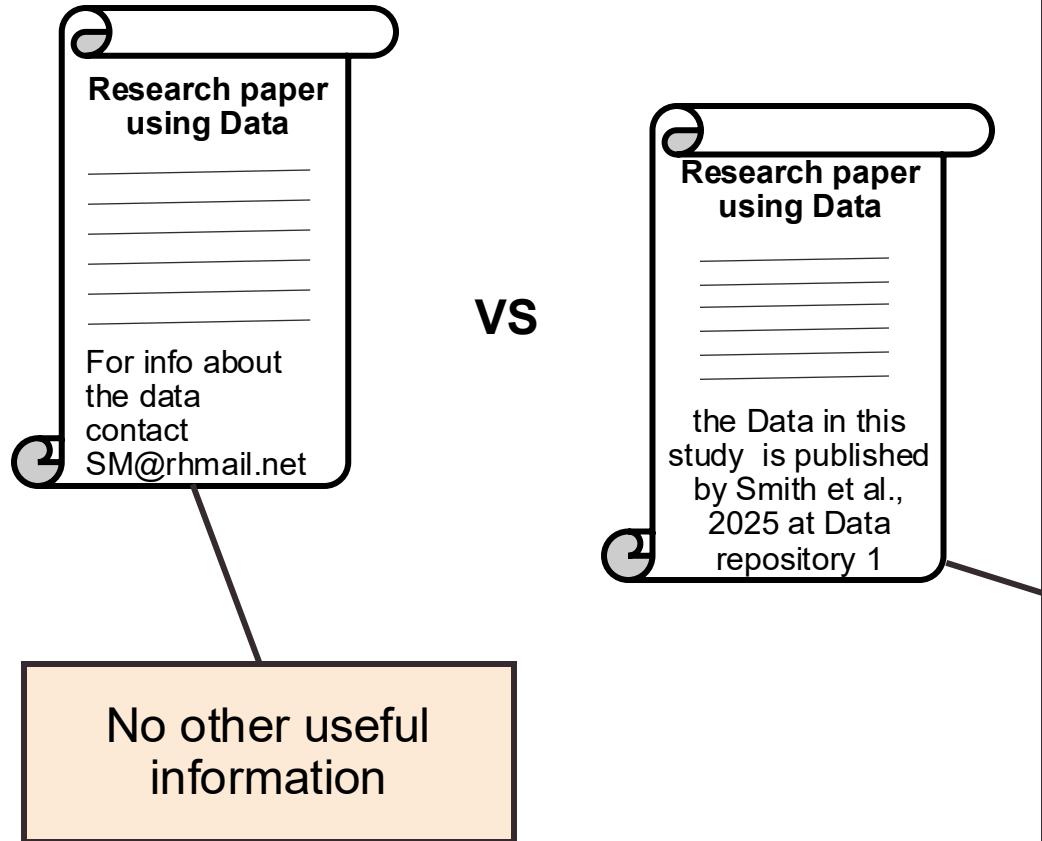
3.1 The data publication process



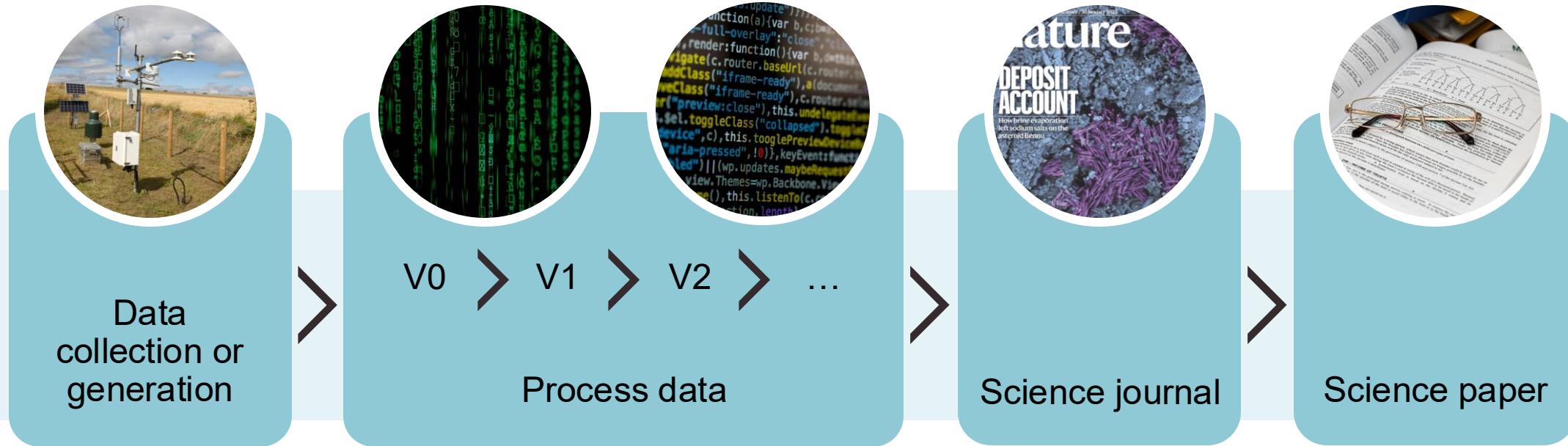
UK Centre for
Ecology & Hydrology



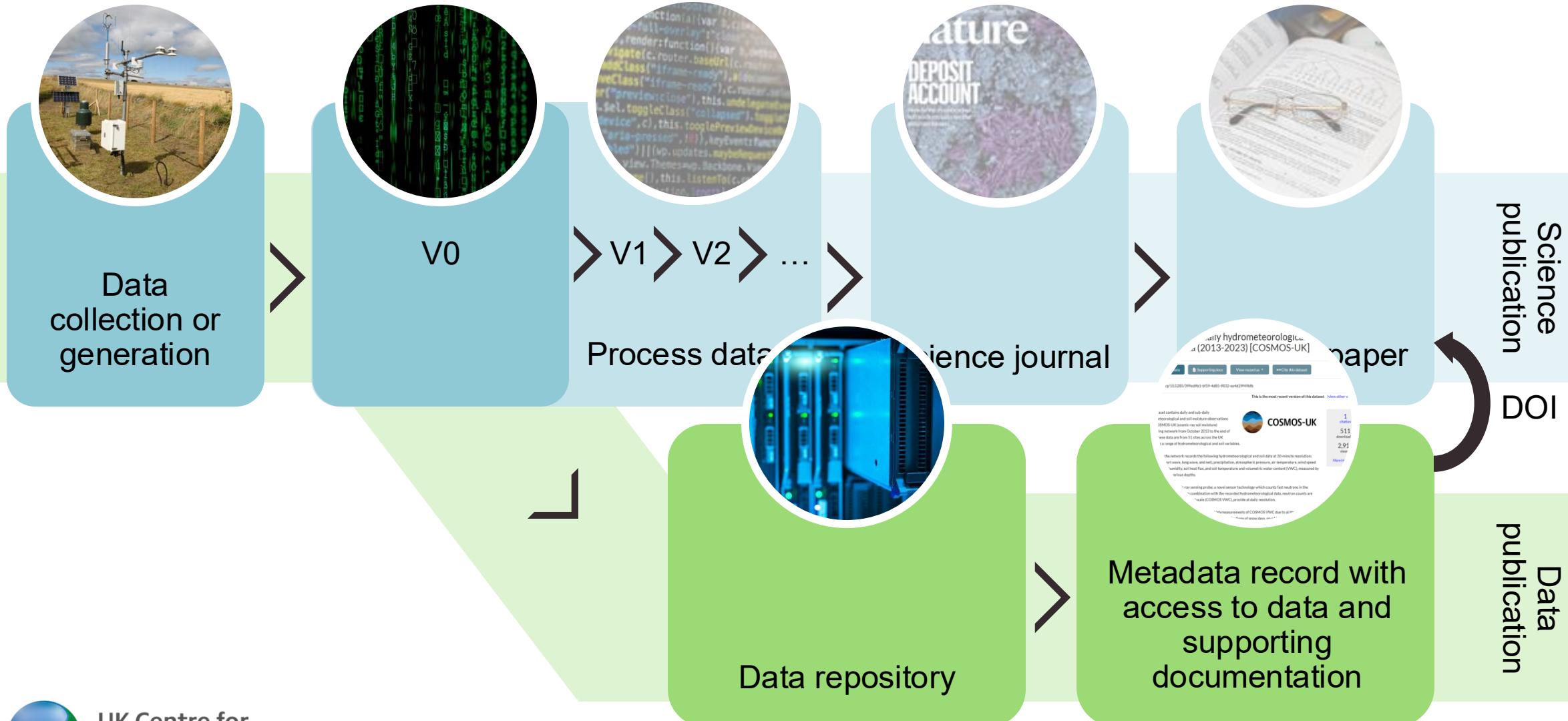
Published data - more than a paragraph in the supplementary!



Data publication



Data publication



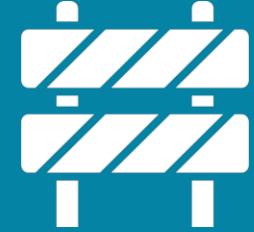
The Power of Persistent identifiers (PIDs)



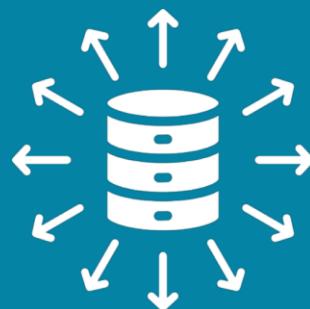
3. Publishing FAIR data

3.2 Why publish data?





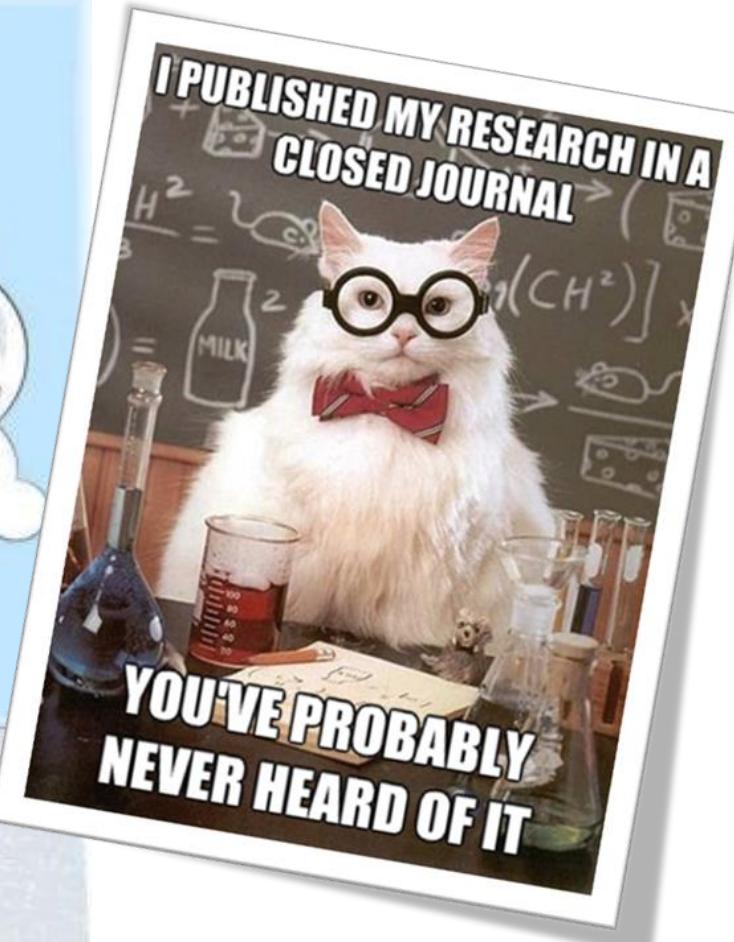
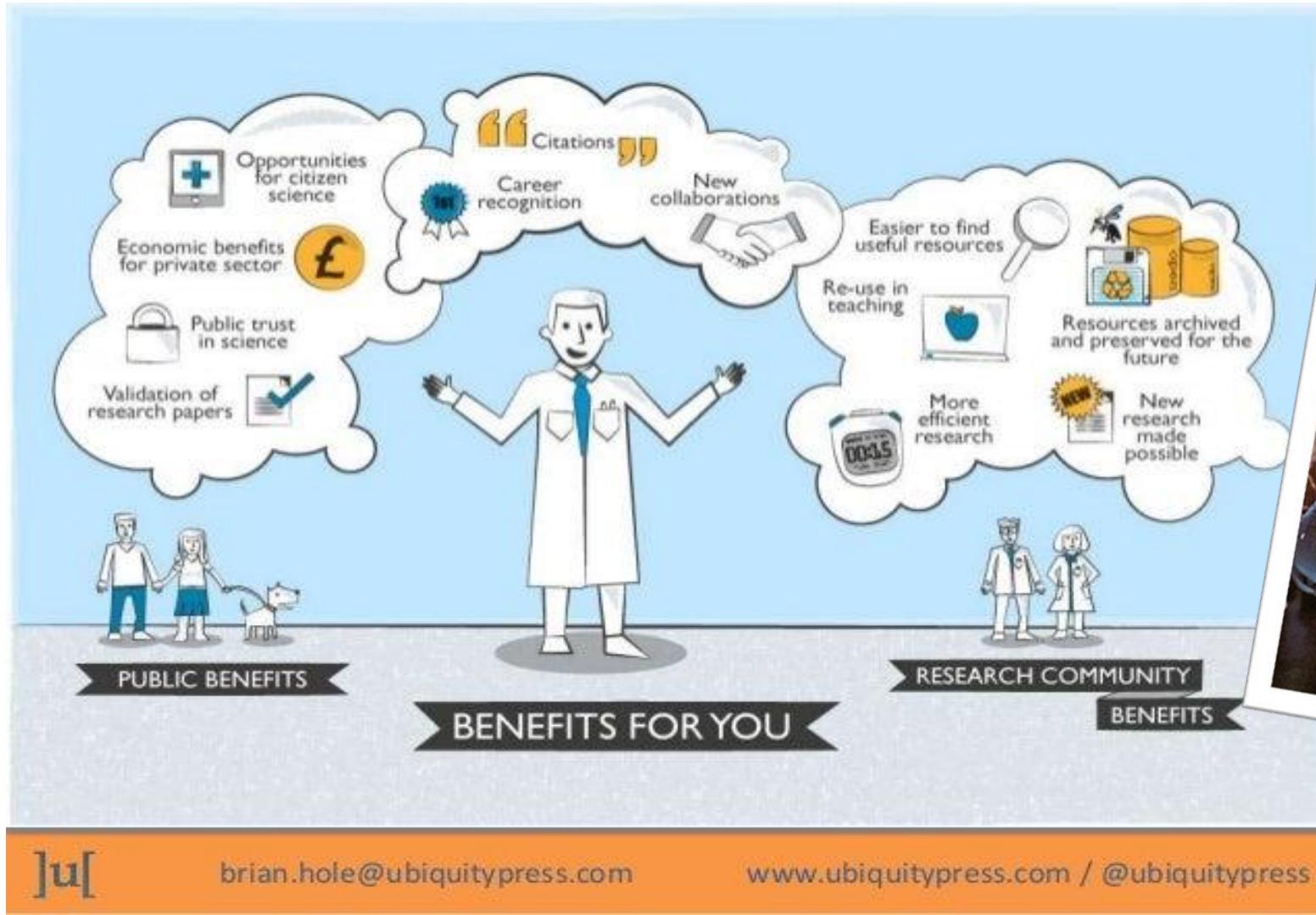
What are some barriers to publishing data?



What are the benefits of publishing data?



Why publish data – Open data = Open science!



brian.hole@ubiquitypress.com

www.ubiquitypress.com / @ubiquitypress



UK Centre for
Ecology & Hydrology



NERC
Environmental
Data Service

ceh.ac.uk

39

Why publish data – recognition of work and outputs



Ó Dochartaigh, B; O'Keeffe, J; Nair, T; Jackson, C; Krishnaswamy, J; Chaudhary, S; MacDonald, A.

Surface and groundwater stable isotope analysis using water samples taken from the Gandak River Basin, Bihar, India, 2017-2019

[Download data](#) [Supporting docs](#) [View record as](#) [Cite this dataset](#)

<https://doi.org/10.5285/09ae86d6-896f-430f-aab4-c5b46c265213>

The dataset contains stable isotope data from surface and groundwater samples collected in the Gandak Basin, north India. The data was collected between March 2017 and February 2019. These measurements were taken to improve understanding of surface and subsurface water interconnections and movement through river and canal networks and underlying aquifers.

The data were collected as part of the NERC sponsored project Coupled Human and Natural Systems Environment (CHANSE), grant number NE/N01670X/1

Publication date: 2020-05-20

Viewing outcome of Professor Richard Pywell

Select the type of research dataset, database or model. *

Dataset/Database/Collection of data

Provide a short name/title for this research dataset, database or model. *

Mutlier archetypes to characterize British landscapes, farmland and farming practices

Briefly describe this research dataset, database or model. *

This dataset consists of landscape and agricultural management archetypes (1 km resolution) at three levels, defined by different opportunities for agri-environment. Tier 1 archetypes quantify broad differences in soil, land cover and population across Great Britain, which cannot be readily influenced by the actions of land managers. Tier 2 archetypes capture more nuanced differences in landscape and management, such as the presence of organic manure spreading or arable land. Tier 3 archetypes are the most detailed, capturing specific localities where land managers can have a significant influence on the landscape. The archetypes are designed to be used in a range of applications, including environmental impact assessments, climate change adaptation planning, and agricultural policy development. The three tiers of archetypes were analysed separately and not as a nested structure (i.e. a single Tier 3 archetype can occur in more than one Tier 2 archetype), predominantly to ensure interarchetype coherence was easily maintained across each tier.

Is this research dataset, database or model published or available to others? *

Yes

Please select the year that this outcome was first published or made available to others. *

2022

If there is a URL which relates to this research dataset, database or model, enter it here.

<https://catalogue.ceh.ac.uk/3244375a-cbd5-458c-9395-41471054609>

Digital Object Identifier

10.5285/3244375a-cbd5-458c-9395-41471054609

Common Outcomes
Why am I asked these? [?](#)



Publications	2
Collaborations & Partnerships	1
Further Funding	0
Next Destination	0
Engagement Activities	6
Influence on Policy, Practice, Patients & the Public	0
Research Tools & Methods	0
Research Datasets, Databases & Models	1
Intellectual Property & Licensing	0
Medical Products, Interventions & Clinical	0



Why publish data – credit to data creators and ECRs

Ó Dochartaigh, B.; O'Keeffe, J.; Nair, T.; Jackson, C.; Krishnaswamy, J.; Chaudhary, S.; MacDonald, A.

Surface and groundwater stable isotope analysis using water samples taken from the Gandak River Basin, Bihar, India, 2017-2019

 Download data

 Supporting docs

 View record as ▾

Ó Dochartaigh, B.; O'Keeffe, J.; Nair, T.; Jackson, C.; Krishnaswamy, J.; Chaudhary, S.; MacDonald, A. (2020). Surface and groundwater stable isotope analysis using water samples taken from the Gandak River Basin, Bihar, India, 2017-2019. NERC Environmental Information Data Centre. <https://doi.org/10.5285/09ae86d6-896f-430f-aab4-c5b46c265213>

<https://doi.org/10.5285/09ae86d6-896f-430f-aab4-c5b46c265213>

The dataset contains stable isotope data from surface and groundwater samples collected in the Gandak Basin, north India. The data was collected between to improve understanding of surface and sub-canal networks and underlying aquifers.

The data were collected as part of the NERC Environment (CHANSE), grant number NE/N

1

Lapworth, D.J., Ó Dochartaigh, B., Nair, T., O'Keeffe, J., Krishnan, G., MacDonald, A.M., Khan, M., Kelkar, N., Choudhary, S., Krishnaswamy, J., & Jackson, C.R.(2021). Characterising groundwater-surface water connectivity in the lower Gandak catchment, a barrage regulated biodiversity hotspot in the mid-Gangetic basin. *Journal of Hydrology*, 594, 125923. <https://doi.org/10.1016/j.jhydrol.2020.125923>

Publication date: 2020-05-20

[More information](#)

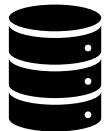


UK Centre for
Ecology & Hydrology

ceh.ac.uk

41

Why publish data – so it can be reused!



Grid-to-grid (G2G)

[EIDC DOI](#), paper DOI



Triple Collocation (TCM)

[EIDC DOI](#), [paper DOI](#)



SMUK

[EIDC DOI](#), [paper DOI](#)



CHESS (land, met and PE)

[EIDC DOI](#), [EIDC DOI](#), [EIDC DOI](#)



COSMOS-UK

[EIDC DOI](#), [paper DOI](#), [paper DOI](#)



Land Cover Map

2015

[EIDC DOI](#)



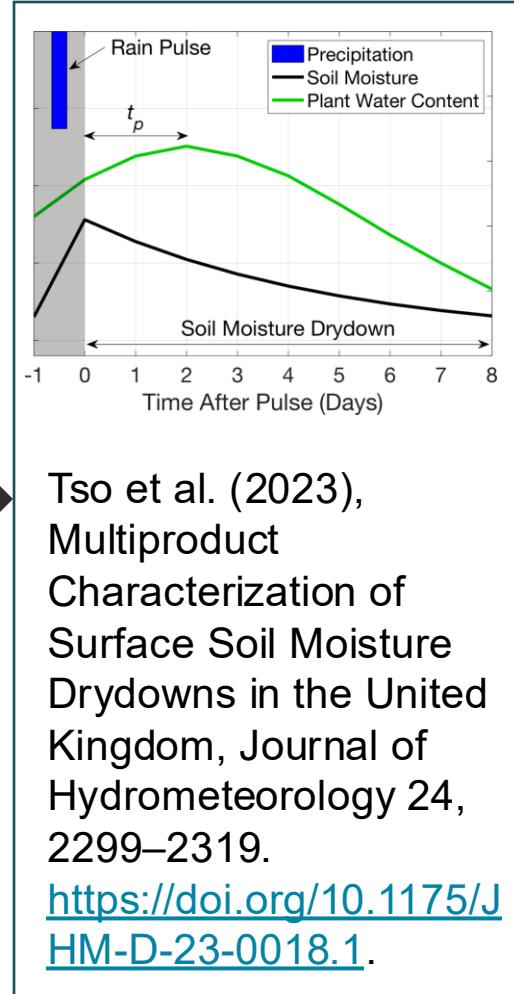
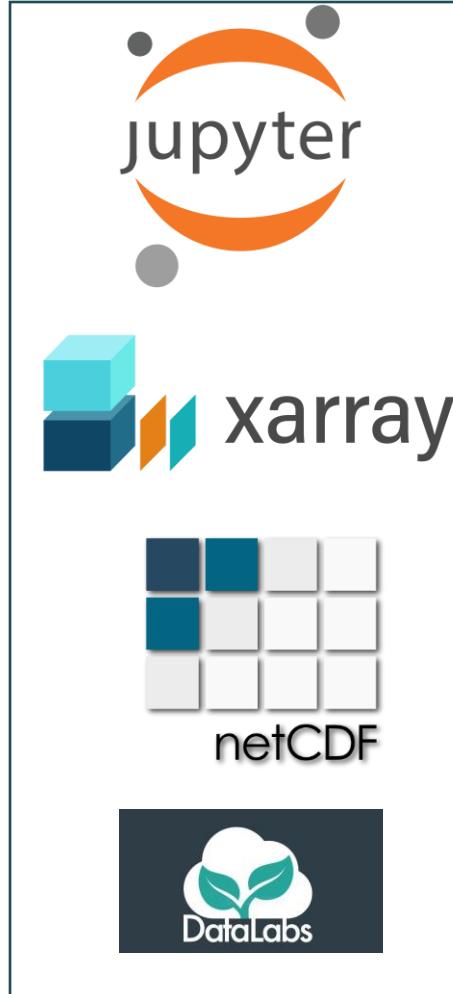
Countryside Survey
topsoil carbon

[EIDC DOI](#)

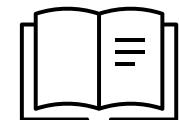


Environmental
Information
Data Centre

See paper
data
availability
statement



UK Centre for
Ecology & Hydrology



Why publish data – meet policy requirements

NERC data policy

Key Principles

The environmental data produced by the activities funded by NERC are considered a public good and they will be made openly available for others to use. NERC is committed to supporting long-term environmental data management to enable continuing access to these data.

Page 1 of 4

NERC will supply the environmental data it holds for free, apart from a few special cases as detailed in the policy.

NERC requires that all environmental data of long-term value generated through NERC-funded activities must be submitted to NERC for long-term management and dissemination.

BES Journals data archiving policy

- Suitable archives
- Third-party data, embargoes and waivers
- FAQs

Data are important products of the scientific enterprise, and they should be preserved and usable for decades in the future. The British Ecological Society thus requires, as a condition for publication, that all data supporting the results in papers published in its journals are archived in an appropriate public archive offering open access and guaranteed preservation. For theoretical papers the underlying model code must be archived.

The data underlying all the results presented in the paper must be archived in a format that allows a third party to reproduce the data. The archived data must allow each result in the published paper to be recreated and the analyses reported in the paper to be replicated in full to support the conclusions made. Authors are welcome to archive more than this, but not less.

Authors may elect to have the data made publicly available at time of publication or, if the technology of the archive allows, may opt to embargo access to the data for a period of up to a year after publication. Exceptions, including longer embargoes, may be granted only in exceptional circumstances at the discretion of the editor, especially for sensitive information such as confidential social data or the location of endangered species.

All papers must have a data archiving statement and data sources must be cited in the

3. Publishing FAIR data

3.3 What to publish?

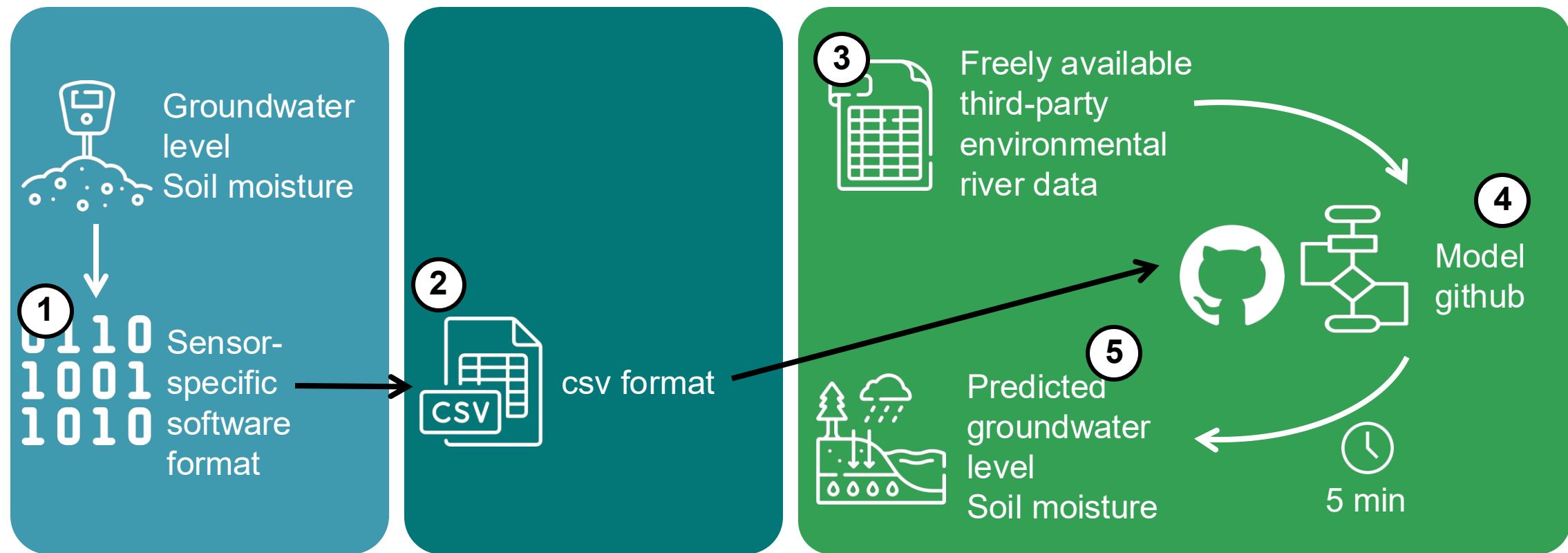


UK Centre for
Ecology & Hydrology



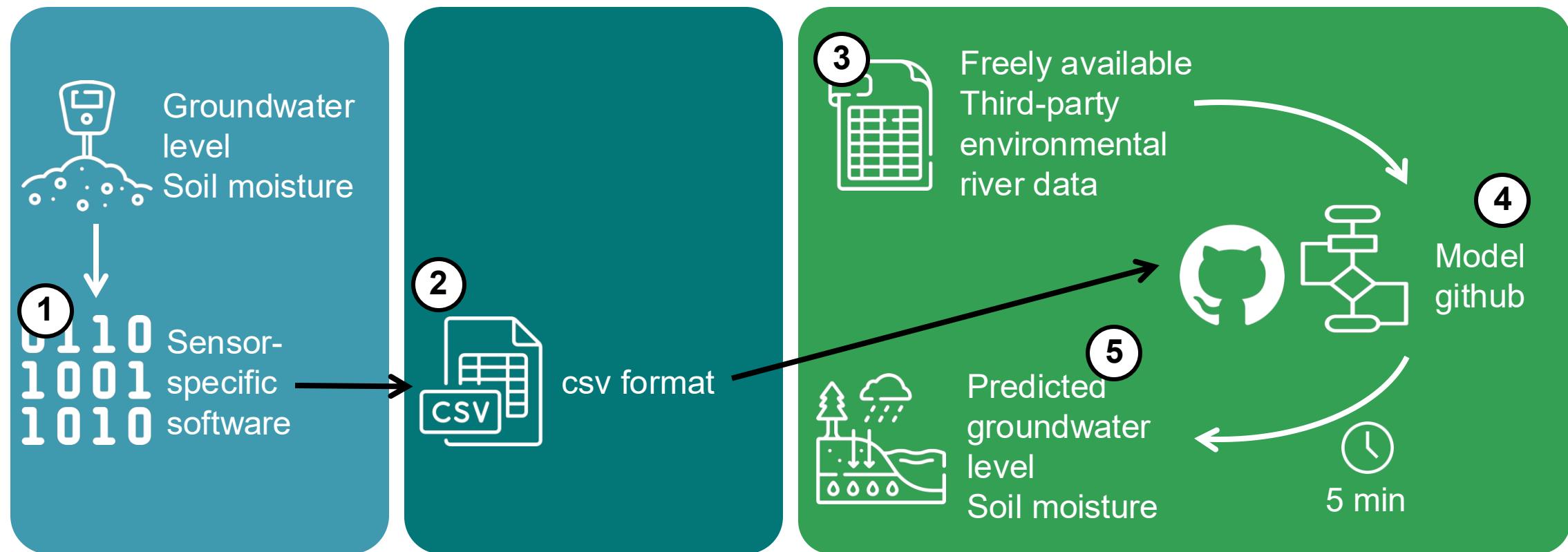
Data of long-term value

Project on groundwater level and soil moisture in the Wear Catchment, north-east England



Data of long-term value

Project on groundwater level and soil moisture in the Wear Catchment, north-east England



Data of long-term value



Not everything! – if can be easily reproduced, it's not of long-term value



Data you have generated – can incorporate third party data, if licence allows



Raw vs Processed – not straightforward!



Complete – or annual snapshot if long-term monitoring



3. Publishing FAIR data

3.4 Where to publish



UK Centre for
Ecology & Hydrology

ceh.ac.uk



What is a data repository?

a centralized place to store, organize, and access data (physical or online)

Purpose

- **Data sharing**
- **Data preservation**
- **Data analysis**
- **Data management**



How to choose a repository

You can search for repositories in your field via
<https://commons.datacite.org/repositories>



Funder data policy requirements



Meets FAIR?



Repository certifications



Intended audience



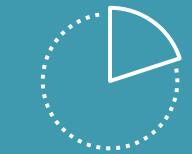
UK Centre for
Ecology & Hydrology

Types of repositories



General Repositories

- Publish data from any field
- Different levels of offerings and certifications



Types of repositories



Specialist repositories

- Repositories focussed on a specific field or data type
- e.g UK Data service – Social datasets
- World Glacier Monitoring Service



Types of repositories



NERC EDS Repositories

- UK data centres run by NERC
- Specific domains
- Required for NERC funded projects



UK Centre for
Ecology & Hydrology

ceh.ac.uk

53

NERC Environmental Data Service

Repositories

We provide a focal point for scientific data and information spanning all environmental science domains. We are a network of distributed data centres, with domain specific expertise:



British
Oceanographic
Data Centre

British Oceanographic Data Centre
Discipline: Marine



NGDC

National Geoscience Data Centre
Discipline: Geoscience

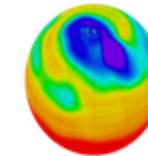


UK Centre for
Ecology & Hydrology



Environmental
Information
Data Centre

Environmental Information Data Centre
Discipline: Terrestrial and
Freshwater



Centre for Environmental
Data Analysis
SCIENCE AND TECHNOLOGY FACILITIES COUNCIL
NATURAL ENVIRONMENT RESEARCH COUNCIL

Centre for Environmental Data Analysis
Discipline: Atmospheric, Earth
Observation, and Solar and
Space Physics



British
Antarctic Survey

UK Polar Data Centre

UK Polar Data Centre
Discipline: Polar and
Cryosphere

Who are the EIDC?



UK's national data centre
for terrestrial and
freshwater sciences.

Certified as a trusted
repository by
CoreTrustSeal



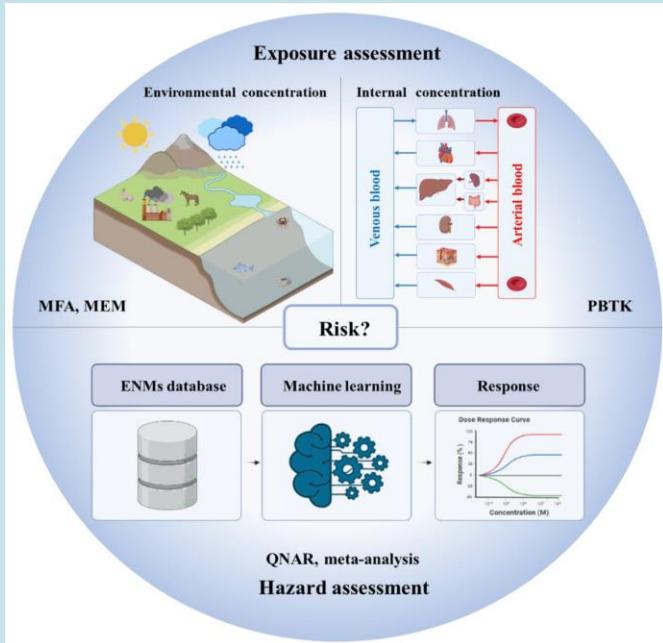
We accept environmental
data in re-usable formats
for curation, including
long-term storage and
public dissemination.



Environmental
Information
Data Centre

What about Models, code, and software?

Models, code, and software



One option to store on
Github, and publish via
Zenodo



UK Centre for
Ecology & Hydrology

scientific data

OPEN

ARTICLE

Introducing the FAIR Principles for research software

Michelle Barker¹✉, Neil P. Chue Hong², Daniel S. Katz³, Anna-Lena Lamprecht⁴, Carlos Martinez-Ortiz⁵, Fotis Psomopoulos⁶, Jennifer Harrow⁷, Leyla Jael Castro⁸, Morane Gruenpeter⁹, Paula Andrea Martinez¹⁰ & Tom Honeyman¹¹

Research software is a fundamental and vital part of research, yet significant challenges to discoverability, productivity, quality, reproducibility, and sustainability exist. Improving the practice of scholarship is a common goal of the open science, open source, and FAIR (Findable, Accessible, Interoperable and Reusable) communities and research software is now being understood as a type of digital object to which FAIR should be applied. This emergence reflects a maturation of the research community to better understand the crucial role of FAIR research software in maximising research value. The FAIR for Research Software (FAIR4RS) Working Group has adapted the FAIR Guiding Principles to create the FAIR Principles for Research Software (FAIR4RS Principles). The contents and context of the FAIR4RS Principles are summarised here to provide the basis for discussion of their adoption. Examples of implementation by organisations are provided to share information on how to maximise the value of research outputs, and to encourage others to amplify the importance and impact of this work.

Introduction

In 2016 the publication of "The FAIR Guiding Principles for scientific data management and stewardship"¹ supported a vision where valuable scientific outputs are made 'FAIR' by becoming more Findable, Accessible, Interoperable and Reusable. From the outset, the FAIR Guiding Principles were intended to be applicable to many kinds of digital assets. Increased understanding of the importance of research software in research has catalysed application of the FAIR Guiding Principles to this type of digital asset.

Community-endorsed FAIR principles for research software were released in 2022 by the FAIR for Research Software (FAIR4RS) Working Group (WG), which was jointly convened by the Research Software Alliance (ReSA), Future of Research Communications and E-Scholarship (FORCE11), and the Research Data Alliance (RDA). This milestone reflects the maturation of the research community in understanding the benefits of having FAIR research software, and coming together as the FAIR4RS WG to achieve this. The FAIR4RS WG is a global and interdisciplinary community whose members share an interest in the application of FAIR principles to research software, such as researchers, software users, developers and maintainers, policy makers, infrastructure support staff, and funders.

The FAIR4RS Principles are relevant to any stakeholder in the research community seeking to increase transparency, reproducibility, and reusability of research. This paper highlights the importance of the FAIR4RS Principles and the positive signals of adoption that demonstrate high levels of community support. It must also be acknowledged that research software and data discoverability is a long-standing challenge and there have

Barker et al., (2022)
<https://doi.org/10.1038/s41597-022-01710-x>

4. FAIR Metadata



UK Centre for
Ecology & Hydrology



What are they?



Metadata: ‘data about data’, structured, key information that provides context about the dataset to allow viewers to understand, interpret, and re-use a dataset.



Supporting documentation: documentation containing information that allows your data to be fully understood.



Why create supporting documentation and metadata?

Never forget what you did or why you did it

Helps researchers interpret the data and derive information from it

Helps researchers to find, retrieve, and reuse the data



Why create supporting documentation and metadata?

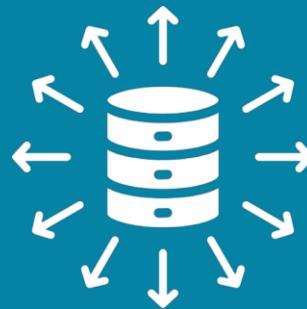


**Ensure proper
attribution of
research outputs**

**Helps comply with
data sharing
policies**

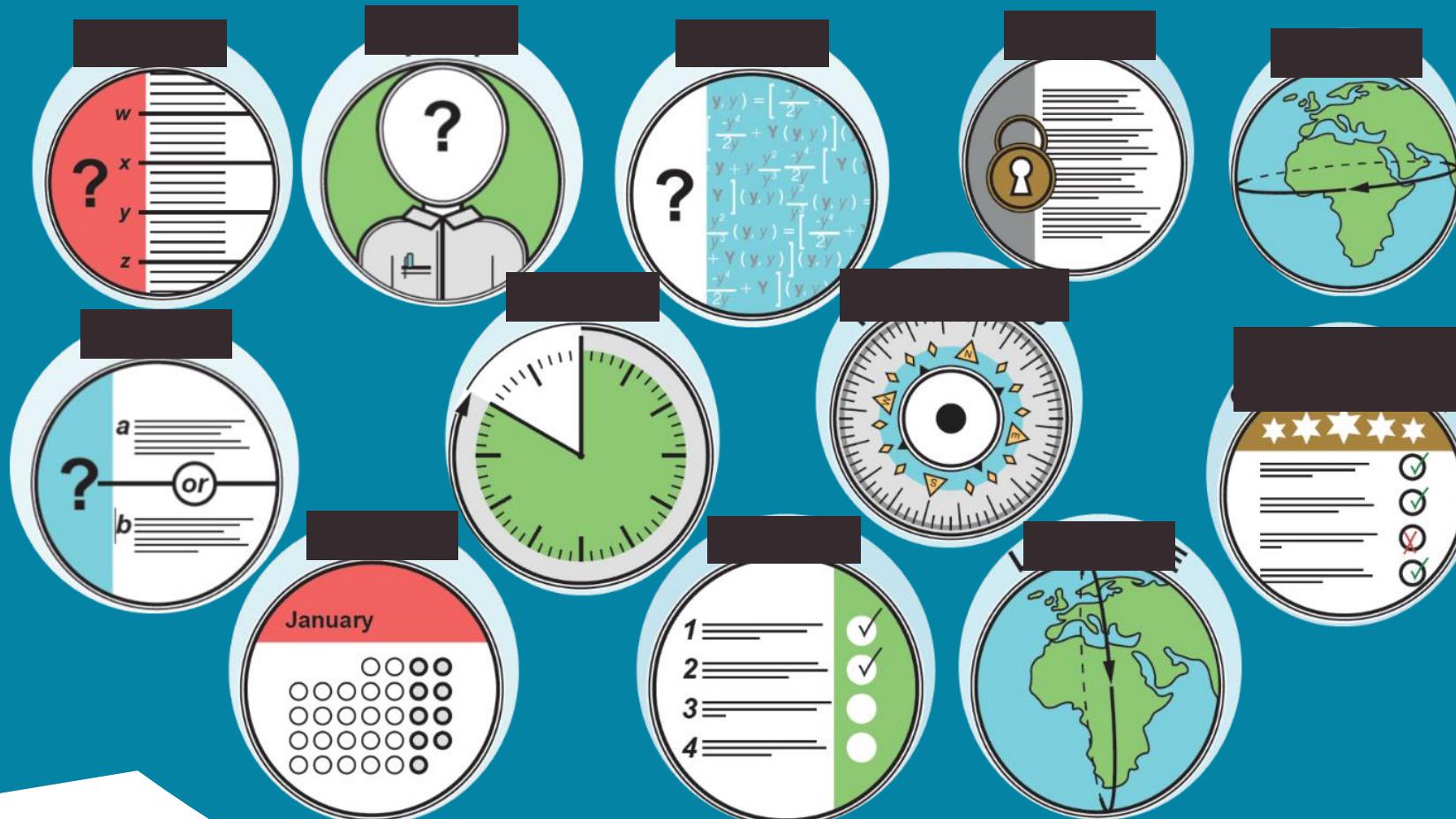
**Avoid
misinterpretation**





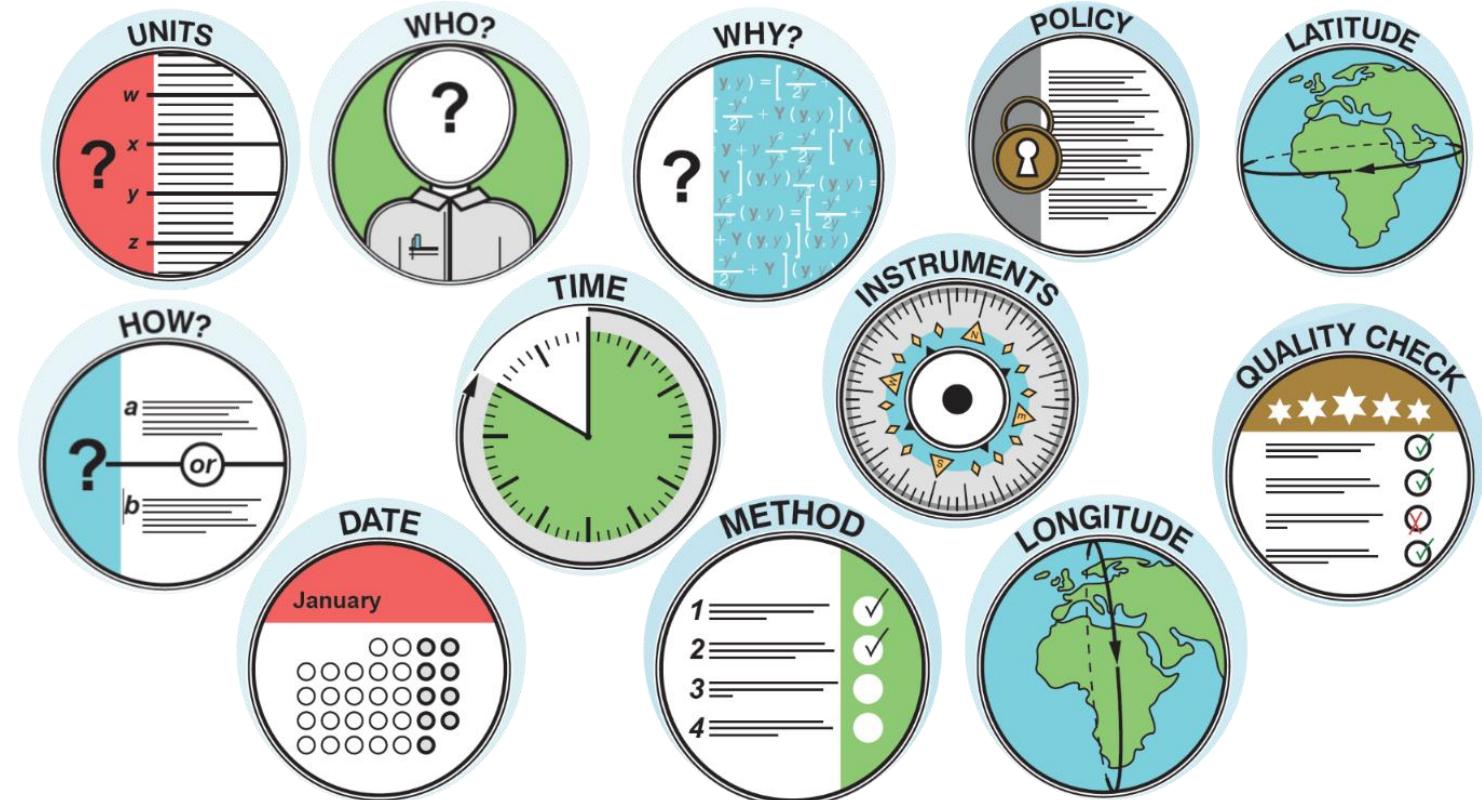
Scenario: You've found the perfect third-party dataset for your research – What information do you need to be able to reuse it effectively?

What information do you need to be able to reuse it effectively?



What information to include in a data publication: the data

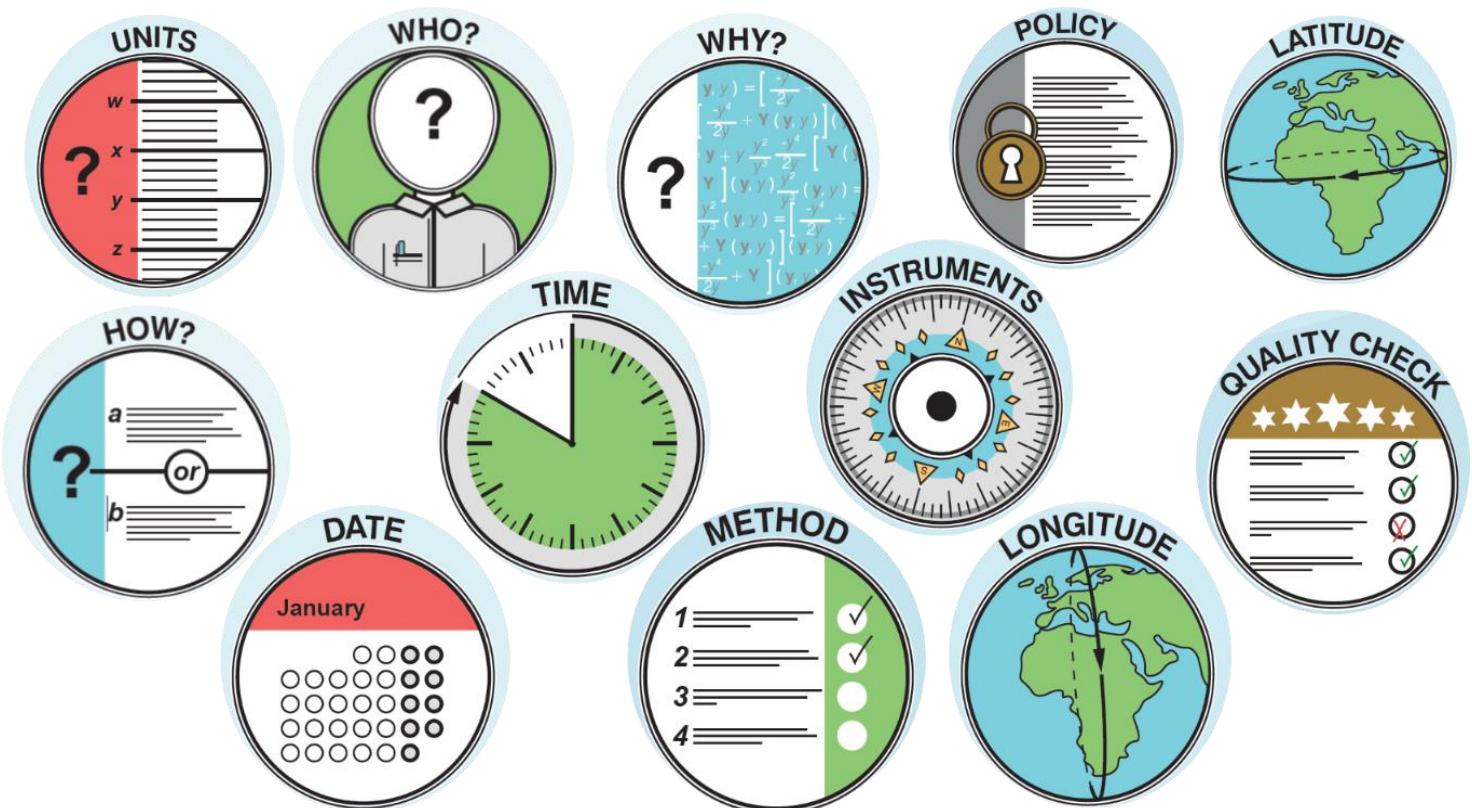
- Data files should be:
- Whole and complete
 - In a reusable format (non-proprietary)
 - QA'd to ensure accuracy and integrity



What information to include in a data publication: Metadata

Metadata – Data about the data

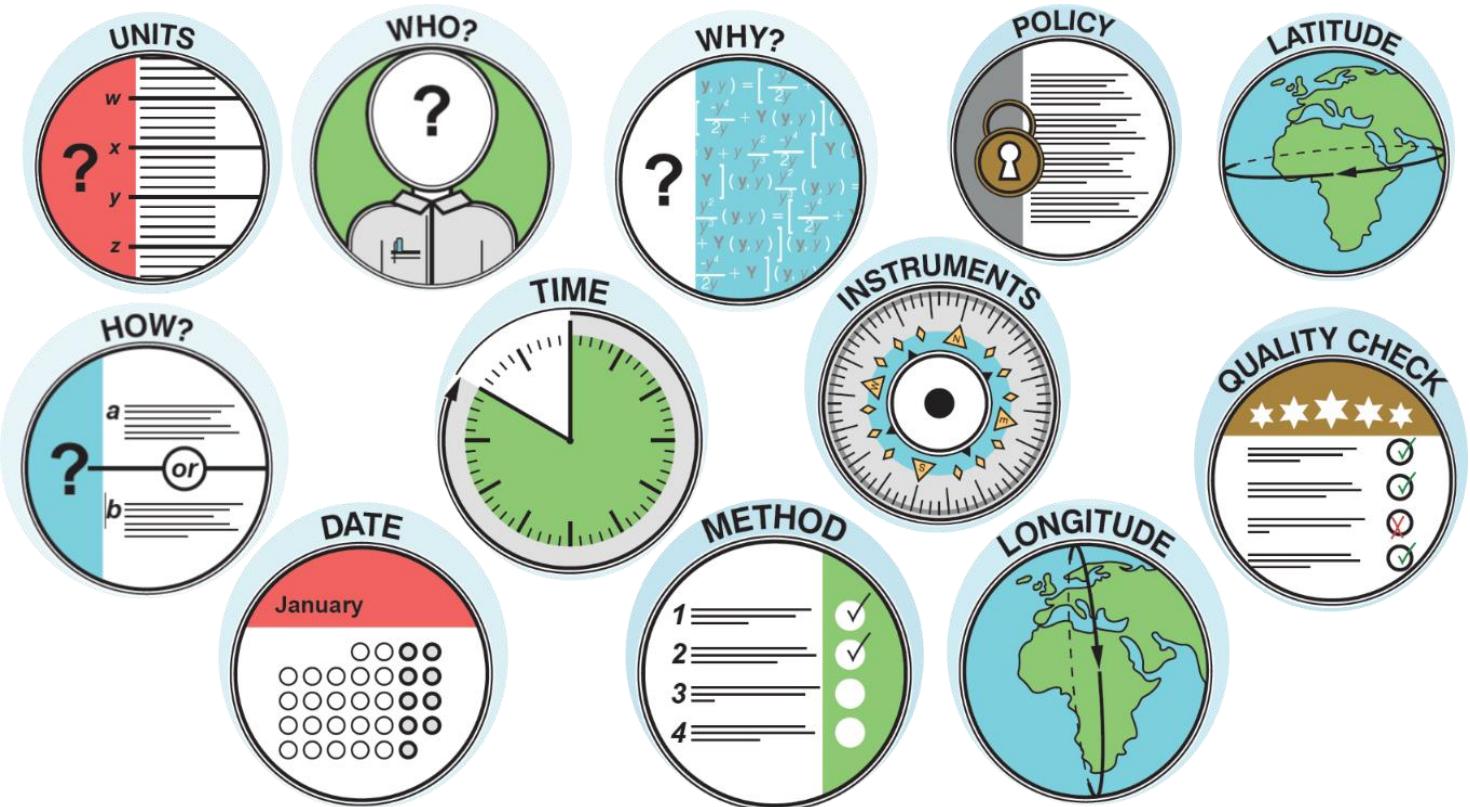
- Descriptive metadata – the WHO, WHAT, WHERE, HOW, and WHY of the dataset



What information included in a data publication: Metadata

Metadata – Data about the data

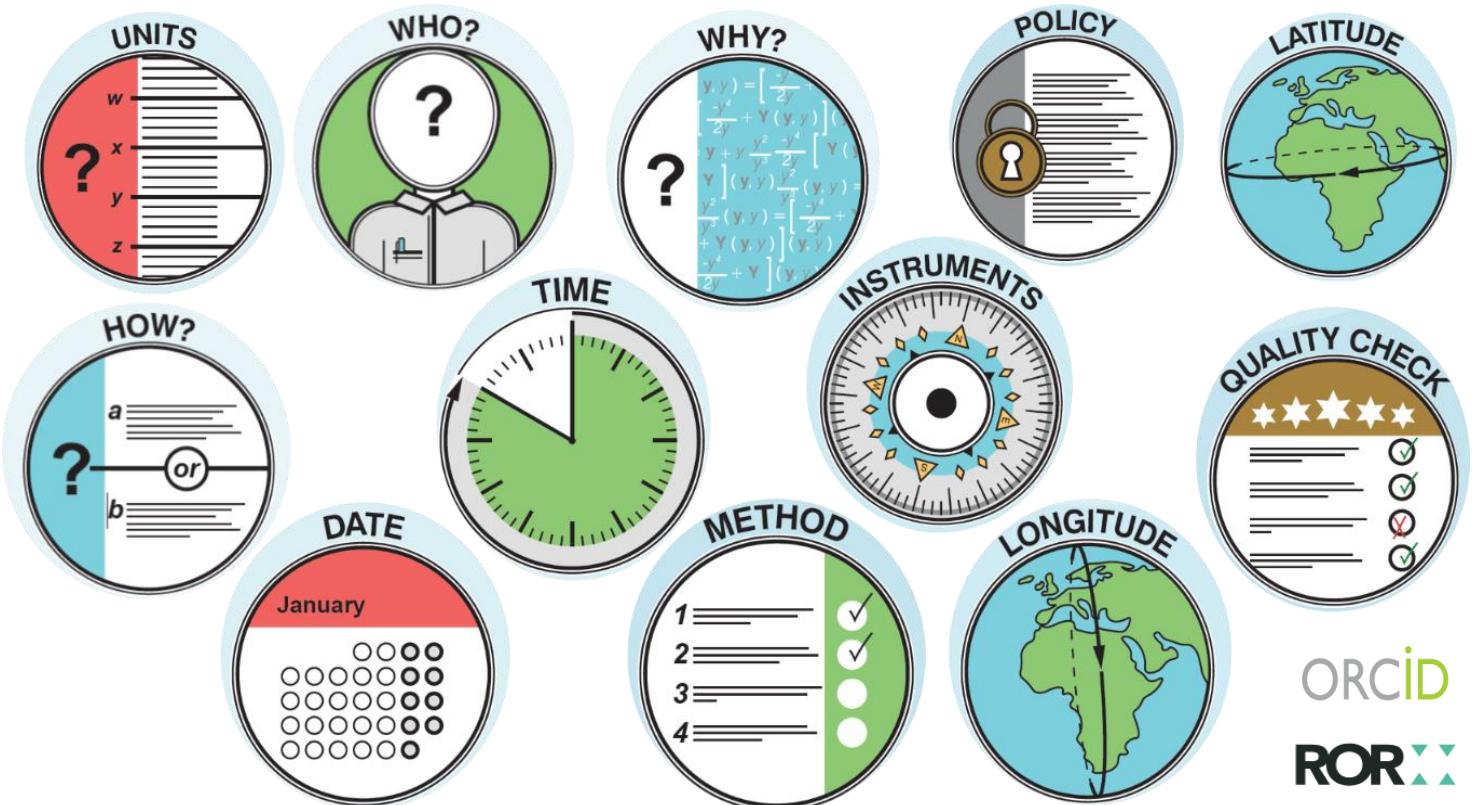
- Structural metadata:
Describes the data format, contents, and how it is organised



What information included in a data publication: Metadata

Metadata – Data about the data

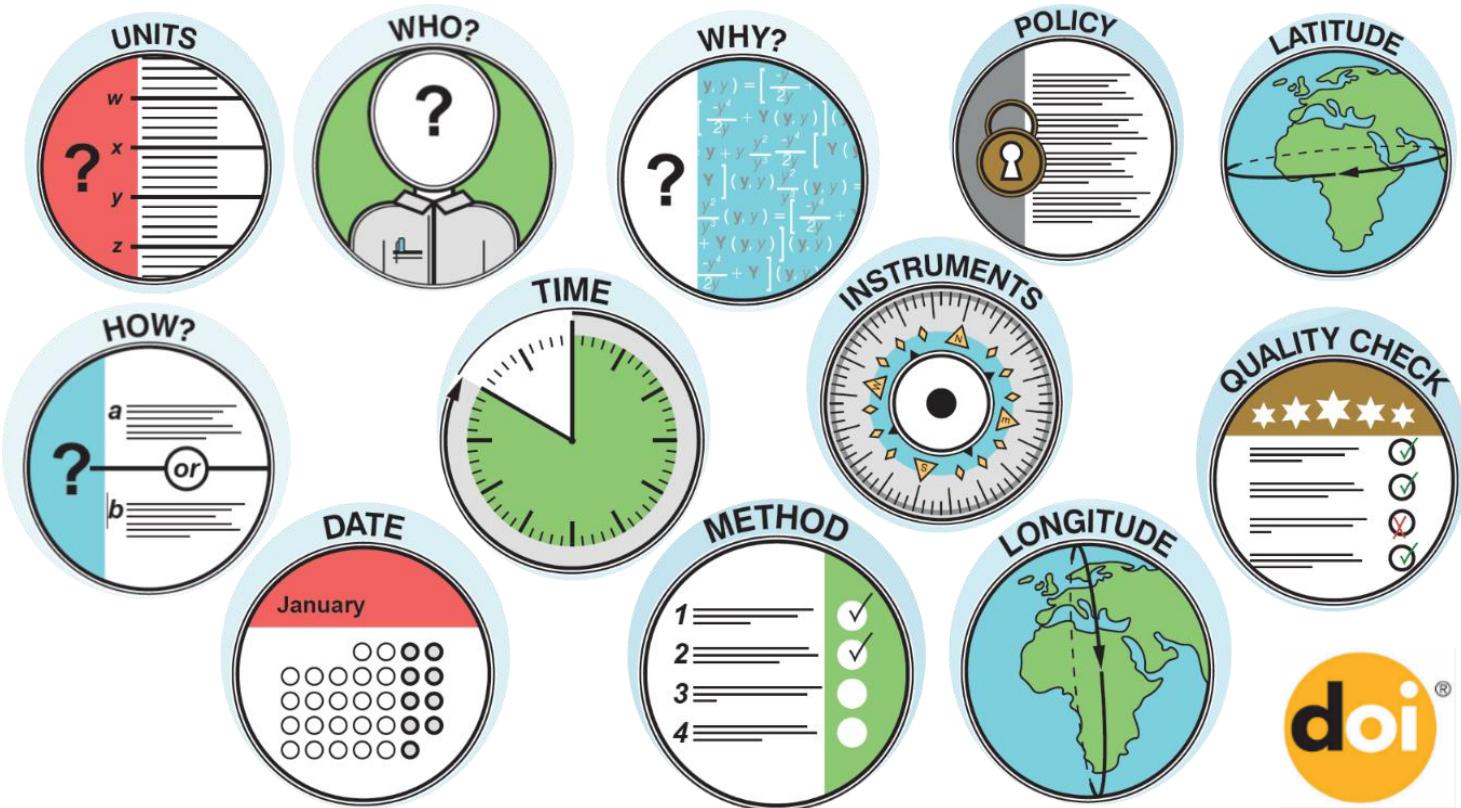
- Administrative metadata – how the data can be used
- Licencing, funding, acknowledgements, authors, institutions
- Project access



What information included in a data publication: Metadata

Metadata – Data about the data

- Relationship metadata: how the data links with other datasets, code, and research
- Versioning
- Source datasets



Example 1 – Supporting document

Working Title

Describe the data resource, not the project that created it or the findings from an experiment. Location and temporal extent should be included if relevant. You may be required to modify the title of the data resource depending on the requirements of the data repository when depositing.

Summary

A short summary about what data are included and why they were collected/generated.

File format

List the format(s) the data file(s) are in e.g., CSV, NetCDF, GeoTIFF. Repositories may only accept non-proprietary formats. Check guidance on suitable formats from the repository you are using.

File names and/or naming convention

List the file names of the data files (including the file extension).

If the files are arranged in a folder (and subfolders), that hierarchy needs to be described here.

If you have a lot of files (e.g., more than 20), you may prefer to specify any file naming conventions used instead.

Example: There are 20 files with the following name convention:

YYYYMMDD_rainfall_StationName.csv

Where YYYYMMDD refers to the date the measurements were recorded and StationName is the name of the station at which the measurements were recorded.

1 | 5

Nature and units of recorded values

Include information about the structure of the data (variable name(s), description, units).

- For tabular data (e.g., spreadsheets or databases) we recommend a structure such as the table below. If you have multiple files with different variables in each, use one table per file.
- For image data, you should include information such as resolution, focal length, etc.
- If applicable, include details of valid range of values for the variables, lowest level of detection, determinants etc.
- If there are missing data, state the value used to represent them (e.g., "Missing values are indicated by 'N/A'")
- If available, include URLs for vocabulary terms

Table 1. Description of variables in [insert filename].

Variable/Column header	Description	Units	Vocabulary URL (if applicable)
Dissolved nitrate	Nitrate measured in soil	mg/l	http://onto.nerc.ac.uk/CAST/16

Spatial coverage

[CONDITIONAL – Required for spatial data resources]

Include the following:

- name and bounding coordinates of the region OR for sampling locations provide the coordinates for each location OR if those aren't possible provide the name of the general region/locations.
- Coordinate reference system and/or grid used e.g., WGS84, OSGB 1936 / British National Grid
- If applicable, relevant geographical information e.g., aspect, elevation, surface area, volume etc.

Temporal coverage and resolution

[CONDITIONAL — Required if applicable]

Include the time period during which the data collection took place e.g., 01 January 2023 to 31



Example 2 – Data dictionaries

The screenshot shows a data dictionary spreadsheet titled "Sheet_1". The left sidebar lists files in the "data/" folder, including "0.1 datadictionary.csv" which is currently selected. The main table has columns for Variable, Variable name, Measurement unit, Allowed values, and Description.

Variable	Variable name	Measurement unit	Allowed values	Description
Participant ID number	ID	Numeric	001-999	ID number assigned to participant in sequential order
Group number	GROUP	Numeric	1-30	Group assigned to participant based on ID number
Age in years	AGE	Numeric	18.0-65.0	Age of participant in years
Date of birth	DOB	mm/dd/yyyy	1-12/1-31/1951-1998	Participant's date of birth
Gender	SEX	Numeric	1 = male 2 = female	Participant's gender
Date of survey	SURVEY	mm/dd/yyyy	01/01/2015 – 01/01/2016	When the participant completed the survey
Self-reported consumer spending	SPEND	Numeric	0-100,000,000	Self-reported average yearly expenditure
Market sentiment	SENTIMENT	Numeric	1 = negative 2 = neutral 3 = positive	Sentiment towards US domestic economy
Actual GDP growth	GDP	Numeric	-5.0-5.0	Average US yearly GDP growth



Example 3 – README files

This readme file was generated on [YYYY-MM-DD] by [NAME]
<help text in angle brackets should be deleted before finalizing your document>
<[text in square brackets should be changed for your specific dataset]>

GENERAL INFORMATION

Title of Dataset:

<provide at least two contacts>

Author/Principal Investigator Information

Name:

ORCID:

Institution:

Address:

Email:

Author/Associate or Co-investigator Information

Name:

ORCID:

Institution:

This template can be found on the
Cornell University website



5. Tools and resources



UK Centre for
Ecology & Hydrology



EIDC – Guidance and Supporting information template

The EIDC provides guidance on data structure, formatting, and metadata contents

Plus a downloadable supporting document template!

The screenshot shows the EIDC website with a header featuring the UKCEH logo and navigation links: Find data, Deposit data, Support, About, Contact us, and Help. A large banner image of a bird's head is present. Below it, a sub-header reads "Supporting documentation". The main content area includes a breadcrumb trail: How to deposit data with the EIDC / Supporting documentation. A text block explains that supporting documentation helps users understand datasets and avoid misuse. Another text block outlines the required information for deposit. A link to a template is provided. A large callout box highlights the "What information is required?" section.

Nature and units of recorded values
Include information about the structure of the data (variable name(s), description, units).

- For tabular data (e.g., spreadsheets or databases) we recommend a structure such as the table below. If you have multiple files with different variables in each, use one table per file.
- For image data, you should include information such as resolution, focal length, etc.
- If applicable, include details of valid range of values for the variables, lowest level of detection, detection limits, etc.
- If there are missing data, state the value used to represent them (e.g., "Missing values are indicated by N/A")
- If available, include URLs for vocabulary terms

Table 1. Description of variables [insert the name]

Variable/Column header	Description	Units	Vocabulary URL [if applicable]
Dissolved nitrate	Nitrate measured in soil	mg/l	http://onto.nerc.ac.uk/CAST/16

Spatial coverage
[CONDITIONAL - Required if spatial data resources]
Include the following:

- name and bounding coordinates of the region OR for sampling locations provide the coordinates for each location OR, if those aren't possible provide the name of the general region/locations.
- Coordinate reference system and/or grid used e.g., WGS84, OSGB1936 / British National Grid
- If applicable, relevant geographical information e.g., aspect, elevation, surface area, volume etc.

Temporal coverage and resolution
[CONDITIONAL - Required if applicable]
Include the time period during which the data collection took place e.g., 01 January 2020 to 31 December 2020.

Include how often the measurements/sampling/experiment/monitoring took place e.g., annually, daily, hourly, every 5 minutes.

Resources

Elixir FAIR Cookbook

The screenshot shows the homepage of the Elixir FAIR Cookbook. At the top, there's a purple header bar with the text "FAIR COOKBOOK" on the left, a GitHub icon, and a search bar on the right. Below the header, there are four main sections: "Findability" (F), "Accessibility" (A), "Interoperability" (I), and "Reusability" (R). Each section has a brief description of exemplar recipes and a "LEARN MORE" button. Below these are two rows of smaller cards: "Infrastructure" (with a server icon), "Assessments" (with a checklist icon), "Applied Examples" (with a document icon), and "Maturity model" (with a bar chart icon).

ARDC

The Turing
way

Resources

Elixir FAIR Cookbook

ARDC

ARDC Resources for FAIR Data and FAIR Digital Research Objects

The ARDC offers a range of best-practice guides, tools and services for making data FAIR.

Besides research data, the FAIR Principles can be useful for other digital research objects. For example, the [FAIR Principles for Research Software \(FAIR4RS\)](#) were published in 2022 to improve the sharing and reuse of research software. We also offer various tools and guides that help make these digital objects FAIR.

Explore our resources for making data and other digital research objects FAIR:



[Tool / Service](#)
[FAIR Data Self-Assessment Tool](#)

Use our FAIR data self assessment tool to assess how FAIR your research dataset is...

[Explore >](#)



[Guide](#)
[FAIR Data Training Resources](#)

Free online FAIR data training resources by the ARDC and providers around the world ...

[Explore >](#)



[Guide](#)
[Metadata](#)

Enabling the discovery and reuse of research data

[Explore >](#)



[Guide](#)
[Community-Endorsed Data Standards](#)

Data standards endorsed by research communities help share data and meet specific needs.

[Explore >](#)

The Turing way



UK Centre for
Ecology & Hydrology

ceh.ac.uk

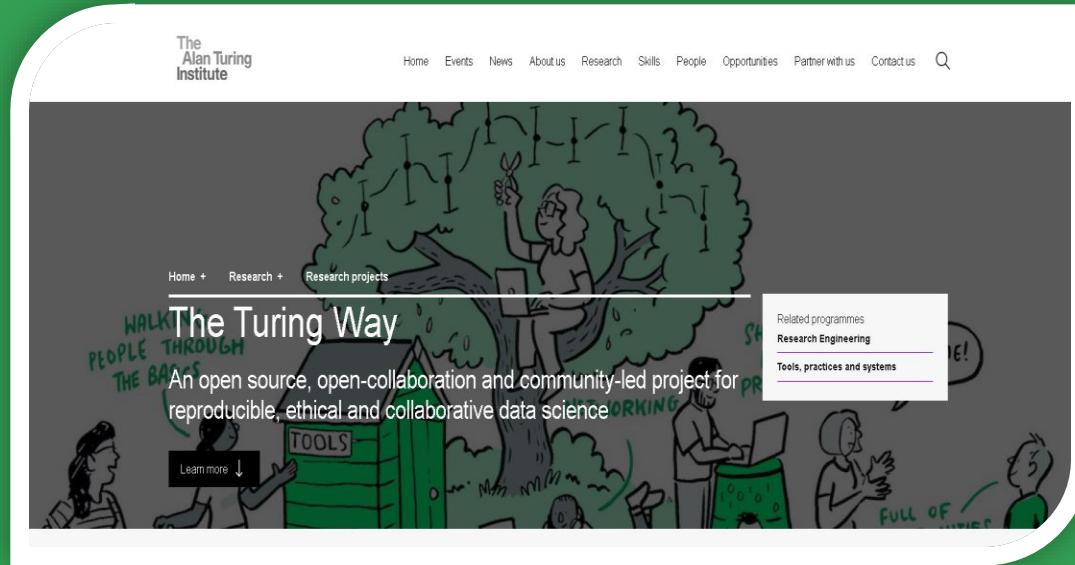
74

Resources

Elixir FAIR
Cookbook

ARDC

The Turing way



UK Centre for
Ecology & Hydrology

ceh.ac.uk

75

Any questions?



UK Centre for
Ecology & Hydrology



Thank You



UK Centre for
Ecology & Hydrology

