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

**ONTWERPVERSLAG**

FOSB WG Metadata & Standardisation

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

Datum: 16/07/2020

1. In depth discussion of the draft metadata model

**Aanwezig**

Niek Van Wettere VUB

Rudi Baccarne UAntwerpen

Roxanne Wyns KU Leuven

Pascale Dengis Departement EWI

Sadia Vancauwenbergh ECOOM-Hasselt

Evy Neyens ECOOM-Hasselt

Dries Moreels U Gent

Dimitri Brosens INBO

Helga Deschrijver IMEC

Francisco Hernandez VLIZ

Katrien Desmet Vlhora, vervangt Ilse Beerland

Anneleen Baerts U Antwerpen

Frederik Coppens VIB -ELIXIR ESFRI

Hanne Vlietinck ECOOM-Hasselt

Hans Polet ILVO

Steven Heylen FWO

Florian Vanlee VUB

Rinaldo Beck VIB

**Verontschuldigd**

Leen Van Campe Departement EWI

Mia Vanstraelen Industrie; the Brain group

Quentin Groom Plantentuin Meise

Alexander Botzki VIB

Ilse Beerland VLHORA

**Verslag Vergadering 16/07/2020**

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

Preliminary agreements based on in-depth discussion of the draft metadata model

Regarding the definition of a dataset, it was agreed that this term only refers to data generated by researchers affiliated to a Flemish knowledge institution and arising from a project or publication.

It was proposed to make a distinction between a dataset and artistic/non-written output. A dataset refers to any form of output resulting from research, whereas artistic/non-written output is defined as

any form of output resulting from creative output. That way the need to model artistic output as a research dataset becomes obsolete. This definition will be further refined based on the consultations with the Schools of Arts.

Regarding the identifier, it was decided to follow the logic of publications. It is mandatory to provide at least 1 globally unique persistent identifier of the dataset, this can be a DOI or handle. This DOI must resolve to a landing page with metadata and access possibilities to the dataset or, at the minimum, information about the access status of the dataset. Preference is given to a concept DOI which presents an overview of all versions of the dataset. In addition, there is the option to provide multiple, alternative identifiers, i.e. a DOI, handle or URL (ex. Link to data repository like Zenodo or institutional repository), in which case the identifier type has to be specified. This arrangement makes the fields ‘link to data repository’ and ‘link to data archive’ obsolete, hence these terms were deleted from the model.

It was decided to leave the abstract *mandatory when applicable*. Providing an abstract is only mandatory when a link to a project or publication is missing since an abstract is already provided there.

The group members decided that providing one or multiple creator name(s) (Names of the main researchers involved in producing the dataset) is mandatory. However, specifying persistent name identifier(s), i.e. an ORCID, is only *mandatory when applicable* since the use of ORCID iD’s are not widespread among the staff of the colleges of higher education. ORCID iD’s are the preferred type of name identifier and their use will be actively encouraged.

Following DataCite, we agreed it is mandatory to provide the creator affiliation: Affiliation to a Flemish Research Producing Organisation of the main (Flemish) researchers involved in producing the dataset, i.e. using the GRID.

We agreed that it is mandatory to provide the title of the dataset in the original language in which the metadata record was published. A title (either the main title or the translation) always coincides with a language field specifying the language of the title (possibly using controlled vocabularies). It is optional to give a Dutch translation of the main, original title of the dataset.

The group members decided to change the name of the term 'archiving organisation' to 'Publisher' in accordance with DataCite/OpenAire. This field refers to the name of the entity that holds, archives, publishes prints, distributes, releases, issues, or produces the resource. Examples include (de)central institutional or generic data repositories, or publishers like Elsevier. Specifying the publisher is mandatory when applicable.

The BOF/IOF decree requires the dataset to be stored in an institutional repository. To avoid that multiple different publishers have to be specified, due to multiple affiliations of researchers, it was agreed that one of the main researchers producing the dataset will take on the task of data manager and deposit the dataset in a (de)central, institutional, or generic data repository (Publisher).

Regarding publication year, we agreed this is a mandatory field which refers to ‘the year when the dataset was or will be available to a community of researchers. It is possible that the metadata record of the dataset has been published with a DOI while the files themselves are not accessible due to an embargo (see Zenodo closed/restricted access with DOI and Publication year).

In accordance with DataCite/OpenAire there was an agreement that, when an embargo period has been in effect, the date when the embargo period ends will be used. If there is no standard publication year value, the date that would be preferred from a citation perspective should be provided.

Furthermore, we settled to remove the fields ‘depositing year’ and ‘archiving year’ and instead follow the logic of publications by providing only the status of the dataset using the field ‘embargo date’. We decided that this field is only mandatory if an embargo date is applicable. We made an agreement that embargo dates relevant to the availability of the dataset can take on the following date type values:

- issued: date used to indicate when the dataset is made available, published or uploaded to a formal database.

- accepted: date used to indicate the start of an embargo period.

- available: date used to indicate the end of an embargo period.

It was proposed to make the field ‘Research discipline’ mandatory. This field allows to classify the dataset on the basis of the disciplines of the Flemish Research Standard. It is unclear whether this field corresponds to the terms Description or Related Identifier (relation Type: Is Described By) in DataCite and OpenAire.

Next meeting, we will resume our discussion from the metadata field ‘keywords’.

|  |
| --- |
| 1. A Doodle has been sent out to arrange a fourth meeting between the 26th of August and the 4th of September 2020. |