

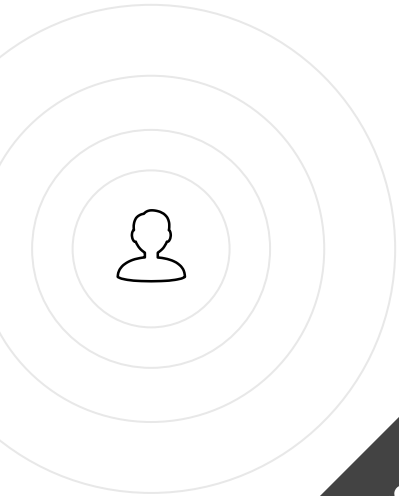
# Plan ahead: practical tools to make your data and software more FAIR

NWO BioPhysics 2022 -  
October 10 2022

[https://tinyurl.com/  
BioPhysics2022FAIR](https://tinyurl.com/BioPhysics2022FAIR)

[https://doi.org/  
10.5281/zenodo.7133499](https://doi.org/10.5281/zenodo.7133499)





**Mateusz Kuzak**

Community Manager  
Netherlands eScience Center  
m.kuzak@esciencecenter.nl



**Jaro Camphuijsen**

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



**Lourens Veen**

Research Software Engineer  
Netherlands eScience Center  
l.veen@esciencecenter.nl



**Kim Ferguson**

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



**Esther Plomp**

Data Steward  
TU Delft, Faculty of Applied Sciences  
e.plomp@tudelft.nl



1

# Introduction

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

Pitches + Break out sessions!



Learn more:  
[The Turing Way](#)



# DMP

- **Tool to 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)

**F**indable 


**A**ccessible 

**I**nteroperable 

**R**eusable 

# Findable

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



**F**



# Findable

- 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

- 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://fairsharing.org)
  - [Research Data Alliance metadata directory](#)
  - [Digital Curation Center](#)



# F

# Evaluation of neodymium isotope analysis of human dental enamel as a provenance indicator using $10^{13} \Omega$ amplifiers (TIMS)

E. Plomp <sup>a</sup>, I.C.C. von Holstein <sup>a</sup>, J.M. Koornneef <sup>a</sup>, R.J. Smeets <sup>a</sup>, J.A. Baart <sup>b, c, 1</sup>, T. Forouzanfar <sup>b, c</sup>, G.R. Davies <sup>a</sup>

Show more

<https://doi.org/10.1016/j.scijus.2019.02.001>

Under a Creative Commons license

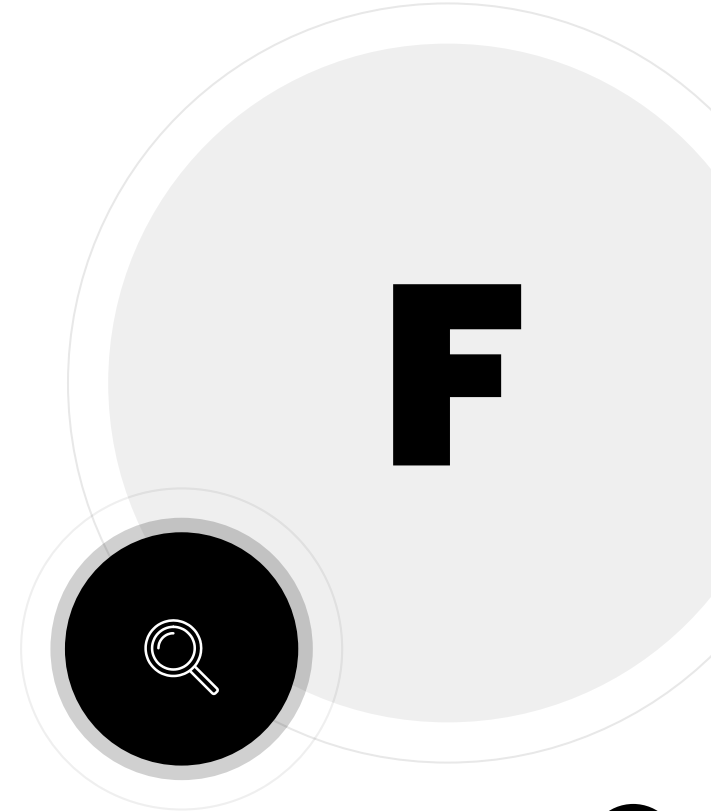
Get rights and content

open access

- 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

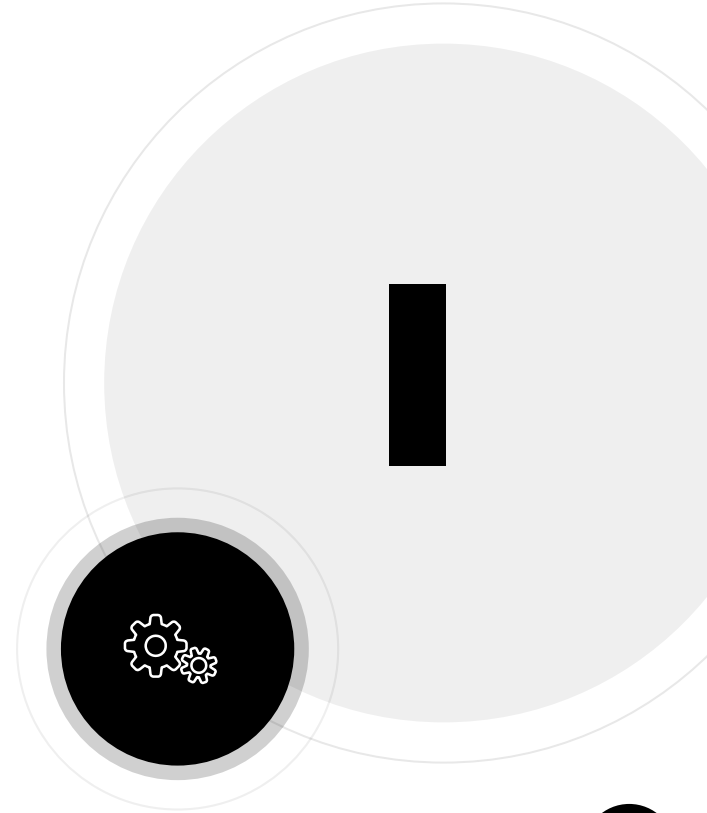
- Consider what will be shared
- Obtain participant consent and perform risk management
- Determine access control
- Share your metadata

**A**



# Interoperable

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



# Reusable

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



R



# Reusable

- Apply a **licence**
- Documentation

Data:

[Creative Commons](#) (Overview)

[Creative Commons License Chooser](#)

Software:

[Choose a License](#)

[tl;dr Legal](#)

## Licences for data

Public Domain Dedication (CC0)

Attribution (CC BY)

Attribution-NoDerivatives (CC BY-ND)

Attribution-NonCommercial (CC BY-NC)

Attribution-NonCommercial-ShareAlike (CC BY-NC-SA)

Attribution-NonCommercial-NoDerivatives (CC BY-NC-ND)

## Licences for software and code

MIT License

Apache License 2

GNU General Public Licence 3 (GNU GPLv3)



# R

Learn more:  
[The Turing Way](#)

**F**indable   
**A**ccessible   
**I**nteroperable   
**R**eusable 

≠

**Standard**

**Open**

**Intrinsic  
quality**

# 2

# Pitches

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





# Software Licensing

tl;drLegal

## Choose an open source license

An open source license protects contributors and users. Businesses and savvy developers won't touch a project without this protection.

Which of the following best describes your situation?



**I need to work in a community.**

Use the [license preferred by the community](#) you're contributing to or depending on. Your project will fit right in.

If you have a dependency that doesn't have a license, ask its maintainers to [add a](#)



**I want it simple and permissive.**

The [MIT License](#) is short and to the point. It lets people do almost anything they want with your project, like making and distributing closed source versions.

[Babel](#), [.NET Core](#), and [Rails](#) use the MIT



**I care about sharing improvements.**

The [GNU GPLv3](#) also lets people do almost anything they want with your project, *except* distributing closed source versions.

[Ansible](#), [Bash](#), and [GIMP](#) use the GNU GPLv3.

# Software citation

How to supply **citation metadata** and how to get a DOI for your software

1. **Why** you should make your software citable
2. Using the **CITATION.cff** file
3. Automatically archiving software releases using **GitHub** and **Zenodo**

- 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](https://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 aware that metadata describing your data(set) should use controlled vocabularies?

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-vandenvelden@tudelft.nl)

**Vrije Universiteit Amsterdam**



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

**VU University Medical Center**



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

**University Medical Center Utrecht**



Data Steward  
Irina van Dijk  
([I.C.vanDijk-9@umcutrecht.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](#))
- Naomi Chrispijn & Renée Calon

