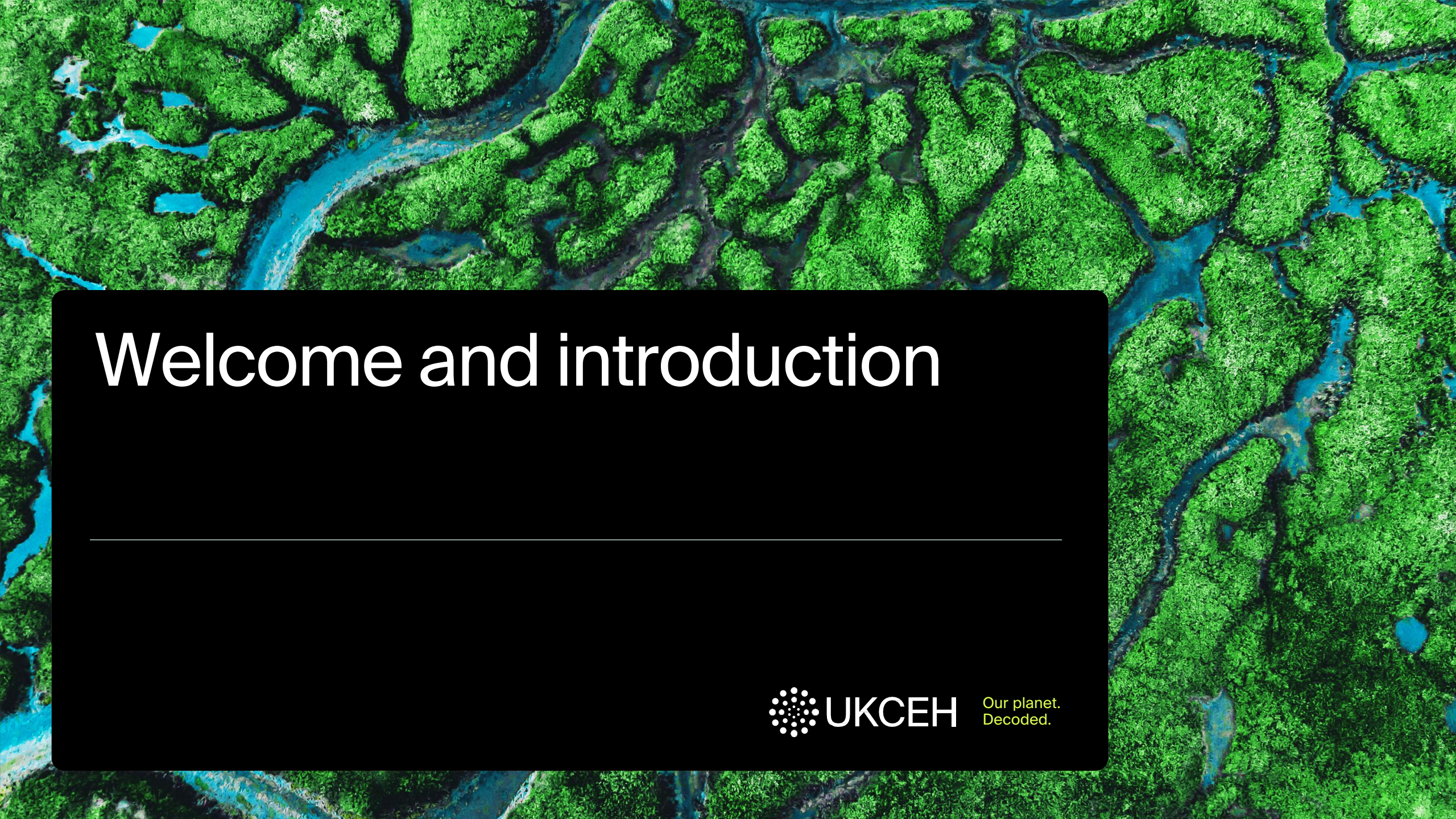


Creating FAIR Outputs

Data Stewardship Team
UK Centre for Ecology & Hydrology





Welcome and introduction

[Introduction]

Introductions



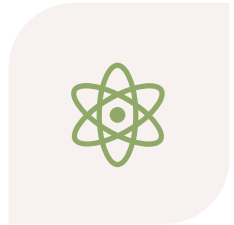
Els Dhiedt: Biogeochemist Data Scientist
at UK Centre for Ecology and Hydrology



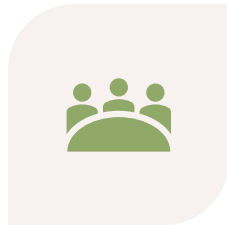
Jasmine Hunter: Data Steward at UK
Centre for Ecology and Hydrology

[Introduction]

Agenda



1. WHAT IS FAIR?



2. PREPARING FAIR
DATA



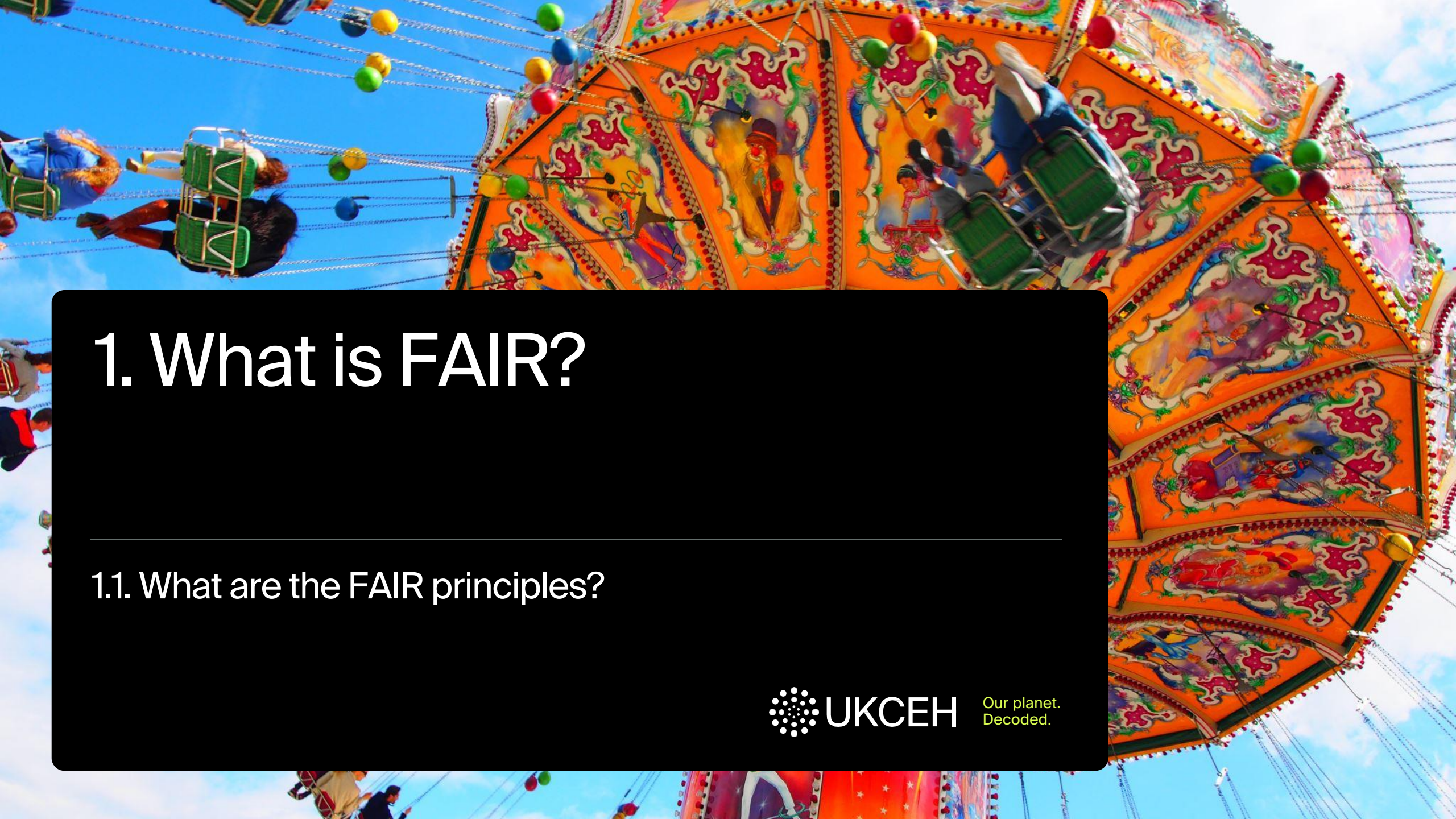
3. PUBLISHING FAIR
DATA



4. SUPPORTING
DOCUMENTATION
AND METADATA



5. TOOLS,
RESOURCES,
QUESTIONS



1. What is FAIR?

1.1. What are the FAIR principles?

What is...?

F

- ☐ Findable
- ☐ Feasible
- ☐ Financeable
- ☐ Functional
- ☐ Friendly

A

- ☐ Automated
- ☐ Aggregated
- ☐ Accessible
- ☐ Accurate
- ☐ Achievable

I

- ☐ Integrated
- ☐ Interoperable
- ☐ Interpreted
- ☐ Intelligent
- ☐ Innovative

R

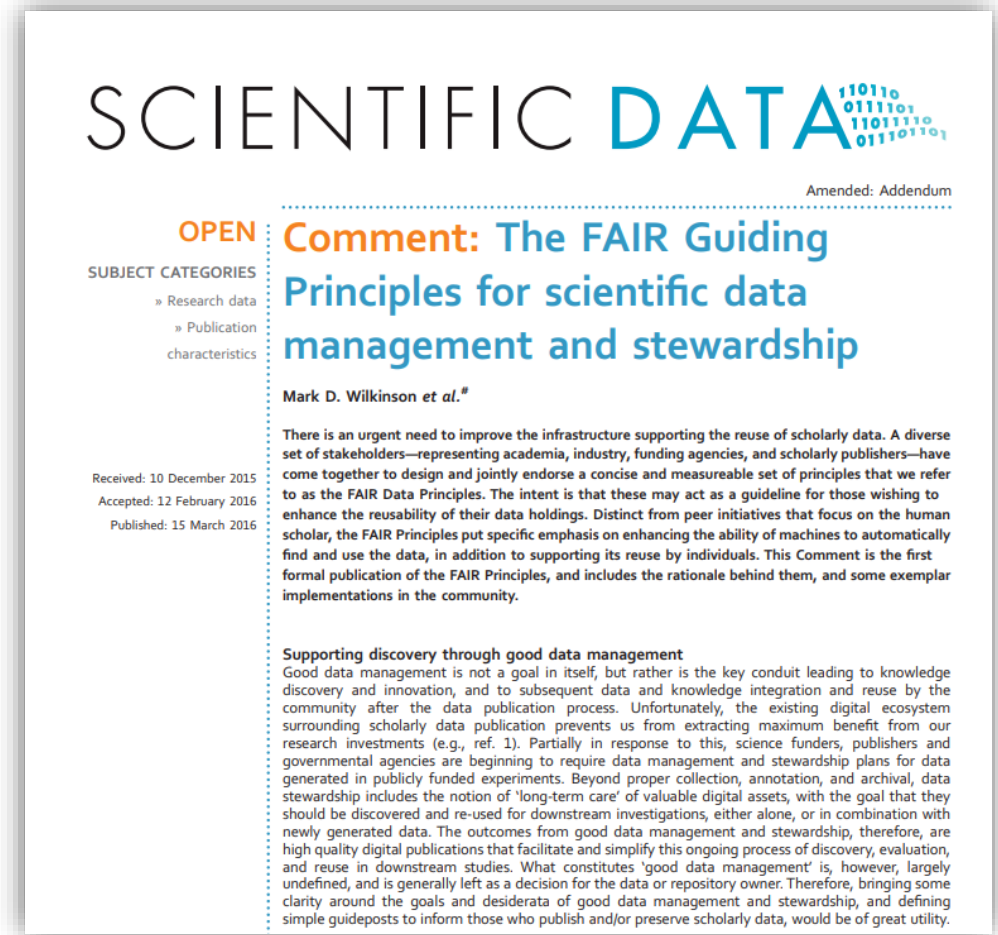
- ☐ Recyclable
- ☐ Reproducible
- ☐ Reliable
- ☐ Relevant
- ☐ Reusable

[What is FAIR?]

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



[What is FAIR?]

What are the FAIR principles?



Findable

Can the data be found easily by humans and machines?



[What is FAIR?]

What are the FAIR principles?



Accessible

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



[What is FAIR?]

What are the FAIR principles?



Interoperable

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



[What is FAIR?]

What are the FAIR principles?



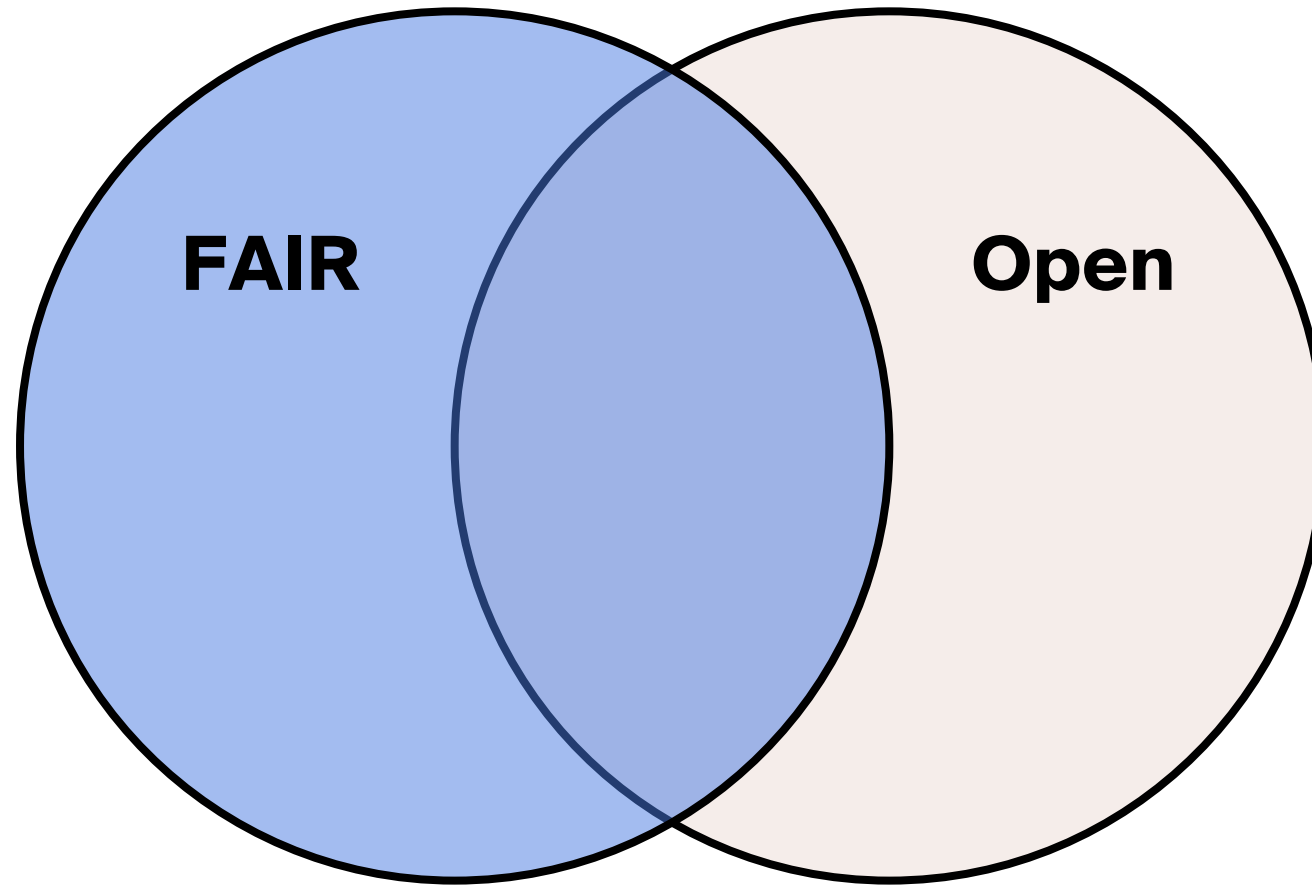
Reusable

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

Is the data provenance fully described in a way that the data can be easily replicated?

[What is FAIR?]

FAIR vs open



[What is FAIR?]

AI-ready data



AI, particularly, Large Language Models (LLMs), is transforming the way people interact with data



Data users at all levels can pose complex questions in natural language to chatbots



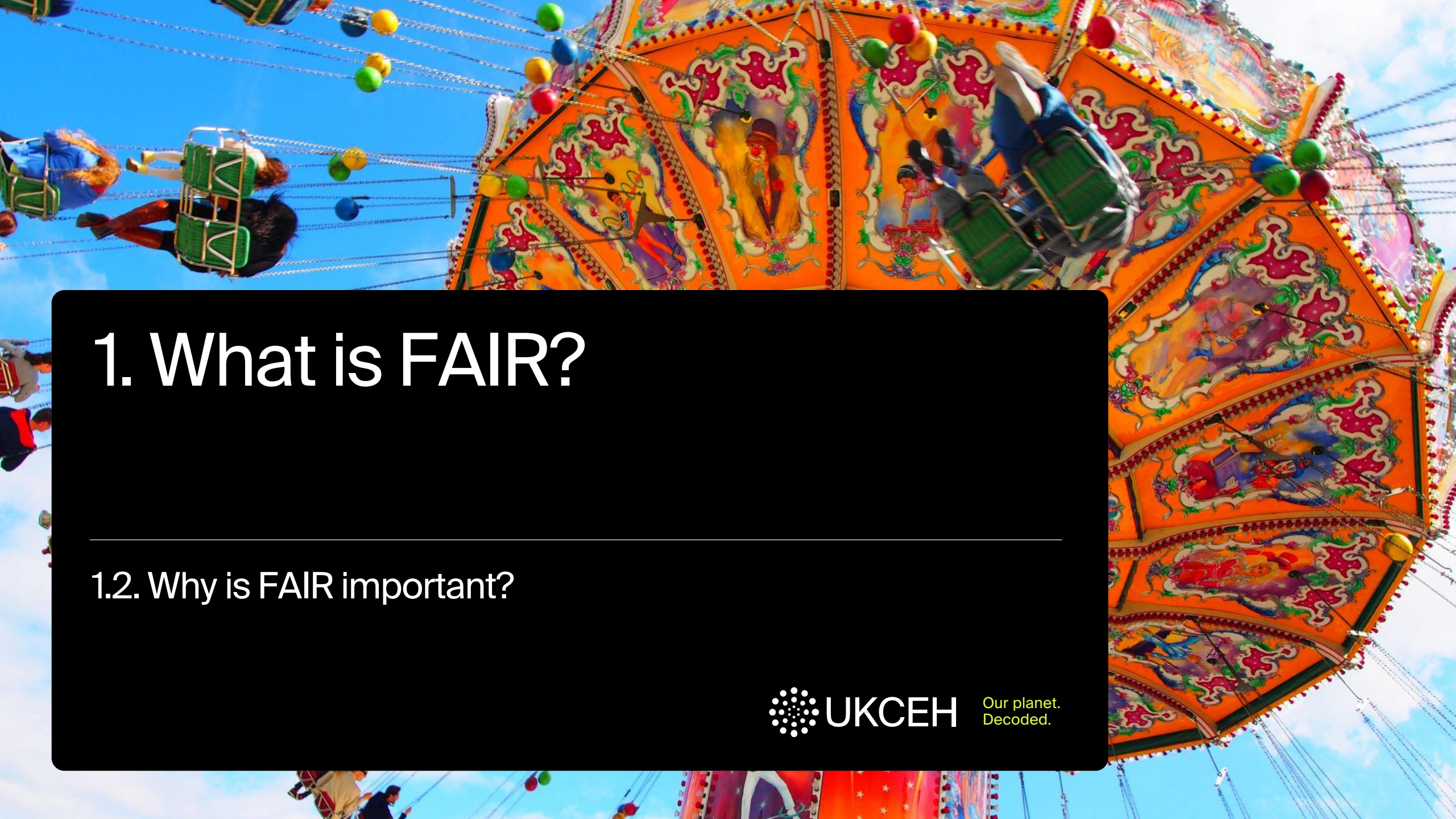
LLMs are only as good as the data they are trained on



Data being accessed and interpreted by AI systems must first be evaluated, validated, structured, governed, and shared in ways that support the responsible and effective use of AI



FAIR will (partly) get you there!



1. What is FAIR?

1.2. Why is FAIR important?

[What is FAIR?]

Obligations



[What is FAIR?]

Obligations



Funders

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



[What is FAIR?]

Obligations



Publishers

- Requirement of published data



[What is FAIR?]

Obligations



Legislation

- GDPR
- Wildlife and Countryside Act
- Freedom of Information Act

[What is FAIR?]

Benefits



[What is FAIR?]

Benefits



Data producer

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



[What is FAIR?]

Benefits



Data user

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



[What is FAIR?]

Benefits



Scientific community

- Building on existing science
- Inspiration



[What is FAIR?]

Benefits



Society

- Progress innovation
- Access to data
- Transparent science



2. Preparing FAIR data

2.1. How to achieve FAIR?

How to achieve FAIR?



[Preparing FAIR data]

How to achieve FAIR?



Findable

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



[Preparing FAIR data]

How to achieve FAIR?



Accessible

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



[Preparing FAIR data]

How to achieve FAIR?



Interoperable

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



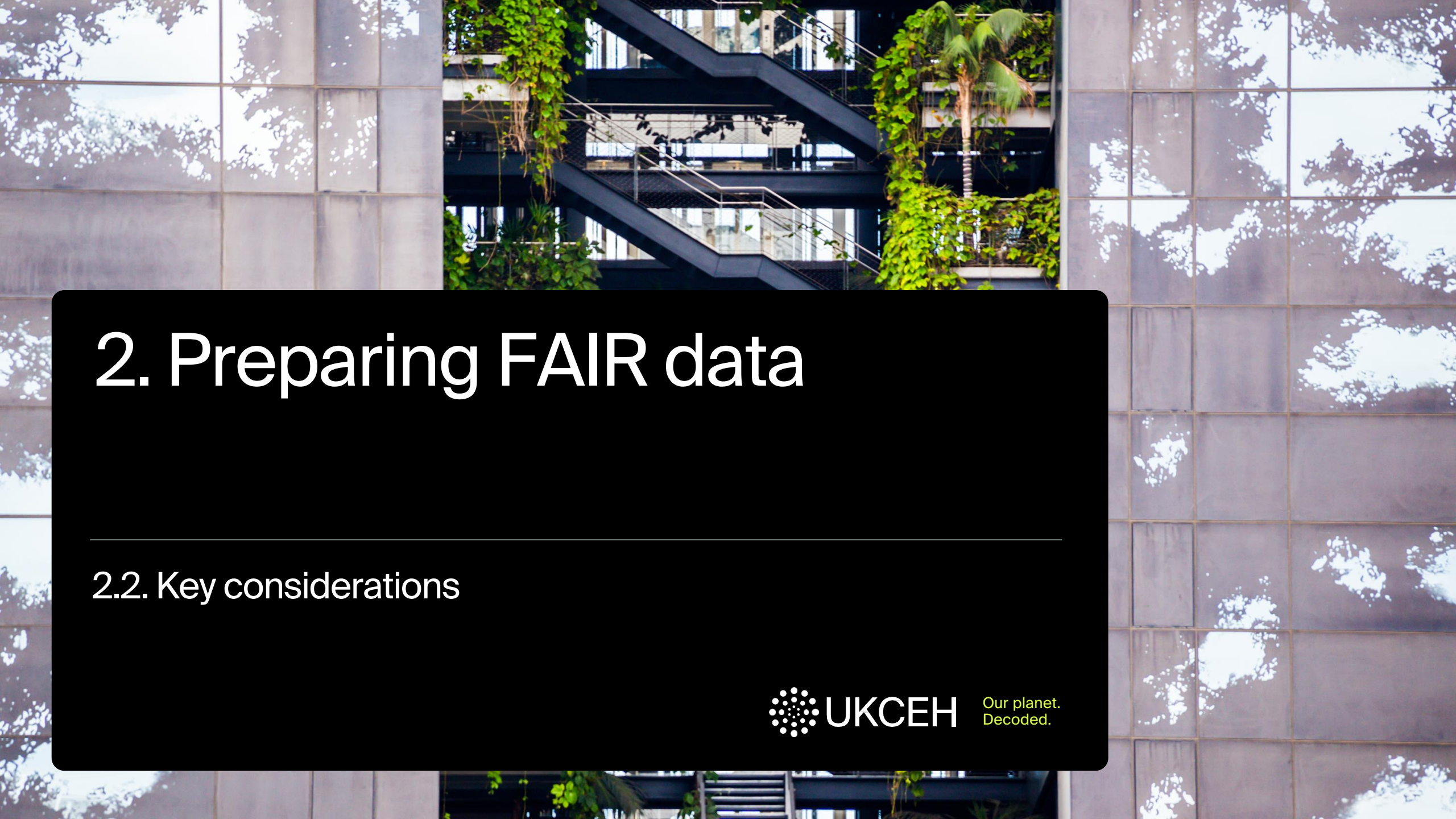
[Preparing FAIR data]

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

[Preparing FAIR Data]

Preparing FAIR data – Community standards

What are the community standards for data in your area?

Techniques and methodologies

- Clear descriptions of sampling methods

Control measurements

- Geochemical reference material and lab standards included in metadata

Uncertainties and error reporting

- Compulsory age model uncertainty ranges
- Spatial resolution included

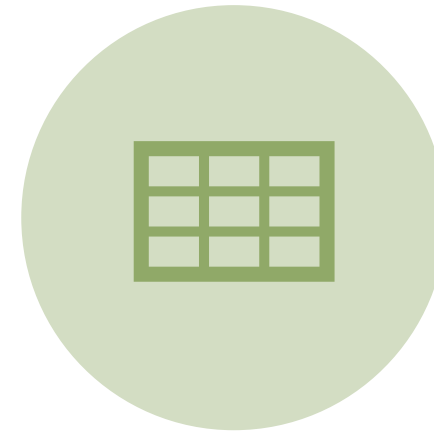
[Preparing FAIR Data]

Community standards

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



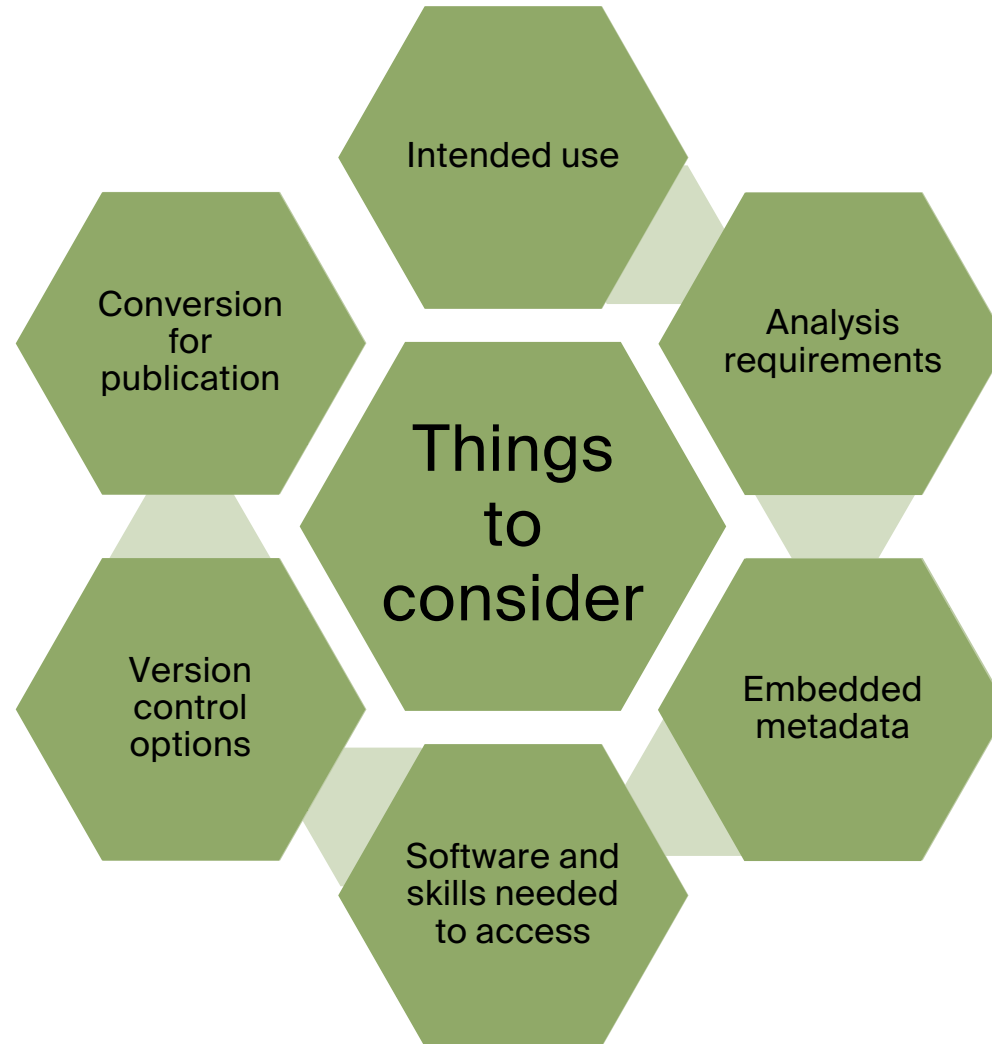
FAIRsharing repository of standards:
ideal for a detailed search. You can apply
filters such as 'subject' (e.g. Ecology).



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

[Preparing FAIR Data]

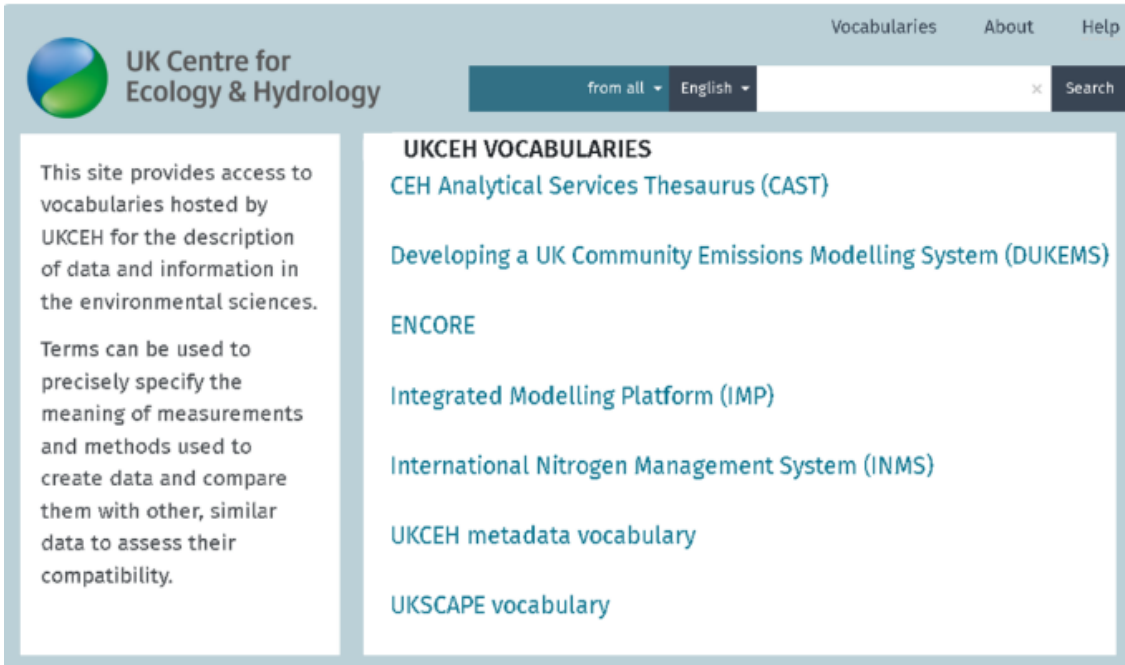
Data formats



[Preparing FAIR Data]

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.



UK Centre for Ecology & Hydrology

Vocabularies About Help

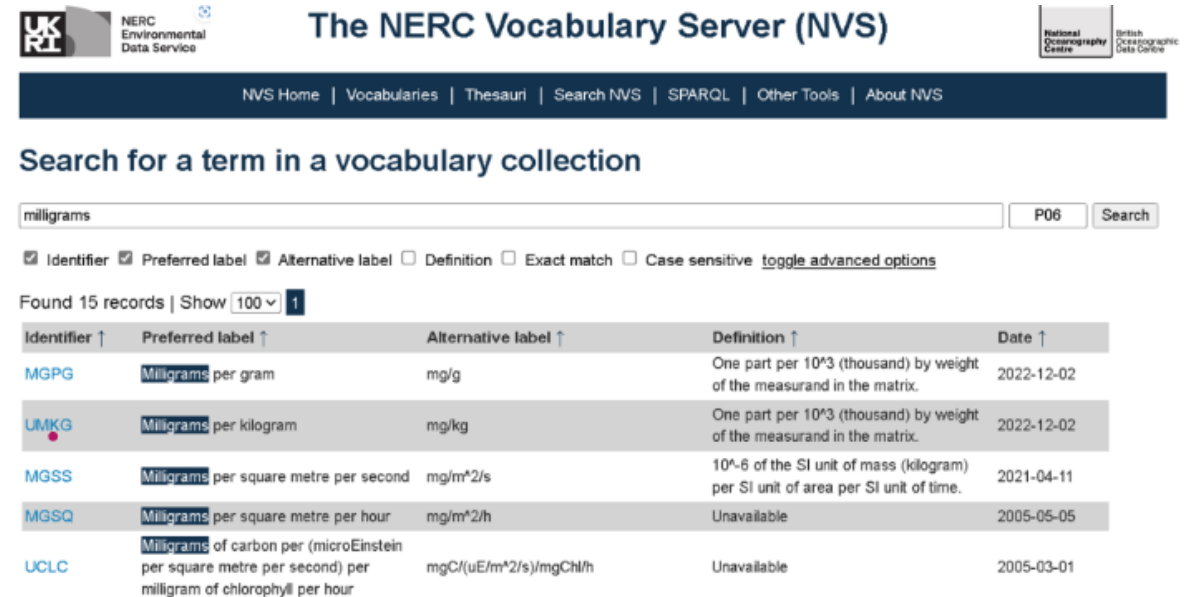
from all English Search

This site provides access to vocabularies hosted by UKCEH for the description of data and information in the environmental sciences.

Terms can be used to precisely specify the meaning of measurements and methods used to create data and compare them with other, similar data to assess their compatibility.

UKCEH VOCABULARIES

- CEH Analytical Services Thesaurus (CAST)
- Developing a UK Community Emissions Modelling System (DUKEMS)
- ENCORE
- Integrated Modelling Platform (IMP)
- International Nitrogen Management System (INMS)
- UKCEH metadata vocabulary
- UKSCAPE vocabulary



NERC Environmental Data Service

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

☒ Identifier ☒ Preferred label ☒ Alternative label ☐ Definition ☐ Exact match ☐ Case sensitive [toggle advanced options](#)

Found 15 records | Show 100 1

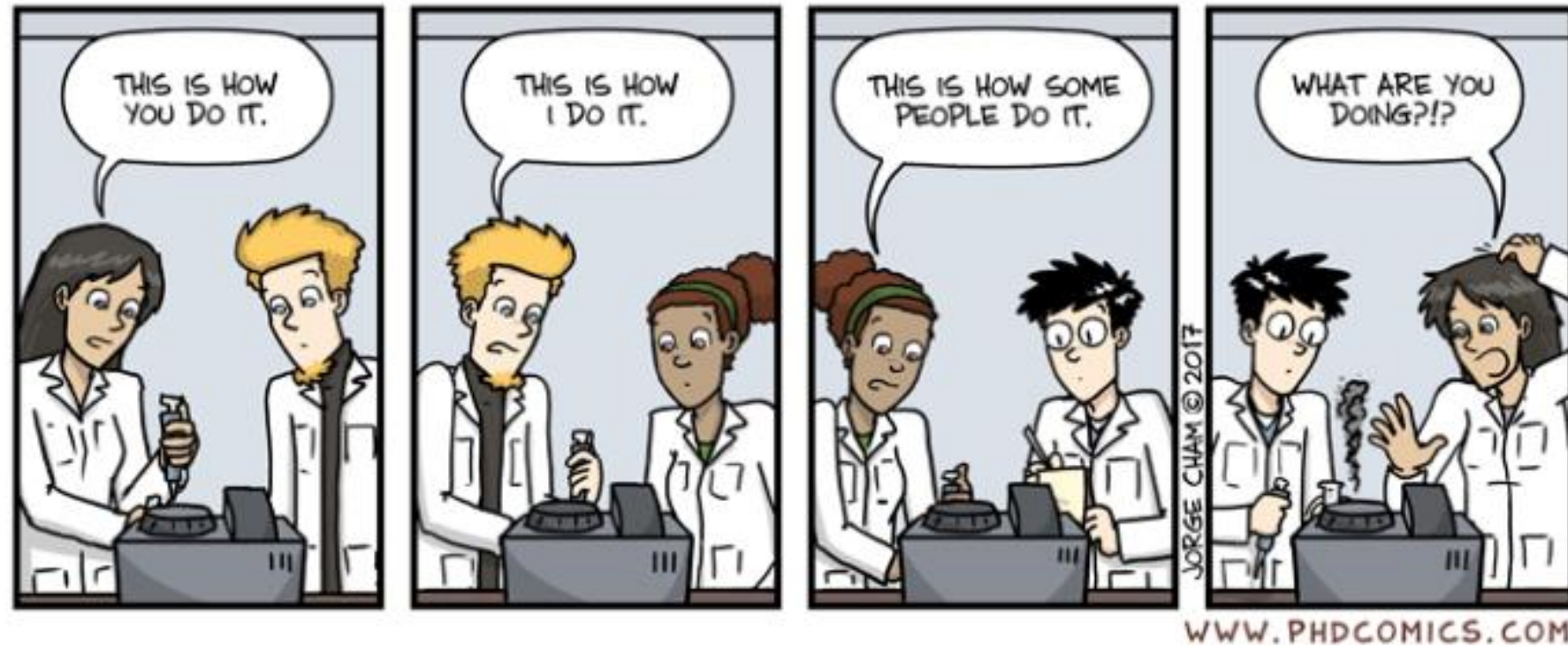
Identifier ↑	Preferred label ↑	Alternative label ↑	Definition ↑	Date ↑
MGPG	Milligrams per gram	mg/g	One part per 10 ³ (thousand) by weight of the measurand in the matrix.	2022-12-02
UMKG	Milligrams per kilogram	mg/kg	One part per 10 ³ (thousand) by weight of the measurand in the matrix.	2022-12-02
MGSS	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/(μE/m ² /s)/mgChl/h	Unavailable	2005-03-01

[Preparing FAIR Data]

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



The background of the slide is a light blue-grey surface covered with numerous small, light-colored wooden blocks. Each block has a black question mark printed on its top face. The blocks are scattered across the entire frame, some overlapping and others slightly offset, creating a textured, patterned effect.

Any Questions?

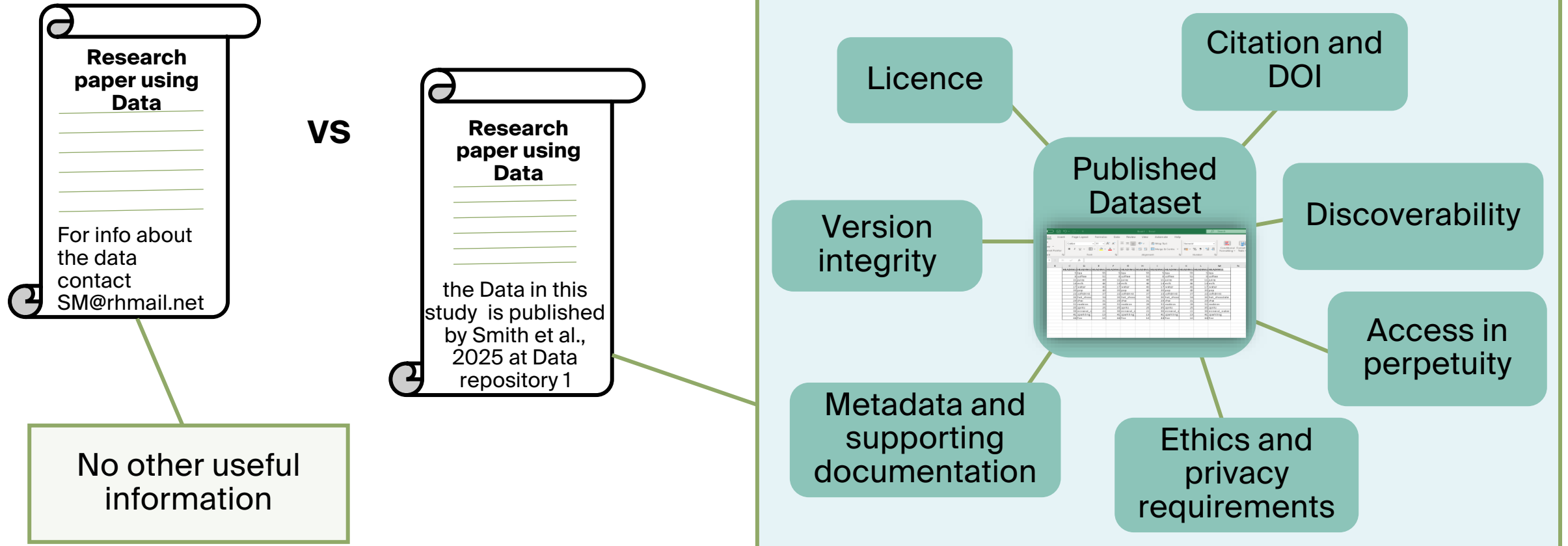


3. Publishing FAIR data

3.1. The data publication process

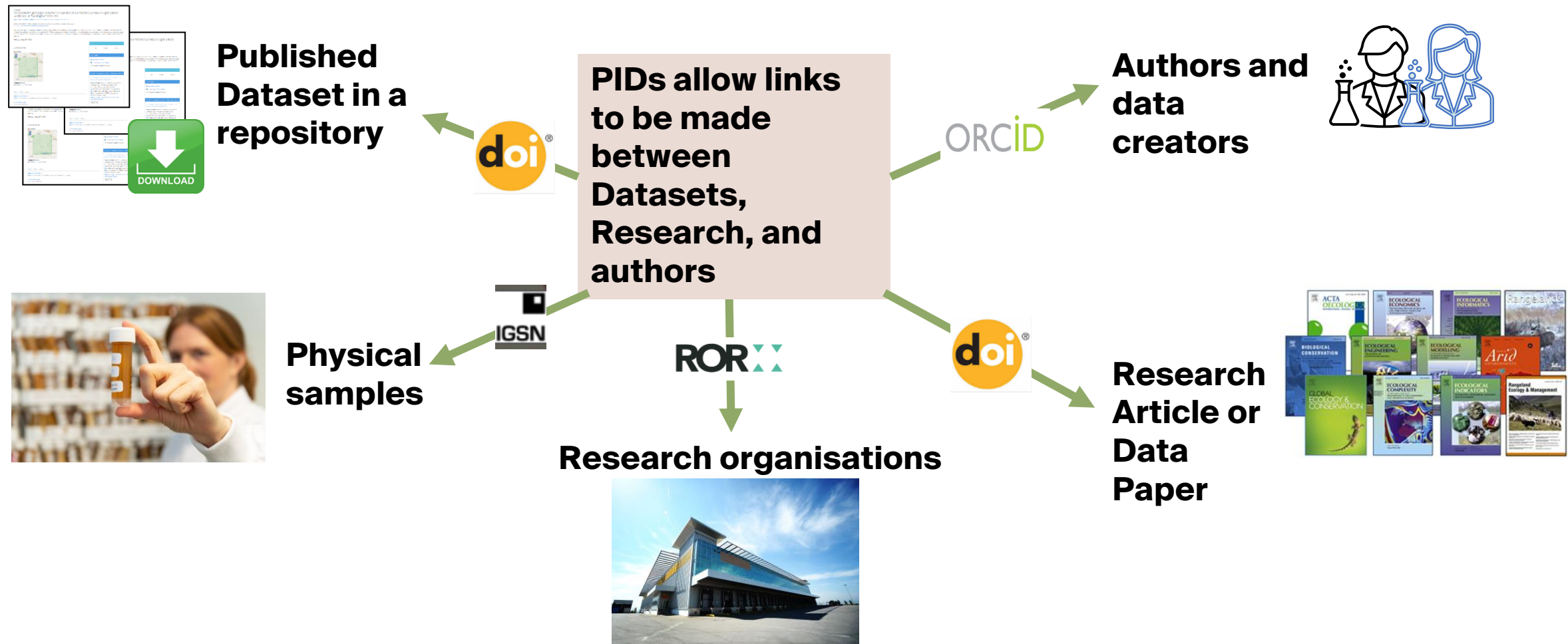
[Publishing FAIR data]

Published data - more than a paragraph in the supplementary!



[Publishing FAIR data]

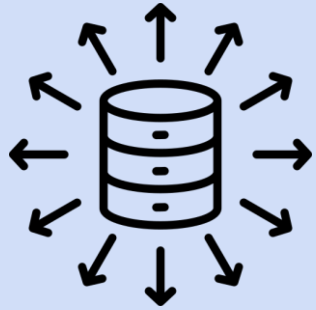
The Power of Persistent identifiers (PIDs)





3. Publishing FAIR data

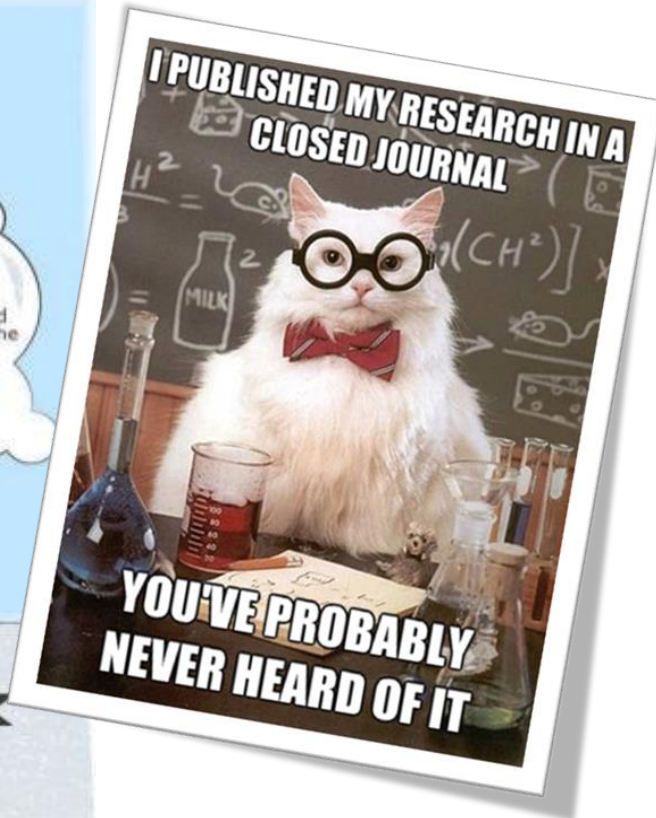
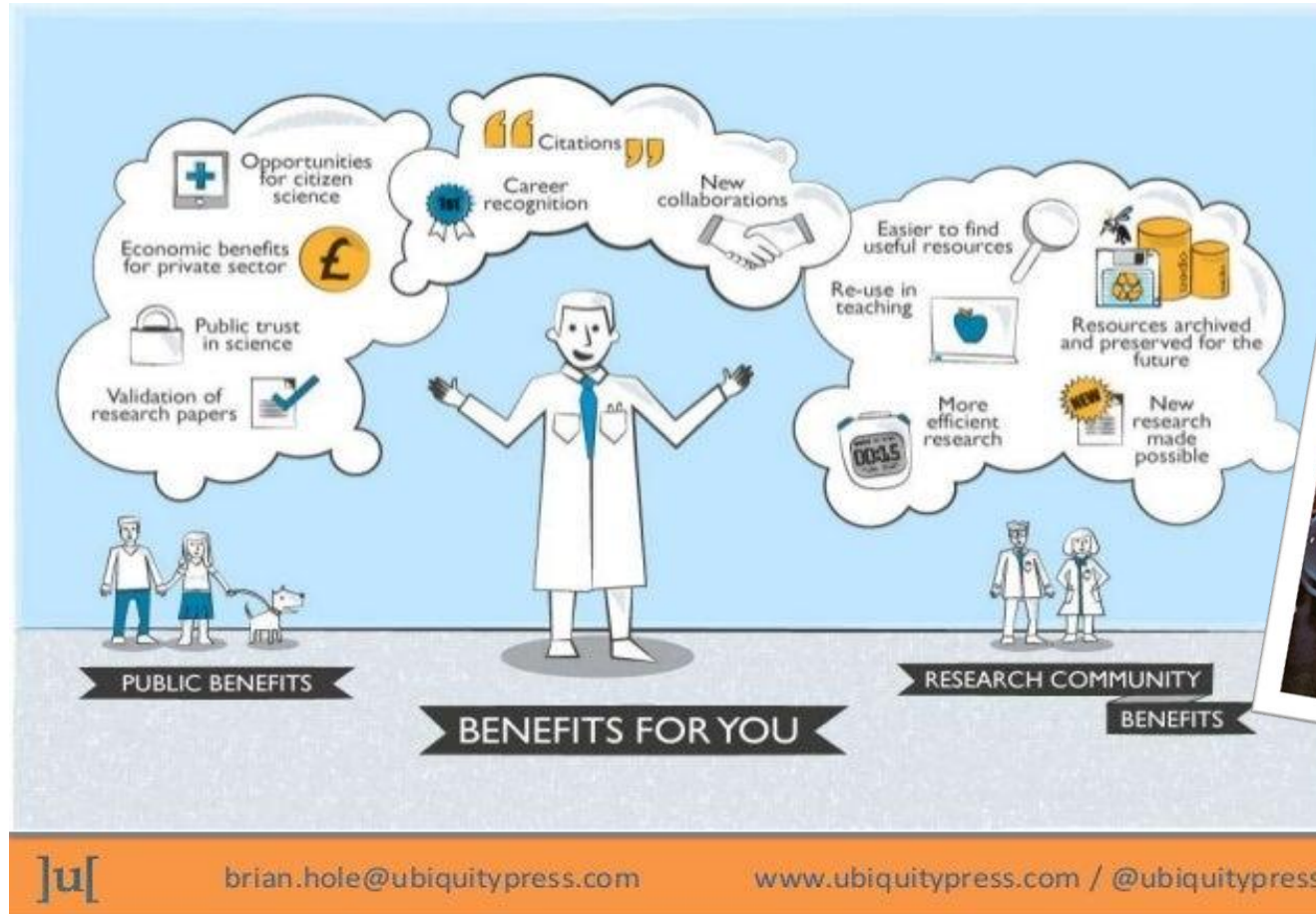
3.2. Why publish data?



What are the benefits of publishing data?

[Publishing FAIR data]

Why publish data – Open data = Open science!



[Publishing FAIR data]

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

1
citation

48
downloads *

1,051
views *

[More information](#)

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

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>

Lapworth, D.J., Ó Dochartaigh, B., Nair, T., O'Keeffe, J., Krishan, 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>

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

1
citation

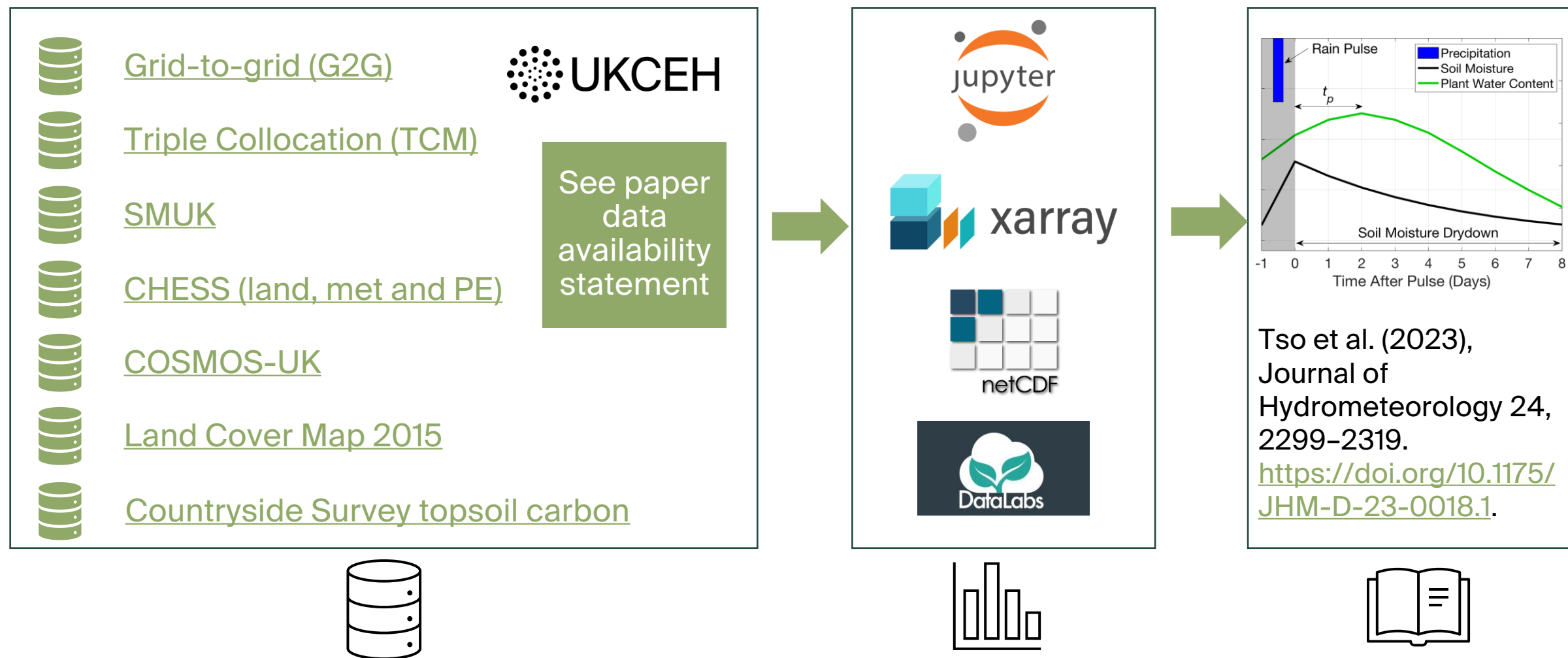
48
downloads *

1,051
views *

[More information](#)

[Publishing FAIR data]

Why publish data – so it can be reused!



[Publishing FAIR data]

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

British Ecological Society 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 reference list (see **submission** guidelines).

[Publishing FAIR data]

Case study: Plynlimon

Uplands of mid-Wales

Research ongoing since 1968

Series of spatial and time series datasets



[Publishing FAIR data]

Case study: Plynlimon – where are the data?

Some data held by the EIDC, earliest data is 2011

How do you access the earlier data?



Environmental
Information
Data Centre[Find data](#)[Deposit data](#)[Support](#)[News & events](#)[About](#)[Contact us](#)[Help](#)

Edit

More ▾

[^ Top](#)[Description](#)[Format](#)[Spatial information](#)[Temporal information](#)[Provenance & quality](#)[Licensing & constraints](#)[Related records](#)[Correspondence](#)[Authors](#)[Additional metadata](#)[Download/Access](#)

Robinson, M.; Kirchner, J.W.; Fry, M.; Cooper, H.

Sub-daily meteorology and river flow observations from a long-term upland catchment experiment in Plynlimon, Wales, 1968-2010

<https://doi.org/10.5285/a815235d-6ae1-4e5a-b5cd-138455607cd1>

Download ↓

Supporting docs 

Cite this dataset “”

[View record as ▾](#)84
downloads1,349
views[How do we collect this information?](#)

This dataset contains hourly automatic weather station (AWS) and 15 minute river flow gauging station data from sites in the Plynlimon (Pumlumon) area of upland mid-Wales, representing the headwaters of the Rivers Severn (Hafren) and Wye (Gwy). A range of instrumentation across 6 AWS sites and 10 gauging stations delivered meteorological and river flow data from 1968 to 2010 to explore changing hydrometeorology as part of the long-term Plynlimon catchment monitoring study.

The dataset includes: air temperature, temperature depression, Vapour Pressure Deficit (VPD), precipitation, solar radiation, net radiation, wind speed, wind direction, and river flow.

[Publishing FAIR data]

Case study: Plynlimon – the impact

1,349 catalogue views
84 downloads

Run a seminar series for the Royal
Geographical Society

FDRI reuse

Application for a small grant to reuse
and explore this dataset

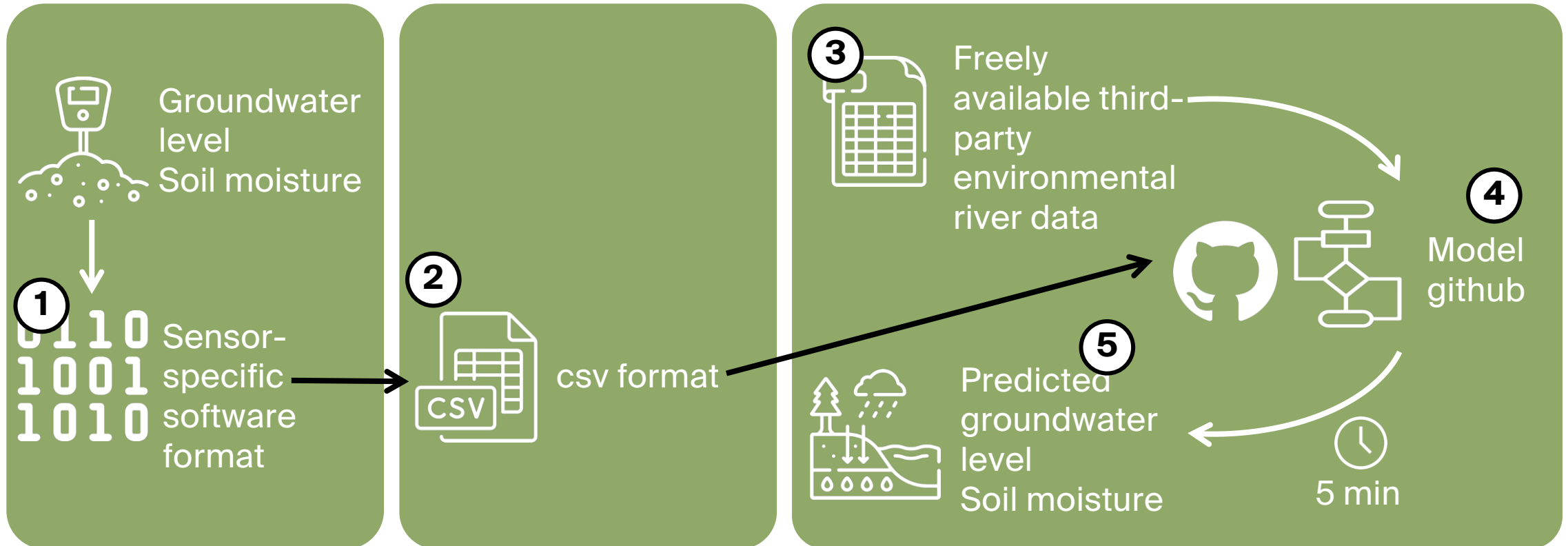
3. Publishing FAIR data

3.3. What to publish?

[Publishing FAIR data]

Data of long-term value

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



[Publishing FAIR data]

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?

[Publishing FAIR data]

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**



[Publishing FAIR data]

How to choose a repository

You can search for repositories
in your field via
[https://commons.datacite.org/
repositories](https://commons.datacite.org/repositories)



Funder data policy
requirements



Meets FAIR?



Repository certifications



Intended audience

[Publishing FAIR data]

NERC Environmental Data Service Repositories

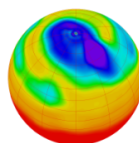
We provide a focal point for scientific data and information spanning all environmental science domains. We are a network of distributed data centres, with domain specific expertise:



British
Oceanographic
Data Centre

British Oceanographic Data Centre

Discipline: Marine



**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



UKCEH | **Environmental
Information
Data Centre**

Environmental Information Data Centre

Discipline: Terrestrial and Freshwater



NGDC

National Geoscience Data Centre

Discipline: Geoscience

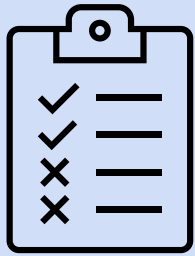


**British
Antarctic Survey**

UK Polar Data Centre

UK Polar Data Centre
Discipline: Polar and Cryosphere

4. Supporting documentation and metadata



Are these data practices FAIR or not?

[Supporting documentation and metadata]

What are they?



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



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 and metadata]

Why create supporting documentation and metadata?



[Supporting documentation and metadata]

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



[Supporting documentation and metadata]

Why create supporting documentation and metadata?



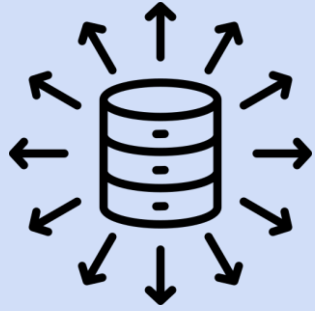
**Ensure proper
attribution of
research outputs**



**Helps comply with
data sharing
policies**



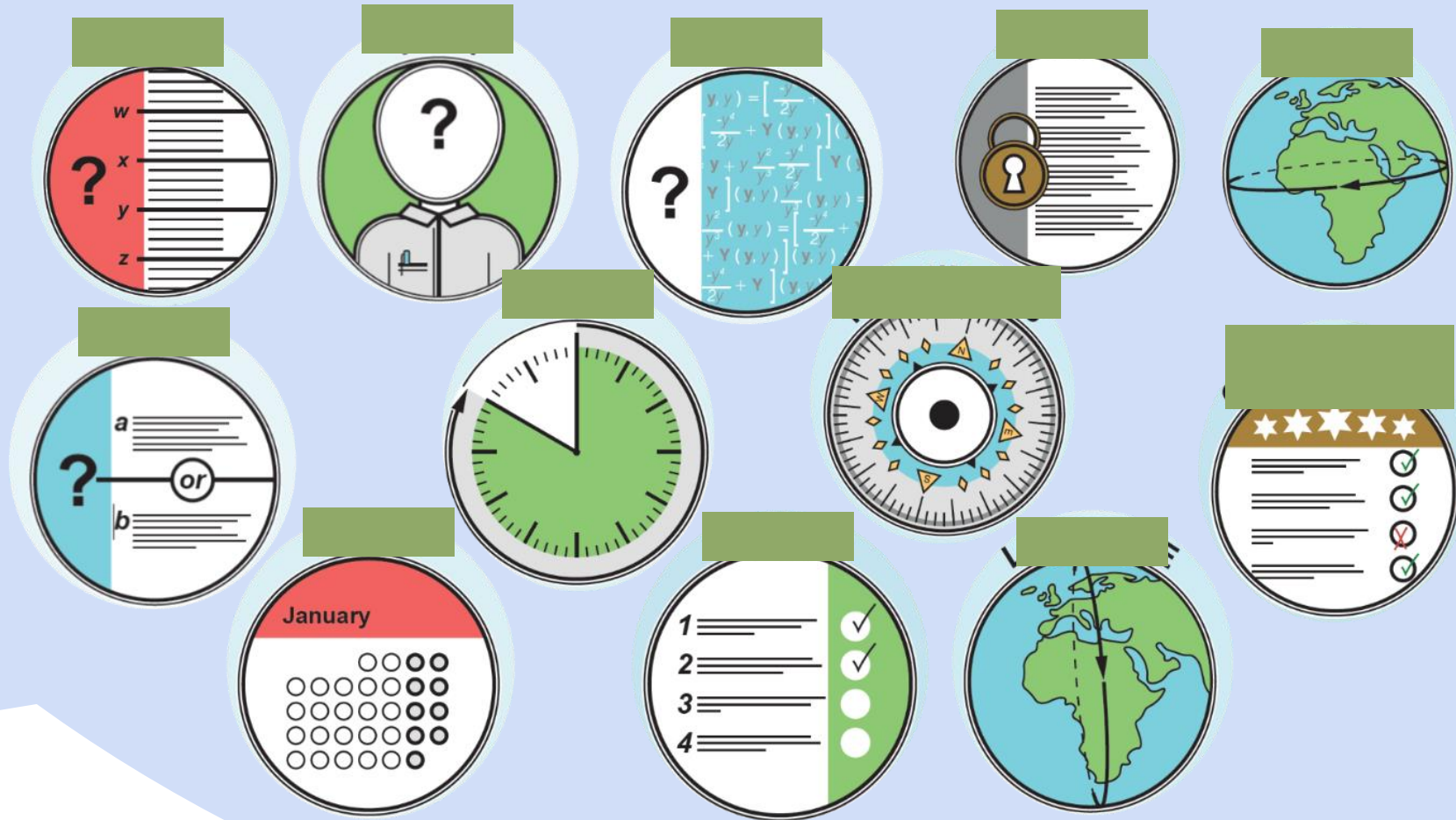
**Avoid
misinterpretation**



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?



[collection and analysis]

What information included 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

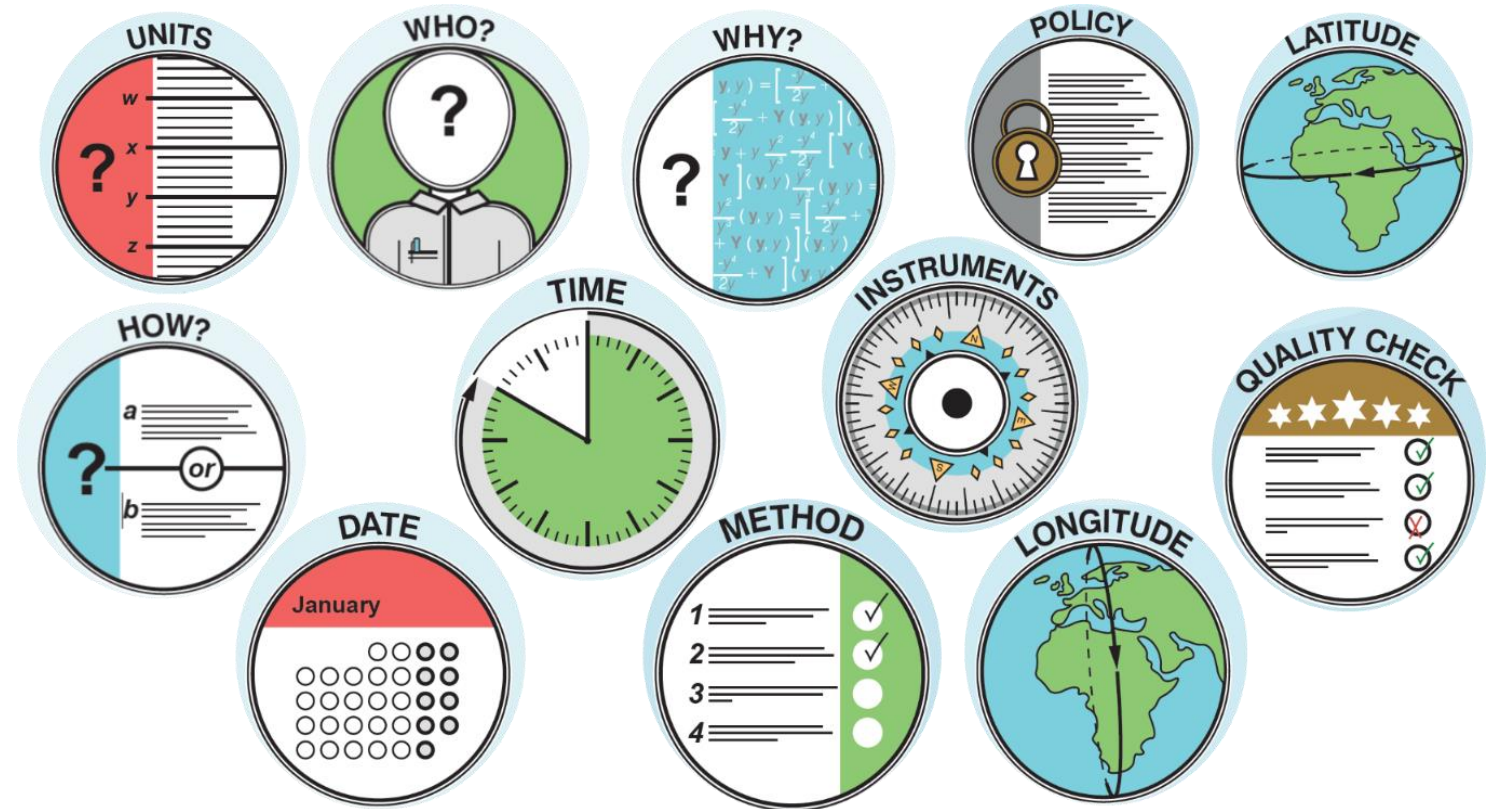


[collection and analysis]

What information included in a data publication: Metadata

Metadata – Data about the data

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

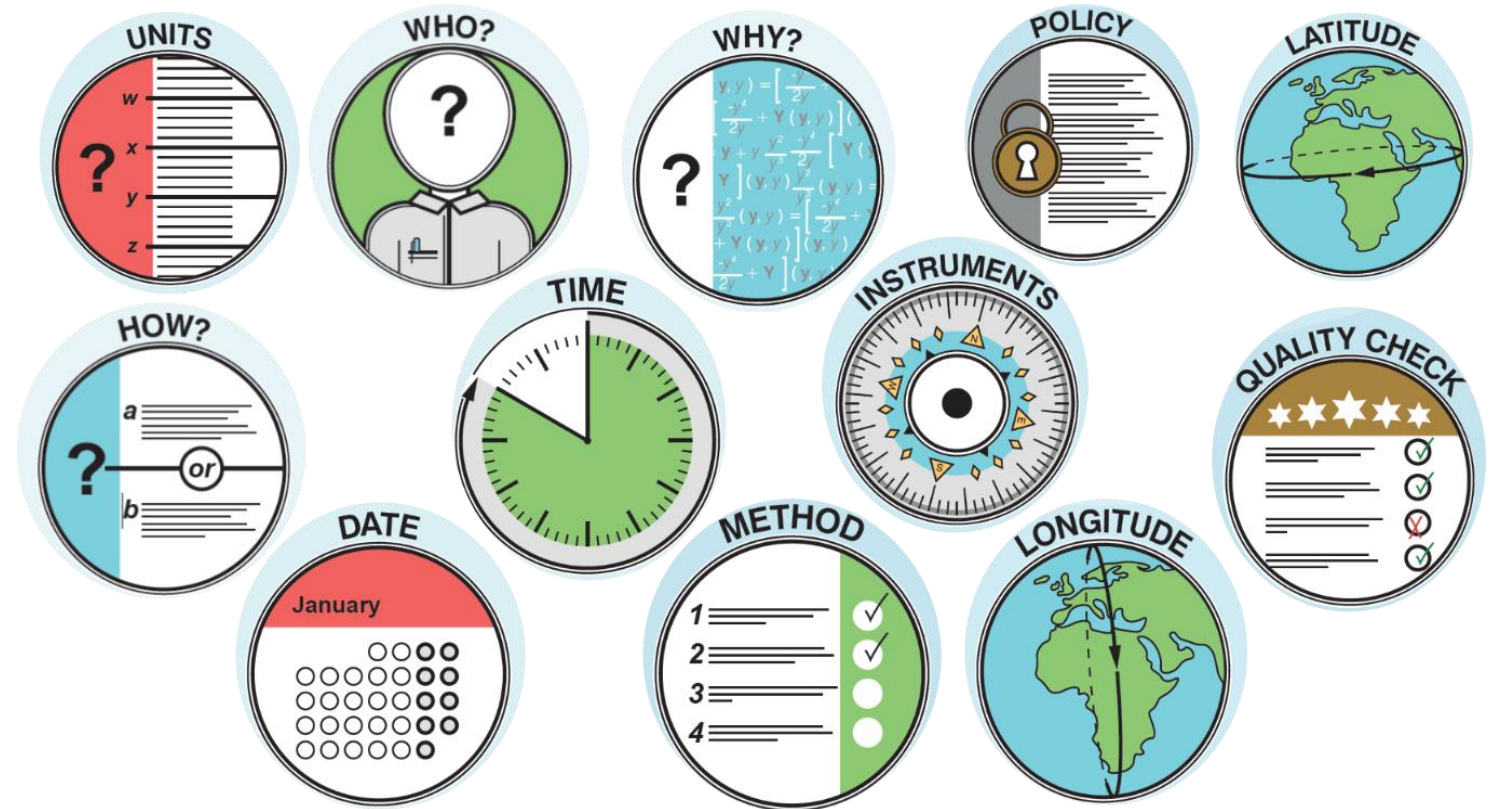


[collection and analysis]

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

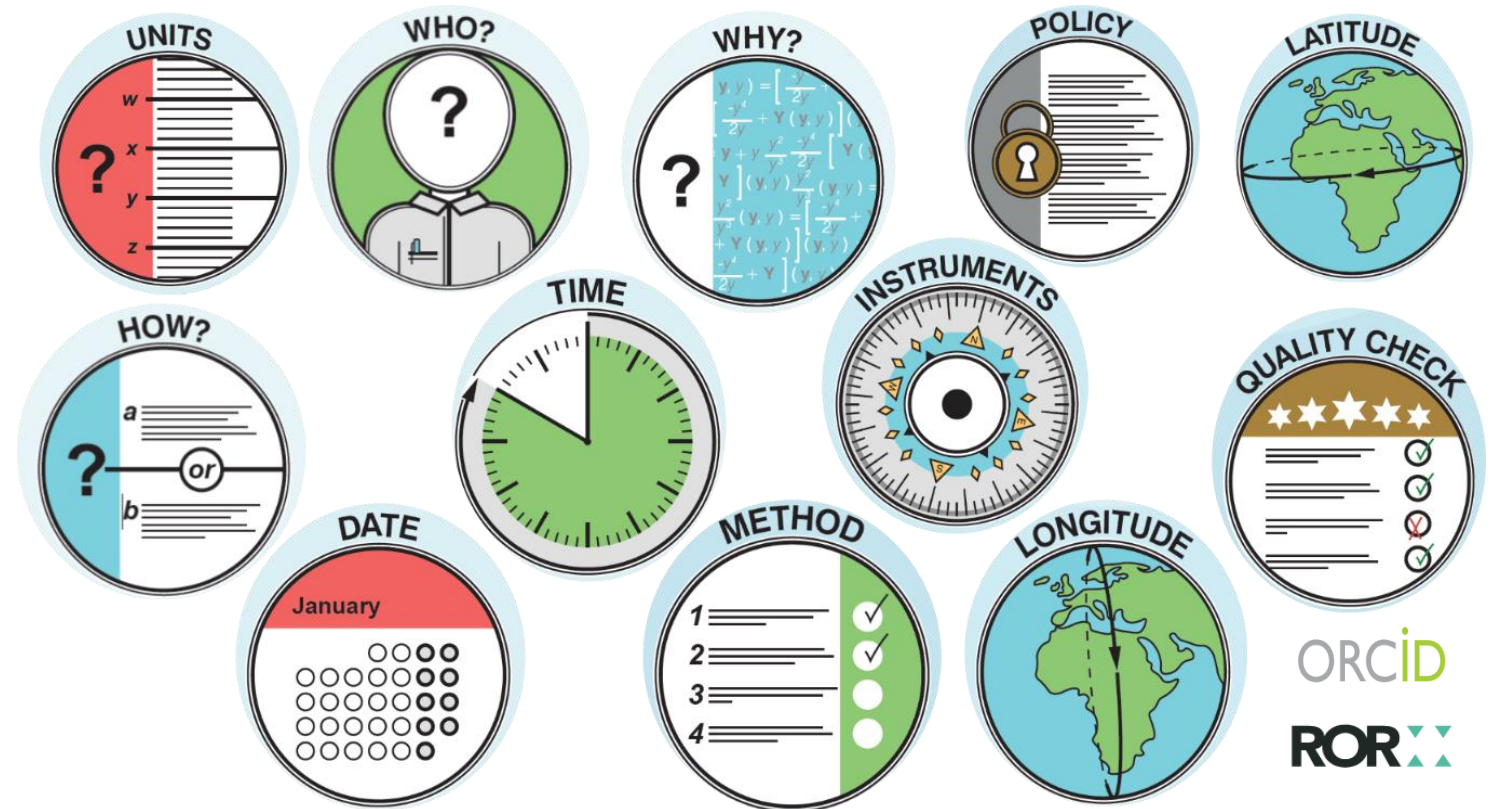


[collection and analysis]

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



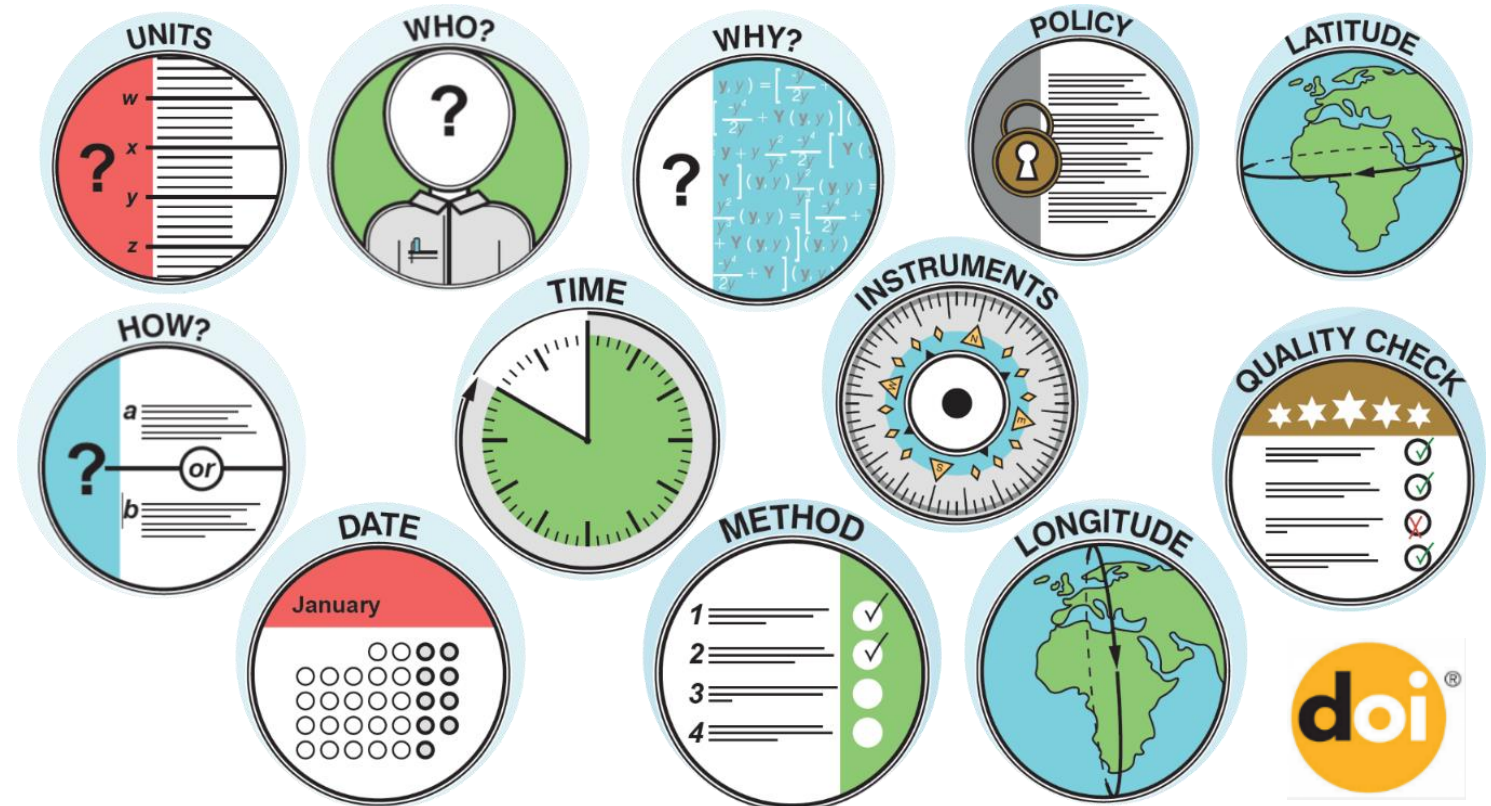
ORCID
ROR

[collection and analysis]

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



A top-down photograph of various gardening tools and items arranged on a light-colored wooden plank surface. On the left is a yellow watering can. Next to it is a small terracotta pot containing soil and several small green seedlings. To the right of the pot is a wooden-handled tool, possibly a trowel or dibble. Further right is a yellow plastic trowel. Below the trowel are several small, reddish-brown seeds scattered on the wood. To the right of the seeds is a pair of blue-handled pruning shears resting on a terracotta plate. Next to the shears is a green plastic plant marker. At the bottom right is a single red tomato.

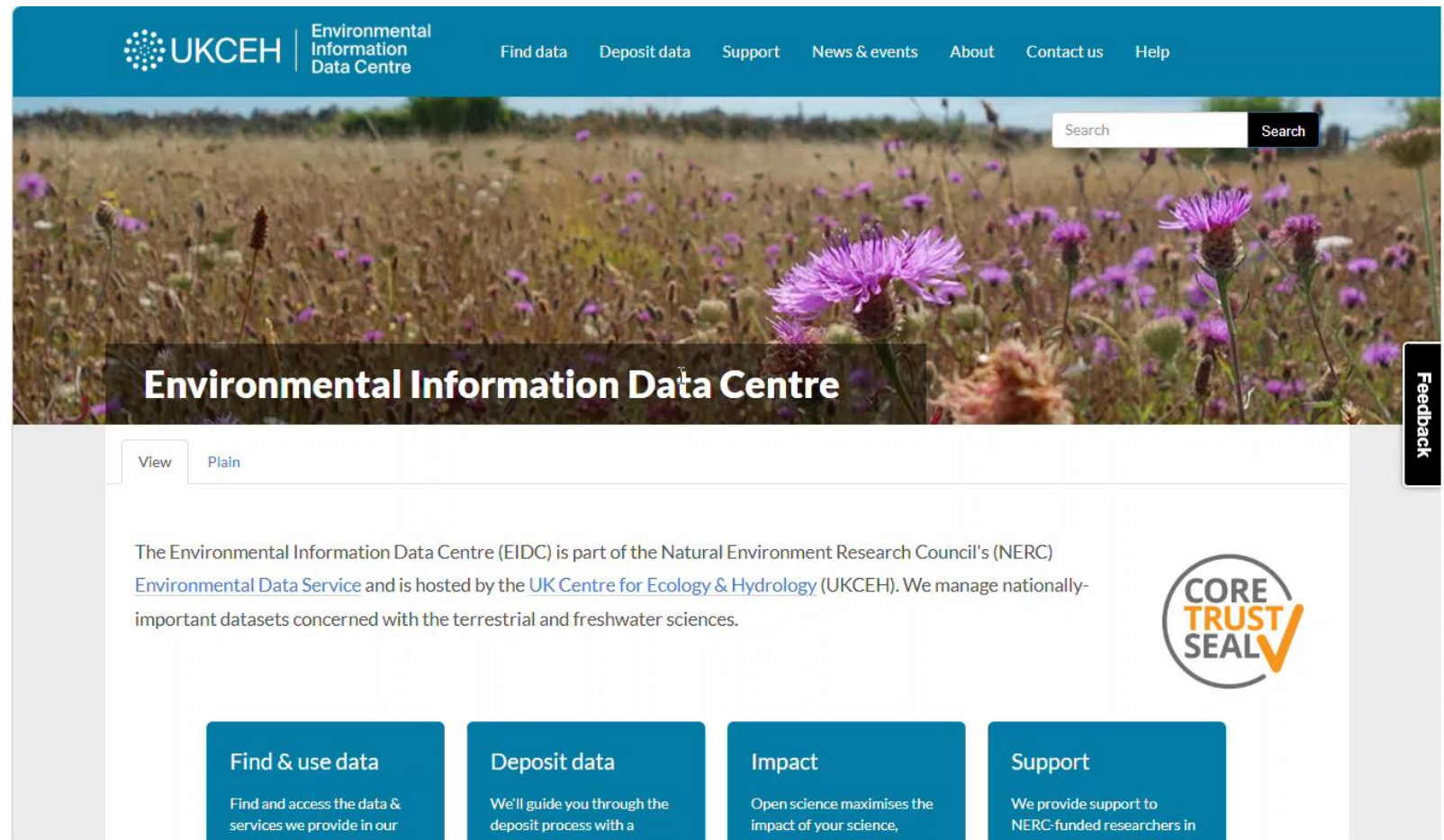
5. Tools and resources

[Tools and resources]

EIDC – Guidance and template

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

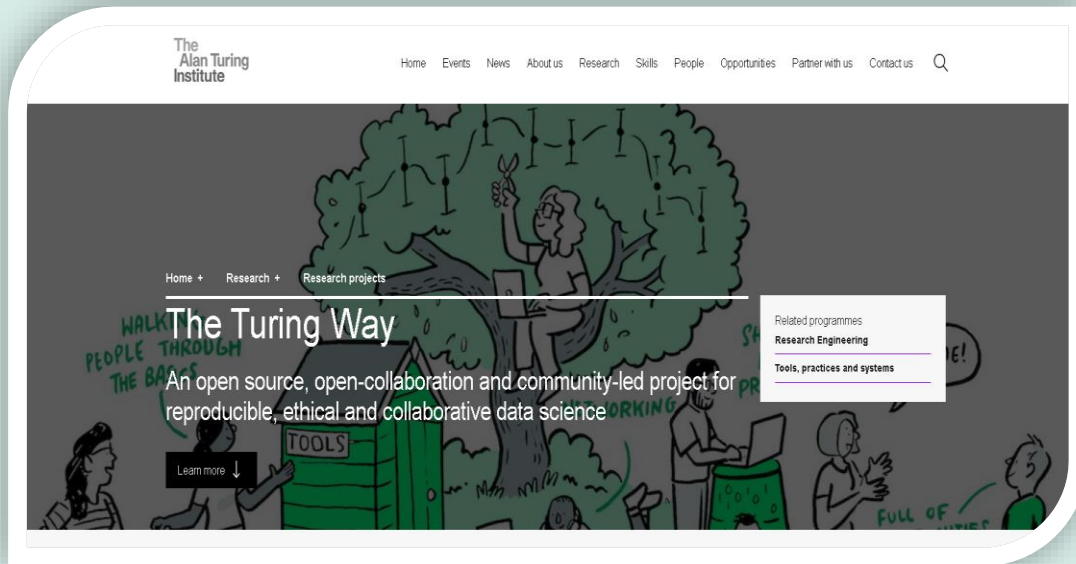
Plus a downloadable supporting document template!



[Tools and resources]

External resources

The Turing way



Elixir RDMkit

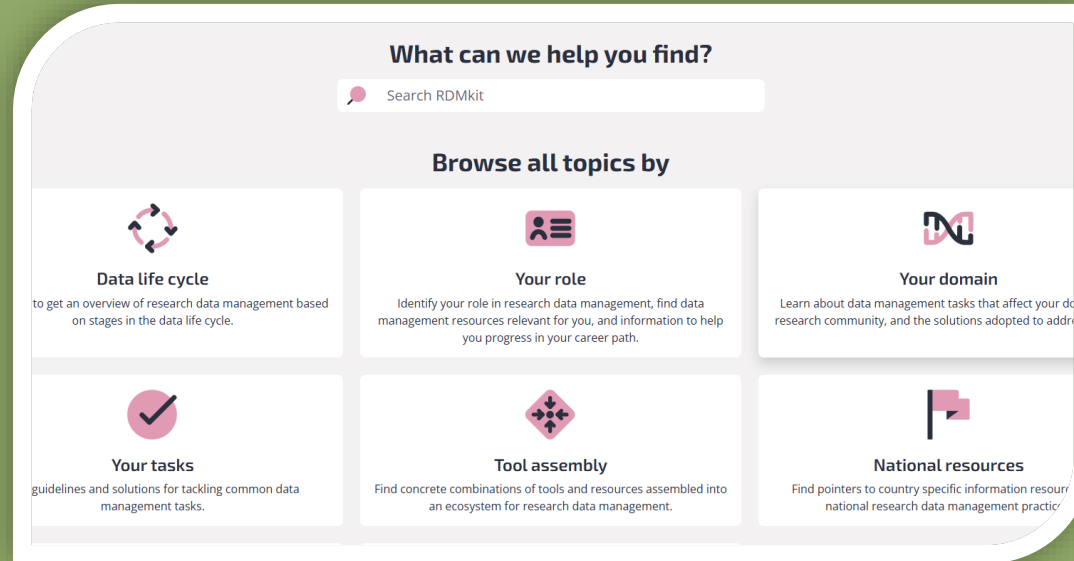
ARDC

[Tools and resources]

External resources

**The Turing
way**

Elixir RDMkit



ARDC

[Tools and resources]

External resources

The Turing
way

Elixir RDMkit

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:



FAIR Data Self-Assessment Tool

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

Explore >



FAIR Data Training Resources

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

Explore >



Metadata

Enabling the discovery and reuse of research data

Explore >



Community-Endorsed Data Standards

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

Explore >

A photograph of a person with grey hair, wearing a green jacket and yellow boots, walking through a rocky stream bed. A yellow bag and a blue bag are on the ground near the stream. The background is a grassy hillside.

Thank You! Any questions?

For more information
please contact:
datastewards@ceh.ac.uk

