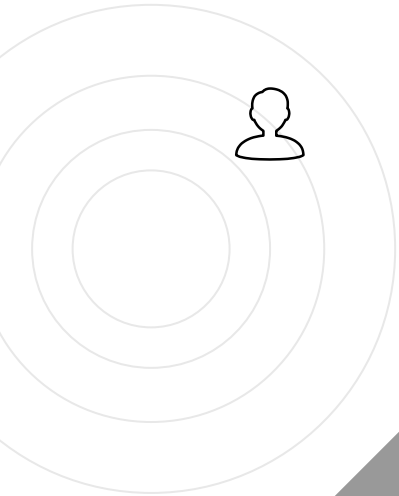


Plan ahead: practical tools to make your data and software more FAIR

NWO Life 2023 – May 24, 2023

[10.5281/zenodo.7962359](https://doi.org/10.5281/zenodo.7962359)





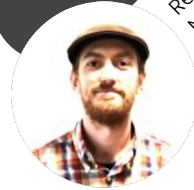
Kim Ferguson

Research Data Management Specialist
DANS
kim.ferguson@dans.knaw.nl



Mateusz Kuzak

Training Programme Lead
Netherlands eScience Center
m.kuzak@esciencecenter.nl



Jaro Camphuijsen

Research Software Engineer
Netherlands eScience Center
j.camphuijsen@esciencecenter.nl



1


Introduction


Findable, Accessible, Interoperable and
Reusable (**FAIR**), Data/Software
Management Plans (**DMP**)


Kim

Pitches + Break out sessions!



Findable 

Accessible 

Interoperable 

Reusable 

Findable

- Deposit your data in a data repository with metadata and a persistent identifier



F

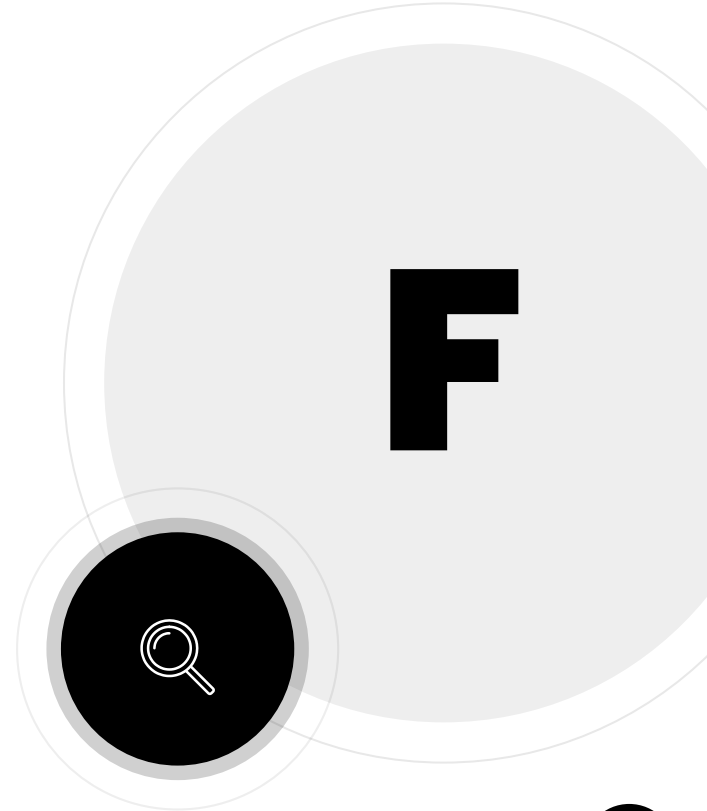


Findable (1)

- Deposit your data in a **data repository** with metadata and a persistent identifier

an online archive that curates research datasets and provides long-term access

- Finalised datasets
- ~10-15 years



Repositories



4TU.Centre for Research Data

European Genome-phenome Archive



[Recommended Repositories](#) (nature)
[Registry of Research Data Repositories](#)
[Fairsharing.org](#)

Findable (2)




- Deposit your data in a data repository with **metadata** and a persistent identifier

Metadata = information about data

- Contextual information
- Title, author, keywords
- When? For what purpose?
- Size? Standards?

- Discipline common metadata standards
 - [FAIRsharing.org](https://www.fairsharing.org/)
 - [Research Data Alliance metadata directory](https://www.researchdataalliance.org/metadata-directory/)
 - [Digital Curation Center](https://www.dcc.ac.uk/)

A large, bold, black letter 'F' is centered within a light gray circular area. The 'F' is composed of three horizontal bars and a vertical stem, all in a sans-serif font.

Original Article |  Open Access |  

Jekyll or Hyde? The genome (and more) of *Nesidiocoris tenuis*, a zoophytophagous predatory bug that is both a biological control agent and a pest

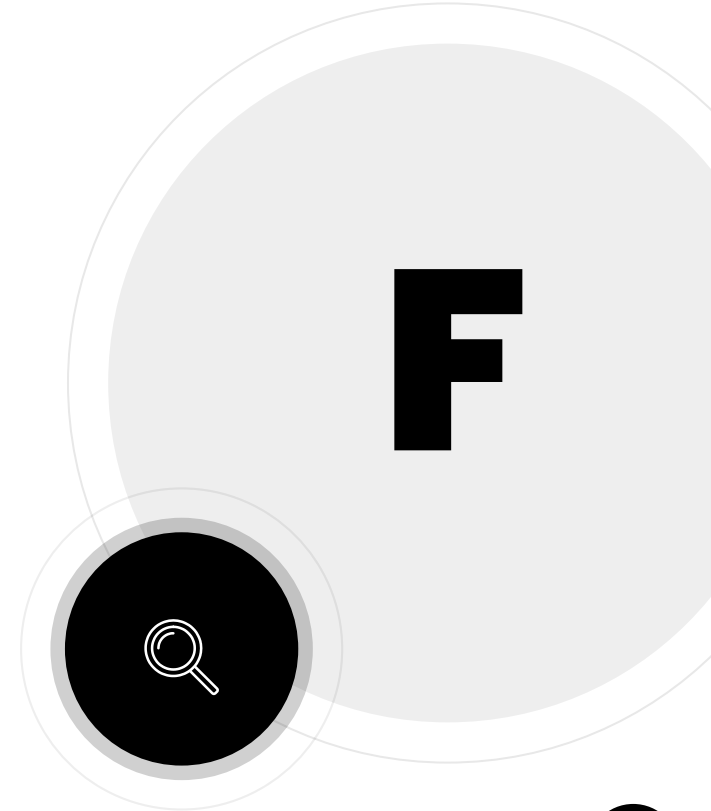
K. B. Ferguson , S. Visser, M. Dalíková, I. Provazníková, A. Urbaneja, M. Pérez-Hedo, F. Marec, J. H. Werren, B. J. Zwaan, B. A. Pannebakker, E. C. Verhulst

First published: 11 December 2020 | <https://doi.org/10.1111/imb.12688> | Citations: 5

- Deposit your data in a data repository with metadata and a **persistent identifier**

A persistent identifier is a long-lasting reference to a file, web page, or other object

ORCID



Accessible

- Consider what will be shared
- Obtain participant consent and perform risk management
- Determine access control
- Share your metadata

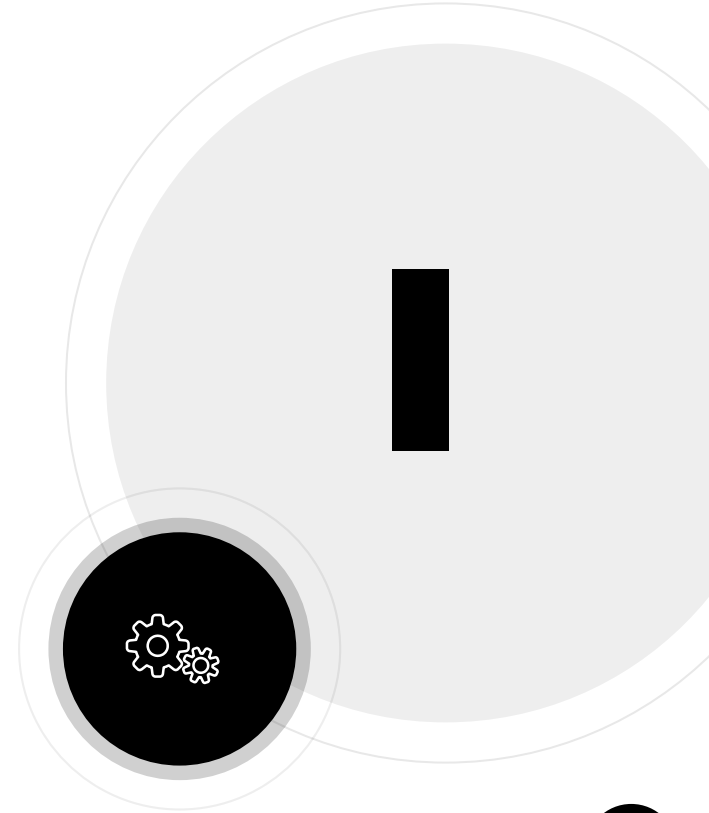
A





Interoperable

- Use open/common format
- Consistent vocabulary
- Discipline specific metadata standards



Reusable (1)

- Apply a licence to specify how others can re-use your data/code
- Documentation

A large, bold, black letter 'R' is centered within a large, light gray circular area. The 'R' is a simple, sans-serif font.

Reusable (2)

- Apply a **licence**
- Documentation

Data:

[Creative Commons](#) (overview)

[Creative Commons License Chooser](#)

Software:

[Choose a License](#)

[tl;dr Legal](#)

Licenses for data

Public Domain Dedication (CC0)
Attribution (CC BY)
Attribution-ShareAlike (CC BY-SA)
Attribution-NoDerivatives (CC BY-ND)
Attribution-NonCommercial (CC BY-NC)
Attribution-NonCommercial-ShareAlike (CC BY-NC-SA)
Attribution-NonCommercial-NoDerivatives (CC BY-NC-ND)
Restrictive licence

<https://data.4tu.nl/info/en/use/publish-cite/upload-your-data-in-our-data-repository/licencing>

Licences for software and code

Public Domain Dedication (CC0)
MIT License
BSD 3-Clause
Apache Licence 2
European Union Public License, version 1.2 (EUPL-1.2)
GNU General Public Licence version 2 (GPL-2.0)
GNU General Public Licence version 3 (GPL 3.0+)
GNU Lesser General Public License (LGPL-3.0)
GNU Affero General Public License (AGPL-3.0)



Learn more:
[The Turing Way](#)

Findable 

Accessible 

Interoperable 

Reusable 

≠

Standard

Open

**Intrinsic
quality**

Learn more:
[The Turing Way](#)

DMP

Data Management Plan

- **Tool**
 - structure the management of your data/software
 - Increase quality of documentation
 - Prevents loss of data
 - No unfindable files through thanks to [file naming conventions](#)
- **How can I make my data/software FAIR?**
- **Data/software sharing**
 - **Validation**
 - **Increased impact ([Citation](#) and collaborations)**

2

Pitches

1. FAIR software management - Jaro Camphuijsen
2. FAIR-Aware tool - Kim Ferguson



FAIR for Research Software

&

Software Management Plans

Focus on Findability and Accessibility

F: Easy for both humans and machines to find.

- example: software published with the DOI

A: Retrievable via standardised protocols.

- example: can be download from github

Article | [Open Access](#) | [Published: 14 October 2022](#)

Introducing the FAIR Principles for research software

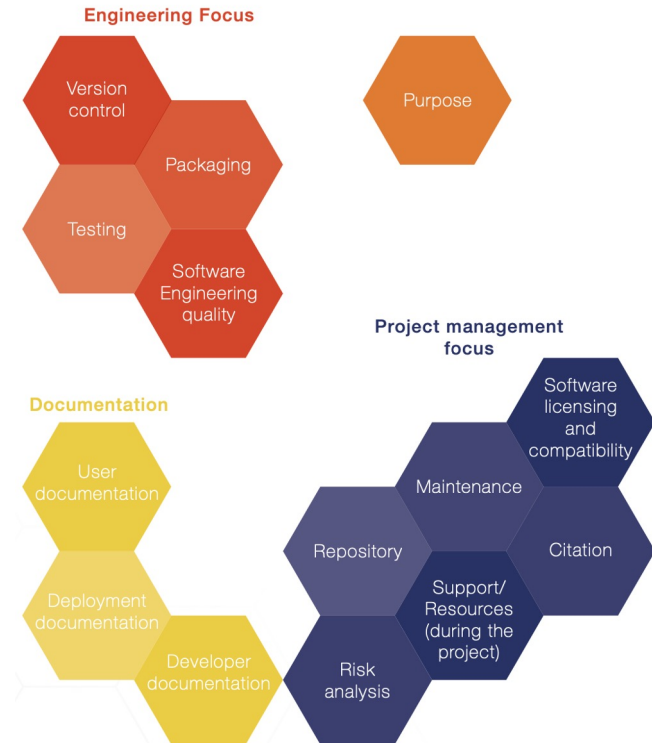
[Michelle Barker](#) , [Neil P. Chue Hong](#), [Daniel S. Katz](#), [Anna-Lena Lamprecht](#), [Carlos Martinez-Ortiz](#), [Fotis Psomopoulos](#), [Jennifer Harrow](#), [Leyla Jael Castro](#), [Morane Gruenpeter](#), [Paula Andrea Martinez](#) & [Tom Honeyman](#)

[Scientific Data](#) **9**, Article number: 622 (2022) | [Cite this article](#)

10k Accesses | **4** Citations | **241** Altmetric | [Metrics](#)

doi.org/10.1038/s41597-022-01710-x

Practical guide to Software Management Plans is an Initiative by NWO and eScience Center.



doi.org/10.5281/zenodo.7038280

- Your first step towards your FAIR data(set).
- What FAIR practices will help you create more FAIR data?
- 10 questions + tips and tricks
- **In this workshop:** working with biological data and elements to keep in mind

fairaware.dans.knaw.nl

FAIR questions

FINDABLE

1. Are you aware that a data(set) should be assigned a globally unique persistent and resolvable identifier when deposited with a data repository? Yes No

2. Are you aware that when you deposit a data(set) in a data repository, you will need to provide discovery metadata in order to make the data(set) findable, understandable and reusable to others? Yes No

3. Are you aware that the data repository providing access to your data(set) should make the metadata describing your data(set) available in a format readable by machines as well as humans? Yes No

ACCESSIBLE

4. Are you aware that access to your data(set) may need to be controlled and that metadata should include licence information under which the data(set) can be reused? Yes No

5. Are you aware that metadata should remain available over time, even if the data(set) is no longer accessible? Yes No

INTEROPERABLE

6. Are you aware that the metadata describing your data(set) should use controlled vocabularies? Yes No

1. Are you aware that a data(set) should be assigned a globally unique persistent and resolvable identifier when deposited with a data repository?

What does this mean?

A **persistent identifier** is a long-lasting reference to a resource. The **data(set)** you deposit in a **data repository** should be assigned a globally unique, persistent and resolvable identifier (PID) so that both humans and machines can find it. Persistent identifiers are maintained and governed so that they remain stable and direct the users to the same relevant object consistently over time. Examples of PIDs include Digital Object Identifier (DOI), Handle, and Archival Resource Key (ARK).

Why is this important?

If your data(set) or metadata does not have a PID, you run the risk of "link rot" (also known as "link death"). When your data(set) or metadata is moved, updated to a new version, or deleted, older hyperlinks will no longer refer to an active page. Without a PID, others will not be able to find or reuse your data(set) or metadata in the long-term.

How to do this?

When you upload your data(set) or metadata to a data repository, the data repository (or other service providers) usually assigns a PID. Repositories ensure that the identifier continues to point to the same data or metadata, according to access terms and conditions you specified.

There are many different types of PIDs, each with their own advantages, disadvantages, and disciplines they are typically used in. Generally speaking, the data repository will have thought about these aspects before deciding which PID type to use. In case you have to choose the PID type yourself, you can visit the Knowledge Hub on the PID Forum for guidance. Some disciplines or organisations also provide tools to help you make this choice, see for example this Persistent Identifier Guide for cultural heritage researchers. Once you have chosen a PID type, you can search for data repositories providing that specific PID in registries such as Re3data or FAIRsharing (see related databases).

Not all data you produce during your research will need a PID. In general, those that underpin published findings or have longer term value are worth assigning a PID. If in doubt about which data should be allocated a PID, speak to your local research data management support team or the data repository.

Want to know more?

Did you know that a PID can refer to any kind of resource? Besides publications or datasets, a PID can also refer to, for example, a person, a scientific sample, a funding body, a set of geographical coordinates, an unpublished report, or a piece of software. Depending on what you find important to link to, you might want to consider using a PID for one or more of these resource types.

Persistent identifiers may point to a data file, a web service response that contains data values, or ideally to an online page that contains metadata for context and the link to access the actual data or details about how to request access. The technical process of translating the identifier to a location is called 'resolving' an identifier.

REUSABLE

7. Are you

Close

3

Break Out Sessions

1. Software licensing - Lourens Veen
2. Software citation - Jaro Camphuijsen
3. FAIR-Aware tool - Kim Ferguson



4

Wrap up



See also the [LCRDM website](#)

TU Delft



[RDM support & Data Stewards & DCC](#)
Esther (e.plomp@tudelft.nl), Yasemin
(y.turkyilmaz-vanderveiden@tudelft.nl)

Vrije Universiteit Amsterdam



[Research Data Services](#)
researchdataservices.ub@vu.nl

**Amsterdam University
Medical Center**



[Data Management](#)
datamanagement@vumc.nl
(Locatie AMC: [Rudy Scholte](#))

University Medical Center Utrecht



[Research Data @ UMC Utrecht](#)
datamanagementumcu@umcutrec
ht.nl

Utrecht University



[RDM support](#)
info.rdm@uu.nl

Leiden University



[Centre for Digital Scholarship](#)
cds@library.leidenuniv.nl

Wageningen University & Research



[Data Management Support](#)
data@wur.nl

University of Amsterdam



[RDM support](#)
rdm-support@uva.nl

University of Groningen



[RDM Support](#)
Researchdata@rug.nl

Radboud University Nijmegen



[RDM Support](#)
rdmsupport@ubn.ru.nl

Radboud University Medical Center



[Data Stewardship](#)
datastewardship@im
@radboudumc.nl



Thanks to:

- Presentation template by SlidesCarnival ([Cymbeline](#))
- Esther Plomp

