Towards FAIR Principles for Research Software

Anna-Lena Lamprecht, @al_lamprecht

Carlos Martinez-Ortiz, @neocarlitos

Anna-Lena Lamprecht, Leyla Garcia, Mateusz Kuzak, Carlos Martinez, Ricardo Arcila, Eva Martin Del Pico, Victoria Dominguez Del Angel, Stephanie van de Sandt, Jon Ison, Paula Andrea Martinez, Peter McQuilton, Alfonso Valencia, Jennifer Harrow, Fotis Psomopoulos, Josep Ll Gelpi, Neil Chue Hong, Carole Goble, Salvador Capella-Gutierrez

Towards FAIR Principles for Research Software

FAIR

Findable
Accessible
Interoperable
Reusable

2016: "**The FAIR guiding principles for scientific data management and stewardship**" (Wilkinson et al., doi:10.1038/sdata.2016.18)

2018: "Central to the realization of FAIR are **FAIR Digital Objects, which may represent data, software or other research resources**." (European Commission)

But: **How do the FAIR Principles relate to software?**

Mismatch between the broad intentions of the **4 foundational FAIR principles** and how the **15 FAIR Guiding Principles** are communicated and perceived.

FAIR and software: an ongoing discussion

This session!

FAIR software at NL-RSE & eScience symposium 2019

"FAIR principles for Software" at 2019 Workshop on Sustainable Software Sustainability (WOSSS19)

"FAIR Software" Birds of a Feather meeting at deRSE 2019

Top 10 FAIR Data & Software Global Sprint, including "10 easy things to make your software FAIR"

"Sharing Your Software – What is FAIR?" at the 2018 American Geophysical Union (AGU) Fall Meeting

"FAIRness assessment for software" at the ELIXIR 2018 BioHackathon

"Making Software FAIR" at the DTL Communities@Work 2018 Conference

TIB Training workshops on FAIR Data and Software

"Applying FAIR Principles to Software" at the 2017 Workshop on Sustainable Software Sustainability (WOSSS17)

CodeMeta Workshop 2016 on The Future of Software Metadata

…

Software is (not) data

```
         ┌─────────────────┐
         │      Data       │
         │  (01101010…)    │
         └────────┬────────┘
      ┌───────────┼───────────┐
┌──────────┐ ┌──────────┐ ┌──────────┐
│Information│ │Instructions/│ │    …    │
│          │ │ Software   │ │          │
└──────────┘ └──────────┘ └──────────┘
```
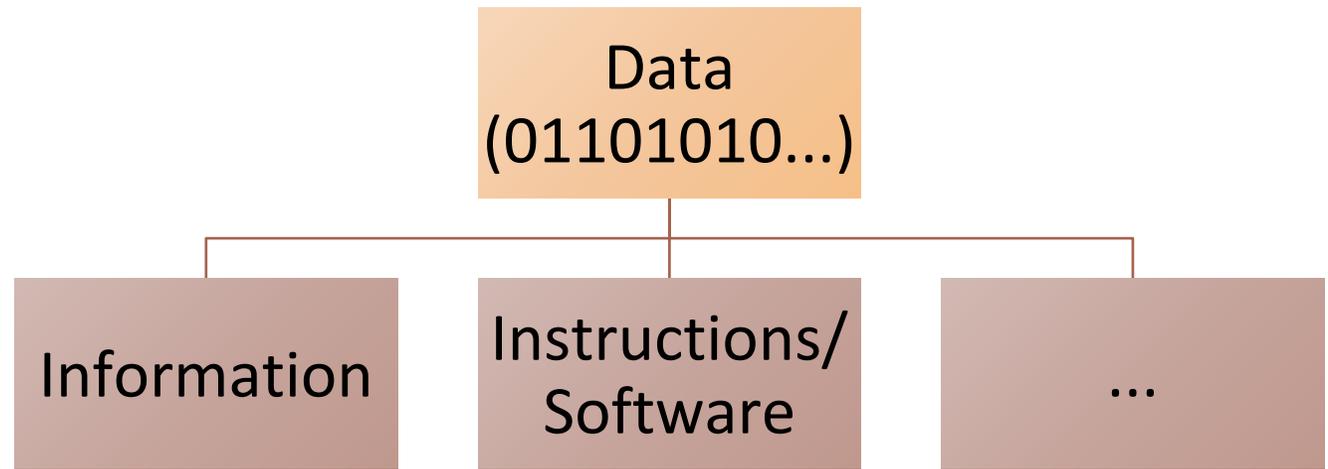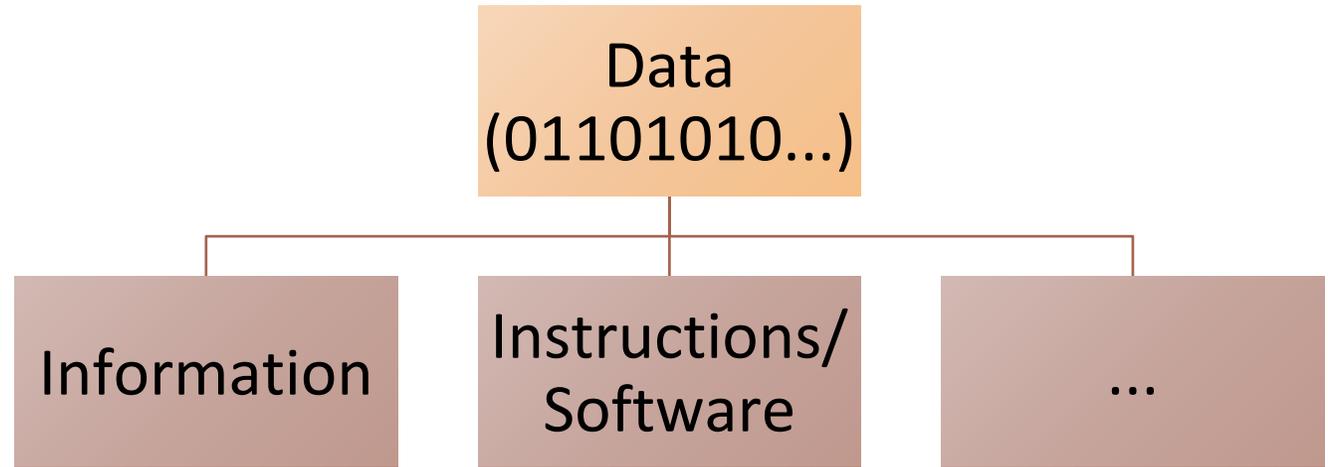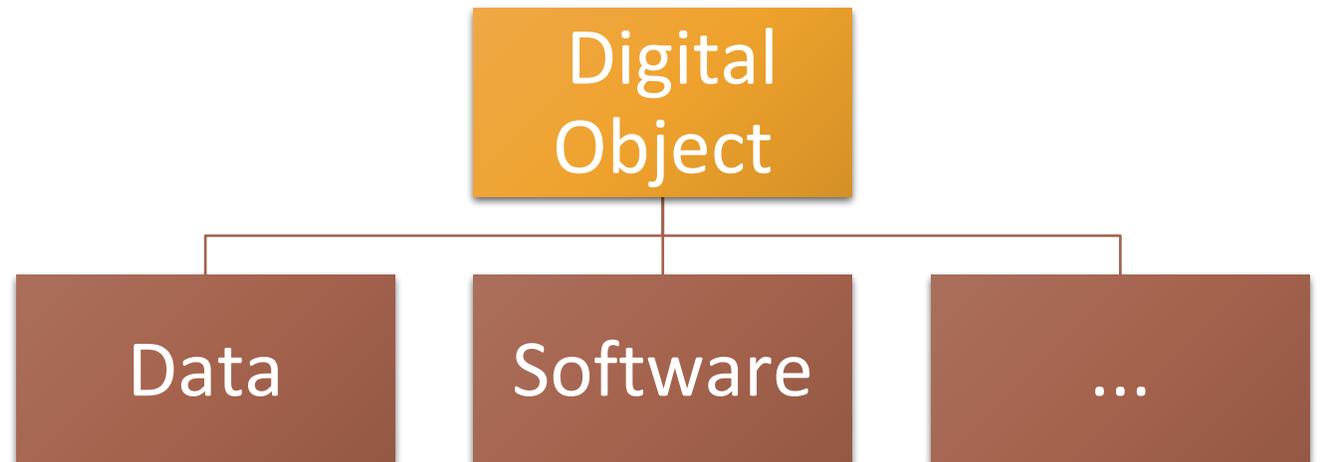
Research software

Research software is "**software that is used to generate, process or analyze results that you intend to appear in a publication**" (Hettrick et al., 2014).

Many forms.

Many purposes.

Many distribution channels.

Traditionally, often created as Free and/or Open Source Software (FOSS).

FAIR and FOSS

Clear overlap of objectives, but not the same.

FOSS: Open source code, open licenses.

FAIR: Open data not a requirement.
- Due to, e.g., privacy and sensitivity concerns with patients' health records.
- Not in the same way valid for research software.
- There is even a demand to make methods available!

Should FAIR software require FOSS?

Ongoing discussion …

FAIR and software quality

Software quality is a major concern in RSE.

Can FAIR meet the expectations?

Distinguish between **form** and **function** of software:

Quality of the **form** of software can be covered by FAIR (code quality, maintainability).

Quality of the **functionality** of software goes beyond FAIR (functional correctness, software security, computational efficiency).

FAIR applied to research software

	FAIR for data	FAIR for software	Operation
F1	(Meta)data are assigned a globally unique and persistent identifier.	Software and its associated metadata have a global, unique and persistent identifier for each released version.	Rephrased
F2	Data are described with rich metadata.	Software is described with rich metadata.	Rephrased
F3	Metadata clearly and explicitly include the identifier of the data it describes.	Metadata clearly and explicitly include identifiers for all the versions of the software it describes.	Rephrased and extended
F4	(Meta)data are registered or indexed in a searchable resource.	Software and its associated metadata are included in a searchable software registry.	Rephrased

	FAIR for data	FAIR for software	Operation
A1	(Meta)data are retrievable by their identifier using a standardized communications protocol.	Software and its associated metadata are accessible by their identifier using a standardized communications protocol.	Rephrased
A1.1	The protocol is open, free, and universally implementable.	The protocol is open, free, and universally implementable.	Remain the same
A1.2	The protocol allows for an authentication and authorization procedure, where necessary.	The protocol allows for an authentication and authorization procedure, where necessary.	Remain the same
A2	Metadata are accessible, even when the data are no longer available.	Software metadata are accessible, even when the software is no longer available.	Rephrased

	FAIR for data	FAIR for software	Operation
I1	(Meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.	Software and its associated metadata use a formal, accessible, shared and broadly applicable language to facilitate machine readability and data exchange.	Rephrased and extended
I2	(Meta)data use vocabularies that follow FAIR principles.	–	Reinterpreted, extended and split
I2S.1	–	Software and its associated metadata are formally described using controlled vocabularies that follow the FAIR principles.	Reinterpreted, extended and split
I2S.2	–	Software use and produce data in types and formats that are formally described using controlled vocabularies that follow the FAIR principles.	Reinterpreted, extended and split
I3	(Meta)data include qualified references to other (meta)data.	–	Discarded
I4S	–	Software dependencies are documented and mechanisms to access them exist.	Newly proposed

	FAIR for data	**FAIR for software**	**Operation**
R1	(Meta)data are richly described with a plurality of accurate and relevant attributes.	Software and its associated metadata are richly described with a plurality of accurate and relevant attributes.	Rephrased
R1.1	(Meta)data are released with a clear and accessible data usage license.	Software and its associated metadata have independent, clear and accessible usage licenses compatible with the software dependencies.	Rephrased and extended
R1.2	(Meta)data are associated with detailed provenance.	Software metadata include detailed provenance, detail level should be community agreed.	Rephrased
R1.3	(Meta)data meet domain-relevant community standards.	Software metadata and documentation meet domain-relevant community standards.	Rephrased

More in the paper

Examples of tooling that enables F, A, I, R software.

Detailed discussion of the challenges around interoperability.

Exemplary FAIRness assessment of two bioinformatics tools (Fastme and ChIPseeker).

…

What's next?

Further discussion!

Agreement on definite FAIR software principles.

Governance model for the FAIR principles.

Metrics and maturity models for FAIR software.

Agreements on the expected degrees of FAIRness.

Acknowledgments

We are grateful to the **numerous people who contributed to the discussions around FAIR research software at different occasions** preceding the work on this paper.

Thank you!