////////////////////////////////////////////////////////////////////////////////////////////////////////////////////

**ONTWERPVERSLAG**

FOSB WG Metadata & Standardisation

////////////////////////////////////////////////////////////////////////////////////////////////////////////////////

Datum: 29/09/2020

1. Joint meeting with the FOSB WG RDM & Open Science- Sub WG KPI’s
2. Presentation on the development of an open data label, FAIR data label and open access label
3. Discussion on implications of KPI’s on the metadata model

**Aanwezig**

Joke Meeus FWO

Niek Van Wettere VUB

Rudi Baccarne UAntwerpen

Sara De Coster KU Leuven (vervangt Roxanne Wyns)

Charlotte D’Eer FWO

Frederik Coppens VIB

Ils de Bal Departement EWI

Pascale Dengis Departement EWI

Sadia Vancauwenbergh ECOOM-UHasselt

Evy Neyens ECOOM-UHasselt

Hanne VLIETINCK ECOOM-UHasselt

Roxanne Wyns KU Leuven

Dimitri Brosens INBO

Helga Deschrijver IMEC

Francisco Hernandez VLIZ

Katrien Desmet VLHORA

Quentin Groom Plantentuin Meise

Hans Polet ILVO

Rinaldo Beck VIB

Dries Moreels UGent

Alexander Botzki VIB

Steven Heylen FWO

Mia Vanstraelen Brain Group

Anneleen Baerts UAntwerpen

Francisco Hernandez VLIZ

Hans Polet ILVO

Steven Heylen FWO

/////////////////////////////////////////////////////////

**Verslag Vergadering 16/09/2020**

Joint meeting with the FOSB WG RDM & Open Science- Sub WG KPI’s

The Flemish government has commissioned the FOSB to develop KPI’s to monitor the achievement of policy objectives regarding open science. The KPIs cover all research that is (at least partially) funded by Flemish public funds (i.e. FWO, BOF, IOF, VlAIO,). The FOSB sub WG on KPI’s was asked to outline KPI’s on Data Management Planning, ORCID’s, open data, FAIR data and open access. The FOSB WG Metadata & Standardization was asked to develop a metadata model for the delivery of research datasets to FRIS and to develop an Open data label and a FAIR data label as part of the model to be able to monitor developments towards the openness and fairness of datasets. Because of the overlap in the objectives of both working groups, and because the KPI’s influence which metadata fields are necessary to keep in the metadata model, a joint meeting was set up with members of the FOSB WG Metadata & Standardization and members of the FOSB sub WG KPI’s.

The discussion started with a presentation that outlined various international solutions for developing and implementing an open data label, a FAIR data label and an open access label.

To development of an Open data label was based on the Open Definition: “*Open research data are data that anyone can freely access, use, modify, and share for any purpose – subject only, at most, to the requirement to attribute and share-alike” (*<http://opendefinition.org/>).

This means that to be truly open the dataset must be accessible (freely available for download) in an open, machine-readable format, and the dataset must be provided under an open license enabling others to freely use, change, and distribute data without any kind of restrictions. Only the requirements to give appropriate credit and share-a-like are allowed.

Based on the work of the Open Definition, the Open Data Monitor, and the Open Data Institute 2 possibilities were presented to create an open data label:

1. A simple YES or NO: If the dataset is completely open access, in an open, machine readable format and with an open license than it gets an open data label.
2. A multidimensional field that expresses the degree of openness of a dataset based on 3 main indicators: availability, open/machine readable format, and open license. By analogy with the 5 stars of open data, this will give a value of 1 to 3 stars for each dataset on FRIS depending on the extent to which the dataset meets the three requirements of open data.

- Availability: Is the dataset available in open access, Yes or No? A dataset is available in open access when the whole dataset is immediately downloadable via the Internet without charge. There is a DOI leading to a landing page with a downloadable link.

- Open and machine-readable format: Is the dataset available in an open/machine-readable format? Yes/No

- Open license: Is the data openly licensed, yes or no? The Open Definition provides a list of conformant licenses (e.g. CC0; CC-BY-4.0; CC-BY-SA-4.0; PDDL; ODC-BY; ODbL). https://opendefinition.org/licenses/

Open ≠ fair data: a dataset can be ‘closed access’ due to privacy/ethical concerns yet still have a high degree of fairness. And vice versa, a dataset can be open but low in quality and low in fairness; open data is accessible but not necessary reusable in the sense of rich metadata.

To develop a FAIR Data Label we referred to the DANS FAIR Data Assessment tool, and the FAIRsFAIR Data Assessment Metrics. However, since the EOSC FAIR WG will release their own FAIR certification in the coming weeks, it was decided to wait for their output.

To develop an Open Access Label, we referred to the previous work of the General Council of the Association of Universities in the Netherlands (VSNU) who make a distinction between no open access, green open access, gold DOAJ classified access, and gold not DOAJ classified access

(<https://www.vsnu.nl/files/documenten/Domeinen/Onderzoek/Open%20access/Definitief%20Definition%20framework%20OA_VSNU-20160217.pdf>).

The comments and feedback on this discussion and the proposed solutions were incorporated in the KPI's for Open Science document version 9 that you can find on GitHub.

Discussion on implications of KPI’s on the metadata model

To accommodate the measurement of these KPI's, some fields were added to the metadata model and other fields were changed from a recommended status to a mandatory status.

**Open Data status (M):** Status on the access possibilities of the dataset: closed due to legitimate opt-out, open, restricted, embargoed.

Based on OpenAire 16. Rights (MA): rightsURI (MA): Use terms from the info:eu-repo-Access-Terms vocabulary. The values are:

info:eu-repo/semantics/closedAccess

info:eu-repo/semantics/embargoedAccess

info:eu-repo/semantics/restrictedAccess

info:eu-repo/semantics/openAccess

For more info on these statuses see: <https://docs.google.com/document/d/1bRQDN_VFSPSMnsADLyzky-sbd6ZPArsHOcYhERdyrL8/edit>

**IP Rights(M):** Intellectual property rights for the dataset with references to open data licenses: free text or link to license (eg. <Http://creativecommons.org/licenses/by/3.0/de.deed.en>). Based on OpenAire see <https://guidelines.readthedocs.io/en/latest/data/use_of_datacite.html#accessrights>.

This changes from recommended to mandatory because the use of an open license is a requirement for a dataset to be open according to the open definition. This allows EWI to monitor how many datasets are provided with an open license.

**Trusted repositories (M):** Is the dataset deposited in a trusted repository? Yes/No? Based on a list of trusted repositories: Core Trust Seal

https://repositoryfinder.datacite.org/

https://www.re3data.org/

**Format (M):** This changes from recommended to mandatory, because the use of an open format is a requirement for a dataset to be open according to the open definition. This allows EWI to monitor how many datasets are available in an open format.

**FAIR Data Label (M):** 2024 - 90%, 60% should have high standard fair label; 2021 2nd half first implementation 2022 first measurement of fair label. roadmap: manual work – automated. Wait for the output of the EOSC FAIR WG.

|  |
| --- |
| An invitation for the next meeting will be sent out. |