ONTO COMMONS

@ontocommons

linkedin.com/company/ontocommons

info@ontocommons.eu

ONTOLOGY-DRIVEN DATA DOCUMENTATION FOR INDUSTRY COMMONS

Towards Materials and Manufacturing Commons - the enablers Digital Marketplaces, FAIR Principles and Ontologies
Berlin – April 4th – 6th 2023

Creating federated FAIR Data Space with the FAIR Data Point

Yann Le Franc – CEO and Sicentific Director – e-Science Data Factory (France)

ylefranc@esciencefactory.com

ORCID: 0000-0003-4631-418X

OntoCommons "Ontology-driven data documentation for Industry Commons" has received funding from the European Union's Horizon Programme call H2020 -NMBP-TO-IND-2020-singlestage, Grant Agreement number 862136

The challenge of data-driven science: working with data!!

Data/Computational Scientist

The challenge of data-driven science: working with data!!

Data/Computational Scientist

Data driven Model

Machine learning algorithms

The challenge of data-driven science: working with data!!

??

Data/Computational Scientist

Data driven Model

Machine learning algorithms

Heterogeneous and Distributed Data Landscape

API 1/METADATA 1

DATA SOURCE 1

API 2/METADATA 2

DATA SOURCE 2

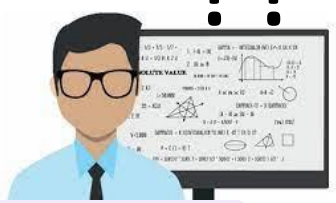
API 3/METADATA 3

DATA SOURCE 3

API 4/METADATA 4

DATA SOURCE 4

The challenge of data-driven science: working with data!!



Data/Computational Scientist

Data driven Model

Machine learning algorithms

Heterogeneous and Distributed Data Landscape

API 1/METADATA 1

- Difficulties to Find the relevant data
- Once found challenging to have access to it
- Data and Metadata are not interoperable
- Time consuming to reuse

DATA SOURCE 4

FAIR principles: a solution to enable data-driven science

nature > scientific data > comment > article

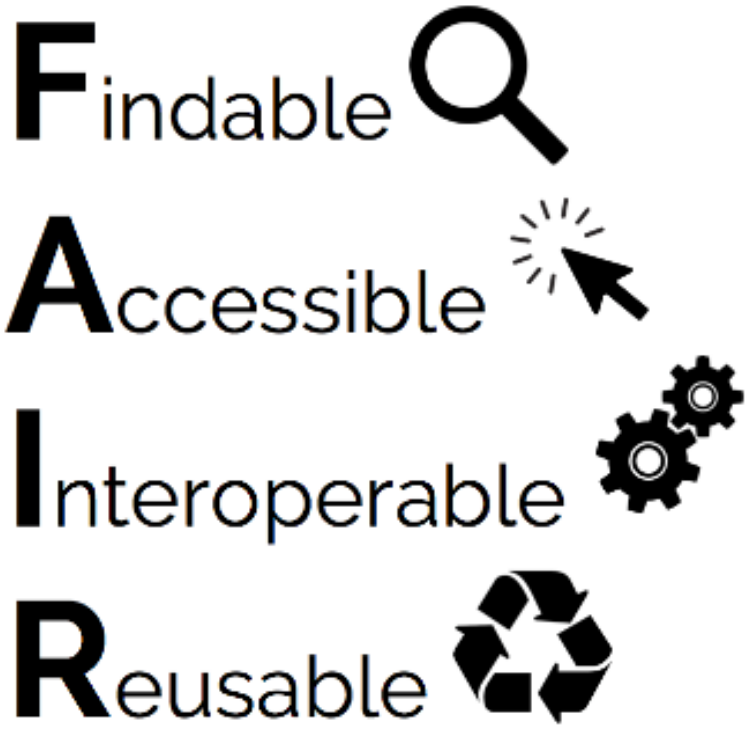
MENU SCIENTIFIC DATA

Comment | OPEN | Published: 15 March 2016

The FAIR Guiding Principles for scientific data management and stewardship

Mark D. Wilkinson, Michel Dumontier, IJsbrand Jan Aalbersberg, Gabrielle Appleton, Myles Axton, Arie Baak, Niklas Blomberg, Jan-Willem Boiten, Luiz Bonino da Silva Santos, Philip E. Bourne, Jildau Bouwman, Anthony J. Brookes, Tim Clark, Mercè Crosas, Ingrid Dillo, Olivier Dumon, Scott Edmunds, Chris T. Evelo, Richard Finkers, Alejandra Gonzalez-Beltran, Alasdair J.G. Gray, Paul Groth, Carole Goble, Jeffrey S. Grethe, Jaap Heringa, Peter A.C 't Hoen, Rob Hooft, Tobias Kuhn, Ruben Kok, Joost Kok, Scott J. Lusher, Maryann E. Martone, Albert Mons, Abel L. Packer, Bengt Persson, Philippe Rocca-Serra, Marco Roos, Rene van Schaik, Susanna-Assunta Sansone, Erik Schultes, Thierry Sengstag, Ted Slater, George Strawn, Morris A. Swertz, Mark Thompson, Johan van der Lei, Erik van Mulligen, Jan Velterop, Andra Waagmeester, Peter Wittenburg, Katherine Wolstencroft, Jun Zhao & Barend Mons - Show fewer authors

Scientific Data **3**, Article number: 160018 (2016) | Download Citation



FAIR Principles have been internationally endorsed (G8, G20, European Commission, NIH, …)

FAIR principles: https://www.go-fair.org/fair-principles/

FAIR principles: a solution to enable data-driven science

nature > scientific data > comment > article

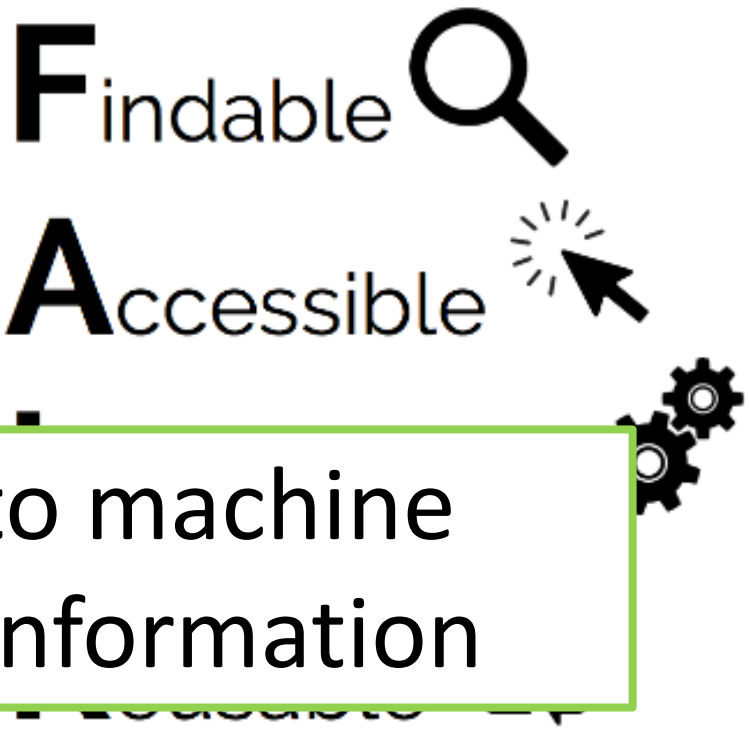
MENU SCIENTIFIC DATA

Comment | OPEN | Published: 15 March 2016

The FAIR Guiding Principles for scientific data management and stewardship

Mark D. Wilkinson, Mi
Arie Baak, Niklas Blom
Bouwman, Anthony J.
Chris T. Evelo, Richard
Goble, Jeffrey S. Greth
Kok, Scott J. Lusher, M
Rocca-Serra, Marco Roos, Rene van Schaik, Susanna-Assunta Sansone, Erik Schultes, Thierry Sengstag, Ted Slater, George Strawn, Morris A. Swertz, Mark Thompson, Johan van der Lei, Erik van Mulligen, Jan Velterop, Andra Waagmeester, Peter Wittenburg, Katherine Wolstencroft, Jun Zhao & Barend Mons - Show fewer authors

Scientific Data **3**, Article number: 160018 (2016) | Download Citation

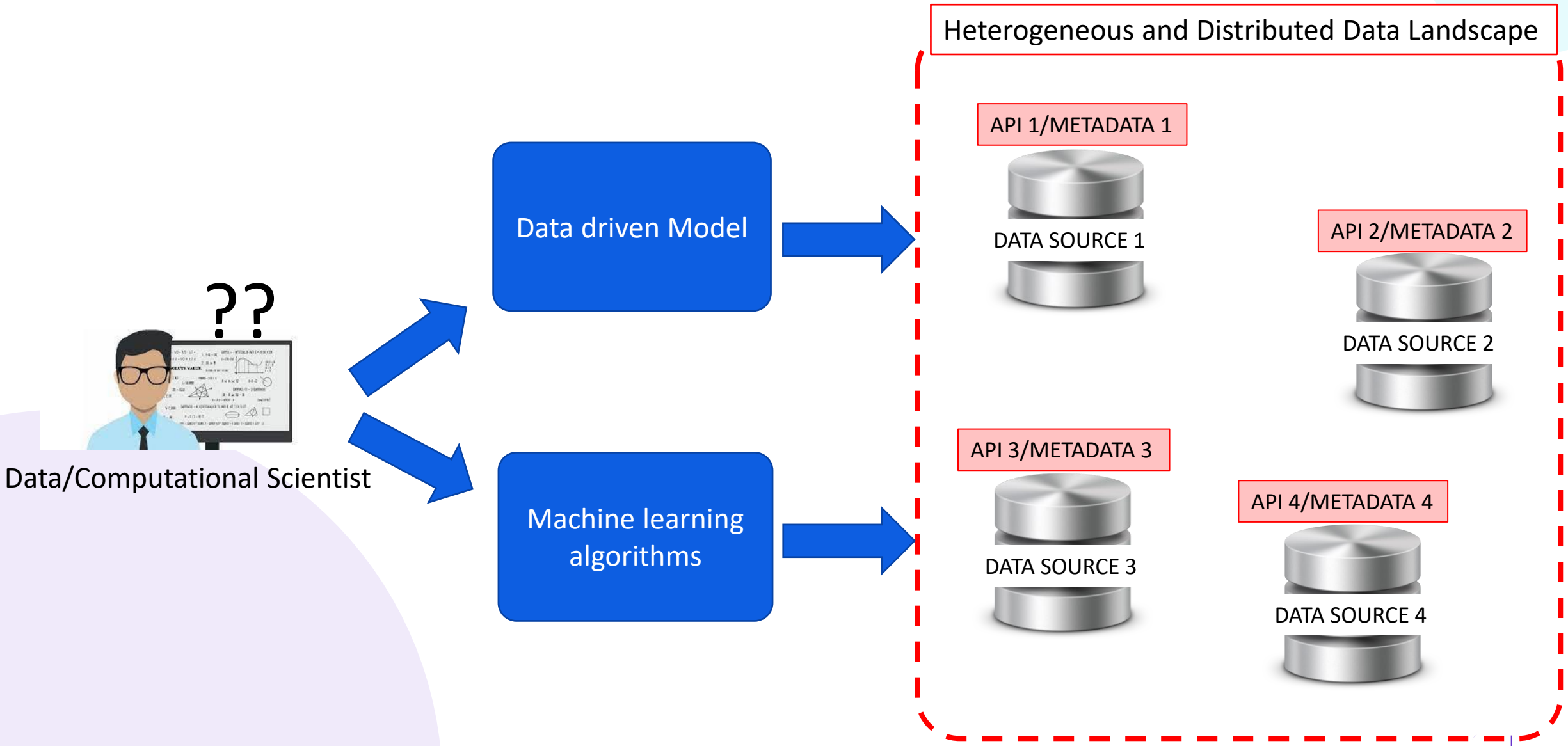


Transform data into machine actionable unit of information

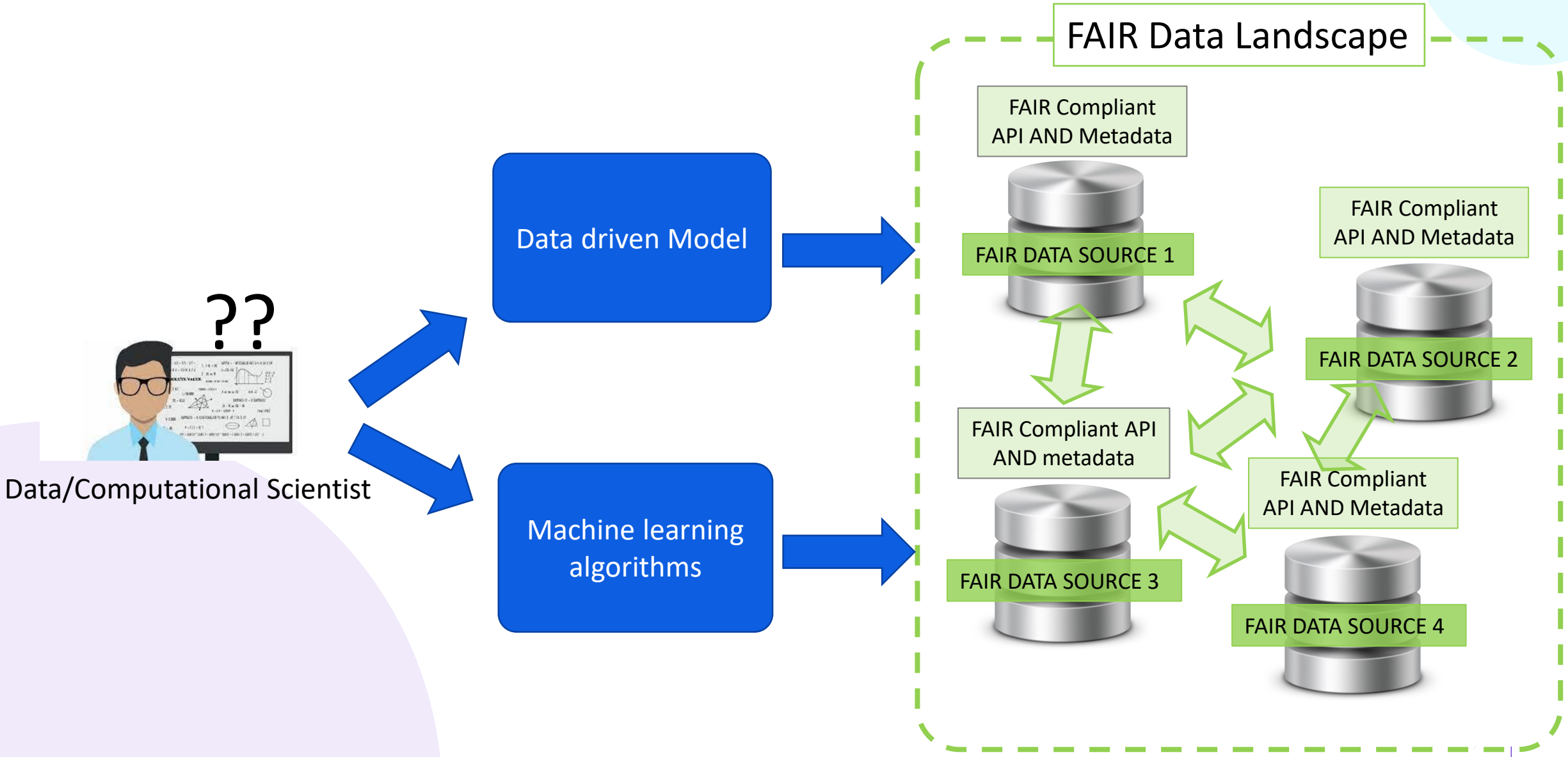
FAIR Principles have been internationally endorsed (G8, G20, European Commission, NIH, …)

FAIR principles: https://www.go-fair.org/fair-principles/

If FAIR principles were implemented everywhere !!



If FAIR principles were implemented everywhere !!



If FAIR principles were implemented everywhere !!

Data/Computational Scientist

Data driven Model

Machine learning algorithms

FAIR Data Landscape

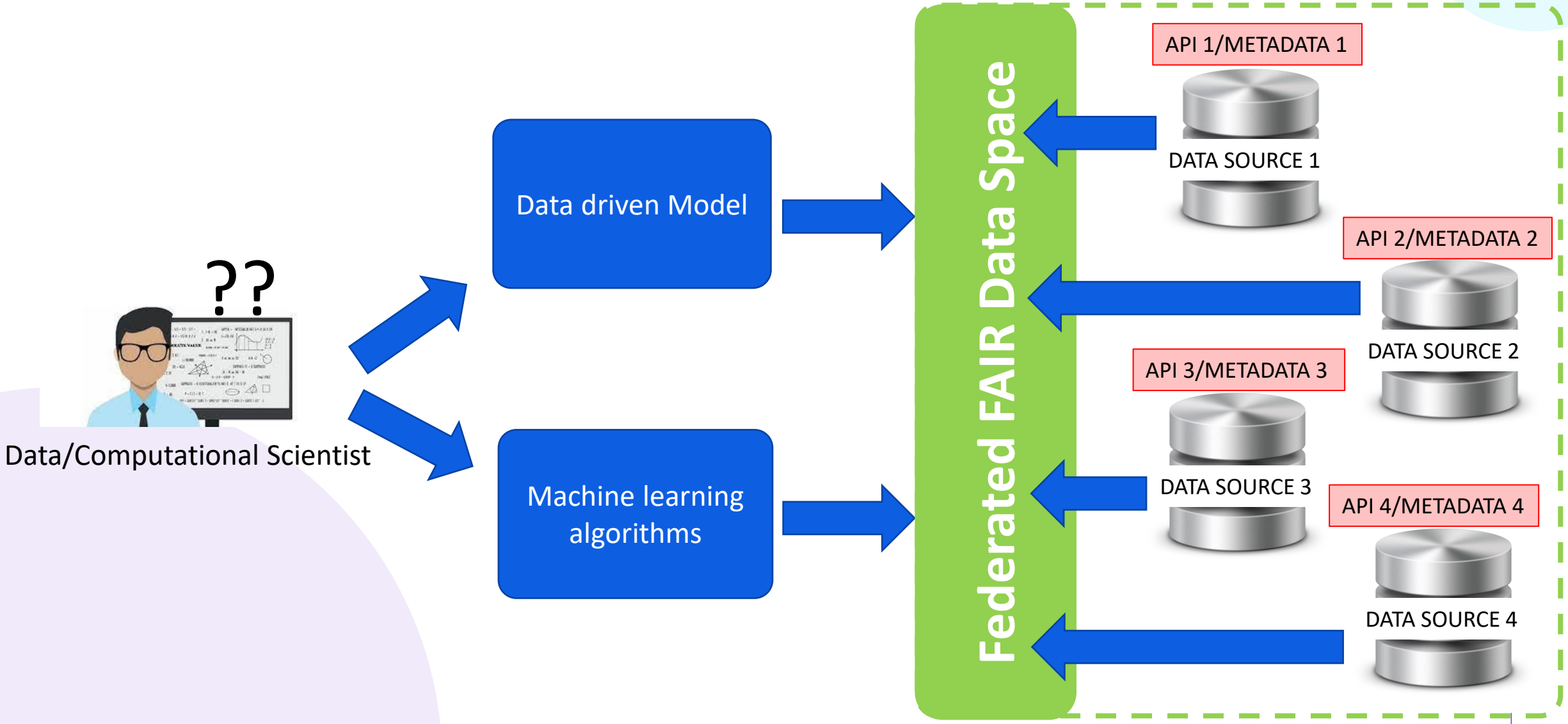
FAIR Compliant API AND Metadata

FAIR Compliant API AND Metadata

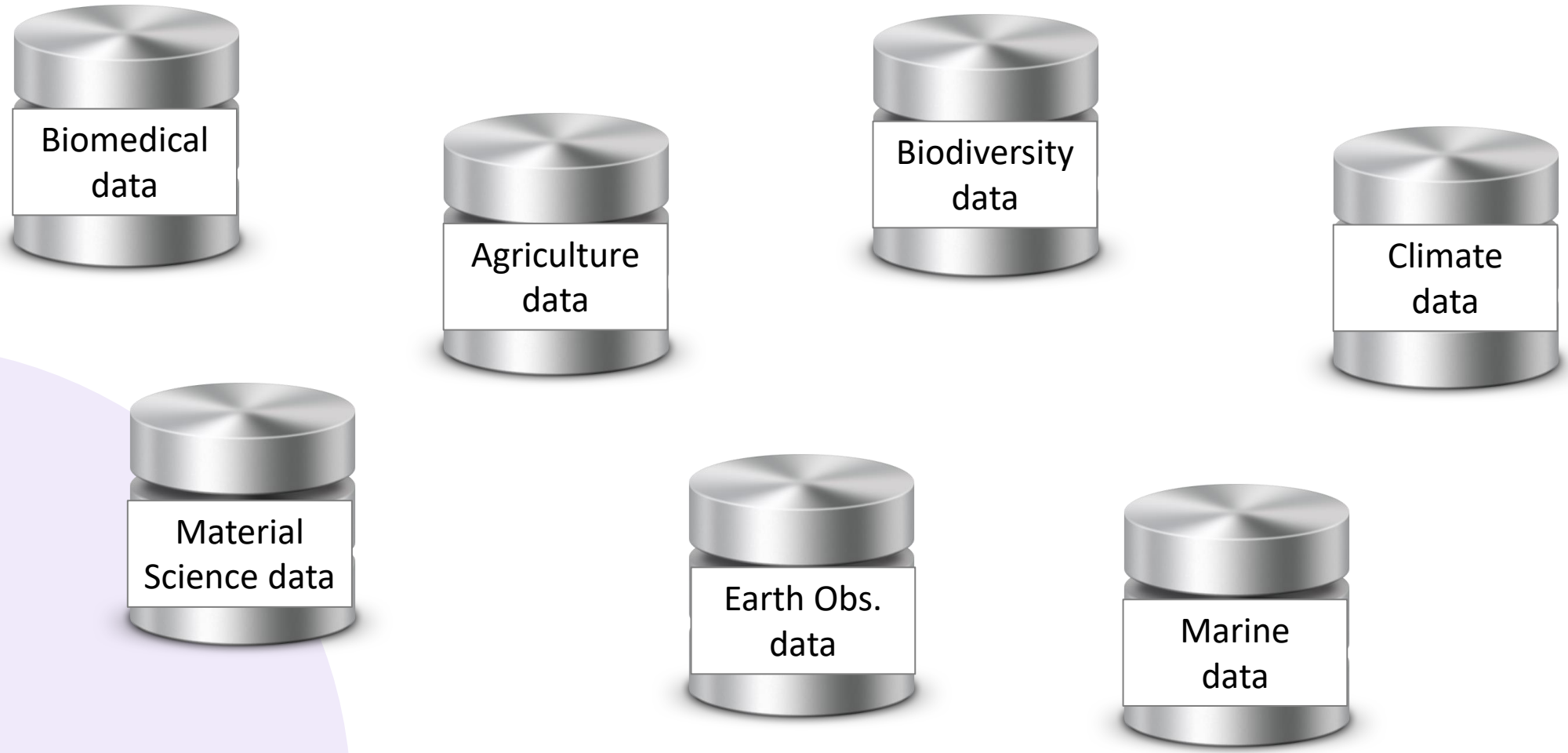
- Requires Changes in Data Sources
- Takes time and resources

Can we accelerate the adoption of the FAIR principles while reducing the cost of change?

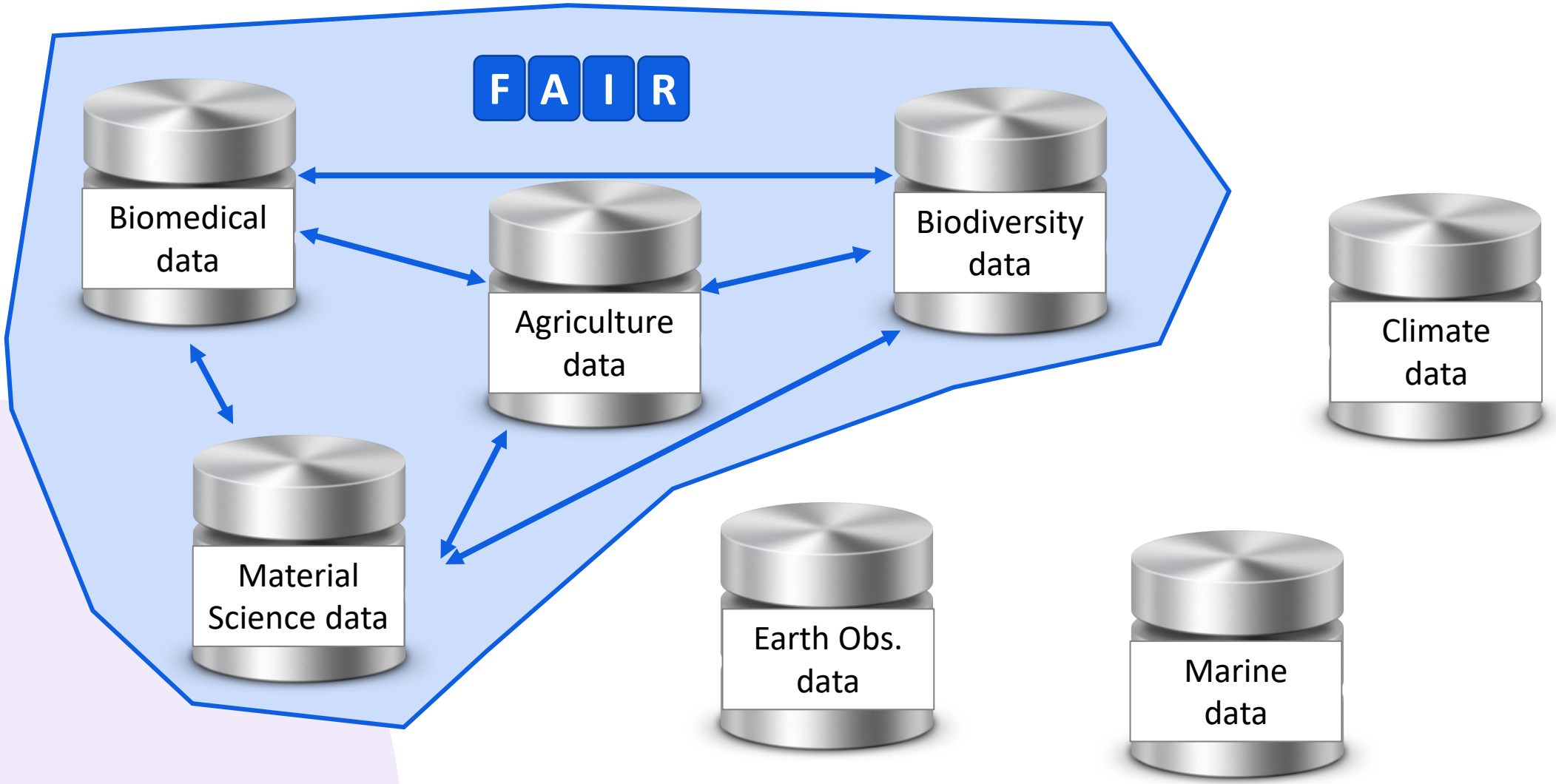
What if we could create FAIR Data Spaces on top of the existing ?!



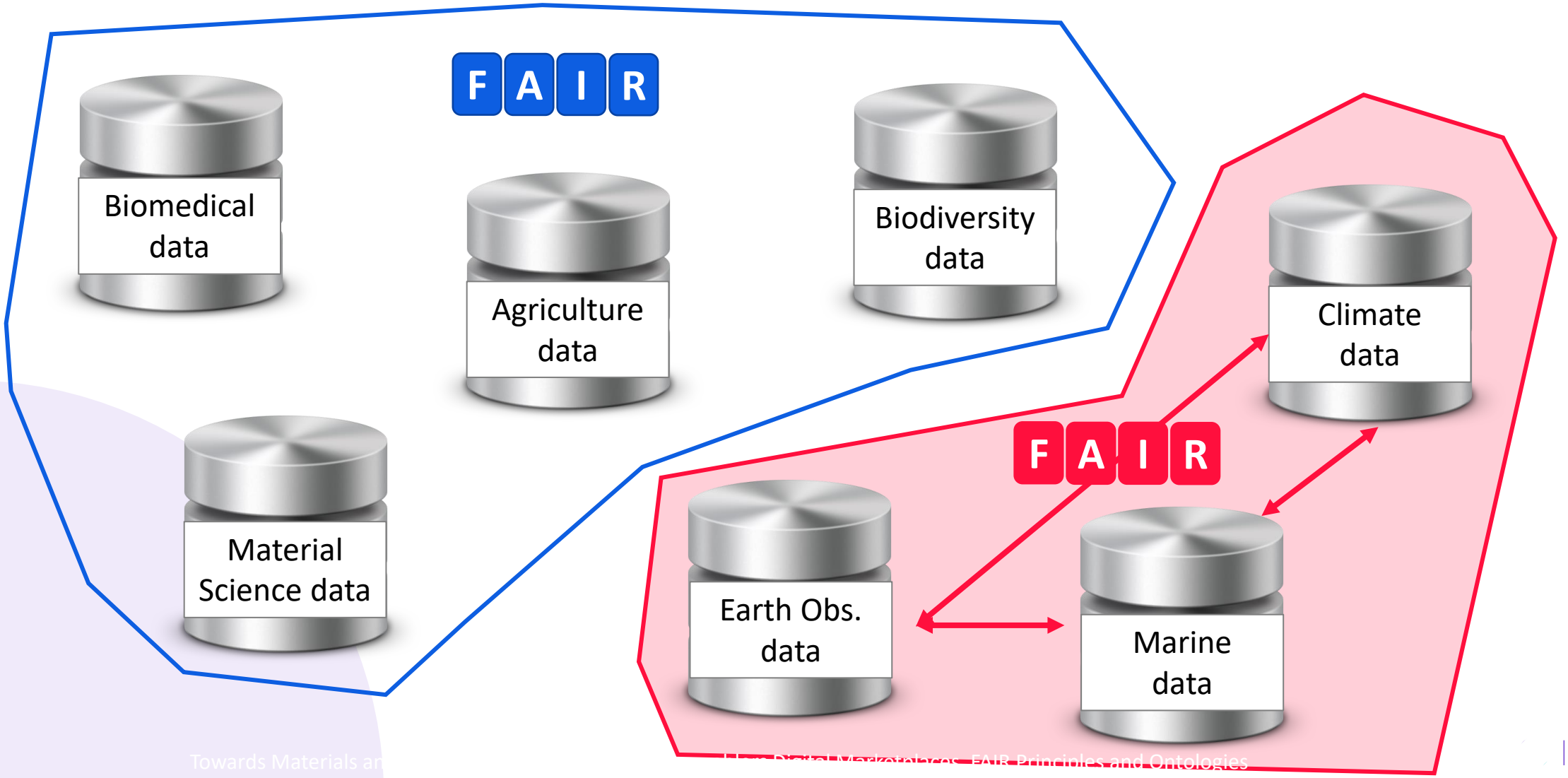
Use-case 1: A cross disciplinary use-case driven project



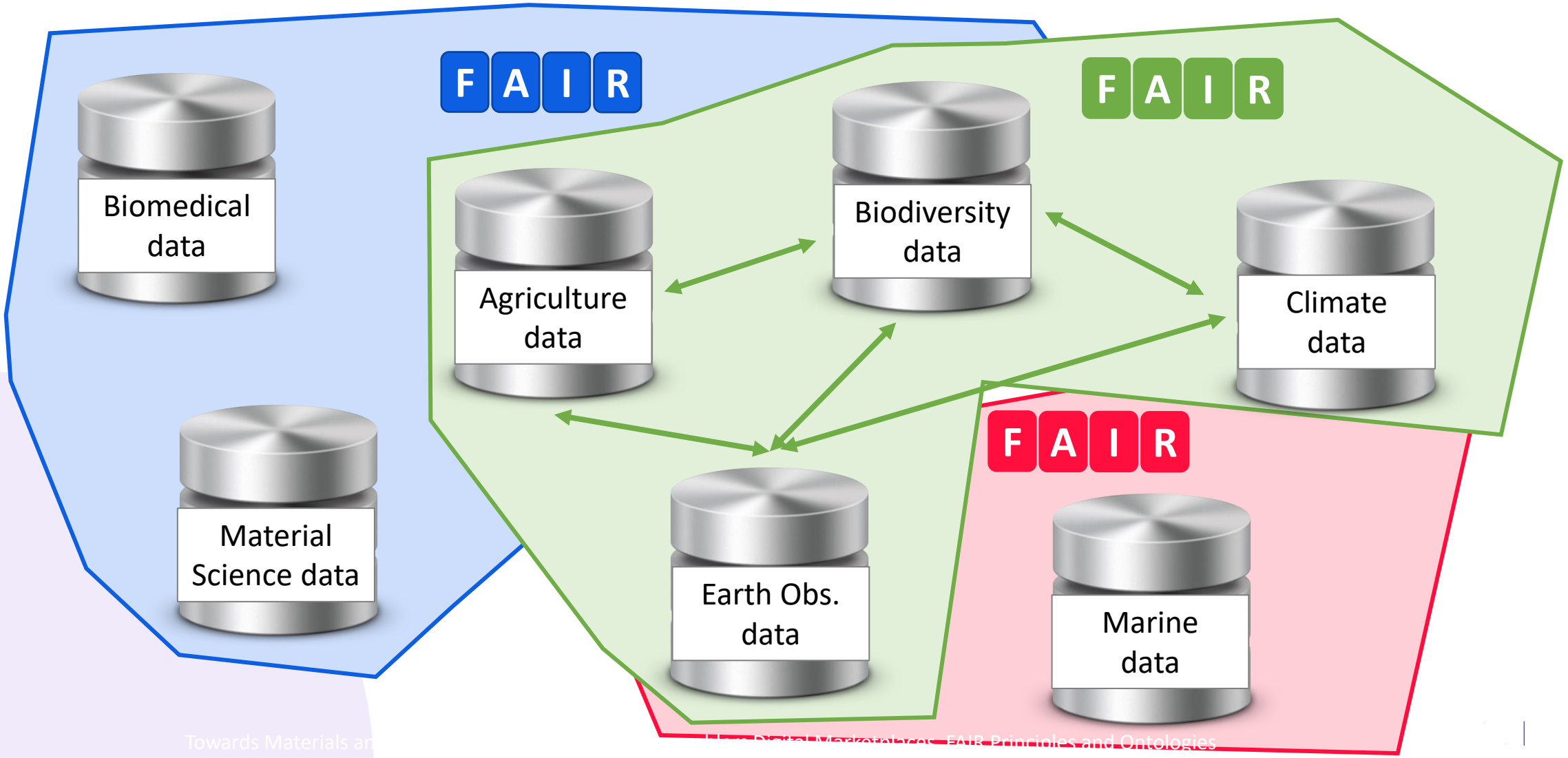
Use-case 1: A cross disciplinary use-case driven project



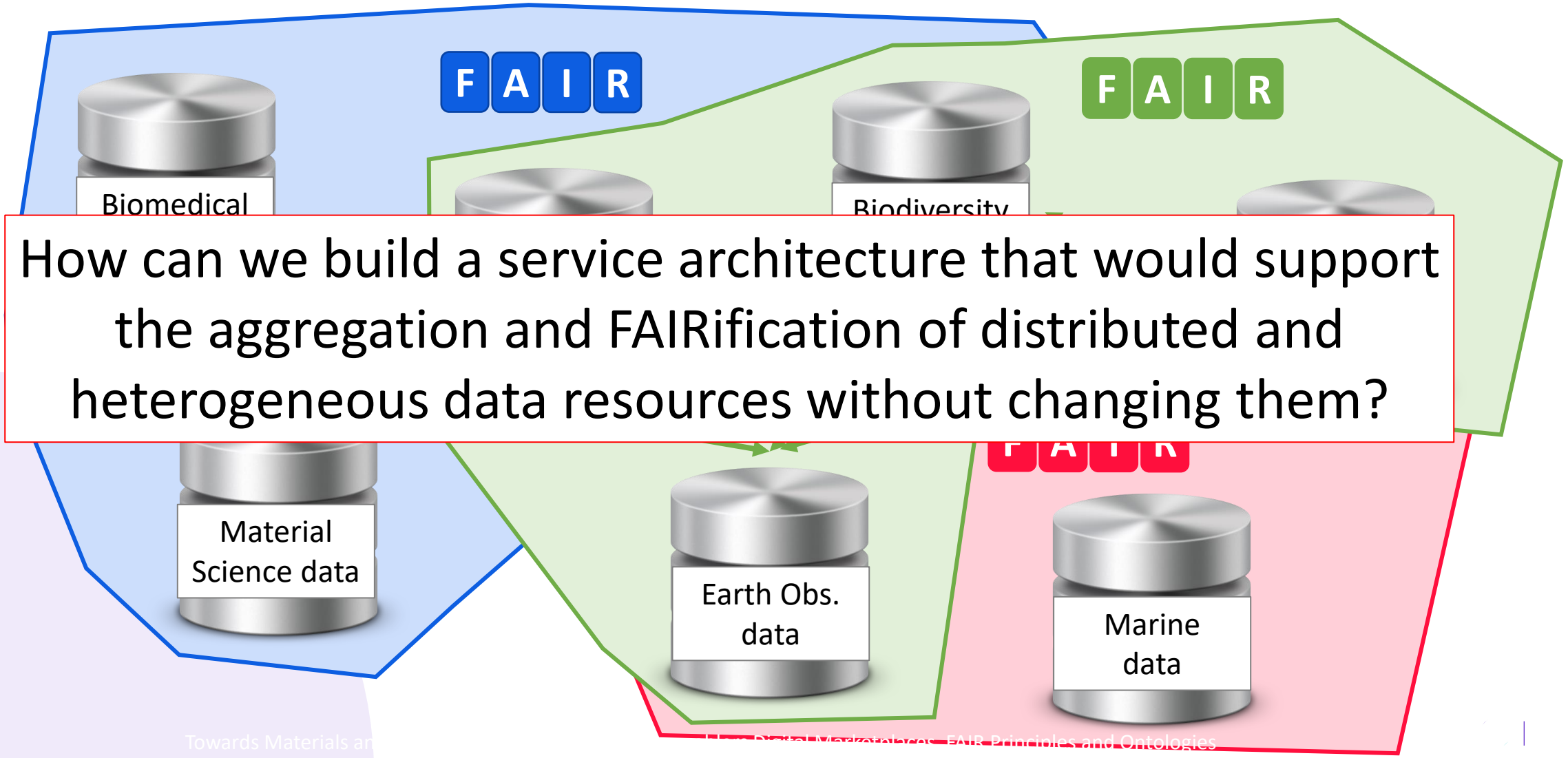
Use-case 1: A cross disciplinary use-case driven project



Use-case 1: A cross disciplinary use-case driven project



Use-case 1: A cross disciplinary use-case driven project



How can we build a service architecture that would support the aggregation and FAIRification of distributed and heterogeneous data resources without changing them?

What does it mean in practice to implement FAIR?

Technical considerations

- Globally Unique Persistent and Resolvable Identifiers
- Metadata
  - Provenance
  - Common format
  - FAIR Vocabularies/Ontologies
  - Leverage domain specific metadata standards
  - Persistent
- Free, Open and Universal Access protocol
- Licences (human and machine readable)

What does it mean in practice to implement FAIR?

Technical considerations

- Globally Unique Persistent and Resolvable Identifiers
- Metadata
  - Provenance
  - Common format
  - FAIR Vocabularies/Ontologies
  - Leverage domain specific metadata standards
  - Persistent
- Free, Open and Universal Access protocol
- Licences (human and machine readable)

What does it mean in practice to implement FAIR?

Technical considerations

- Globally Unique Persistent and Resolvable Identifiers
- Metadata
  - Provenance
  - Common format
  - FAIR Vocabularies/Ontologies
  - Leverage domain specific metadata standards
  - Persistent
- Free, Open and Universal Access protocol
- Licences (human and machine readable)

Existing tooling

Publish FAIR data

Enrich (meta)data with semantics

Evaluate FAIRness

ONTO COMMONS
ONTOLOGY-DRIVEN DATA DOCUMENTATION FOR INDUSTRY COMMONS

FAIR Specifications

TABLE OF CONTENTS

1 **Introduction**

1.1 Purpose
1.2 Document Conventions

2 **Overall Description**

2.1 Usage scenarios
2.1.1 Data discovery
2.1.2 Data access
2.1.3 (Meta)Data publication
2.1.4 Publishing other types of content
2.2 Goals
2.2 Product Perspective

3 **Architecture**

4 **Metadata**

4.1 Navigation information
4.2 Metadata schemas
4.2.1 Metadata Service metadata
4.2.2 Catalog metadata
4.3 Resource extensions

5 **Application Programming Interface (API)**

6 **FDP Specification Compliance**

FAIR Data Point

Working Draft, 20 March 2023

FAIR DATA POINT

This version:
https://specs.fairdatapoint.org/v1.1

Latest version:
https://specs.fairdatapoint.org

Previous Versions:
FDP specs v1.0

Feedback:
l.o.boninodasilvasantos@utwente.nl

Issue Tracking of the specifications:
GitHub

Reference Implementation:
https://github.com/FAIRDataTeam/FAIRDataPoint/

Issue Tracking of the reference implementation:
GitHub

Editors:
Luiz Olavo Bonino (University of Twente, Leiden University Medical Center, GO FAIR International Support and Coordination Office)
Kees Burger (Leiden University Medical Center)
Rajaram Kaliyaperumal (Leiden University Medical Center)

Luiz Olavo Bonino da Silva Santos, Kees Burger, Rajaram Kaliyaperumal, Mark D. Wilkinson; FAIR Data Point: A FAIR-Oriented Approach for Metadata Publication. *Data Intelligence* 2023; 5 (1): 163–183. doi: https://doi.org/10.1162/dint_a_00160

ONTO COMMONS
ONTOLOGY-DRIVEN DATA DOCUMENTATION FOR INDUSTRY COMMONS

FAIR Specifications

W3C Recommendation

TABLE OF CONTENTS

1
1.1
1.2
2
2.1
2.1.
2.1.
2.1.
2.1.
2.1.
2.2
2.2
3
4
4.1
4.2
4.2.
4.2.
4.3
5 Application Programming Interface (API)
6 EDP Specification Compliance

Data Catalog Vocabulary (DCAT) - Version 2

W3C Recommendation 04 February 2020

This version:
https://www.w3.org/TR/2020/REC-vocab-dcat-2-20200204/

Latest published version:
https://www.w3.org/TR/vocab-dcat-2/

Latest editor's draft:
https://w3c.github.io/dxwg/dcat/

Implementation report:
https://w3c.github.io/dxwg/dcat-implementation-report/

Previous version:
https://www.w3.org/TR/2019/PR-vocab-dcat-2-20191119/

Previous Recommendation:
https://www.w3.org/TR/2014/REC-vocab-dcat-20140116/

Editors:
Riccardo Albertoni (CNR - Consiglio Nazionale delle Ricerche, Italy)
David Browning (Refinitiv)
Simon Cox (CSIRO)
Alejandra Gonzalez Beltran (Scientific Computing Department, Science and Technology Facilities Council, UK) (Previously at the University of Oxford)
Andrea Perego (European Commission, Joint Research Centre)
Peter Winstanley (Scottish Government)

FAIR DATA POINT

iO FAIR International Support

ONTO COMMONS
ONTOLOGY-DRIVEN DATA DOCUMENTATION FOR INDUSTRY COMMONS

FAIR Specifications

W3C Recommendation

TABLE OF CONTENTS

1
1.1
1.2
2
2.1
2.1.
2.1.
2.1.
2.1.
2.2
2.2
3
4
4.1
4.2
4.2.
4.2.
4.3
5 Application Programming Interface (API)
6 FDP Specification Compliance

Data Catalog Vocabulary (DCAT) - 2

W3C Recommendation 04 February 2020

This version:
https://www.w3.org/TR/2020/REC-vocab-dcat-2-20200204/

Latest published version:
https://www.w3.org/TR/vocab-dcat-2/

Latest editor's draft:
https://w3c.github.io/dxwg/dcat/

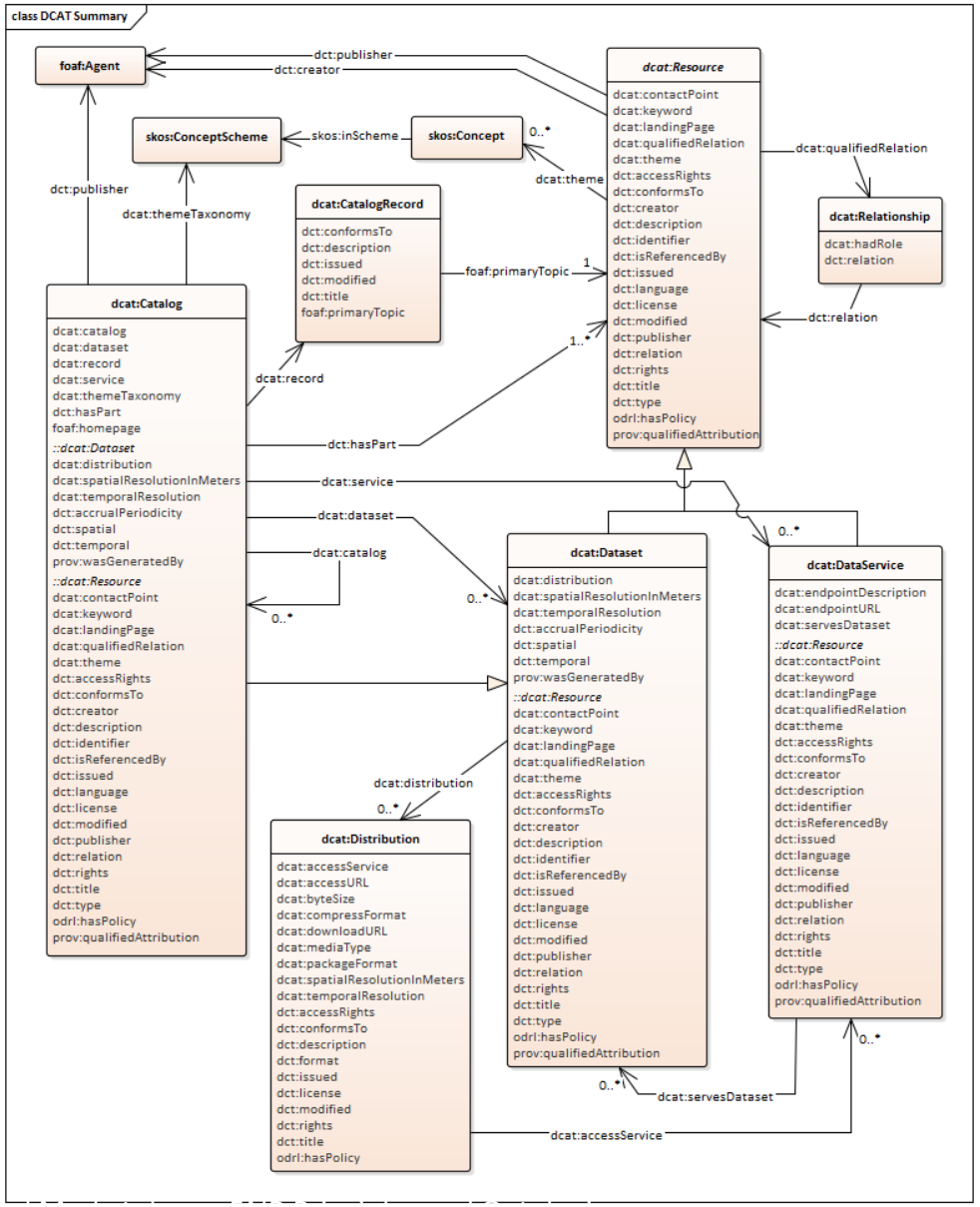
Implementation report:
https://w3c.github.io/dxwg/dcat-implementation-report/

Previous version:
https://www.w3.org/TR/2019/PR-vocab-dcat-2-20191119/

Previous Recommendation:
https://www.w3.org/TR/2014/REC-vocab-dcat-20140116/

Editors:

Riccardo Albertoni (CNR - Consiglio Nazionale delle Ricerche, Italy)
David Browning (Refinitiv)
Simon Cox (CSIRO)
Alejandra Gonzalez Beltran (Scientific Computing Department, Science Council, UK) (Previously at the University of Oxford)
Andrea Perego (European Commission, Joint Research Centre)
Peter Winstanley (Scottish Government)



Our other sources of inspiration

Enhancing the Discoverability and Interoperability of Multi-disciplinary Semantic Repositories

Doron Goldfarb[1][0000-0003-1183-6041] and Yann Le Franc[2][0000-0003-4631-418X]

[1] Environment Agency Austria, Vienna, Austria
[2] e-Science Data Factory, Paris, France
doron.goldfarb@umweltbundesamt.at, ylefranc@esciencefactory.com

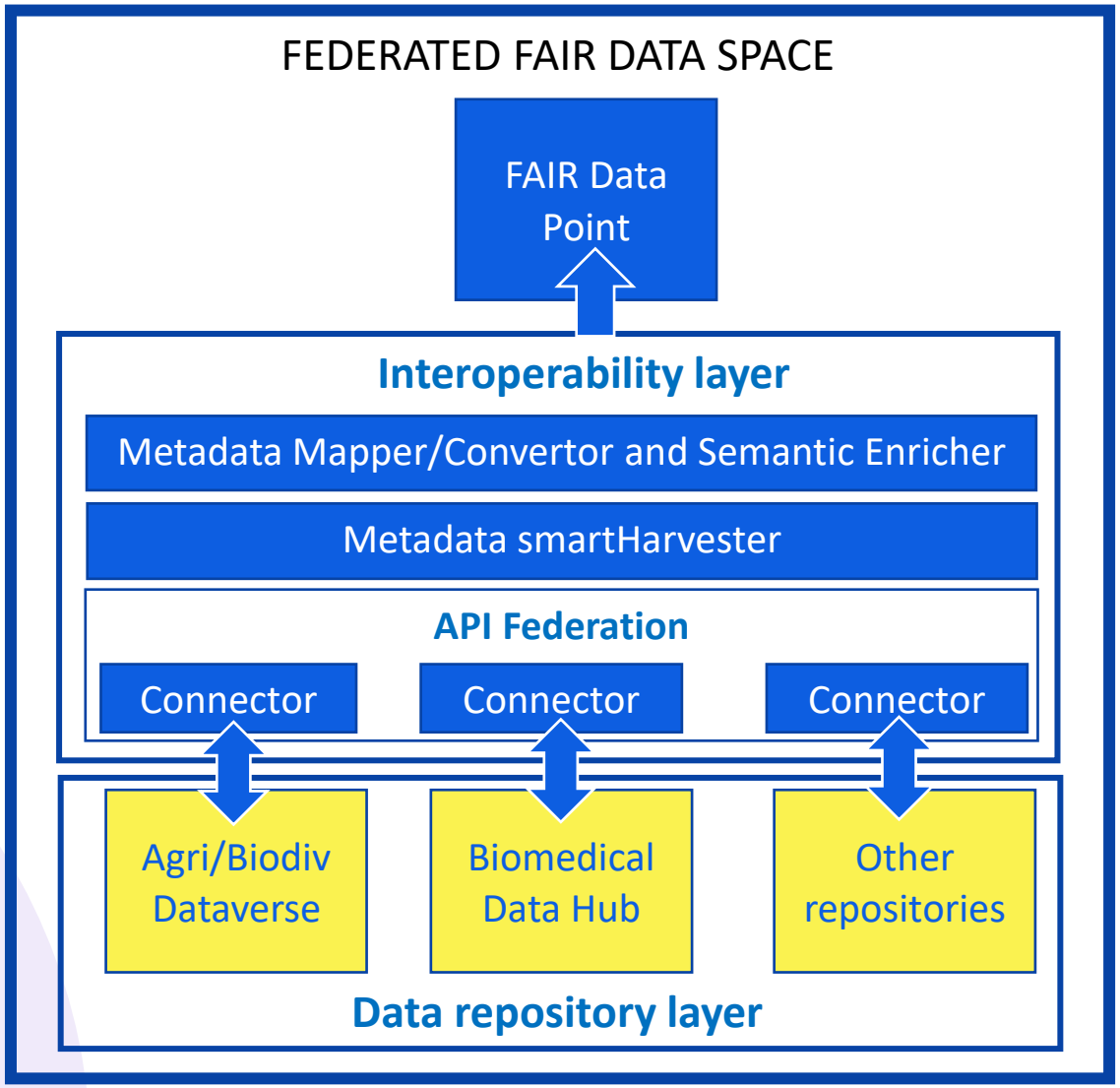
Abstract. The aggregation of multi-disciplinary information is a challenge faced by large-scale data infrastructures serving scientific domains such as biodiversity, agronomy or ecology. This requires the integration of ontologies or thesauri from different domains. These semantic resources are often hosted within domain specific repositories which can be harvested for that purpose. The lack of discoverability, the technical and metadata heterogeneity of the semantic repositories pose a challenge for their effective integration. In this context, we argue that there is a need for a semantic lookup-service to access and use this heterogeneous landscape. We then present a proof-of-concept design and implementation for harvesting different ontology repositories (BioPortal, AgroPortal and EBI-OLS). We show some preliminary analytics and discuss technical issues regarding aggregation. Finally, we conclude with an open call for collaboration to address the issues hampering such initiatives.

Keywords: Ontology libraries, Semantic annotation, Ontology lookup service, EUDAT.

1 Introduction

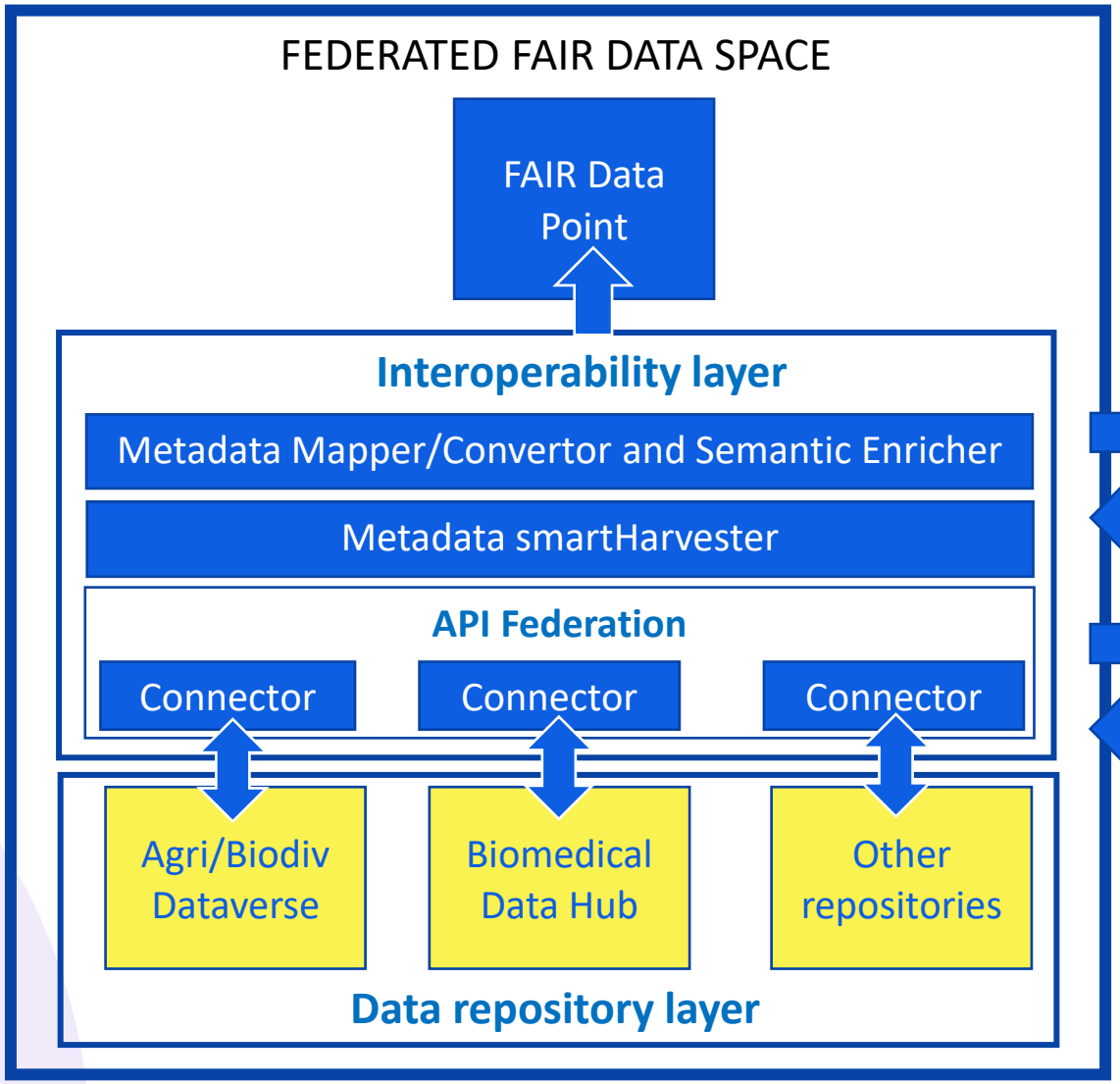
Semantic technologies are increasingly used by domain-specific Research Infrastructures (RIs) and large-scale multi-disciplinary infrastructures such as EUDAT[1]. Semantically-enabled services offer a framework to aggregate data from multiple sources, enhancing discoverability and interoperability. The EUDAT pilot service B2Note[2] is one such service, allowing the creation of semantic annotations of datasets within and outside of the EUDAT infrastructure. The process of annotation is about "attach[ing] data to some other piece of data" [1]. In the scope of the Semantic Web, this usually refers to the contextualisation of information within a wider knowledge graph in order to support discovery and, eventually, automated reasoning. Such a

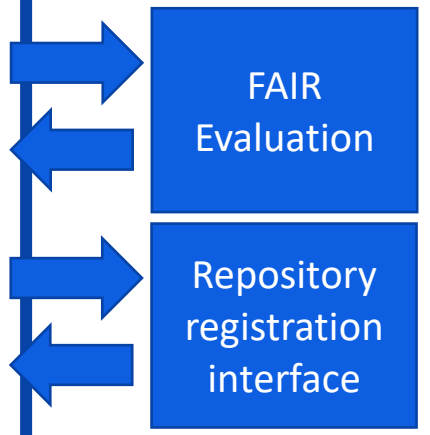
A simple architecture for Federated FAIR Data Spaces (F2DS)



Concept, Design: Y. Le Franc, N. Cazenave

A simple architecture for Federated FAIR Data Spaces (F2DS)

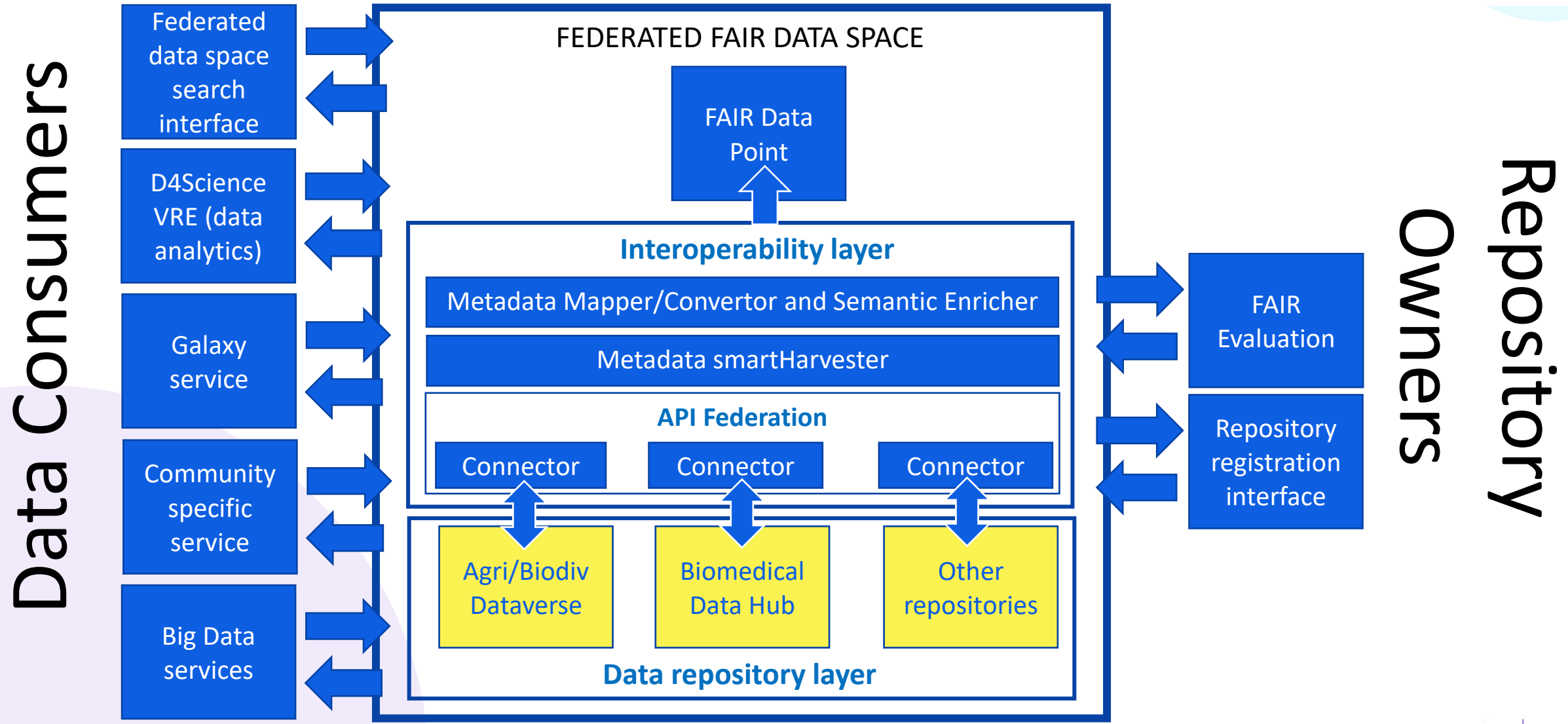




Repository Owners

Concept, Design: Y. Le Franc, N. Cazenave

A simple architecture for Federated FAIR Data Spaces (F2DS)



Concept, Design: Y. Le Franc, N. Cazenave

Easy deployment and scalability

Bundle of open-source services easily deployed on any cloud system using Kubernetes

Create F2DS on the fly for various purposes

Register your repository in several F2DS without re-entering all the information

Benefits

Create in no-time data spaces across internal resources, institutions and domains

No changes required for repositories

Make repository content more FAIR

Multiple APIs for accesing and processing F2DS content (SparQL, REST API and web UI)

Connect easily any processing services (Galaxy, VREs, PANGEO, HPC/HTC, ...)

User-centric workflow for repository owner

Register Data Repository

Provide metadata about your repository

User-centric workflow for repository owner

Register Data Repository

Provide metadata about your repository

Register API

Provide API description

User-centric workflow for repository owner

Register Data Repository

Provide metadata about your repository

Register API

Provide API description

Map Metadata

W3C

Data Catalog Vocabulary (DCAT) Version 2

W3C Recommendation 04 February 2020

User-centric workflow for repository owner

Register Data Repository

Provide metadata about your repository

Register API

Provide API description

Map Metadata

Data Catalog Vocabulary (DCAT) Version 2

W3C Recommendation 04 February 2020

Enrich Metadata

Semantic Web

User-centric workflow for repository owner

Register Data Repository

Provide metadata about your repository

Register API

Provide API description

Map Metadata

Data Catalog Vocabulary (DCAT) Version 2

W3C Recommendation 04 February 2020

Enrich Metadata

Semantic Web

Publish Metadata

FAIR DATA POINT

User-centric workflow for repository owner

Register Data Repository

Provide metadata about your repository

Register API

Provide API description

Map Metadata

Data Catalog Vocabulary (DCAT) Version 2

W3C Recommendation 04 February 2020

Enrich Metadata

Semantic Web

Publish Metadata

FAIR DATA POINT

Search and Reuse

Evaluate FAIRness

User-centric workflow for repository owner

Register Data Repository

Provide metadata about your repository

Register API

Provide API description

Map Metadata

Data Catalog Vocabulary (DCAT) Version 2
W3C Recommendation 04 February 2020

Enrich Metadata

Semantic Web

Publish Metadata

FAIR DATA POINT

Need for a harmonised semantic space to access content for (meta)data enrichment

User-centric workflow for repository owner

Register Data Rep	Register API	Map	Enrich	Publish
Provid… about yo…		Version 2 W3C Recommendation 04 February 2020	Semantic Web	

How can we make semantic artefacts FAIR to leverage the wealth of existing semantic artefacts to improve FAIRness of data?

Need for a harmonised semantic space to access content for (meta)data enrichment

FAIR Semantics recommendations

- Developed with the contribution of a community of experts from various domains (earth science, geospatial, biodiversity, ecology, biomedicine,…)
- 17 general recommendations/12 best practices
- Aligned with existing community specific recommendations (OBO Foundry, IOF principles,…)
- Third iteration of the FAIR Semantics recommendations aligned with RFC 2119 and including a minimum metadata schema for semantic artefacts https://zenodo.org/record/6276577

Project Title	Fostering FAIR Data Practices in Europe
Project Acronym	FAIRsFAIR
Grant Agreement No	831558
Instrument	H2020-INFRAEOSC-2018-4
Topic	INFRAEOSC-05-2018-2019 Support to the EOSC Governance
Start Date of Project	1st March 2019
Duration of Project	36 months
Project Website	www.fairsfair.eu

D2.8 FAIR Semantics Recommendations

Third Iteration

Work Package	WP2
Lead Author (Org)	Yann Le Franc (eSDF)
Contributing Author(s) (Org)	Luiz Bonino (LUMC - DTL), Hanna Koivula (CSC), Jessica Parland-von Essen (CSC), Robert Pergl (eSDF)
Due Date	28.02.2022
Date	25.02.2022
Version	1.0 Draft not yet approved by the European Commission
DOI	10.5281/zenodo.6276576

Dissemination Level

- ☐ PU: Public
- ☐ PP: Restricted to other programme participants (including the Commission)
- ☐ RE: Restricted to a group specified by the consortium (including the Commission)
- ☐ CO: Confidential, only for members of the consortium (including the Commission)

Why defining a minimum metadata for FAIR?

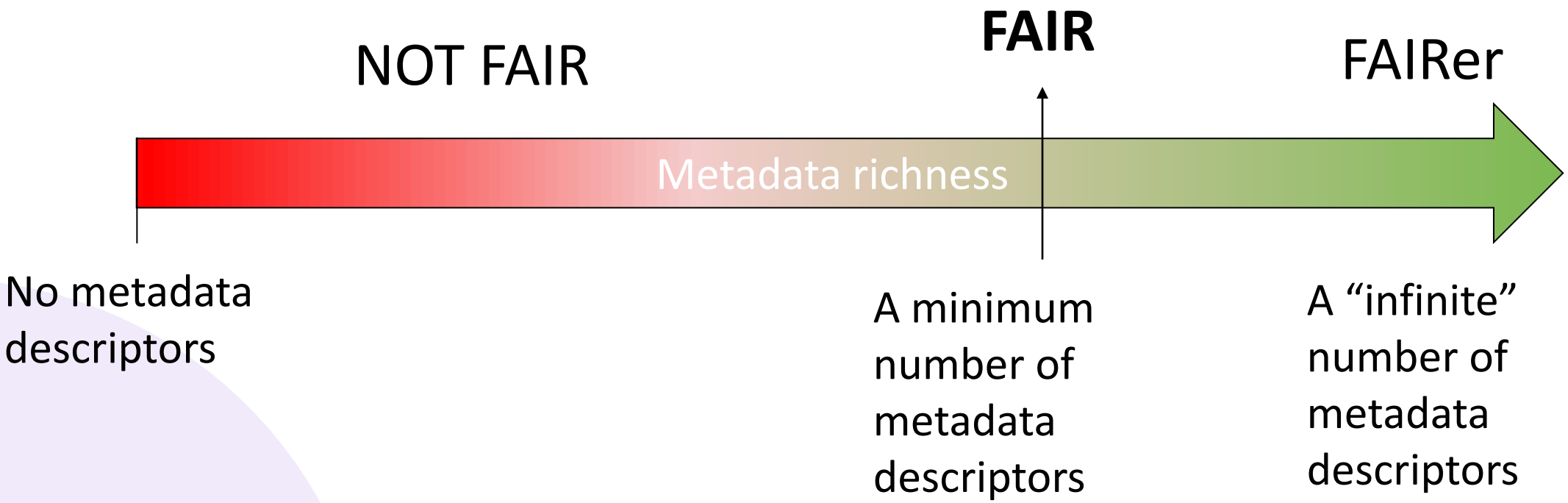
NOT FAIR

FAIR?

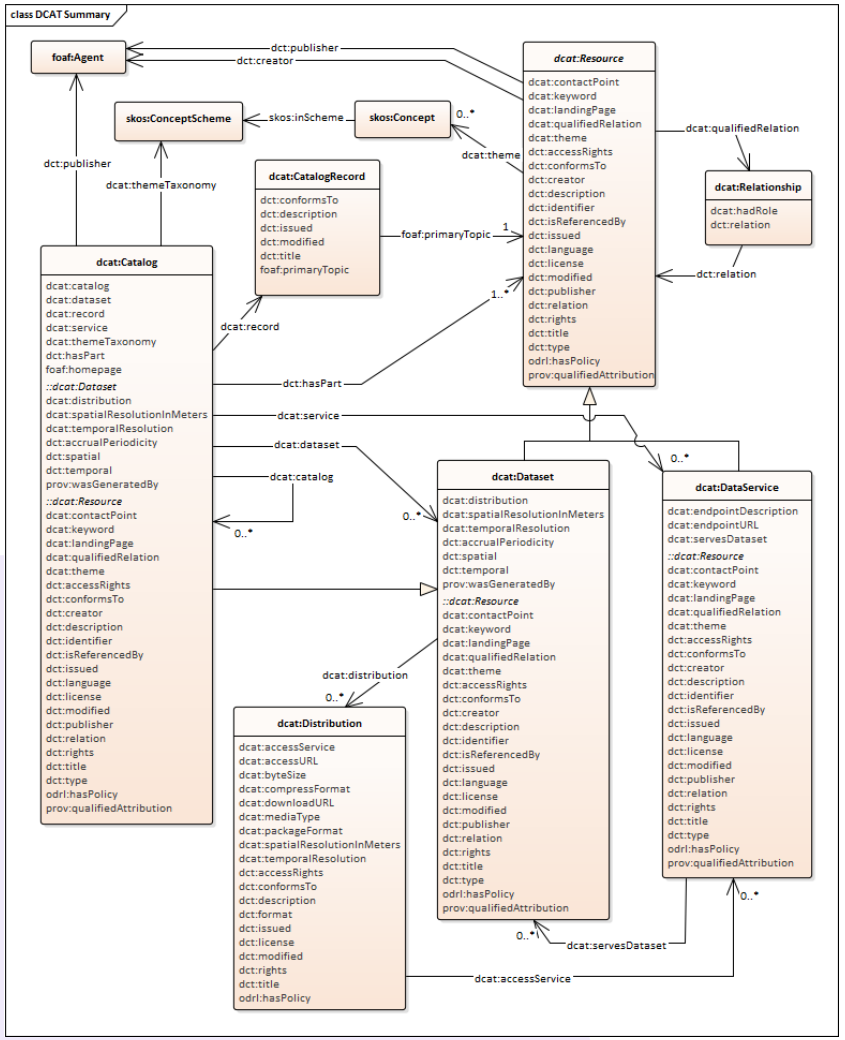
No metadata descriptors

A “inifinite” number of metadata descriptors

Why defining a minimum metadata for FAIR?



Minimum metadata for FAIR Semantic artefacts



Stable

MOD: Metadata for Ontology Description and publication

Release August 2, 2018

This version:
http://www.isibang.ac.in/ns/mod/1.4

Latest version:
http://www.isibang.ac.in/ns/mod/1.4

Previous version:
http://www.isibang.ac.in/ns/mod/1.2
https://www.isibang.ac.in/ns/mod/1.1
https://www.isibang.ac.in/ns/mod/1.0

Revision:
1.4

Authors:
Biswanath Dutta, (Indian Statistical Institute)
Clement Jonquet, (University of Montpellier)

Contributors:
Anne Toulet, (University of Montpellier)
Udaya Varadarajan, (Indian Statistical Institute)

Publisher:
http://www.isibang.ac.in/

Download serialization:
Format JSON LD Format RDF/XML Format N Triples Format TTL

License:
License Creative Commons Attribution 4.0

Cite as:
Dutta, B., Toulet, A., Emonet, V. and Jonquet, C. (2017). New Generation Metadata vocabulary Description and Publication. In E. Garoufallou, S. Virkus, R. Siatri and D. Koutso Communications in Computer and Information Science (CCIS) 755, proceedings of 11th M Semantics Research Conference (MTSR 2017), November 28 - December 1, 2017, Talli Springer Nature, pp. 173-185.

A minimum metadata schema for FAIR Semantic Artefacts

Workshop jointly organised with RDA VSSIG: community driven minimal metadata model and DCAT profile for Semantic Artefact

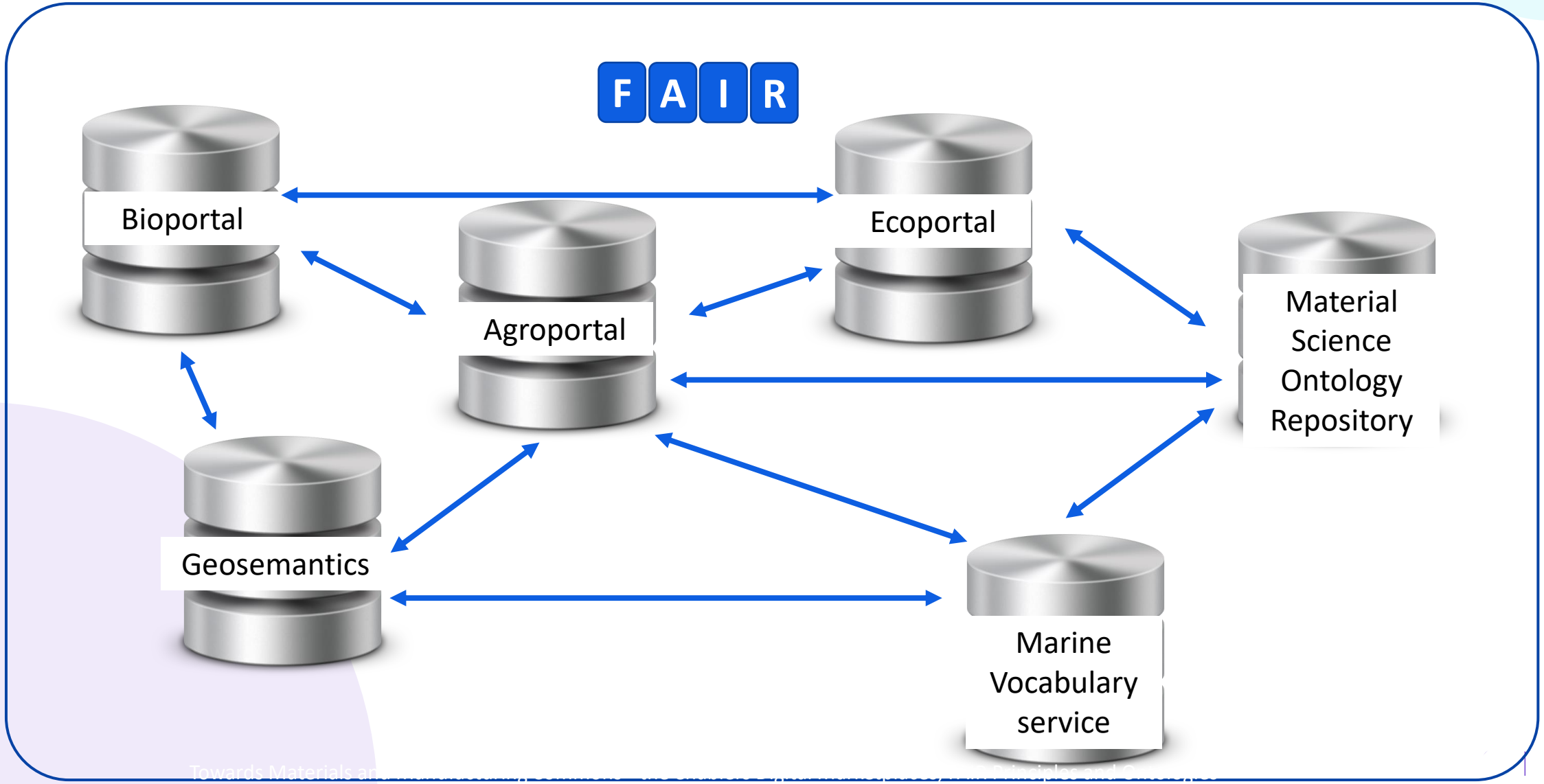
- 76 participants from 17 different communities
- Define a common model for publishing Semantic Artefacts based on DCAT
- Define a minimum set of descriptive metadata for FAIR Semantic Artefact using MOD (Dutta et al., 2015; Dutta et al., 2017)

⇒ Initial DCAT profile using standards metadata properties and enriched with 12 properties from MOD

⇒ Profile formalisation (OWL/SHACL)

https://github.com/FAIRsFAIR/SemanticDCAT-AP

Use-case 2: Creating a FAIR Semantic Space



Using the FAIR Semantic space

- a PoC search engine for semantic artefacts across multiple community driven semantic artefact repository

- a PoC cross-disciplinary Semantic index for fast access to existing concepts and relations to enrich data and metadata

ONTO COMMONS

@ontocommons

linkedin.com/company/ontocommons

info@ontocommons.eu

ONTOLOGY-DRIVEN DATA DOCUMENTATION FOR INDUSTRY COMMONS

Thank you for your attention !

OntoCommons “Ontology-driven data documentation for Industry Commons” has received funding from the European Union’s Horizon Programme call H2020 -NMBP-TO-IND-2020-singlestage, Grant Agreement number 862136

Join our community

Follow us on Twitter:
@ontocommons

Follow us on LinkedIn:
linkedin.com/company/ontocommons

Subscribe to our Newsletter:
ontocommons.eu/newsletter