

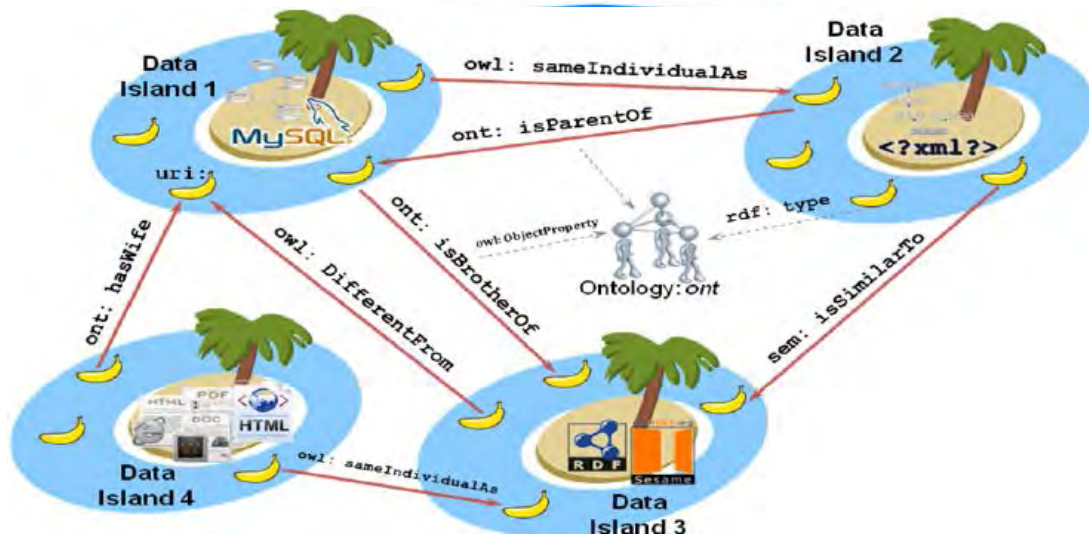


Data interoperability in Europe: where we are and where we are going?

Milan Ojsteršek, University of Maribor, Slovenia

milan.ojstersek@um.si

How to achieve interoperability between data islands?



Meta)data Interoperability principles:

- (Meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.
- (Meta)data use vocabularies that follow FAIR principles.
- (Meta)data include qualified references to other (meta)data.

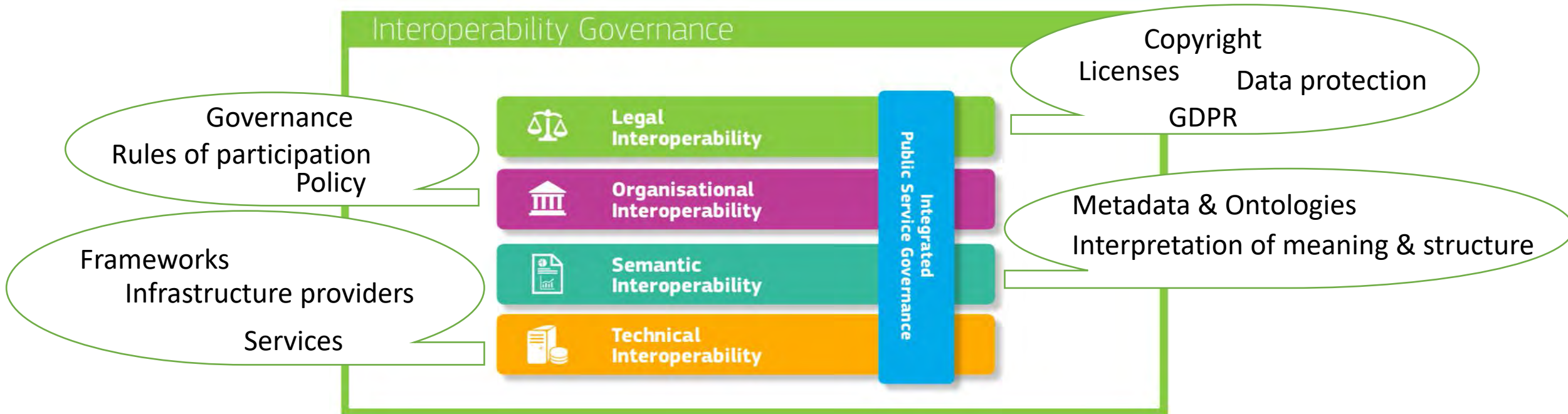
Source: https://commons.wikimedia.org/wiki/File:Islands_of_Data.svg

Source: https://www.researchgate.net/publication/267692879_Towards_Executable_Reality_Business_Intelligence_on_Top_of_Linked_Data/figures?lo=1

Source: Wilkinson, M. D. et al. The FAIR Guiding Principles for scientific data management and stewardship. Sci. Data 3:160018 doi:10.1038/sdata.2016.18 (2016)



Layers of interoperability



Source: [The European Interoperability Framework four levels of interoperability](#)

Who influences data interoperability in Europe?

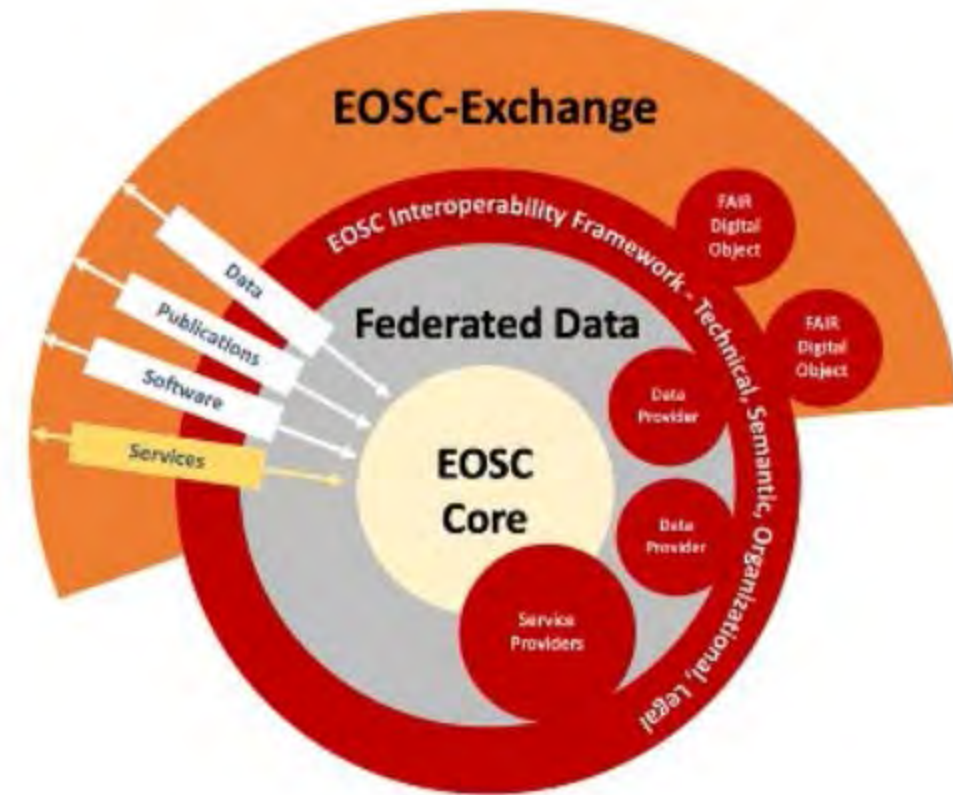


- [European Open Science Cloud \(EOSC\)](#)
- [EU research infrastructures](#) ([ELIXIR](#), [CLARIN](#), [CESSDA](#), [DARIAH](#), [BBMRI](#), ...)
 - [European Strategy Forum on Research Infrastructures \(ESFRI\)](#)
 - [European Research Infrastructure Consortium \(ERIC\)](#)
- [Pan-European e-Infrastructures](#) ([GEANT](#), [EUDAT](#), [OpenAire](#), [EGI](#), [PRACE](#)...)
- EU and EOSC projects ([FAIRsFAIR](#), [EOSC Enhance](#), [EOSC Pilot](#), [EOSC hub](#), [SSHOC](#), [ESCAPE](#), [EOSC Future](#), [DICE](#), [EGI-ACE](#), [OpenAIRE Advance](#), [OpenAire-Nexus](#), [ByCOVID](#), [ENVRI-FAIR](#), [PaNOSC](#), [ExPaNDS](#), [EOSC-Life](#), ...)
- EU and global initiatives ([RDA](#), [CODATA](#), [FAIR digital object forum](#), [GoFAIR](#), [DDI](#), [COAR](#), [IVOA](#)...).
- [EOSC regional projects](#) ([EOSC Nodic](#), [NI4OS-Europe](#), [EOSC- Pillar](#), [EOSC Synergy](#)).

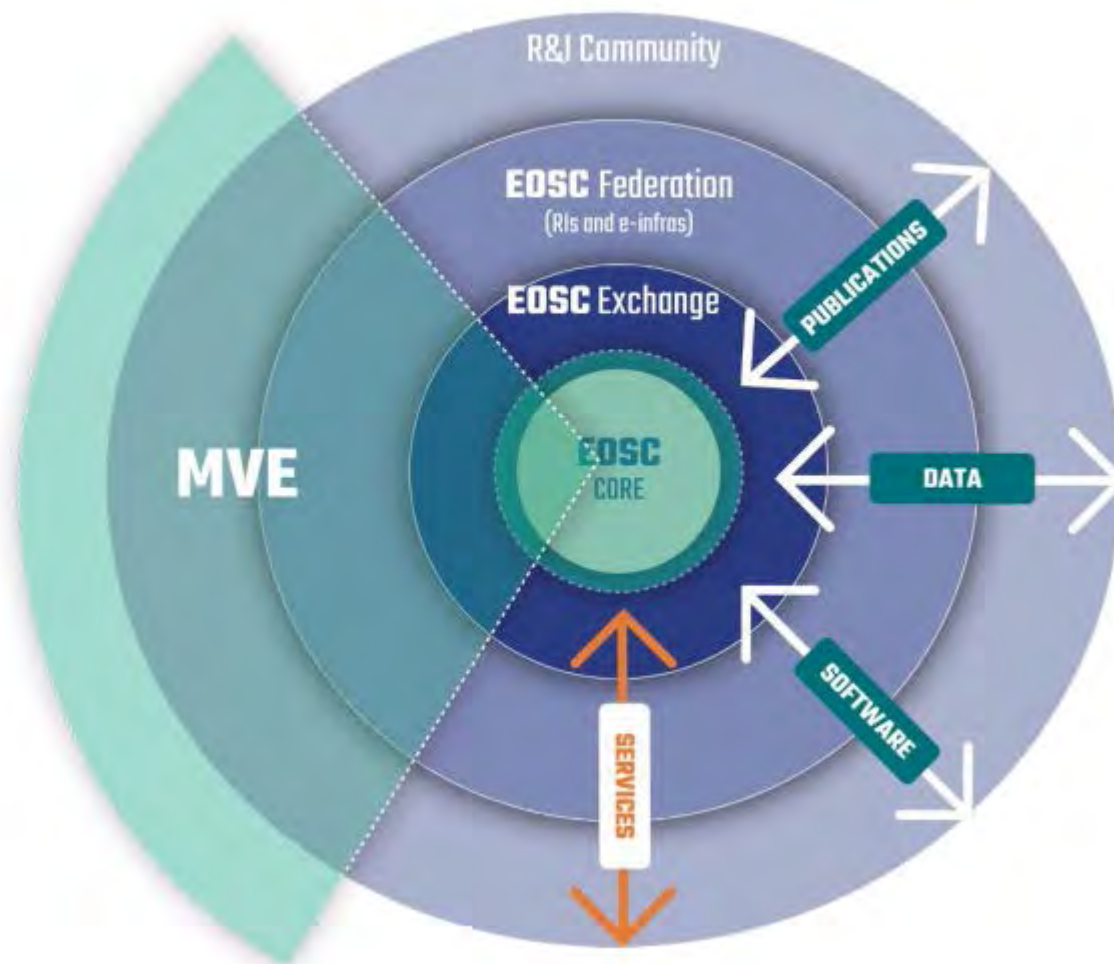
European Open Science Cloud vision



- A web of FAIR data and services.
- Federation of eInfra and Research Infrastructures (RIs).
- Environment in which data can be brought together with services



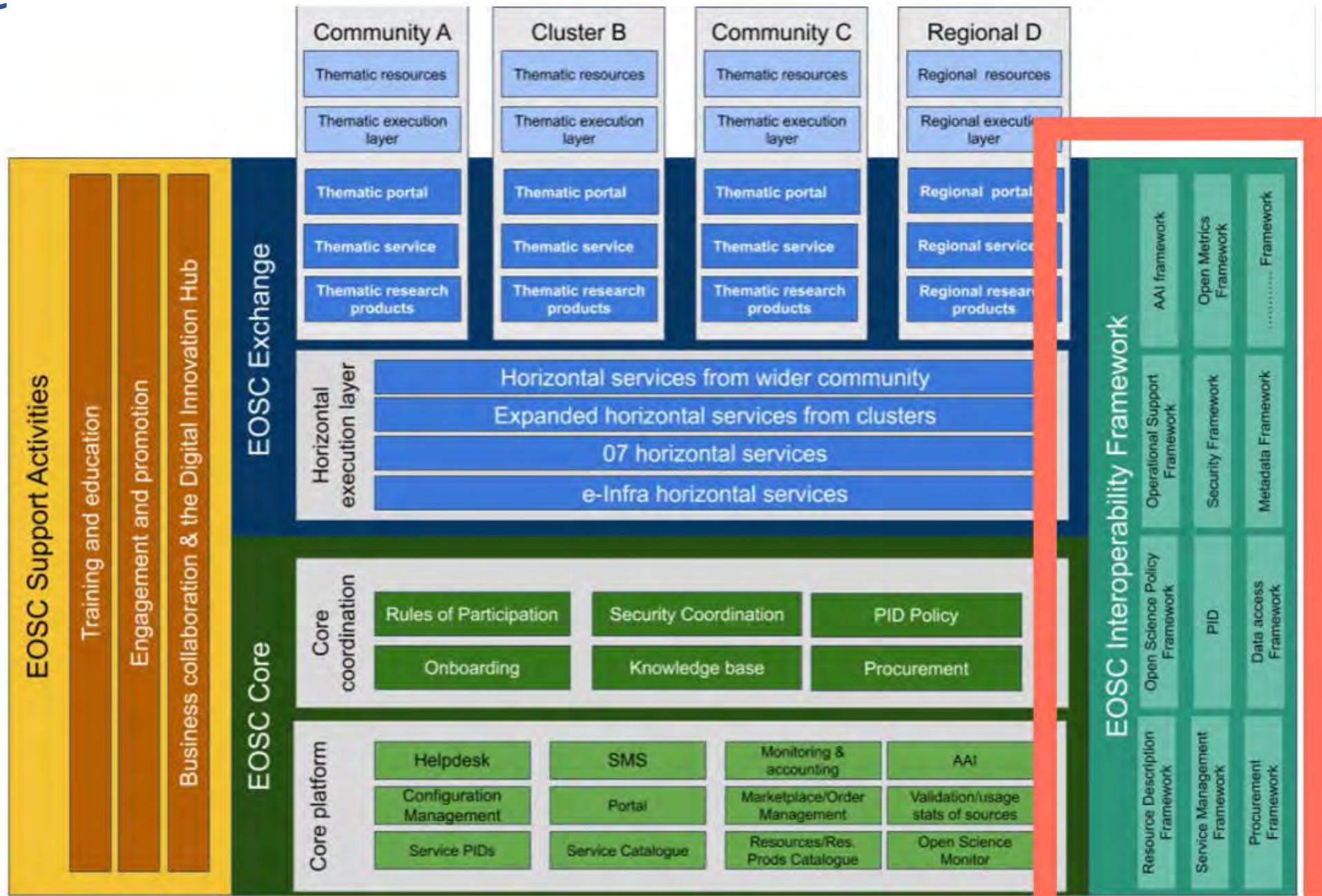
Minimal Viable EOOSC



Source: [EOOSC Core](#)

- Core Services
- EOOSC Portal & Market Place
- Interoperability Framework
- Support Service

EOSC Architecture – EOSC interoperability framework



Source: [EOSC Future](#)

EOSC providers



**International
 Science Council**



e-Infrastructures

Ex: EUDAT, EGI, national e-Infrastructures (e.g. national research clouds), OpenAIRE, PRACE, FENIX, D4Science

- ✓ Software & other research products
- ✓ Storage
- ✓ Computing
- ✓ Services
- ✓ Training



Research Infrastructures

Ex: ERICs, ESFRIs, EMODnet, Copernicus, DIAS (Data and Information Access Services), clusters

- ✓ Software & other research products
- ✓ Storage
- ✓ Datasets
- ✓ Computing
- ✓ Services
- ✓ Training



Higher Education Institutes (HEIs)

- ✓ Software & other research products
- ✓ Publications
- ✓ Services
- ✓ Storage
- ✓ Datasets



Research Institutes

Ex: Research performing organisations, universities, research centres

- ✓ Software & other research products
- ✓ Datasets
- ✓ Publications



Archives & Repositories

Ex: Institutional thematic and national repositories, archives and aggregators

- ✓ Software & other research products
- ✓ Datasets
- ✓ Publications



Computing & Data Centers

Ex: BCS, CSC, PSNC, CERN

- ✓ Software & other research products
- ✓ Storage
- ✓ Computing
- ✓ Services
- ✓ Training



Publishers & Journals

- ✓ Publications



Libraries

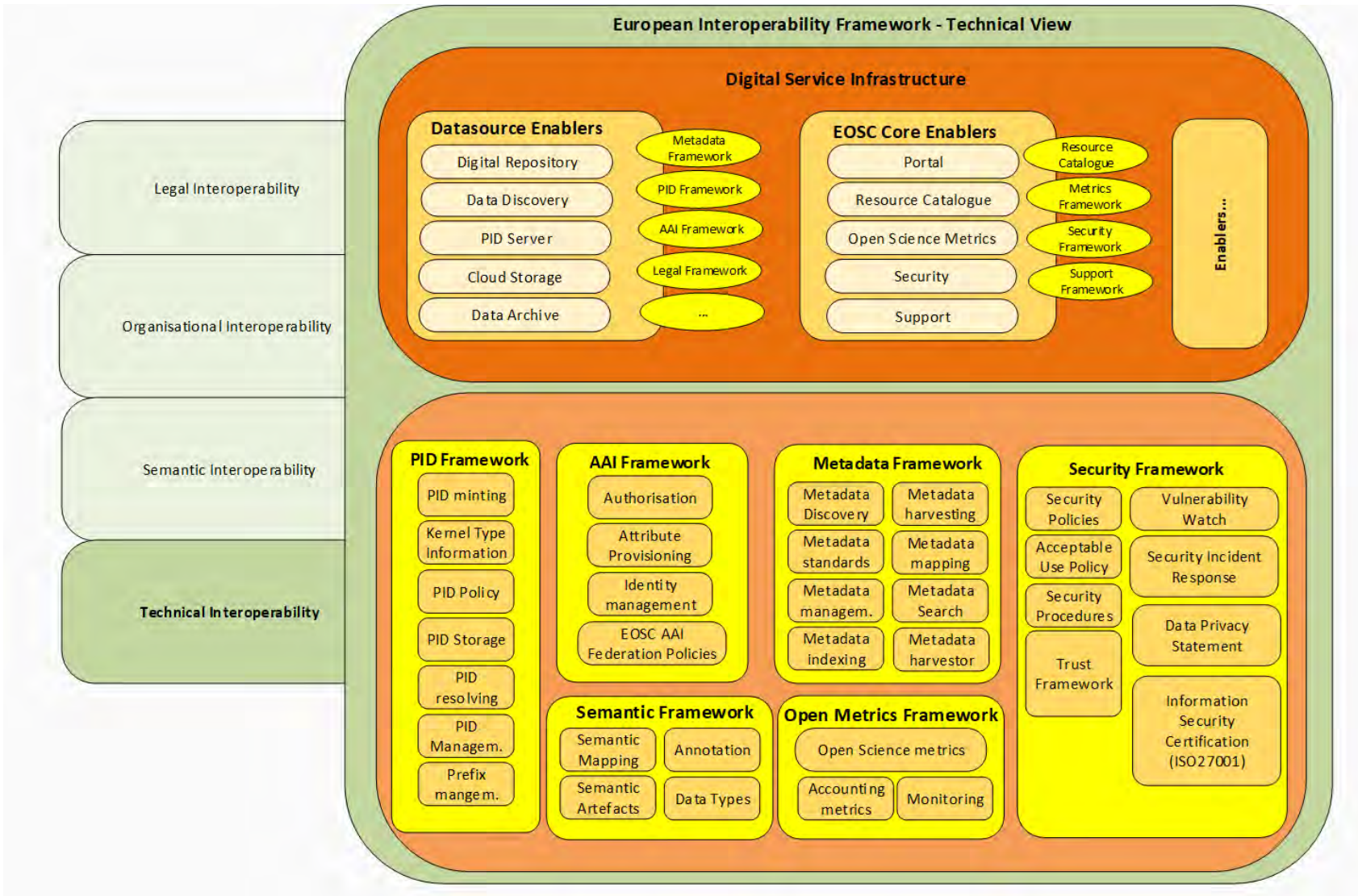
Ex: University libraries, institutional libraries, national libraries, LIBER

- ✓ Publications
- ✓ Datasets
- ✓ Training

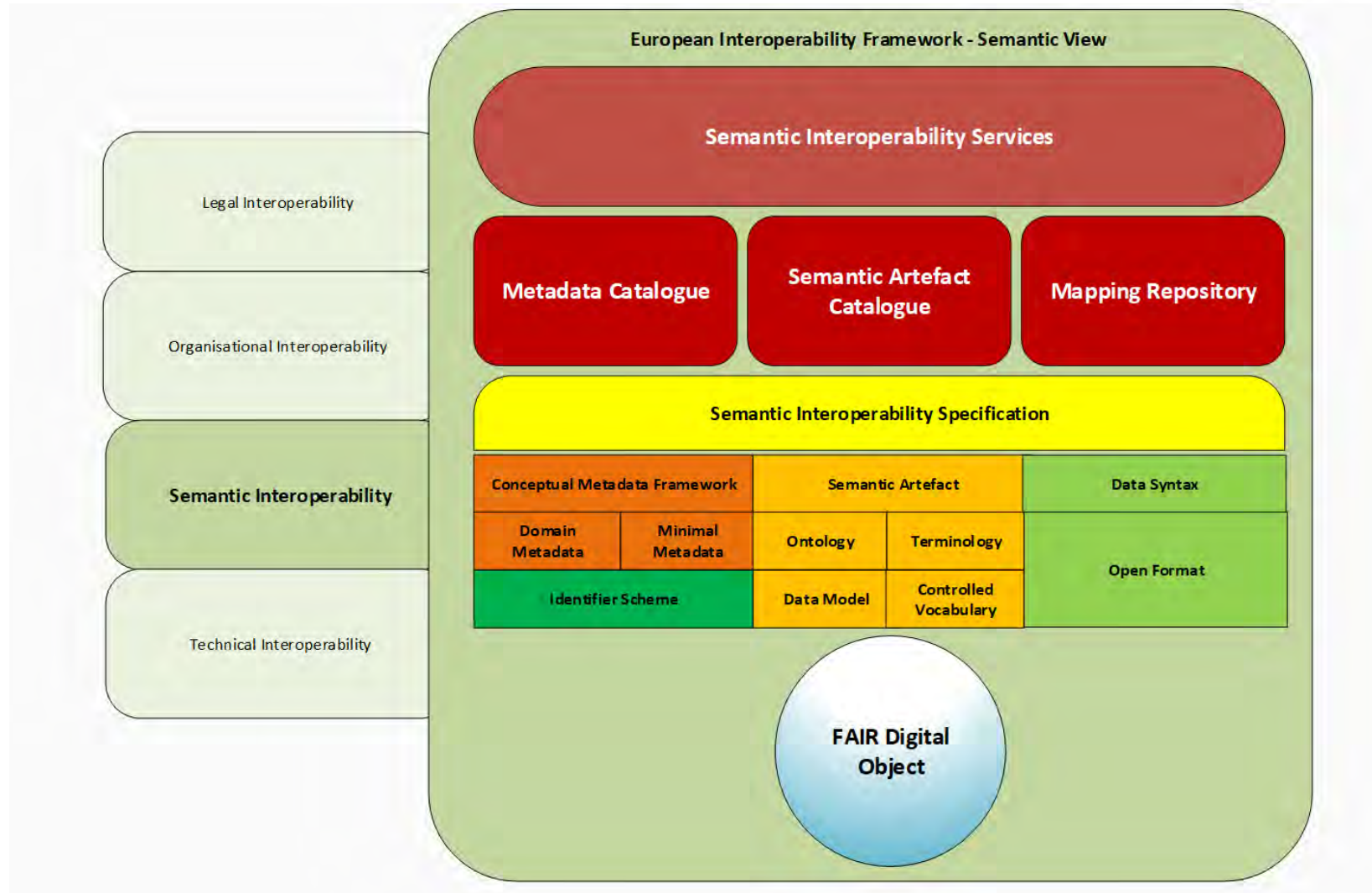


Commercial Providers

- ✓ Software & other research products
- ✓ Storage
- ✓ Datasets
- ✓ Computing
- ✓ Services
- ✓ Training
- ✓ Publications



Source: [The EOSC Interoperability Technical view](#)



Source: [The EOSC Interoperability Framework](#) Semantic view



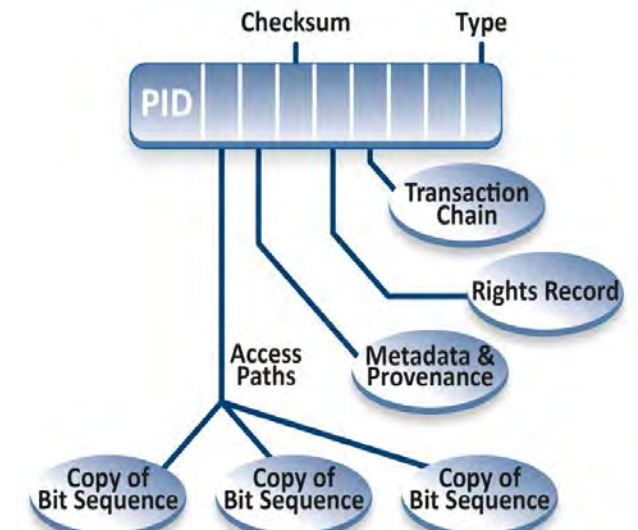
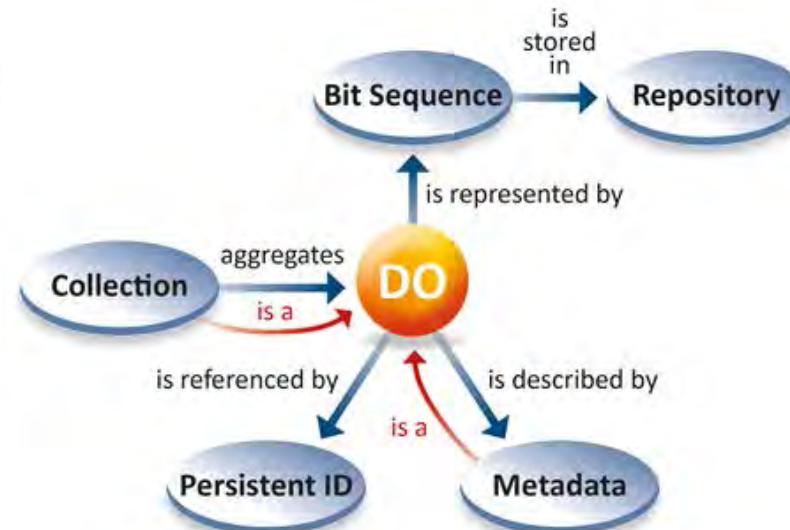
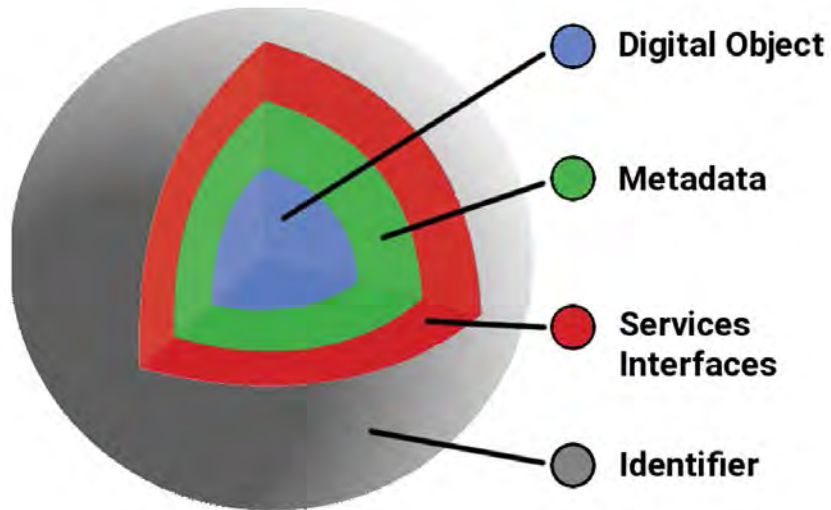
- EOSC Strategic Research and innovation agenda: <https://eosc.eu/sria>
- EOSC interoperability framework: <https://op.europa.eu/sl/publication-detail/-/publication/d787ea54-6a87-11eb-aeb5-01aa75ed71a1> and <https://doi.org/10.2777/620649>
- EOSC architecture working group view on the minimum viable EOSC. Report from the EOSC Executive Board Working Group (WG) Architecture: <http://doi.org/10.2777/492370>
- Eriksson, van de Sanden, Kurowski, Coppens, Corcho, & Ojsteršek. (2021, January 5). EOSC Interoperability Framework Reference Architecture (Version 1.0). Zenodo. <http://doi.org/10.5281/zenodo.4420096>
- Graber-Soudry, Ohad, Minssen, Timo, Nilsson, Daniel, Corrales, Marcelo, Wested, Jakob, & Illien, Bénédicte. (2021, January 27). Legal Interoperability and the FAIR Data Principles (Version 1.0). Zenodo. <http://doi.org/10.5281/zenodo.4471312>
- [Licentia](#) a web service application with the aim to support users in licensing data. Their goal is to provide a full suite of services to help in the process of choosing the most suitable license depending on the data to be licensed.
- [ODRL representation of public licences](#)
- EOSC rules of participation: <http://doi.org/10.2777/30541>
- Wilkinson, M. D. *et al.* The FAIR Guiding Principles for scientific data management and stewardship. *Sci. Data* 3:160018 doi: 10.1038/sdata.2016.18 (2016).

Important documents



- FAIRSFair deliverables: <https://www.fairsfair.eu/reports-deliverables>
- SSHOC deliverables: <https://sshopencloud.eu/publications/deliverables>
- PID architecture for EOSC: <https://op.europa.eu/sl/publication-detail/-/publication/3136c3e6-4f07-11eb-b59f-01aa75ed71a1/language-en/format-PDF/source-search> and <http://doi.org/10.2777/525581>
- A Persistent Identifier (PID) policy for the European Open Science Cloud (EOSC): <http://doi.org/10.2777/926037>
- EOSC Authentication and Authorization infrastructure (AAI): <https://op.europa.eu/sl/publication-detail/-/publication/d1bc3702-61e5-11eb-aeb5-01aa75ed71a1/language-en/format-PDF/source-search> and <http://doi.org/10.2777/8702>
- AARC Blueprint architecture: <https://aarc-project.eu/architecture/>
- EOSC Future provider days: <https://eoscfuture.eu/eventsfuture/provider-days/>
- WG DataIO: Data interoperability landscape references - <https://docs.google.com/document/d/1q4L-nxDYQYES5qKaB3z-sRTRI4diKAPY/edit>

FAIR digital object

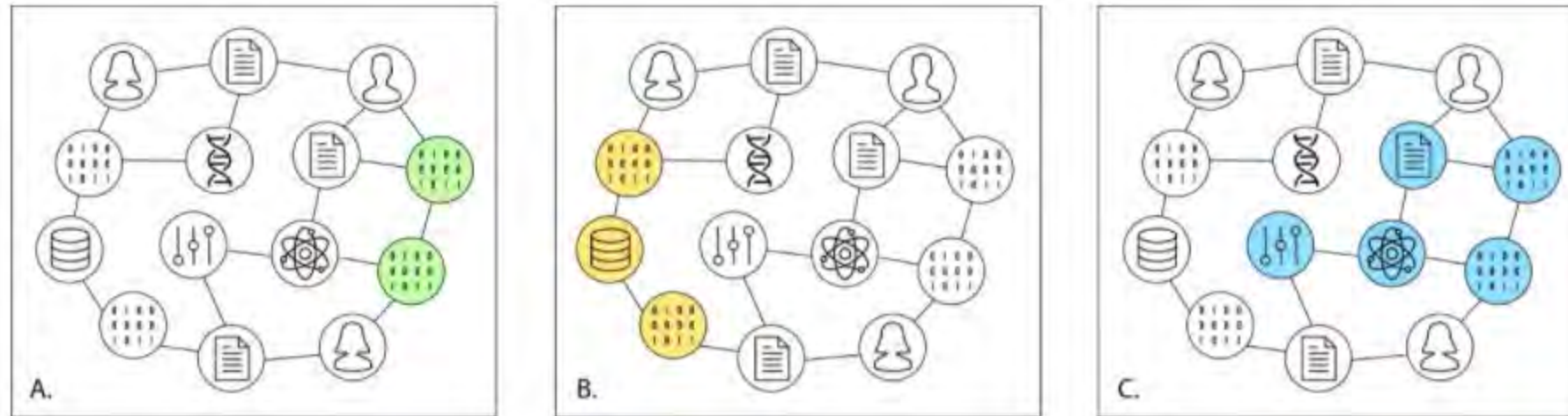


Digital Object Interface Protocol

Source: RDA's Data Foundation & Terminology Group (DFT) 2014:
 Core Model

Source: Schwardmann, U., 2020. Digital Objects – FAIR Digital Objects: Which Services Are Required?. *Data Science Journal*, 19(1), p.15. DOI: <http://doi.org/10.5334/dsj-2020-015>

PID graph



A – linkage of different version of software

B – datasets stored in the repository

C – linkage of different digital objects of the research project

Manghi, P., et al.: Openaire research graph dump (2019). <https://doi.org/10.5281/zenodo.3516918> and <https://graph.openaire.eu/>

Source: Martin Fenner and Amir Aryani. Introducing the PID Graph, Datacite blog - <https://doi.org/10.5438/jwvf-8a66>

Motivation for metadata crosswalk



- Many metadata schemas, application profiles, and guidelines from metadata aggregators.
- Poor quality of metadata in repositories.
- Controlled vocabularies are not used in metadata elements.
- No appropriate mappings were made between different versions of metadata schemas, application profiles, controlled vocabularies, and other semantic artifacts.
- Metadata schemas allow the entering of character strings for most metadata elements.
- Many repository platforms don't use authority control for persons, places, corporate names, projects, and research groups.
- Metadata elements in different schemas are in different granularity levels.
- Metadata schemas and application profiles do not have all metadata elements defined that are needed for citation, findability, reusability, accessibility, and interoperability.
- EOSC recommendations of minimum metadata set for repositories and data archives.



- Crosswalks from schemas to schema.org:
<https://docs.google.com/spreadsheets/d/1P6WH8h4OnIVR9UJj3FcOebNUpLnKNBCuvEp3NsLRh04/edit#gid=1789151191> and <https://rd-alliance.org/group/research-metadata-schemas-wg/outcomes/collection-crosswalks-fifteen-research-data-schemas>
- Dataverse Metadata Crosswalk: <https://goo.gl/yN2f9V>
- Schema.org Dataset Mapping:
<https://docs.google.com/spreadsheets/d/16HNJVKUdueVIPedcp3x2HXI0RJ4zrlpQWrTIkAf-IB4/edit#gid=0>
- EDM I Metadata properties, use cases and mappings:
<https://docs.google.com/spreadsheets/d/1dtHpbp5cVaooVdqhvDjLHKM5Y8IfC-iRSU6OA6BLSUg/edit#gid=1110916251>
- Mapping onto EUDAT-B2FIND Metadata Schema:
<http://b2find.eudat.eu/guidelines/mapping.html>

Crosswalk of most used metadata schemes and guidelines for metadata interoperability



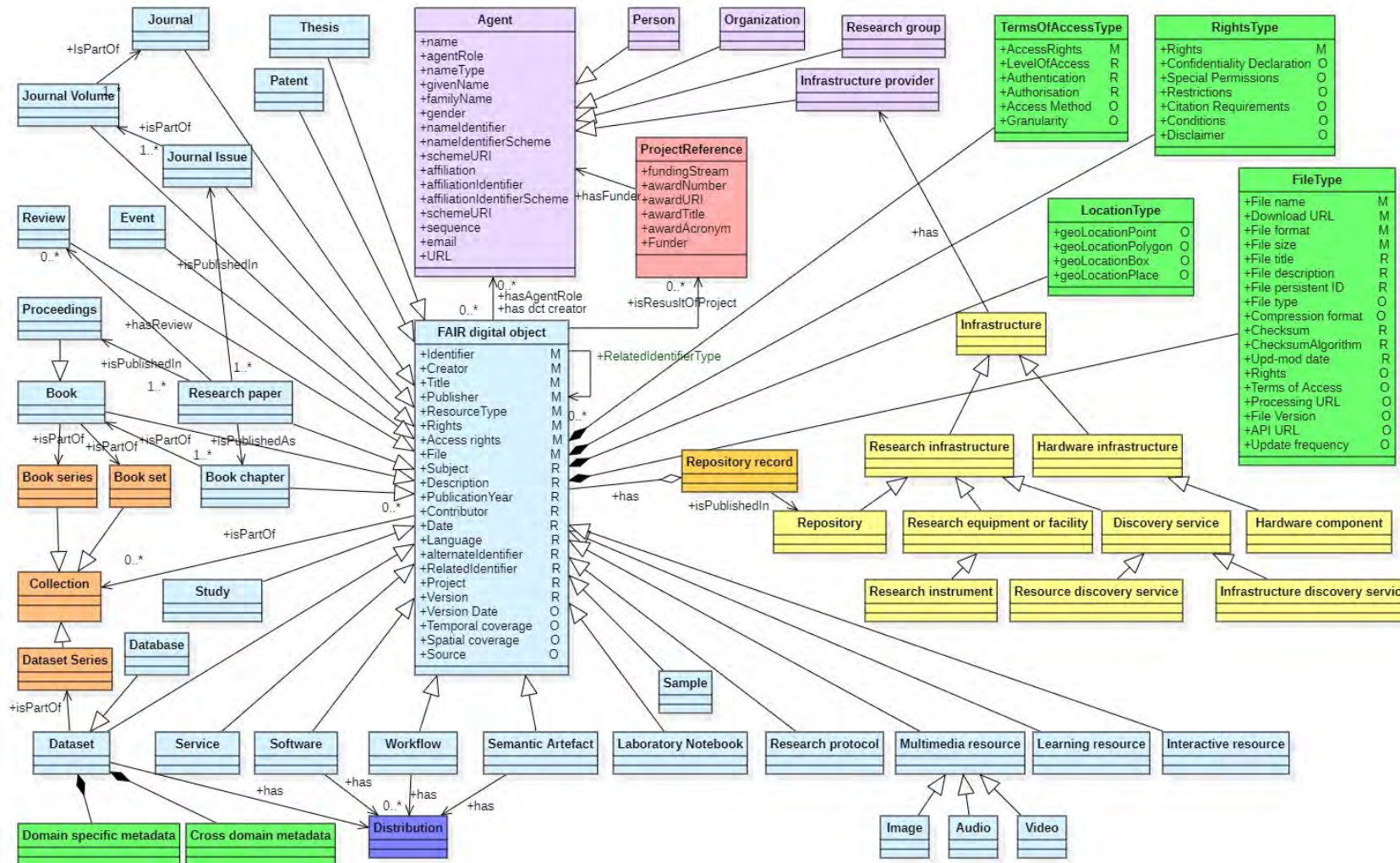
- [RDA metadata IG recommendation of the metadata element set,](#)
- [EOSC Pilot - EDM metadata set,](#)
- [Dublin CORE Metadata Terms,](#)
- [Datacite 4.3 metadata schema,](#)
- [DCAT 2.0 metadata schema and DCAT 2.0 application profile,](#)
- [EUDAT B2Find metadata recommendation,](#)
- [OpenAIRE Guidelines for Data Archives,](#)
- [OpenAire Guidelines for literature repositories 4.0,](#)
- [OpenAIRE Guidelines for Other Research Products,](#)
- [OpenAIRE Guidelines for Software Repository Managers,](#)
- [OpenAIRE Guidelines for CRIS Managers,](#)
- [Crossref 4.4.2 metadata XML schema,](#)
- [Harvard Dataverse metadata schema,](#)
- [DDI Codebook 2.5 metadata XML schema,](#)
- [Europeana EDM metadata schema,](#)
- [Schema.org and RDA crosswalks from schemas to schema.org,](#)
- [Bioschemas and Schema.org Dataset Mapping and](#)
- [The PROV Ontology](#)

Most commonly used controlled vocabularies in the crosswalk



- Datacite vocabularies
- Crossref vocabularies
- Vocabularies used in OpenAIRE
- Other vocabularies (MARC, Dublin Core, COAR, Schema.org)

EOSC minimal metadata Application profile



Catalogues

- [OpenAire \(guidelines\)](#)
- [B2Find \(mapping onto EUDAT-B2FIND Metadata Schema\)](#)
- [CESSDA data catalogue](#)
- [NI4OS Europe catalogue](#)
- [FAIRsharing](#)
- [Agroportal](#)
- [Bartoc](#)
- [EOSC portal and marketplace](#)
- [re3data.org](#)



**International
Science Council**

- [ECRIN Clinical Research Metadata Repository](#)
- [BBMRI-ERIC Sample Directory](#)
- [WorkflowHub](#)
- [SSHOC service catalogue](#)
- [Social Sciences & Humanities Open Marketplace](#)
- [PaNOSC federated dataset catalogue](#)
- [IVOA Registry of Registries](#)
- [ELIXIR Core data resources](#)



**International
Science Council**

