FOSB WG Metadata & Standardization

UHASSELT
KNOWLEDGE IN ACTION

Sadia Vancauwenbergh, 2020-10-26

Scope – generisch metadata model

- Generisch metadata model (based on DataCite; Model_v1.0):
  - Connectie met FRIS en decentrale systemen op technische en semantisch eenduidige wijze
  - Rekening houdend met
    - FRIS
    - Vlaamse onderzoeksinstellingen
    - FOSB KPI document:
      - ORCID, DMP, FAIR data, Open data
      - Meetmethodiek in afstemming met FOSB WG RDM & OS
    - FOSB architectuurplan
    - In lijn met EOSC, en andere internationale standaarden
    - Coherent met ISA², W3C DCAT-AP
  - Timing: Q4 2020

Scope – generisch metadata model

- Definitie onderzoeksdata

  In afstemming met FOSB WG'en en consultatie

  Data and objects generated by researchers affiliated with a Flemish research institution in the course of their investigations, regardless of their form or method. This includes the whole range of data: raw, unprocessed datasets, proprietary generated and processed data and secondary data obtained from third parties. Examples of research data include, but are not limited to, notes, surveys, images, objects, audiovisual files, spreadsheets, databases, statistical data, geographical data, simulations, software developed for research purposes, samples of any kind including of biological material, personal data, patient data, …

  Rekening houdend met KPI ORCID

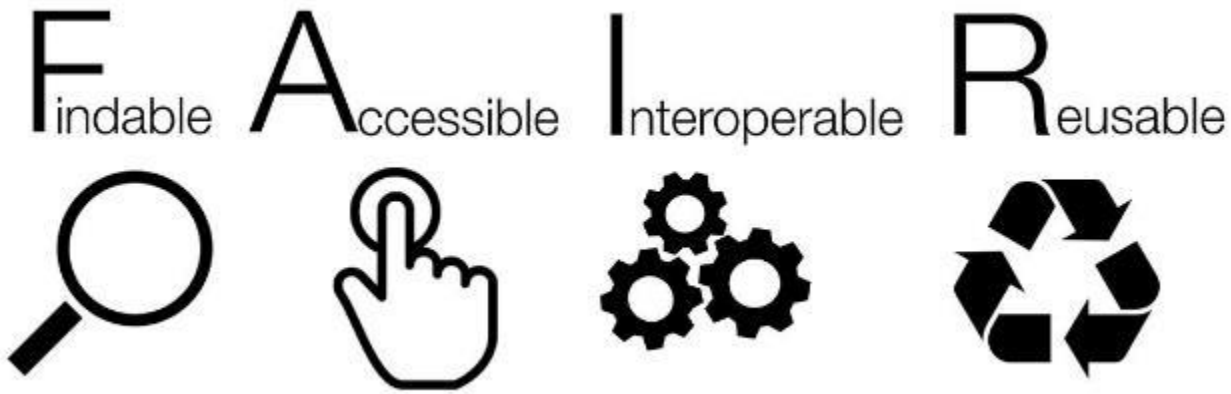
Scope – generisch metadata model

Term	Definition
Dataset	Data and objects generated or collected by researchers affiliated by a Flemish
Identifier	The Identifier is a unique string that identifies the landing page of a resource, i.e. a
Alternative	The alternative identifier is a unique string that identifies a resources, i.e. a DOI,
Identifier type	The type of the Identifier.
Abstract	Description of the dataset. Mandatory if no link to a project or publication is
Creator Name	Name of the main researchers involved in producing the dataset.
Creator Identifier	Persistent identifier(s): an ORCID for the main researchers involved in producing the
Creator Affiliation	Affiliation to a Flemish Research Producing Organisation of the main (Flemish)
Title	The name by which the research dataset is known, has to be provided in the original
Language	The primary language of the resource.
Publisher	The name of the entity that holds, archives, publishes prints, distributes, releases,
Publication year	The year when the dataset was or will be available to a community of researchers.
Embargo date	Different embargo dates relevant to the availability of the dataset. Yes/No value
Date type	Data type values: issued, accepted, available.
Research discipline	Classification of the dataset on the basis of the disciplines of the Flemish Research
Keywords	Main keywords describing the dataset. (free text field)
Contributor	The institution or person responsible for collecting, managing, distributing, or
Size	Unstructured size information about the dataset.
Format	Technical format of the dataset.
Technical	Technical description of the dataset.
Version	The version number of the dataset.
IP rights	Intellectual property rights for the dataset
Restrictions	
Open data status	Status on the access possibilities of the dataset: OpenAire 16. Rights (MA: rightsURI
Trusted repository	Core Trust Seal
Data - link to	List of the project/grant agreement identifier(s).
Data - link to	
Data - link to DMP	Through the DOI Fabrica web interface or the API of the DataCite Metadata
Data - link to	
Fair Data Label	

Scope

- Generisch metadata model (based on DataCite; Model_v1.0):
  - Connectie met FRIS en decentrale systemen op technische en semantisch eenduidige wijze (standaard, proces, procedure)
  - Rekening houdend met
    - FRIS, Vlaamse onderzoeksinstellingen
    - FOSB KPI document:
      - ORCID, DMP, **FAIR data**, **Open data**
      - **Meetmethodiek in afstemming met FOSB WG RDM & OS**
      - **Vastleggen van kwaliteitsregels FA/IR**
    - FOSB architectuurplan
    - In lijn met EOSC, en andere internationale standaarden
  - Timing:
    - Q1 2021 (impact FAIR), Q3 2021 (Open Data), Q4 2021 (rules FA), 2022 (rules FAIR), 2023 implementatie FAIR

FAIR Principes



- RDA FAIR Data Maturity Model WG
- FAIRsFAIR project: data assessment metrics
- DANS: FAIR Stars
- GO-FAIR
- EOSC FAIR WG → results end December 2020

Source: Wilkinson et al., 2016; https://www.rd-alliance.org/groups/fair-data-maturity-model-wg; https://www.go-fair.org/; https://www.fairsfair.eu/

FAIR Metrics

DANS

- F, A, & I as separate dimensions of data quality
- Score datasets on each dimension (1-5)
- Reusability = average FAIRness
  ⇒ indicator of data quality
- Make scoring automatic

FAIRSFAIR
Fostering Fair Data Practices in Europe

- 15 metrics built on existing work
  - RDA FAIR Data Maturity Model
  - DANS Fairdat/FAIREnough
  - WDS/RDA Data Fitness
- Iteratively improve via pilot tests
- v0.3 since August 2020

FIELD	DESCRIPTION
Metric Identifier	FsF-F1-01D
Metric Name	**Data** is assigned with a globally unique identifier.
Description	A data object may be assigned with a globally unique identifier such that it can be referenced unambiguously by humans or machines. Globally unique means an identifier should be associated with only one resource at any time. Examples of unique identifiers of data are Uniform Resource Identifier (URI) such as URL and URN, Digital Object Identifier (DOI), the Handle System, identifiers.org, w3id.org and Archival Resource Key (ARK). A data repository may assign a globally unique identifier to your data or metadata when you publish and make it available through their services.
FAIR Principle	F1. (Meta) data are assigned globally unique and persistent identifiers
CoreTrustSeal Alignment	R13. The repository enables users to discover the data and refer to them in a persistent way through proper citation.
ASSESSMENT	
Requirement(s)	• Data identifier (IRI, URL) • List of globally unique identifier schemes
Method	Check if the data identifier is specified based on a globally unique identifier scheme.
COMMENTS	
Related Resources:	

Open Data

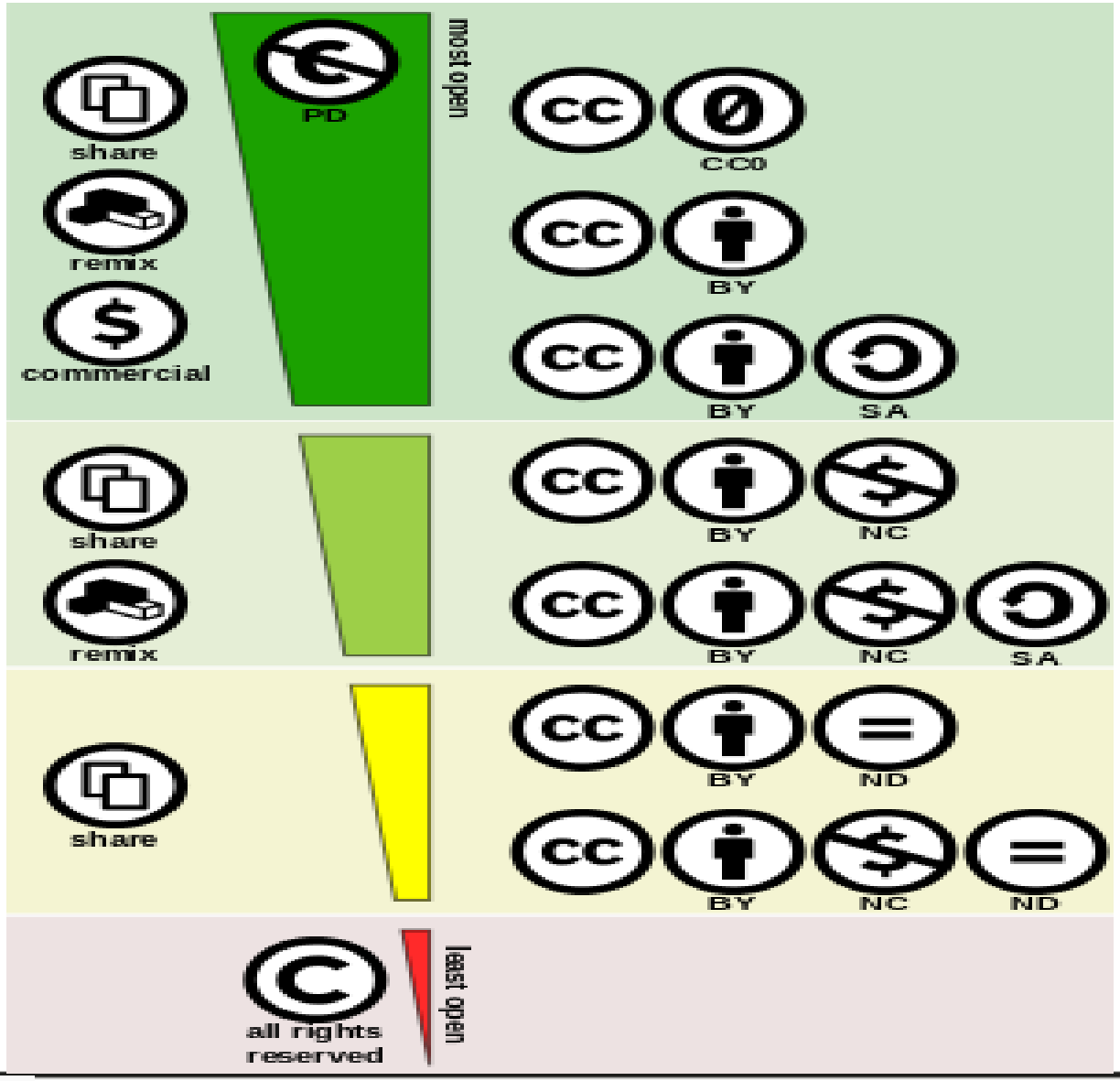
The Open Definition: *Open data can be **freely used, modified, and shared** by **anyone** for **any purpose***.

- Open & machine-readable format
- Open license

License	Domain	By	SA	Comments
Creative Commons CCZero (CC0)	Content, Data	N	N	Dedicate to the Public Domain (all rights waived)
Open Data Commons Public Domain Dedication and Licence (PDDL)	Data	N	N	Dedicate to the Public Domain (all rights waived)
Creative Commons Attribution 4.0 (CC-BY-4.0)	Content, Data	Y	N	
Open Data Commons Attribution License (ODC-BY)	Data	Y	N	Attribution for data(bases)
Creative Commons Attribution Share-Alike 4.0 (CC-BY-SA-4.0)	Content, Data	Y	Y	
Open Data Commons Open Database License (ODbL)	Data	Y	Y	Attribution-ShareAlike for data(bases)

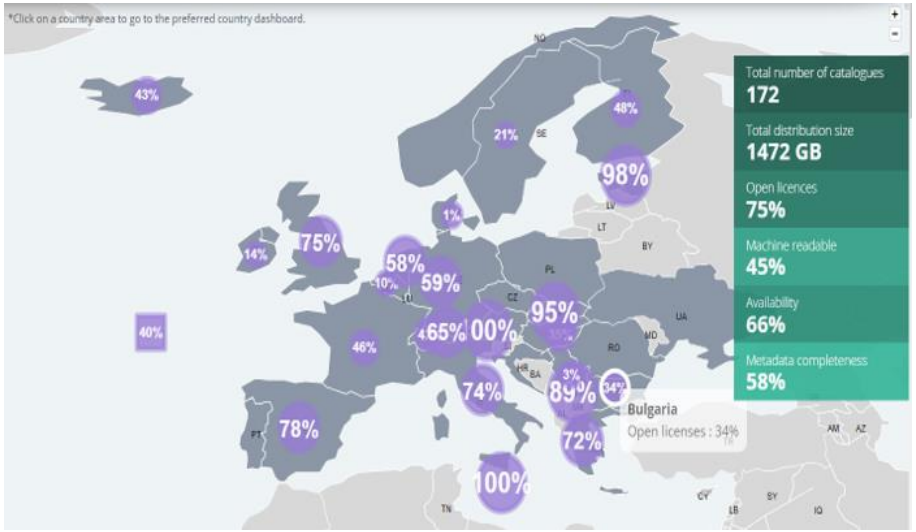
Source: Open Knowledge Foundation, http://opendefinition.org/

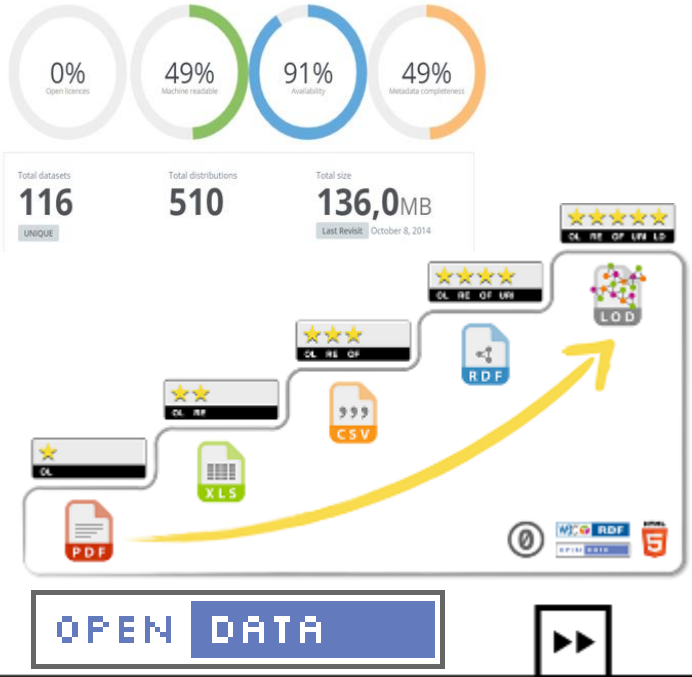
Creative Commons Licenses (Green = open)



Open Data Metrics

- The Open Data Monitor
  - **Availability**
  - **Open license**
  - Open formats
  - **Machine-**
  - **Metadata completeness**
  - Discoverability
  - Overall Quality
- The Open Data Institute
  - 5 stars of Linked Open Data
  - Certificates
- Open Knowledge Foundation
  - Yes/No vs Degree of Openness (stars)





Scope – Disciplinary standards

- Disciplinaire standaarden – task force

- Task force:
  - Geclusterd rond disciplinaire onderzoeksthema's
  - Rekening houdend met de beschikbare praktijken aanwezig in ondermeer de ESFRI's
  - Gebaseerd op internationale standaarden, bv. fairsharing.org
  - Rekening houdend met werk in ISA², DCAT-AP

  - Ontwikkeling/inclusie van een standaard, geharmoniseerd proces en protocol voor disciplines (cfr. Doelstellingen generisch datamodel)

  - Timing:
    - Q4 2020 – Q4 2021

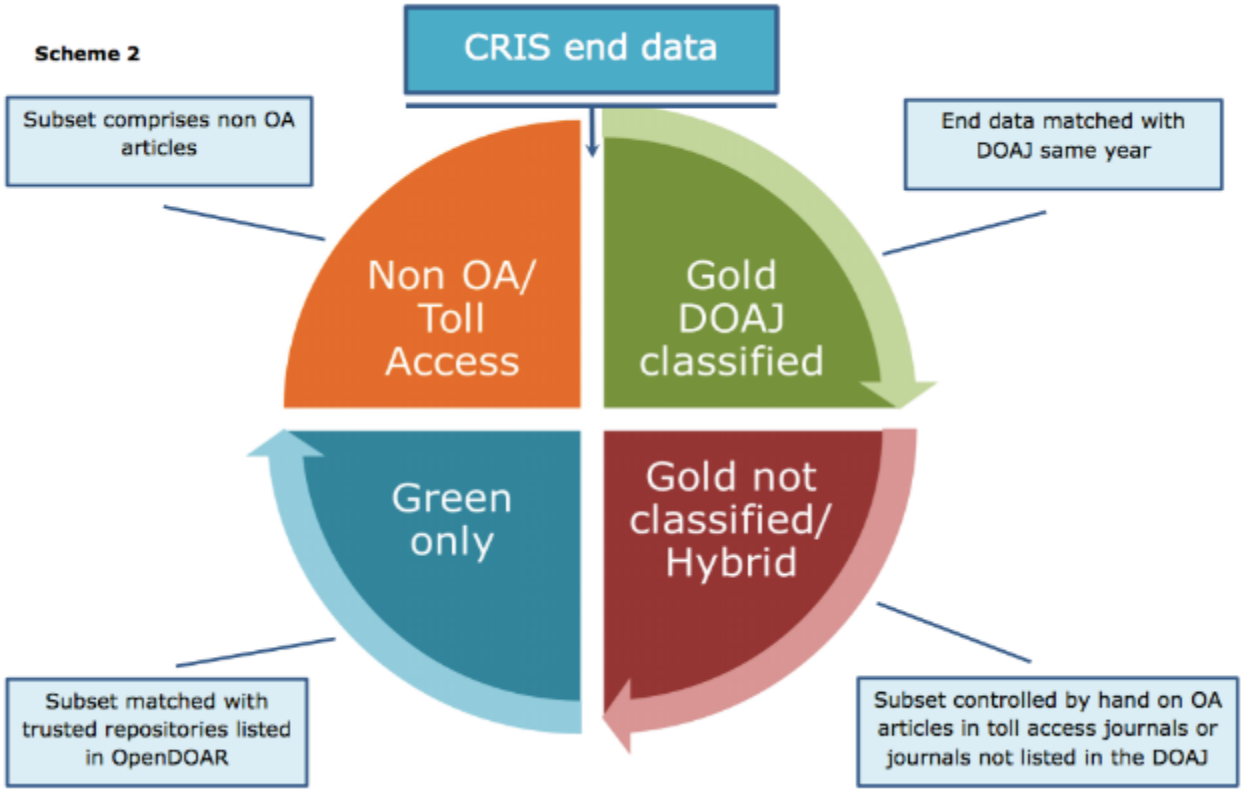
Scope – Data management plan

- Data management plan – task force

  - Bepaling van een geharmoniseerde DMP-template vragen ism FOSB WG RDM & OS (Q1 2021)
  - Definiëren van procesflow voor efficiënt samenwerken op 1 DMP
  - Ontwikkelen van standaard voor machine actionable DMPs:
    - Informatie uitwisseling tussen systemen namens de stakeholders betrokken in de research data life cycle (onderzoekers, RDM-stewards, ICT, …)
    - Faciliteert automatische uitwisseling, integratie en validatie van de informatie in DMPS en reduceert hierdoor workload voor de stakeholders
    - Rekening houdend met:
      - DMP workflows
      - Gebruik van PIDs en controlled vocabularies
      - Standaard data model voor DMPs, beschrijving van componenten,
      - Business en validation rules, automatisatie namens stakeholders
      - Beschikbaar maken van DMPS voor geautomatiseerd gebruik
      - Faciliteren van DMP evaluatie en monitoring
      - Faciliteren van DMP als levend document in termen van update, versioning
      - Faciliteren om DMPs publiek beschikbaar te maken waar relevant
  - Timing
    - Opmaken SvZ (Q4 2021)
    - Uitwerken stappenplan (Q4 2022)

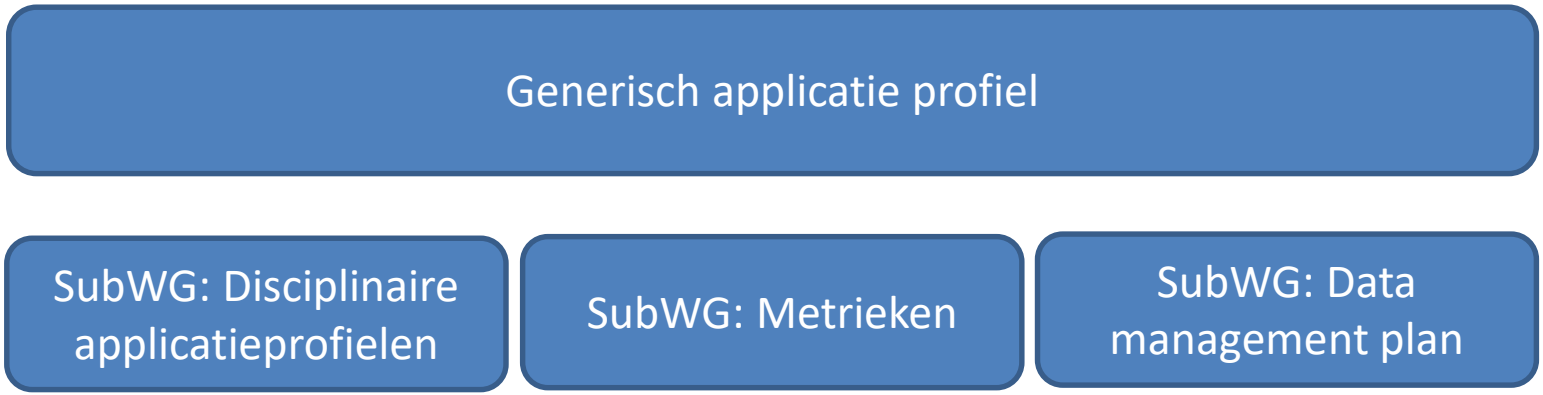
Scope – Open Access metrics

- Ontwikkeling van gedifferentieerd OA label en monitoringsmethodiek
- Analyse van verrijkingsprocessen bv. door beschikbaar stellen van metadata in repositories via Unpaywall, DOAJ (Q4 2021)



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

FOSB WG Metadata en Standaardisatie



UHASSELT
KNOWLEDGE IN ACTION

Questions? Suggestions?

Sadia.Vancauwenbergh@uhasselt.be

UHASSELT

KNOWLEDGE IN ACTION