FAIR software Management

netherlands
eScience center

Jaro Camphuijsen

Research Software Engineer

These slides: 10.5281/zenodo.7962359

FAIR for research software & Software Management Plans

netherlands eScience center

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.

-

Article | Open Access | Published: 14 October 2022
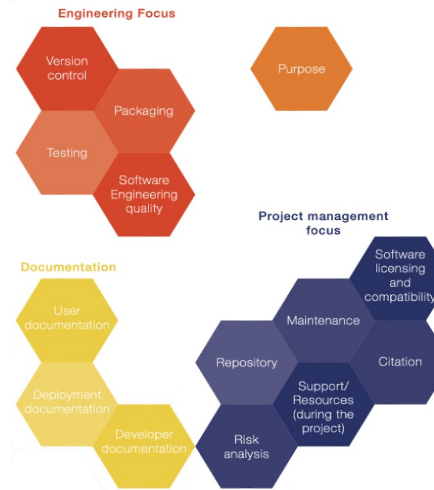### Introducing the FAIR Principles for research software
Michelle Barker ✉, Neil P. Chue Hong, Daniel S. Katz, Anna-Lena Lamprecht, Carlos Martinez-Ortiz, Fotis Psomopoulos, Jennifer Harrow, Leyla Jael Castro, Morane Gruenpeter, Paula Andrea Martinez & Tom Honeyman

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.

Engineering Focus
Version control, Packaging, Testing, Software Engineering quality

Purpose

Documentation
User documentation, Deployment documentation, Developer documentation

Project management focus
Software licensing and compatibility, Maintenance, Repository, Citation, Support/Resources (during the project), Risk analysis

doi.org/10.5281/zenodo.7038280

Why FAIR4RS?

- More and more research done by/with/through software
- Safeguard scientific values
- Recognizing the increasing value of research software
- Not all FAIR (data) principles can be easily applied
- Identifying shared challenges of the research software community → future work

Exercise: identify FAIR4RS principles

F

Software, and its associated metadata, is easy for both humans and machines to find.
- **F1**. Software is assigned a globally unique and persistent identifier.
  - **F1.1**. Components of the software representing levels of granularity are assigned distinct identifiers.
  - **F1.2**. Different versions of the software are assigned distinct identifiers.
- **F2**. Software is described with rich metadata.
- **F3**. Metadata clearly and explicitly include the identifier of the software they describe.
- **F4**. Metadata are FAIR, searchable and indexable.

A

Software, and its metadata, is retrievable via standardised protocols.
- **A1**. Software is retrievable by its identifier using a standardised communications protocol.
  - **A1.1**. The protocol is open, free, and universally implementable.
  - **A1.2**. The protocol allows for an authentication and authorization procedure, where necessary.
- **A2**. Metadata are accessible, even when the software is no longer available.

I

Software interoperates with other software by exchanging data and/or metadata, and/or through interaction via application programming interfaces (APIs), described through standards.
- **I1**. Software reads, writes and exchanges data in a way that meets domain-relevant community standards.
- **I2**. Software includes qualified references to other objects.

R

Software is both usable (can be executed) and reusable (can be understood, modified, built upon, or incorporated into other software).
- **R1**. Software is described with a plurality of accurate and relevant attributes.
  - **R1.1**. Software is given a clear and accessible license.
  - **R1.2**. Software is associated with detailed provenance.
- **R2**. Software includes qualified references to other software.
- **R3**. Software meets domain-relevant community standards.

Exercise: identify FAIR4RS principles

Comet is a command-line tool and desktop application for tandem mass spectrometry sequence database search. The software is licensed under the Apache 2.0 open source licence, and the publicly accessible project repository on GitHub includes detailed information about its development (**??**).Comet uses standard data types from the proteomics domain for its input and output data (**??**). It is registered in the bio.tools catalogue of bioinformatics tools, where it has a globally unique and persistent identifier (**??**), and rich metadata (**??**). The datatypes are documented in the metadata as functional annotations (**??**). Comet can be downloaded via the browser following the links provided in the metadata using https (**??**). The code includes dependencies to external software packages, such as Thermo Scientific's MSFileReader library (**??**). The metadata in bio.tools is independent from the Comet repository, and will stay accessible should the software itself become inaccessible (**??**). The metadata includes the persistent identifier (**??**) and is searchable and indexable (**??**).

Comet

https://uwpr.github.io/Comet/

Exercise: identify FAIR4RS principles

F

Software, and its associated metadata, is easy for both humans and machines to find.

F1. Software is assigned a globally unique and persistent identifier.

F1.1. Components of the software representing levels of granularity are assigned distinct identifiers.

F1.2. Different versions of the software are assigned distinct identifiers.

F2. Software is described with rich metadata.

F3. Metadata clearly and explicitly include the identifier of the software they describe.

F4. Metadata are FAIR, searchable and indexable.

A

Software, and its metadata, is retrievable via standardised protocols.

A1. Software is retrievable by its identifier using a standardised communications protocol.

A1.1. The protocol is open, free, and universally implementable.

A1.2. The protocol allows for an authentication and authorization procedure, where necessary.

A2. Metadata are accessible, even when the software is no longer available.

I

Software interoperates with other software by exchanging data and/or metadata, and/or through interaction via application programming interfaces (APIs), described through standards.

I1. Software reads, writes and exchanges data in a way that meets domain-relevant community standards.

I2. Software includes qualified references to other objects.

R

Software is both usable (can be executed) and reusable (can be understood, modified, built upon, or incorporated into other software).

R1. Software is described with a plurality of accurate and relevant attributes.

R1.1. Software is given a clear and accessible license.

R1.2. Software is associated with detailed provenance.

R2. Software includes qualified references to other software.

R3. Software meets domain-relevant community standards.

Exercise: identify FAIR4RS principles

Comet is a command-line tool and desktop application for tandem mass spectrometry sequence database search. The software is licensed under the Apache 2.0 open source licence, and the publicly accessible project repository on GitHub includes detailed information about its development (**R1**).Comet uses standard data types from the proteomics domain for its input and output data (**I1**). It is registered in the bio.tools catalogue of bioinformatics tools, where it has a globally unique and persistent identifier (**F1**), and rich metadata (**F2**). The datatypes are documented in the metadata as functional annotations (**I2**). Comet can be downloaded via the browser following the links provided in the metadata using https (**A1**). The code includes dependencies to external software packages, such as Thermo Scientific's MSFileReader library (**R2**). The metadata in bio.tools is independent from the Comet repository, and will stay accessible should the software itself become inaccessible (**A2**). The metadata includes the persistent identifier (**F3**) and is searchable and indexable (**F4**).
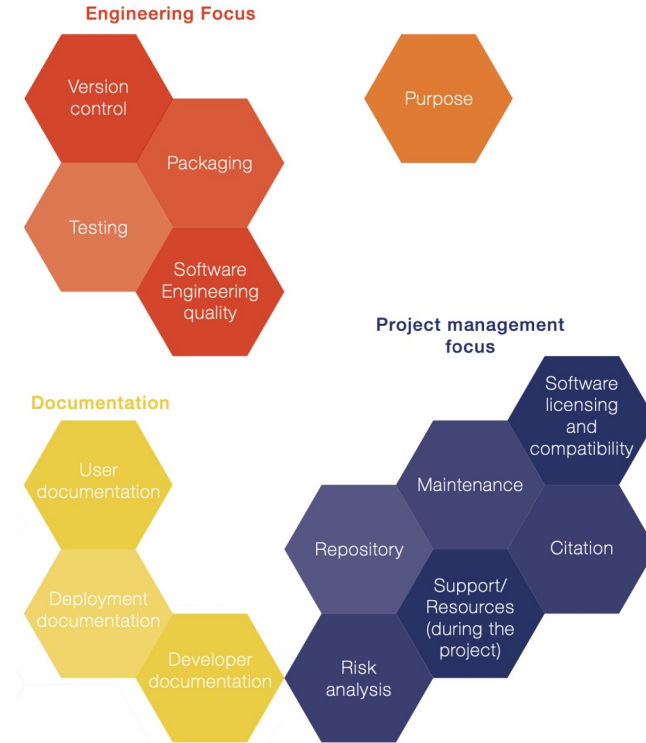
https://uwpr.github.io/Comet/

Why Software Management Plans?

It's not hard!

A software management plan (SMP) can help to:

- Explain why developing new software is necessary
- Make the research software reusable and sustainable
- Plan for necessary resources
- Allow for verification of work that went into implementation

Engineering Focus

Version control

Packaging

Testing

Software Engineering quality

Purpose

Project management focus

Software licensing and compatibility

Maintenance

Repository

Citation

Support/ Resources (during the project)

Risk analysis

Documentation

User documentation

Deployment documentation

Developer documentation

doi.org/10.5281/zenodo.7038280

Exercise: SMP templates

1. Pick a software project
2. Go to: tinyurl.com/SMP-guide
3. Select the core requirements (*section 5*) you think are relevant for this type of software
4. Create an SMP template by formulating question (*examples: section 6*)
5. Bonus: fill in the template for your chosen project

Digital Skills Programme

Our Digital Skills Workshops are **free and open** to researchers in the Netherlands.

Custom (paid) workshops dedicated to your organisation?

For more info contact: training@esciencecenter.nl

Check upcoming workshops: www.esciencecenter.nl/events

Upcoming workshop notification?

Subscribe to the newsletter: tinyurl.com/nl-escience-news