FAIRifying the B Centre Checklist

Lene Offersgaard, CST, NorS,
University ofCopenhagen

SCCTC Utrecht 2019-05-21

UNIVERSITY OF COPENHAGEN

Formal checks and the CoreTrustSeal

B Centre Checklist

**CTS (earlier DSA) asks for a lot of details and documentation**

Compliance to CLARIN mission/vision/community

FAIR

- In 2014 the term FAIR was launched at a Lorentz workshop
- In 2016 the FAIR principles were published
- More detailed info: https://www.go-fair.org

A set of guiding principles to make data:

- **F**indable, **A**ccessible, **I**nteroperable, and **R**eusable

Findable   Accessible   Interoperable   Reusable
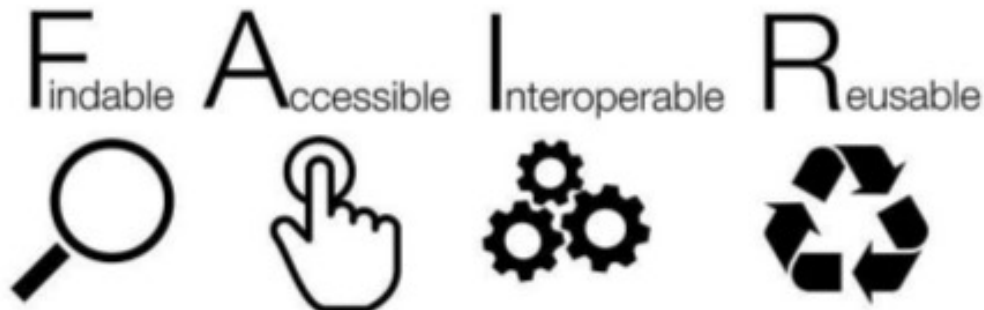
Let's talk about FAIR data...

"Making your data FAIR is a gradual process with small steps one at a time. In the end, it will become best practice in your research field"

www.vidensportal.deic.dk/FAIR

Findable

F1. (Meta)data are assigned a globally unique and persistent identifier

F2. Data are described with rich metadata (defined by R1 below)

F3. (Meta)data are registered or indexed in a searchable resource

F4. Metadata clearly and explicitly include the identifier of the data it describes

CLARIN and Findable

CLARIN provides:

✓ Persistent Identifiers(PID) on metadata

✓ PID's on data files

• Rich Metadata, the researchers provides metadata in a form, metadata curation/support to the researchers

✓ Metadata are searchable at individual repositories and at vlo.clarin.eu

✓ Citations are provided

Accessible

A1. (Meta)data are retrievable by their identifier using a standardized 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 data are no longer available

CLARIN and Accessible

CLARIN provides:

✓ You can get to the metadata of a resource using the PID (example for vlo.clarin.dk)

✓ You can get to the data files using the PID's

✓ Data can be retrieved by http protocol or by curl

✓ SPF (where are you from-service) for authentication

✓ Rights given by home institution through WAYF for authorisation as academic, if necessary

✓ Metadata is always open and accessible, even if data are restricted as aca or further restricted

Interoperability

I1. (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.

I2. (meta)data use vocabularies that follow FAIR principles
- Ontologies
- Vocabularies
- Taxonomies

I3. (meta)data include qualified references to other (meta)data

CLARIN and Interoperable

✓ Metadata follows OLAC, DC and CMDI standards

▪ Researchers are expected to use open formats and/or formats commonly understood in the specific research area of the Humanities – curation step in deposit gives an option for guidance if chosen formats are not commonly understood – but the researcher is free to choose the formats

• The CLARIN Dspace Repository system has the option to add some controlled vocabularies e.g. currently for resource types···

✓ References can be given to related data

Reusable

R1. meta(data) are richly described with a plurality of accurate and relevant attributes

R1.1. (meta)data are released with a clear and accessible data usage license

R1.2. (meta)data are associated with detailed provenance

R1.3. (meta)data meet domain-relevant community standards

CLARIN and Reusable

✓ Data have a clear license attached
   ➢ Most common open licenses currently in the list.
   ➢ We can add more licenses, if the researcher community have a good reason to use yet another license, and when new licenses are widely accepted (Dspace repos)

✓ Clear who have provided the data

▪ The researcher has to document the creation and processing of data in the metadata description, by attaching documentation files, or by adding links to documentation as related files

▪ Documentation of why data are created/processed can perhaps be checked in curation step

▪ Data hopefully meet domain standards, but formats are chosen by the researcher – curation/support option

✓ Metadata meet domain standards

Compliance to FAIR -1

- To comply with the FAIR principle *"R1.1. (meta)data are released with a clear and accessible data usage license"*,

It could be an explicit requirement to attach licenses to data (and metadata)

an issue about metadata that the CLIC can discuss and then SCCTC can take action on it.

Compliance to FAIR -2

- Checklist item 6.9b try to comply with FAIR *"I2. (meta)data use vocabularies that follow FAIR principles"*.

Item 6.9b requires that the used CMDI profile schemas have valid ConceptLinks to CCR. Currently this is not taken as an requirement, as almost all CMDI profiles miss some CCR links.

Compliance to FAIR -3

- To fulfil FAIR "F2 data are described with rich metadata."

might be a larger task for data providers. We could consider to add a requirement for at least providing metadata including *description* and *data license information*.

Compliance to FAIR -4

- To show even more awareness to FAIR "A2 metadata are accessible, even when the data are no longer available"

we could ask the repositories about how they will treat deleted data, and if they will keep the metadata accessible.

They can perhaps provide their answer to CTS *"R7. The repository guarantees the integrity and authenticity of the data"*, it can also be left to the CTS reviewers to check this, or we can decide not to go for a change here.

Concerning 'reusable' data

- Repositories are **depending on researchers** providing enough documentation of the data -> for other researchers to re-use the data

- Sharing data between research communities are depending on communities sharing/understanding used stds.

Promoting use of existing std's are still relevant!