



A list of recommendations to improve the workflow of dataset metadata in the Flemish research landscape.



FRDN.
VLAANDEREN

UHASSELT

KNOWLEDGE IN ACTION

Evvy Neyens, Evelien Dhollander, Dieuwertje Bloemen,
Kevin Leonard, Werner Brebels, Tom Wuyts,
Pascale Dengis, Marc Portier, & Hanne Poelmans

ecoom

Agenda

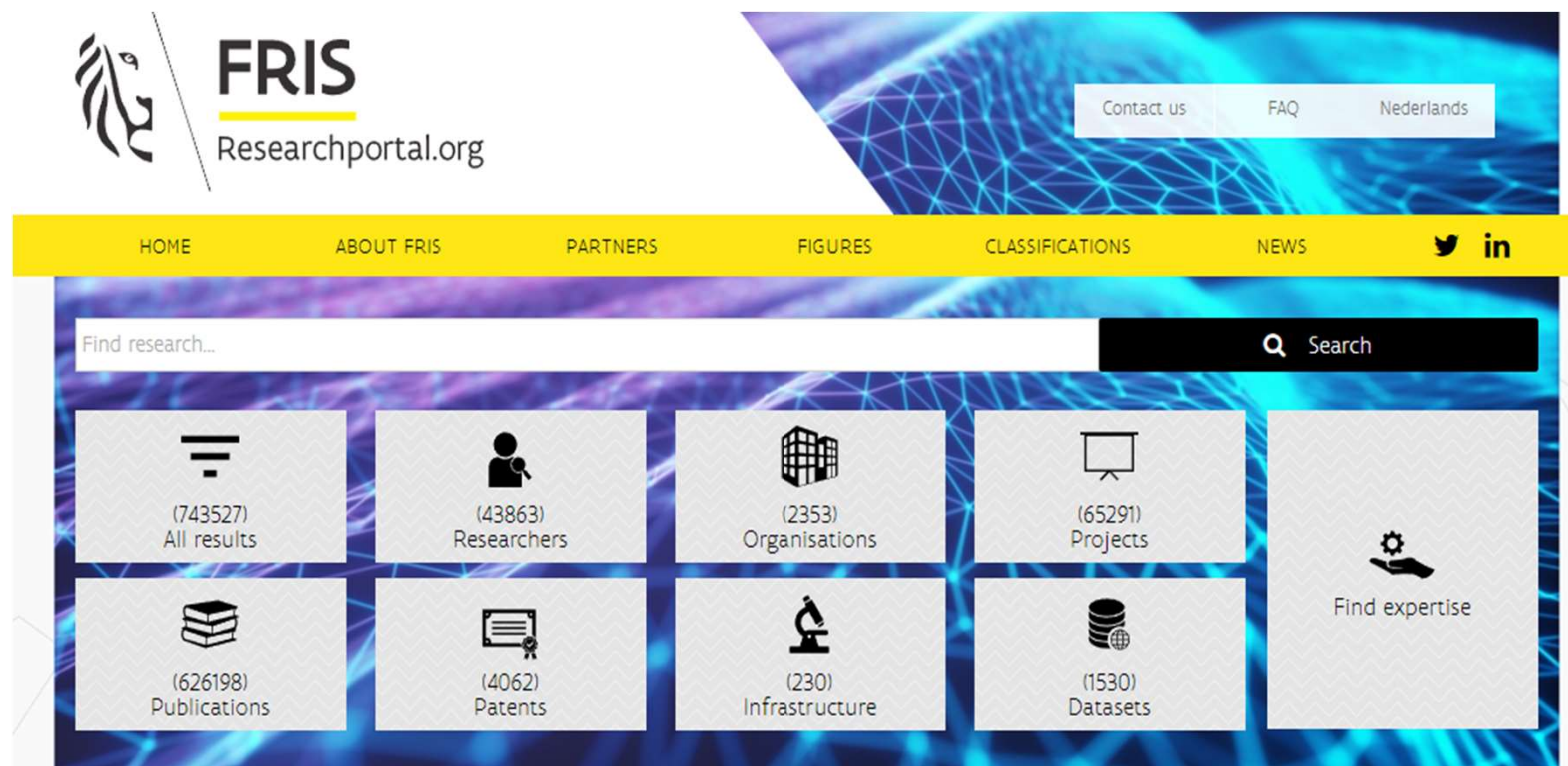
Dataset metadata in the Flemish Research Landscape

1. Background & legislation
2. Scope of the Project Group Enriching Metadata
3. Method
4. Current workflow & limitations
5. Recommendations for improvement

Flanders Research Information Space

<https://researchportal.be/en>

- Regional CRIS-system & BI-tool
 - monitor R&D, Innovation & Open Science KPIs
- Research Discovery platform: linked to EOSC via OpenAIRE
- RPOs deliver research information in real-time using CERIF-XML
- FRIS Portal built and hosted by Department EWI, Flemish Government supported by Expert Centre R&D monitoring (ECOOM-UHasselt)
 - classifications & information models used for visualizing & reporting



Background & legislation

- Legal obligation for RPOs to deliver metadata on research objects to FRIS
 - researchers, organisations, projects, publications, datasets, infrastructure, & patents
 - Special & industrial research fund decrees (BOF/IOF), decrees of science & innovation, Open Science Policy etc.
- **Flemish Open Science Policy (2019)**
 - EU guideline: make publicly financed research data 'As open as possible, as closed as necessary'
 - Flemish Open Science Board (FOSB) & Flemish Research Data Network (FRDN) coordinated by Research Foundation Flanders (FWO)
- **Open Science Roadmap:**
 - 5 KPIs: ORCID, DMP, FAIR data, Open data, Open Access
 - KPIs monitored via FRIS-portal

Open data KPI & FOSB Application Profile for research dataset metadata

Open data KPI

- Datasets linked to publicly funded, peer-reviewed journal articles should become openly available: 'as open as possible principle'
- mandatory registration of dataset metadata and transmission to FRIS for all datasets underlying peer-reviewed, Flemish-funded publications from 2019 onwards

WG Metadata & standardization chaired by ECOOM-UHasselt

⇒ FOSB Application Profile for dataset metadata

- ⇒ 31 properties: 16 originating from DataCite
- ⇒ Requirements (M, MA, R, O), Values, Definition & example
- ⇒ Monitor open data KPI: Access rights & licenses (Open definition)

Term	DataCite		Mandatory (M), Mandatory if Applicable (MA), Recommended (R), Optional (O)	KPI implications	Value	Definition
Data - Identifier	identifier	https://support.datacite.org/docs/schema-mandatory-properties-v43#1-identifier	M	Mandatory for KPI	value: DOI, handle, ARK, PURL, URN, URL	The Identifier is a persistent a resource, i.e. a concept. A record has a minimum of 1 a version DOI.
Data - Identifier type	identifier: identifierType	https://support.datacite.org/docs/schema-mandatory-properties-v43#1-identifier	M		Controlled list: Identifier: DOI, handle, ARK, PURL, URN, URL	The type of the Identifier.
Data - Alternative identifier	alternateIdentifier	https://support.datacite.org/docs/schema-optional-properties-v43#1-alternateIdentifier	O		Free text field	The alternative identifier is handle, URL or other (ex. Li repository, a local accession)
Data - Identifier type	identifier: alternateIdentifierType	https://support.datacite.org/docs/schema-mandatory-properties-v43#1-identifier	O		Free text field	The type of the Identifier.

FRDN Project group Enriching Metadata

- Map current flow of dataset metadata
⇒ optimised, more efficient flow

- Sep. 2022 – Dec. 2023

- Co-chairs:

Evy Neyens (ECOOM-UHasselt)

Evelien Dhollander (UGent)

- Members:

Dieuwertje Bloemen (KU Leuven)

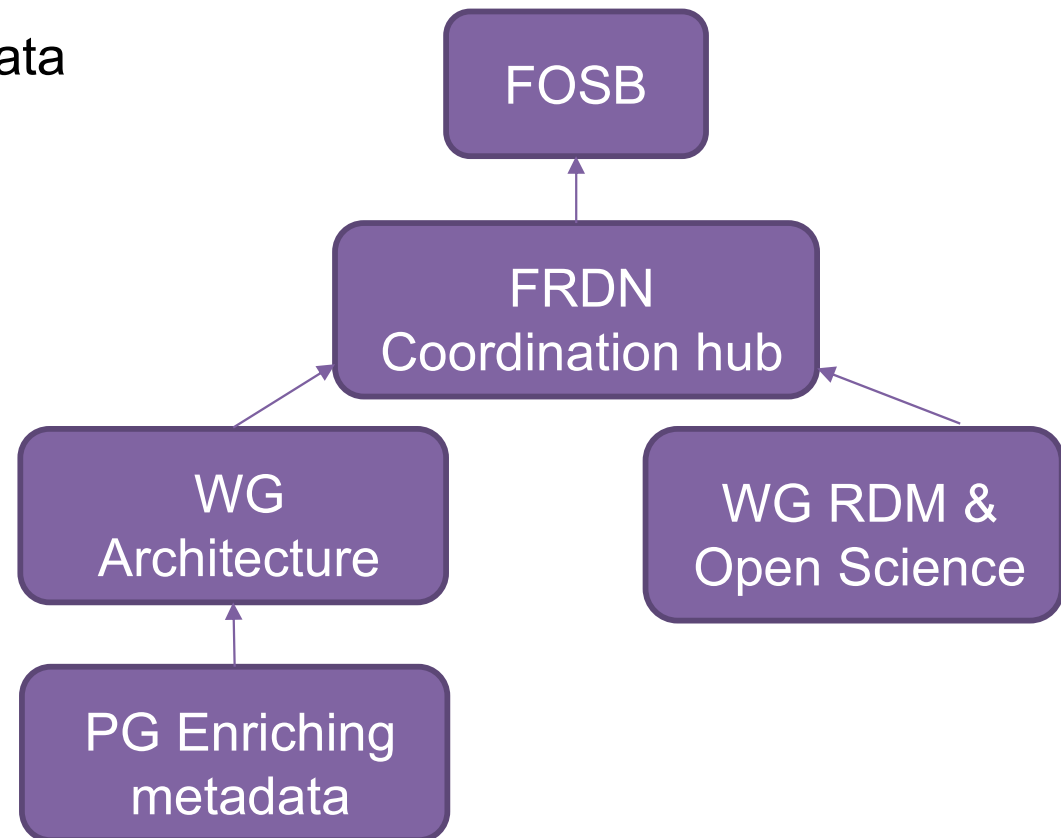
Kevin Leonard (UGent)

Werner Brebels (UHasselt)

Pascale Dengis (EWI-FRIS)

Tom Wuyts (EWI-FRIS)

Marc Portier (VLIZ)



PG Enriching metadata

Scope & deliverables



1. Identify points of improvement + optimize metadata flow
2. Solution to harvest dataset metadata from external repositories
3. Save time & money
4. Improve quality & findability



Key deliverable 1:

- Business Process Model of current metadata flow
- A list of points of improvement within the current ecosystem

Key deliverable 2:

- Proof of concept for harvesting metadata from external repositories



Principles:

- enter once, reuse often
- As simple as possible
- Decentralized & future proof solution



Method

Survey



Part 1: current approaches to dataset metadata collection & publishing

Part 2: identify points of improvement & ideal metadata flow



Bilateral structured interviews focusing on part 2



12 RPOs: 5 universities, 3 strategic research centres,
4 scientific institutions

9 of which participated in the follow-up interviews



Survey + interview responses: qualitative thematic analysis

Points of improvement identified + translated to recommendations



Results

Current approach dataset metadata collection & publishing



6 RPOs: record dataset metadata via CRIS/metadata catalogue
6 RPOs: dataset metadata catalogue in process



9 RPOs: no obligation to register dataset metadata in RDM policy
2 RPOs: obligation to register dataset metadata for
Flemish-funded research
1 RPO: data sharing policy



6 RPOs use FOSB Application Profile for dataset metadata
3 RPOs use several schemes
3 RPOs do not record any dataset metadata yet



Current flow dataset metadata

Step 1: Researcher publishes dataset in external repository

Minimal descriptive metadata + PID recording

Notify KI/register metadata in CRIS/IR

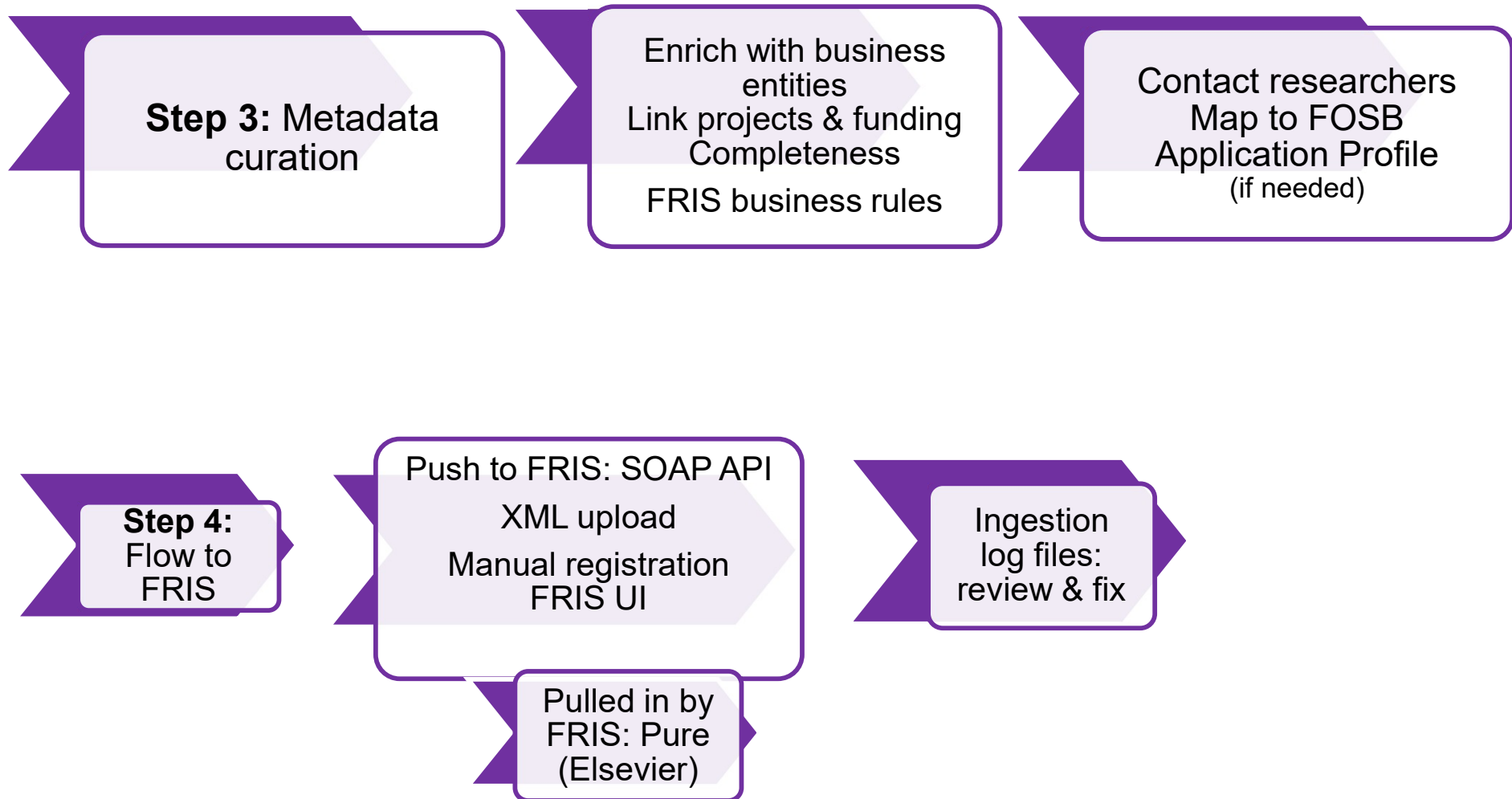
Step 2: Metadata registration in CRIS/metadata repository

Template FOSB standard/other standard

Autocomplete metadata fields via PID

Metadata enrichment: add extra fields FOSB standard

Current flow dataset metadata



Current flow dataset metadata

Retroactive harvesting of dataset metadata (< 2019)

- 3 out of 5 Flemish universities have a procedure for retroactive harvesting of dataset metadata
- 2 out of 7 participating Strategic Research Centres & Scientific institutions

Data stewards:

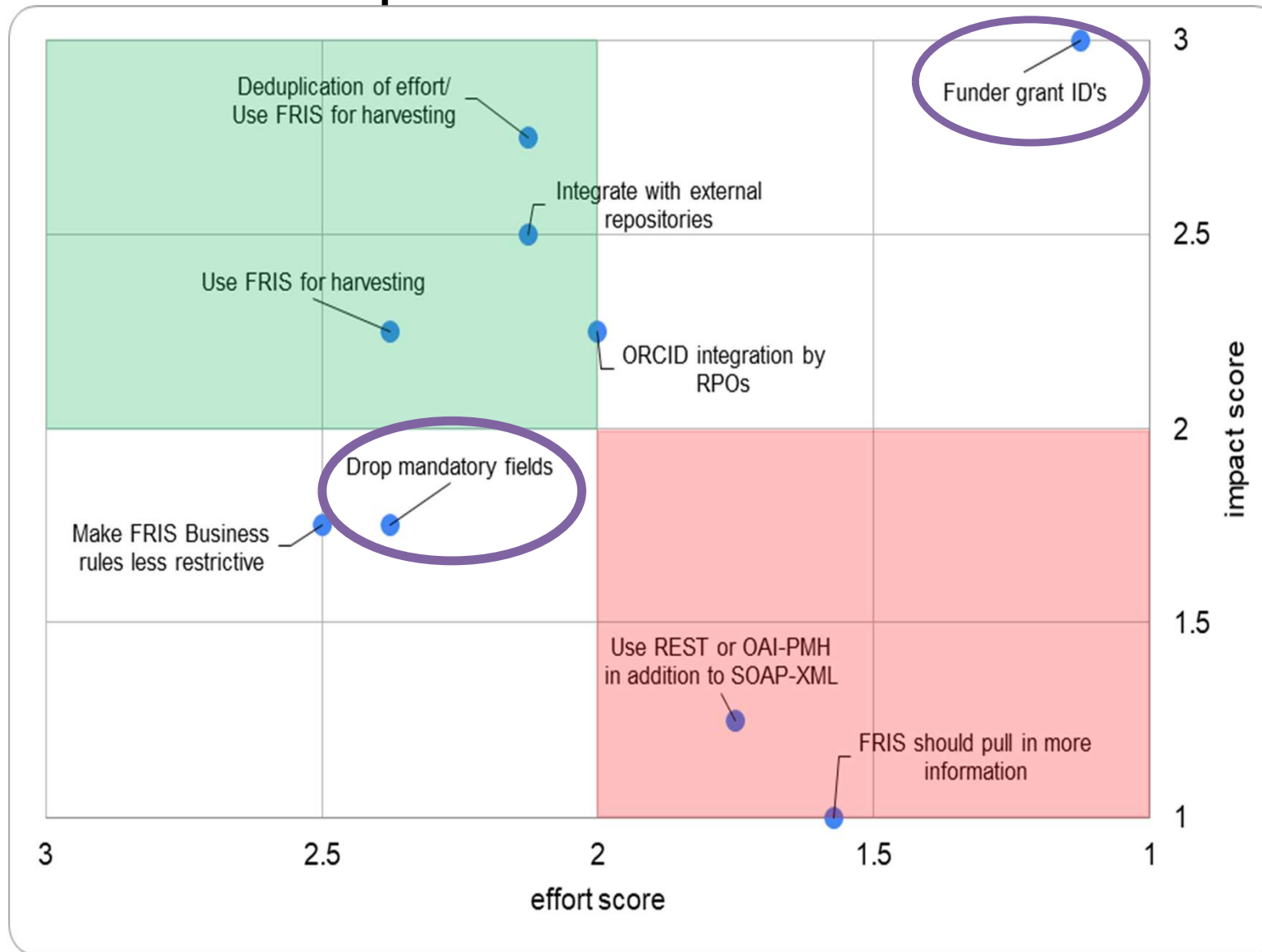
- Query external sources via SoP (OpenAIRE Explore, DataCite, Zenodo etc.)
- DataCite & OpenAIRE APIs + specific DOIs
- Data Monitor (Elsevier)
- RDR Dataverse API (KU Leuven)

⇒ Pull in metadata + contact researchers to complete the records

- DOI registration if none is available

Prioritisation exercise

Impact vs easiness matrix



Results

Key deliverable 1: List of recommendations



1. Deduplication of effort: use FRIS for harvesting
2. Integrate with external repositories
⇒ Key dev. 2: Zenodo report
3. ORCID integration by RPOs
4. Adapt specific FOSB Application Profile metadata fields from mandatory to recommended

NEYENS, E., Dhollander, E., Bloemen, D., Leonard, K., Brebels, W., Wuyts, T., Dengis, P., & Portier, M. (2024). Dataset Metadata in the Flemish Research Landscape. Zenodo.
<https://doi.org/10.5281/zenodo.10634664>



Results

Funder Grant IDs



5. Flemish strategy for Funder Grant Identifiers

⇒ Establish FRDN WG to develop a Flemish PID strategy

Recommendation with the highest impact
However ... not easy to implement



ECOOM@uhasselt.be

Thank you.

Evy Neyens

Evy.neyens@UHasselt.be

