Towards a Semantic Interoperable Flemish Research Information Space: Development and Implementation of a Flemish Application Profile for Research Datasets

Evy Neyens
UHasselt, Data Science Institute &
ECOOM

Sadia Vancauwenbergh
UHasselt, Data Science Institute &
ECOOM

Abstract

In Flanders, Research Performing Organizations (RPO) are required to provide information on publicly financed research to the Flemish Research Information Space (FRIS), a current research information system and research discovery platform hosted by the Flemish Department of Economics, Science and Innovation. FRIS currently discloses information onresearchers, research institutions, publications, and projects. Flemish decrees on Special and Industrial research funding, and the Flemish Open Science policy require RPOs to also provide metadata on research datasets to FRIS. To ensure accurate and uniform delivery of information across all information providing institutions on research datasets to FRIS, it isnecessary to develop a common application profile for research datasets. This article outlines the development of the Flemish application profile for research datasets that was developed by the Flemish Open Science Board (FOSB) Working Group Metadata and Standardization. The main challenge was to achieve interoperability among stakeholders, which in part had existing metadata schemes and research information infrastructures in place, while others were still in the early stages of development.

Introduction

There is broad consensus that data sharing enhances scientific transparency and is a key driver of scientific progress and technological innovation. The cornerstone of the EU's Open Science policy is to make scientific information funded by public resources openly available, according to the principle "as open as possible, as closed as necessary" (EU Digital Agenda, 2014; COM (2016) 178; Directive (EU) 2019/1024). A crucial step towards Open Science at the European level has been the initiation of the European Open Science Cloud (EOSC) by European Commissioner Carlos Moedas in 2016 (EU Digital Agenda, 2014; COM (2016) 178).

The scope of EOSC is to connect existing European research infrastructures and data centres into one virtual research hub where researchers can deposit, curate, share and reuse their research data (e.g. publications, data, software, etc.) across disciplines and countries (EOSC Strategic Implementation Roadmap 2018-2020). At present, access to research data in Europe has been fragmented across many different institutions, research centres and disciplines each using their own research infrastructures and governance models. These data silos lack interoperability and complicate Open Access to and reuse of publicly funded research data (EOSC Strategic Implementation Roadmap 2018-2020). EOSC thus represents an evolution towards a federated model that unites existing and future research infrastructures, with universal access to data and common standards for all federated resources (EOSC Declaration Brussels, 2017). Member states are required to align their national science policies in order to connect with EOSC.

In Flanders, the research landscape is quite diverse in terms of the identity and characteristics of research performing organizations (RPOs), and the maturity of each organization in respect to research data management. Currently, Flemish RPOs (e.g. universities, higher education colleges, strategic research centres) are required to provide metadata on research information to the Flemish Research Information Space (FRIS) (www.researchportal.be) by means of decrees and agreements.

The FRIS-portal is a current research information system (CRIS) and a research discovery platform governed by the Department Economy, Sciences and Innovation (EWI) of the Flemish government. The portal discloses information on publicly funded research (e.g. researchers, research institutions, publicly funded projects and publications) from over 40 data providers (research performing & research funding organisations). To ensure a harmonized and uniform delivery of information, the various stakeholders mutually agree on the underlying information model, the exchange format and business and validation rules. For instance, the data providers deliver metadata on research information (persons, organizations, projects, publications, etc.) in real-time directly to FRIS via CERIF-XML (Common European Research Information Format).[1]

Since 2018, two Flemish decrees, the Special Research Fund (BOF) Decree and the Industrial Research Fund (IOF) Decree, require Flemish universities to provide metadata on research datasets to FRIS by the end of 2021. Universities are formally obliged to report on datasets resulting from research financed by BOF/IOF decrees with a minimum of information on the content of the dataset, its accessibility, the technical format, a DOI, an open data label, the link to the open data repository, and the link to the project.

In addition, the Flemish Department of Economy, Science, and Innovation (EWI) has established the Flemish Open Science Board (FOSB) in December 2019 to create a roadmap for the development and implementation of an Open Science policy in Flanders (Flemish government Note, VR 2019 2012 DOC.1265/1BIS). Within the FOSB, three working groups were installed, focusing on 'Research data management & Open science', 'Architecture' and ' Metadata & standardization'. The working groups consist of representatives of the Flemish research institutions, who will shape the Flemish Open Science policy through collaboration and in mutual consultation with all

---

[1] https://www.eurocris.org/eurocris_archive/cerifsupport.org/category/cerif/cerif-xml/index.html

stakeholders. To realize and monitor the Flemish Open Science policy, five key performance indicators (KPIs) for open science were established (Flemish government Note, VR 2020 0412 MED.0379/1):

1. Researchers affiliated with a Flemish RPO are registered in FRIS using a globally unique and persistent identifier (ORCID).[2]

2. Publicly financed research projects must maintain a Data Management Plan (DMP).

3. Publicly funded peer-reviewed scientific publications should become available in Open Access.

4. Research data underlying publicly funded scientific publications should become more FAIR: Findable, Accessible, Interoperable, and Reusable (Wilkinson et al., 2016). For this purpose, a FAIR Data Label for datasets will be developed.

5. Research data underlying publicly funded scientific publications should become openly available (with the exception of legitimate opt-outs). For this purpose, an Open Data label will be developed.

To ensure a uniform and accurate delivery of the information on research data by the various stakeholders, there is a need for a semantically harmonised standard metadata model. In order to display the metadata information on research data in an unambiguous manner, it is paramount that all RPOs converge with all stakeholders on conventionalized, unambiguously interpretable metadata fields. This allows for a perfect understanding of the information contained on the FRIS portal, and indirectly, to a higher potential visibility and re-use of research data from the RPOs. In addition, the development of reliable metrics for the Open Science indicators (e.g. Open data label, FAIR data label) requires semantically harmonized metadata.

Therefore, the working group on 'Metadata & standardization' (hereafter referred to as WG M&S) was assigned to create a common information model for describing research data. This information model must ensure the connection with EOSC through FRIS (see Figure 1) and should align with related international initiatives, such as OpenAIRE, ESFRI, EOSC-Hub, and the FAIR indicators (FAIRsFAIR.eu,[3] GO-FAIR[4]), and international standards for data representation, such as the Research Data Alliance (RDA).

The scope of the WG M&S was to develop a common information model on research data to semantically streamline the delivery of information from various data providers to the Flemish context and to FRIS, while being flexible enough to incorporate additional metadata properties to allow the Flemish department of Economy, Science and Innovation (EWI) to measure and monitor the degree of FAIRness and openness of datasets through FRIS. This prompted the working group to create a Flemish application profile for research datasets rather than merely adopting an already existing metadata standard.

The WG M&S consists of experts delegated by the data providers to represent all FRIS stakeholders equally. The working group assembled bi-monthly for an online meeting, due to COVID-19, and communication took place via mail and GitHub. This article outlines the development of a semantically harmonized Flemish application profile for research datasets amongst all Flemish stakeholders, and the implementation thereof in an inter-organisational context.

---

[2] https://orcid.org/
[3] https://www.FAIRsFAIR.eu
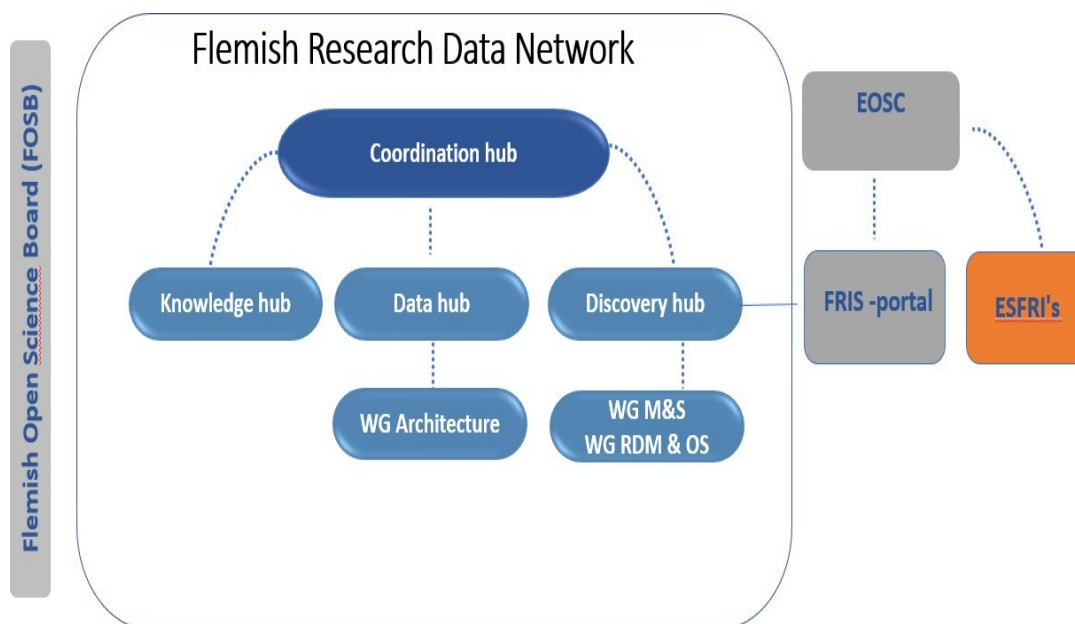[4] https://www.go-fair.org

Figure 1. Connecting the Flemish Research Data Network with EOSC

Development of a Standard Application Profile for Research Datasets

The overall process of developing the Flemish application profile for research data sets can be divided in several steps. First, the FOSB WG M&S reviewed existing generic metadata standards for research data based on their purpose, adaptability and adoption within the European research landscape (e.g. re3data,[5] DataCite[6]). Then, a discipline -agnostic metadata standard that could serve as a model for general metadata information on research data was identified and selected. Metadata provide descriptive, structural and administrative information on research output (e.g. publications, datasets) to support its discovery, retrieval, management and long-term preservation (Hurley et al., 1999; Day, 2005). Metadata standards provide semantic interoperability by describing and linking each entity of the data model to a shared, controlled vocabulary. This ensures a uniform interpretation of metadata terms and allows for exchange and re-use in a standardised manner by both humans and machines.

At present, there are numerous generic and discipline-specific ontologies and metadata standards that are not fully interoperable with each other (Azeroual et al., 2018; Johnston, 2003, Day, 2005). In Flanders, most institutional repositories' metadata standards are based on DataCite, OpenAire, and Dublin Core. Since the DataCite metadata model appeared to be common ground for many institutional repositories in Flanders, and because it is already being used within OpenAire, this model was suggested as the most appropriate sample model by the working group members.

DataCite is a non-profit organisation that was founded with the aim of making datasets citable and discoverable by assigning globally unique and persistent Digital Object Identifiers (DOI) to datasets (DataCite). To enhance discoverability, sharing and reuse of datasets, DataCite created a generic metadata model that is suitable to describe datasets from all types of research disciplines.

---

[5] https://www.re3data.org/

[6] https://datacite.org/

The DataCite Metadata Schema version 4.3 is the most recent version and includes 6 Mandatory and 13 recommended and optional (sub)properties to identify and retrieve data objects such as datasets (DataCite Metadata Schema v4.3, 2019).

Since the FOSB WG M&S was assigned to develop an application profile that ensures a uniform semantic understanding by all Flemish data providers, and that also enables monitoring of the Open Science indicators on FAIR and Open data, the WG decided to base the application profile on the DataCite standard. Next, through intensive consultations with the members of the working group in several meetings, it was decided which metadata properties and subproperties were retained or not, which properties had to be modified to accommodate institution-specific needs and the FRIS portal. Furthermore, in-depth discussions were held to find agreements on the metadata semantics accompanying the metadata properties and subproperties, and what properties were missing from the DataCite model that were required to meet the request of the Flemish government to monitor KPIs on Open Science via FRIS.

This consultation procedure resulted in a Flemish application profile for research datasets consisting of 32 metadata fields on 21 properties (see Table 1, Annex), out of which 15 originated from DataCite. For each metadata property, a definition and allowed values were added, and it was indicated whether the field is Mandatory (M), Mandatory if applicable (MA), Recommended or Optional (O). To enable the monitoring of the Open Science indicators, three additional properties were defined, i.e. Open format, Legitimate opt-out and a FAIR data label. Moreover, the semantic descriptions of the metadata properties as defined by DataCite were adapted to the Flemish context where appropriate.

This application profile pertains to datasets, defined by the working group as: *"Data and objects generated or collected by researchers affiliated by a Flemish research institution in the course of their investigations, regardless of their form or method, that form the object on which researchers test a hypothesis. This includes the full range of data: raw, unprocessed datasets, proprietary generated and processed data and secondary data obtained from third parties."* At present, the application profile has already been formally endorsed by the FOSB and is being implemented at the side of FRIS and the Flemish RPOs who are providing information to FRIS.

Perceived challenges and obstacles

During the consultation process, several challenges emerged that complicated the implementation of the Flemish application profile. First, there were institution-specific challenges that were mainly related to the different CRIS-systems and metadata standards that were already implemented within the stakeholders' institutions, and the different level of RDM and Open Science policies. Second, there were legal obligations related to the monitoring of the Open Science indicators (Flemish Government KPI Note Open Science, VR 2020 0412 MED.0379/1), that required the creation of additional properties.

Institutional differences

On the FRIS-portal all research information can be searched and retrieved by means of the research disciplines of the *Flemish Research Discipline Standard* (VODL, Vancauwenbergh & Poelmans, 2019). Hence, the Flemish department EWI requested to add a new metadata property 'Research discipline' to the information model for research data, to be able to classify datasets based on the Flemish Research Discipline Standard (VODL, Vancauwenbergh & Poelmans, 2019). Some group members suggested that adding this property would be redundant since classification codes can also be accommodated in the 'Subject' field of DataCite. However, the 'Subject' property of DataCite is not specific enough in the context of FRIS, where research information is typically searchable by research discipline codes. Therefore, to enable end-users to use research disciplines as a search filter

on the FRIS-portal to find and retrieve datasets more easily, the WG M&S decided to use the more specific property 'Research discipline' instead of the more generic property 'Subject'.

A similar discussion was held about the 'Keywords' field. Although in DataCite, keywords can be included within the recommended property 'Subject', it was opted to create a new mandatory element 'Keywords'. Since the term 'Keywords' is more specific and more informative, this increases the findability and searchability of datasets to the end-users of the FRIS-portal, thus contributing to its role as a discovery hub.

Political and legal obligations

In Flanders, the decrees on Special and Industrial research funding (BOF/IOF) legally require universities to deliver information on datasets to the FRIS-portal (B.S.22/03/2013). Metadata of datasets that form the basis of scientific publications resulting from research projects that are financed by the Special or Industrial Research Fund (BOF/IOF) have to be reported to the FRIS-portal, with at least information on: the content of the dataset, its accessibility, the technical format, a DOI, an open data label, the link to the open data repository, and the link to the project. In addition, the Flemish government's KPI Note on Open Science outlines five KPIs (ORCIDs, DMP, Open Access, FAIR data, & Open data) to make research data openly available and FAIR (VR 2020 0412 MED.0379/1). These KPIs relate to all research that is (fully or partially) funded by Flemish public resources (i.e. FWO, BOF, IOF, VlAIO,...), with the exception of research funded with private means (VR 2019 2012 DOC.1265/1BIS).

In order to accurately monitor adoption and progress of the Open Science indicators of the Flemish government, it was essential to make certain metadata fields mandatory, as well as to incorporate new fields. For instance, Flemish legal decrees and convenants obligate the inclusion of an Open Researcher and Contributor ID (ORCID) in FRIS. In addition, this ORCID is used as one of the indicators to monitor Open Science in Flanders. The Flemish government's KPI Note on Open Science (VR 2020 0412 MED.0379/1) requires that 95% of researchers that are affiliated to a Flemish university, should register an ORCID by 2026. In brief, this urged the incorporation of the ORCID for the main researchers involved in producing the dataset using the property 'Creator Identifier', the development of rule sets in order to monitor ORCIDs and policies to stimulate the creation of ORCID iDs among the staff of higher education institutions and research institutes and the registration in their repositories.

Another example concerns the adding or mandating of metadata properties to allow the Flemish department EWI to measure and monitor the FAIRness and openness of datasets. The KPI on Open Data states that research data underlying publicly financed peer-reviewed scientific publications should be made openly available, with the exception of legitimate opt- outs (VR 2020 0412 MED.0379/1). 'Open Data' was defined as: "*Digital research data that are accessible to anyone and can be used, modified and/or shared by anyone for any purpose*" (the Open definition). [7] This implies the occurrence of open licenses that give explicit permission to (re)use the data without any restrictions, except for the condition to share-alike and to give attribution (e.g. CC0, CC-by-4.0, ODC-By-1.0). Additionally, datasets are only considered to be completely open when they are available using free, open source formats (e.g. XML, JSON, CSV). Given that many researchers are not yet aware of user licenses, and because open formats are not available for certain types of research data, the Flemish government decided to work in two stages. Up to 2022, datasets are considered to be open if they are accessible to everyone. From 2023, an open license should be added to the dataset.

Monitoring the open data indicator on the FRIS-portal requires metadata information on: the persistent identifier that provides immediate access to the landing page of the dataset, the access possibilities of the dataset (open, embargoed, restricted or closed), a link to the publication that made use of the dataset, and the reasons behind legitimate opt-outs (e.g., privacy, intellectual property rights, ethical aspects, aspects of dual use, other). In addition, information on the license that

---

[7] https://opendefinition.org/

specifies any usage rights is required to measure the more extensive definition of open data that will be in use from 2023.

To be able to apply metrics for the open data indicator, the optional metadata elements 'Access rights' and 'Licenses' had to be made mandatory, and a few properties had to be added on top of the DataCite scheme: 'Open format', 'Link to publication', & 'Legitimate opt-out'. Since open formats (free, non-proprietary formats such as XML, JSON, CSV, etc.) also contribute to the openness of datasets, the WG decided to add an optional property 'Open format' in the form of a toggle check (Yes/No). Adding a new property 'Link to publication' was required to check whether datasets underpinning publicly financed peer-reviewed publications are indeed openly available. The field 'Legitimate opt-out' was added to give data providers the option to specify why the data are not open. Furthermore, the decree on Special Research funding (BOF) obliges universities to also report on the technical format of the dataset, as well as provide a link to the project that produced the dataset. Hence, the property 'Format' that is optional in DataCite, had to be set to 'mandatory' for the universities. In addition, a property

'Link to project' was added to the information model to link datasets to the projects they are resulting from using the project identifier assigned by the funder.

Finally, the Flemish coalition agreement on Open Science states that a FAIR data label has to be assigned to each dataset reflecting its degree of FAIRness. Such a label is not included in the DataCite schema and thus requires an extra property in the Flemish application profile. The FAIR principles - 'Findable', 'Accessible', 'Interoperable' and 'Reusable- were defined in 2016 (Wilkinson et al., 2016) and were later further refined by various initiatives such as GO FAIR[8] and FORCE11.[9] Although the RDA working group "FAIR data Maturity Model"[10] has recently developed a detailed set of FAIR indicators, no measuring instrument has yet been developed at the international level. In anticipation of a European FAIR metric, the FOSB taskforce Metrics is preparing a preliminary FAIR metric that will be used to complete the FAIR data label in the Flemish application profile.

Using the metadata fields from this Flemish application profile for datasets, it is possible to verify whether the objective criteria for the 'F' and 'A' of FAIR have already been fulfilled. For example, "Is there a DOI leading to the landing page of the dataset?".

According to the FAIR principles,[11] Findability minimally requires a PID and rich metadata (at minimum: title, creator(s), contributor(s), description, publication year, keywords, research discipline(s), format, access rights, licenses, links to related publications/projects). However, the FAIR principles do not define 'rich' metadata and this concept is interpreted differently in different FAIR metrics. Accessibility requires that: 1) the (meta)data are publicly available using standardized communication protocols that allow for authentication and authorization processes, and 2) the metadata should persist after the data are no longer available.

To automatically assess the accessibility of the dataset, it is possible to verify whether the metadata provides information on: 1) the access rights of the dataset, and 2) a user license that specifies the conditions for (re)use. Therefore, to automatically score the datasets for findability and accessibility, the information of the following metadata fields of the Flemish application profile for datasets (see Table 1) can be considered:

- for findability: Identifier, Title, Abstract, Creator Name, Creator Identifier, Creator Affiliation, Publication year (and/or Embargo date if applicable), Research discipline, Keyword(s), Contributor type & Contributor name, Format, Version, Link to project(s), Link to publication.

- for accessibility: Access rights, and Licenses.

[8] https://www.go-fair.org
[9] https://www.force11.org/
[10] https://www.rd-alliance.org/groups/fair-data-maturity-model-wg
[11] https://www.go-fair.org/fair-principles/

The metadata property 'FAIR Data Label' (see Table 1) can then be used to provide a score for interoperability and reusability of datasets. Although this metric has yet to be designed within the Task Force Metrics, a project group within the FOSB WG M&S, one possibility is to have the researchers complete a FAIR self-assessment at the end of their research project, and then deliver this score to FRIS using the metadata property 'FAIR Data Label'.

Conclusion

This paper describes the development of a Flemish application profile for research data that will be used by all Flemish information providers to deliver information on datasets to the FRIS-portal. The regulatory framework and the Open Science policy in Flanders require RPOs to provide metadata on datasets to the FRIS-portal, a platform of the Flemish government that discloses research information to the general public. To disclose information on Flemish research data in an unambiguous manner and to perform accurate monitoring on the Flemish Open Science KPIs, the various FRIS data providers have to deliver this information using standardized procedures. This necessitates the creation of a semantically interoperable information model with a common understanding of the meaning of the metadata elements, and the implementation of business and validation rules. To meet the Open Science indicators imposed by the Flemish government, there was a need for a flexible metadata standard that was adaptable to the regional context in Flanders. In consultation with the FRIS data providers and the stakeholders, a Flemish application profile for the delivery of metadata on research datasets to the FRIS-portal was developed. This Flemish application profile for datasets enables the use of automated metrics for the Open Science indicators related to FAIR and Open data.

The paper details the steps in the development process and highlights the influence of various internal and external factors. The development of an application profile, based on a metadata standard, is a complex matter that is influenced by various external factors such as the use of different CRIS-systems by the various stakeholders, political influences, and developments at the European and international level.

Acknowledgements

References

Azeroual, O., Saake, G., & Schallehn, E. (2018). Analyzing data quality issues in research information systems via data profiling. International Journal of Information Management, 41, 50-56.

DataCite Metadata Schema 4.3 [Internet]. 2019. Available from: https://schema.datacite.org/ [Accessed: 2020-12-22].

Day, M. (2005). DCC Digital Curation Manual: Instalment on Metadata. HATII, University of Glasgow; University of Edinburgh; UKOLN, University of Bath; Council for the Central Laboratory of the Research Councils.

Decree of the Flemish Government concerning the support of the Industrial Research Funds and the interface activities of the associations in the Flemish Community [Internet]. 2009, last updated 2019-08-31. Available from: https://codex.vlaanderen.be/PrintDocument.ashx?id=1018147&geann oteerd=false [Accessed: 2020-12-23].

Decree of the Flemish Government concerning the financing of the Special Research Funds at the universities of the Flemish Community [Internet]. 2013, last updated 2019-08-18. Available from: https://data-onderwijs.vlaanderen.be/edulex/document.aspx?docid=14492 [Accessed: 2020-12-23]

Directive (EU) 2019/1024 of the European Parliament and of the Council of 20 June 2019 on open data and the re-use of public sector information, in OJ L 172, 26.6.2019, 56– 83,ELI: http://data.europa.eu/eli/dir/2019/1024/oj.

European Commission. (2014). Digital Agenda for Europe. Retrieved from https://ec.europa.eu/digital-single-market/en/europe-2020-strategy

European Commission. Directorate-General for Research and Innovation. (2016a). COM (2016)178 European Cloud Initiative - Building a competitive data and knowledge economy in Europe. Retrieved from https://eur- lex.europa.eu/legalcontent/EN/TXT/PDF/?uri=CELEX:52016DC0178&from=EN.

European Commission. (2016b). Open Innovation, Open Science, Open to the World. A vision for Europe. Retrieved from https://ec.europa.eu/digital-single-market/en/news/open- innovation-open-science-open-world-vision-europe

European Commission. (2016c). Open access & Data management – H2020 Online Manual. Retrieved from http:// ec.europa.eu/research/participants/docs/h2020-funding- guide/cross-cutting-issues/open-access-dissemination_en.htm.

European Commission. (2017). European Open Science Cloud. New research and innovation opportunities. Retrieved from https://eosc- portal.eu/sites/default/files/eosc_declaration.pdf

European Commission. Directorate-General for Research and Innovation. (2018). EOSC Strategic Implementation Roadmap 2018-2020. Retrieved from https://ec.europa.eu/info/publications/european-open-science-cloud-eosc-strategic-implementation-plan_en

European Commission. (2019). Factsheet Open Science. Retrieved from https://ec.europa.eu/info/sites/info/files/research_and_innovation/knowledge_publications_tools_and_data/documents/ec_rtd_factsheet-open- science_2019.pdf

Flemish application profile for research data [Internet]. 2020. Available from: https://github.com/EvyNeyens1988/FOSB-WG-Metadata-and-Standardisation/blob/master/2020-11-20_FOSB_metadatamodel_V1.7ter_FRIS%20implementation%20guidelines%20v2.xlsx [Accessed: 2020-12-24].

Flemish Government. (2019a). Nota aan de Vlaamse regering. Open Science beleid voor Vlaanderen en de oprichting van de Flemish Open Science Board (FOSB). Retrieved from https://www.ewi-vlaanderen.be/sites/default/files/bestanden/nota_aan_de_vlaamse_regering_-_open_science_beleid_voor_vlaanderen_en_de_oprichting_van_de_flemish_open_science_board_fosb.pdf

Flemish Government. (2019b). Ontwerpbesluit van de Vlaamse Regering houdende de herverdeling van begrotingsartikelen van de algemene uitgavenbegroting van de Vlaamse Gemeenschap voor het begrotingsjaar 2019 in het kader van het Open Science beleid. Retrieved from https://codex.vlaanderen.be/Zoeken/Document.aspx?NUMAC=2020 030067

Flemish Government. (2020). Nota aan de Vlaamse regering. Open Science beleid in Vlaanderen: roadmap, KPI's en meerjarenplan. Retrieved from https://www.ewi-vlaanderen.be/sites/default/files/bestanden/5fc5f514b328e9000c0007f7.pdf

FORCE11 (n.d) Guiding Principles for Findable, Accessible, Interoperable and Re-usable Data Publishing, version b1.0, Retrieved from https://www.force11.org/fairprinciples

GO FAIR (n.d) FAIR Principles. Retrieved from https://www.go-fair.org/fair- principles/.

Hong, N.C. Cozzino, S., Genova, F. Hoffman-Sommer, Marta; Hooft, R., Lembinen, L., Marttila, J., Teperek, M., Ball, M., Barker, M., Berezko, Oleksandr et al. (2020). Six Recommendations for Implementation of FAIR Practice. European Commission. https://doi.org/10.2777/986252

Hurley, B. J. (1999). The Making of America II Testbed Project: a digital library service model. Digital Library Federation.

Johnston, P. (2003). Metadata and interoperability in a complex world. Ariadne, 37. Available: http://www.ariadne.ac.uk/issue37/dc-2003-rpt/

re3data.org Schema 3.0 XML Schema [Internet]. (2021) Available from: https://www.re3data.org/schema [Accessed: 2021-02-21]

Vancauwenbergh, S., & Poelmans, H. (2019). The creation of the Flemish research discipline list, an important step forward in harmonising research information (systems). Procedia computer science, 146, 265-278.

Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., ... & Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. Scientific data, 3(1), 1-9

Annex

Table 1. Flemish Application Profile for Research Data

Term	Value	Definition
Dataset		Data and objects generated or collected by researchers affiliated by a Flemish research institution in the course of their investigations, regardless of their form or method, that form the object on which researchers test a hypothesis. This includes the full range of data: raw, unprocessed datasets, proprietary generated and processed data and secondary data obtained from third parties.
Identifier (M)	value: DOI, handle, ARK, PURL, URN, URL	The Identifier is a persistent and unique string that identifies the landing page of a resource, i.e. a concept DOI, handle, ARK, PURL, URN, URL. Each metadata record has a minimum of 1 persistent identifier. A concept DOI is preferred over a version DOI.
Identifier type (M)	Controlled list: Identifier: DOI, handle, ARK, PURL, URN, URL	The type of the Identifier.
Alternative identifier (O)	Free text field	The alternative identifier is a unique string that identifies a resource, i.e. a DOI, handle, URL or other (ex. Link to data repository like Zenodo or institutional repository, a local accession number). Multiple identifiers can be supplied.

Term	Value	Definition
Identifier type (O)	Free text field	The type of the Identifier.
Abstract (MA)	Free text field	Description of the dataset. Mandatory if no link to a project or publication is provided.
Description (O)	Free text field	Methods; Series Information; Table of Contents; Technical Info; Other.
Creator Name (M)	Free text field	Name of the main researchers involved in producing the dataset.
Creator Identifier (MA)	ORCID	Persistent identifier(s): an ORCID for the main researchers involved in producing the dataset.
Creator Affiliation (M)	GRID	Affiliation to a Flemish Research Performing Organisation of the main (Flemish) researchers involved in producing the dataset.
Title (M)	Free text field	The name by which the research dataset is known, has to be provided in the original language in which the metadata record was published together with the language field. English is preferred. The translation is optional.
Language (O)	controlled values list: ISO 639- 1 norm	The primary language of the resource (=dataset).

Term	Value	Definition
Publisher (MA)	Free text field	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, publisher like Elsevier.
Publication year (M)	format: YYYY	The year when the dataset was or will be available to a community of researchers.
Embargo date (MA)	format: yyyy-mm-dd	Different embargo dates relevant to the availability of the dataset.
Date type (MA)	Data type value: - available	Data type value: - available: date used to indicate the end of an embargo period, sometimes specific conditions are applying (GDPR, ...)
Research discipline (R)	Controlled vocabulary: VODS	Classification of the dataset on the basis of the disciplines of the Flemish Research Standard. (where the research dataset originated, subfield L3)
Keyword(s) (M)	Free text field	Main keyword(s) describing the dataset.
Contributor type (R)	https://schema.datacite.org/meta/kernel-4.3/doc/DataCite-MetadataKernel_v4.3.pdf	The institution or person responsible for collecting, managing, distributing, or otherwise contributing to the development of the resource. To supply multiple contributors, repeat this property. The type of contributor of the resource.

Term	Value	Definition
Contributor name (MA)	Free text field	The full name of the contributor.
Contributor NameIdentifier (R)	value: identifier	Uniquely identifies an individual or legal entity, according to various schemes.
Contributor NameIdentifierScheme (MA)	ORCID GRID	The name of the name identifier
Size (O)	Free text field	Unstructured size information about the dataset. Size (e.g. bytes, pages, inches, etc.) or duration (extent), e.g. hours, minutes, days, etc., of a resource.
Format (M universities / O)	Free text field	Technical format of the dataset.
Open Format (O)	Toggle check	Fully open for all parts of the data and it should be possible to select it if there is an open format in addition to a non- propriety format.
Version (O)		The version number of the dataset.

Term	Value	Definition
Licenses (R / M from 2023)	Examples: no license, Creative Commons CCZero (CC0) Open Data Commons Public Domain Dedication and Licence (PDDL) Creative Commons Attribution 4.0 (CC-BY-4.0),	Any rights information for this resource, commonly expressed through licenses.
Access Rights (M)	values: 'open', 'embargoed' (with indication of expiry date) and 'restricted' or 'closed'	Status on the access possibilities of the dataset. Access rights are specified using the OpenAire 16. Rights (MA) Within this property you have 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

Term	Value	Definition
Legitimate opt-out (R)	values: privacy intellectual property rights ethical aspects aspects of dual use other	Specification on the legitimate opt-out indicated in the field access rights

Term	Value	Definition
Legitimate opt-out clarification (O)	Free text field	Clarification on the reasons behind the legitimate opt-out value indicated in the field legitimate opt-out.
Link to Project(s) (MA)		Via FRIS object 'project' attached with a project identifier. The code assigned by the funder to a sponsored award (grant).
Link to Publication (MA)	values: DOI, handle,….	Via FRIS object 'data' attached with an identifier for publications (DOI, handle).
Fair Data Label (M)		To be determined in Task Force metrics. Align with EOSC FAIR WG. Based on RDA FAIR Data Maturity Model