EOSC-Pillar

Coordination and Harmonisation of National &Thematic Initiatives to support EOSC

WP5 – The Data Layer

Des services de données FAIR au niveau national français

Yann Le Franc, PhD – Technical Manager - CINES

EOSC-Pillar has received funding from the European Union's Horizon 2020 research and innovation Programme under Grant Agreement No. 857650.

Soutenir l'intégration des principes FAIR dans les communautés nationales du projet.

TECHNOLOGIE

- Tâche 5.1/Tâche 5.2
- Tâche 5.5

FORMATION/SUPPORT

- Tâche 5.3
- Tâche 5.4

Soutenir l'intégration des principes FAIR dans les communautés nationales du projet.

TECHNOLOGIE

Tâche 5.1/Tâche 5.2

Des services innovants pour créer et utiliser des espaces fédérés de données FAIR

Tâche 5.5

Intéropérabilité sémantique pluridisciplinaire

FORMATION/SUPPORT

Tâche 5.3

Un catalogue de ressources sur les bonnes pratiques de gestion des données recherches FAIR

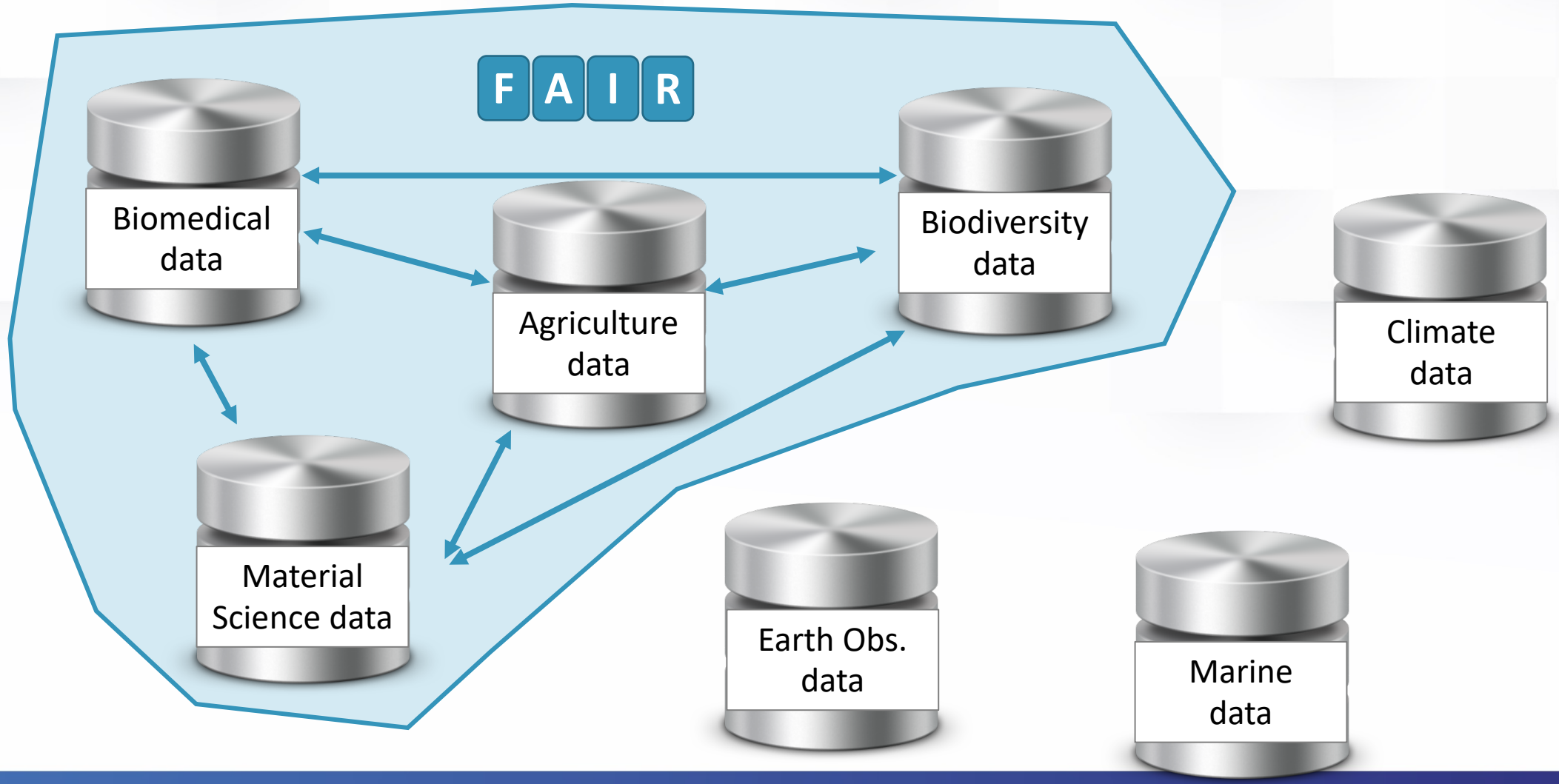
Tâche 5.4

Former les data stewards et les chercheurs d'aujourd'hui et de demain

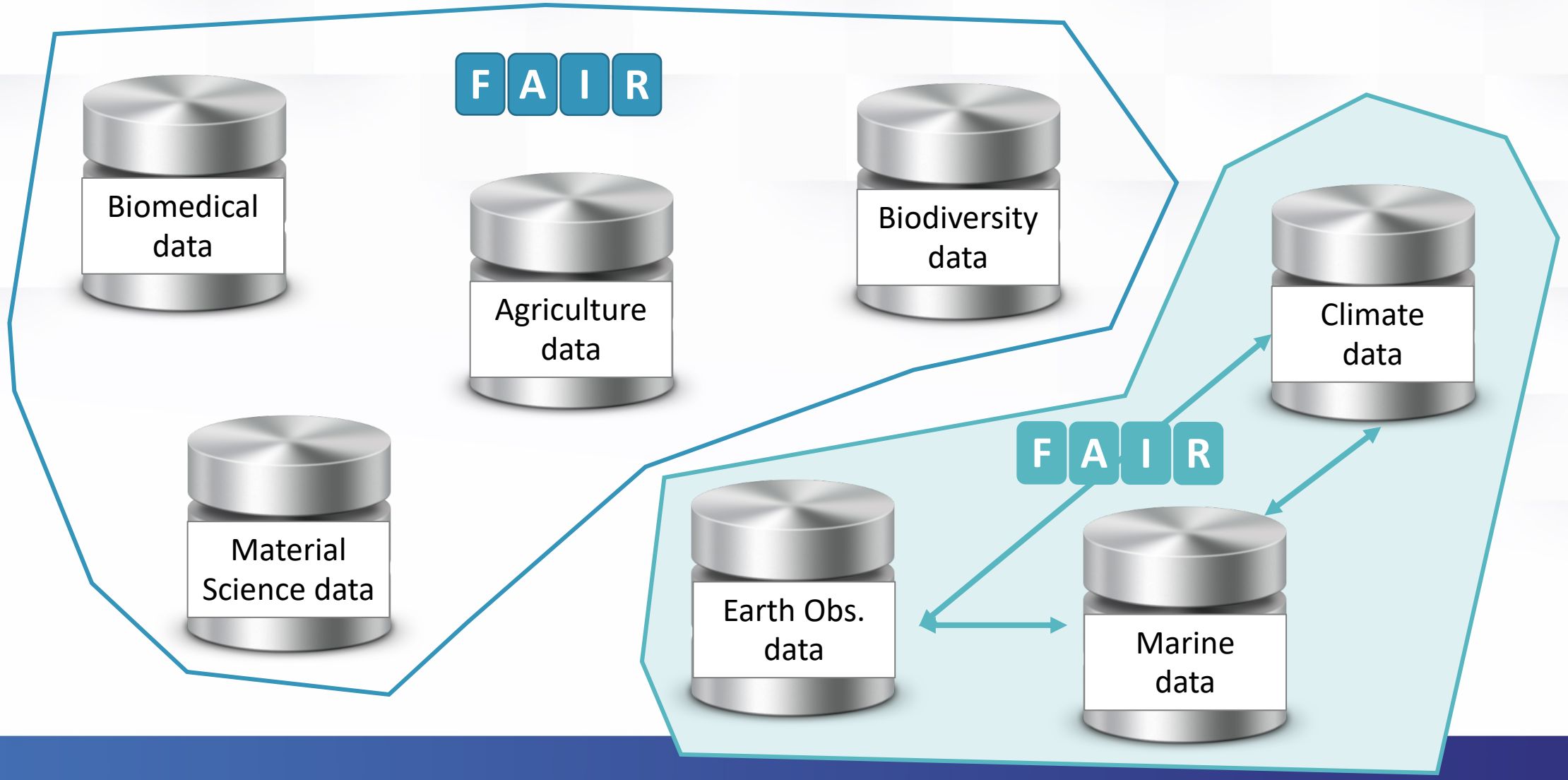
Technologie

Pourquoi construire un espace fédéré de données FAIR ?
EOSC-Pillar: un projet pluridisciplinaire

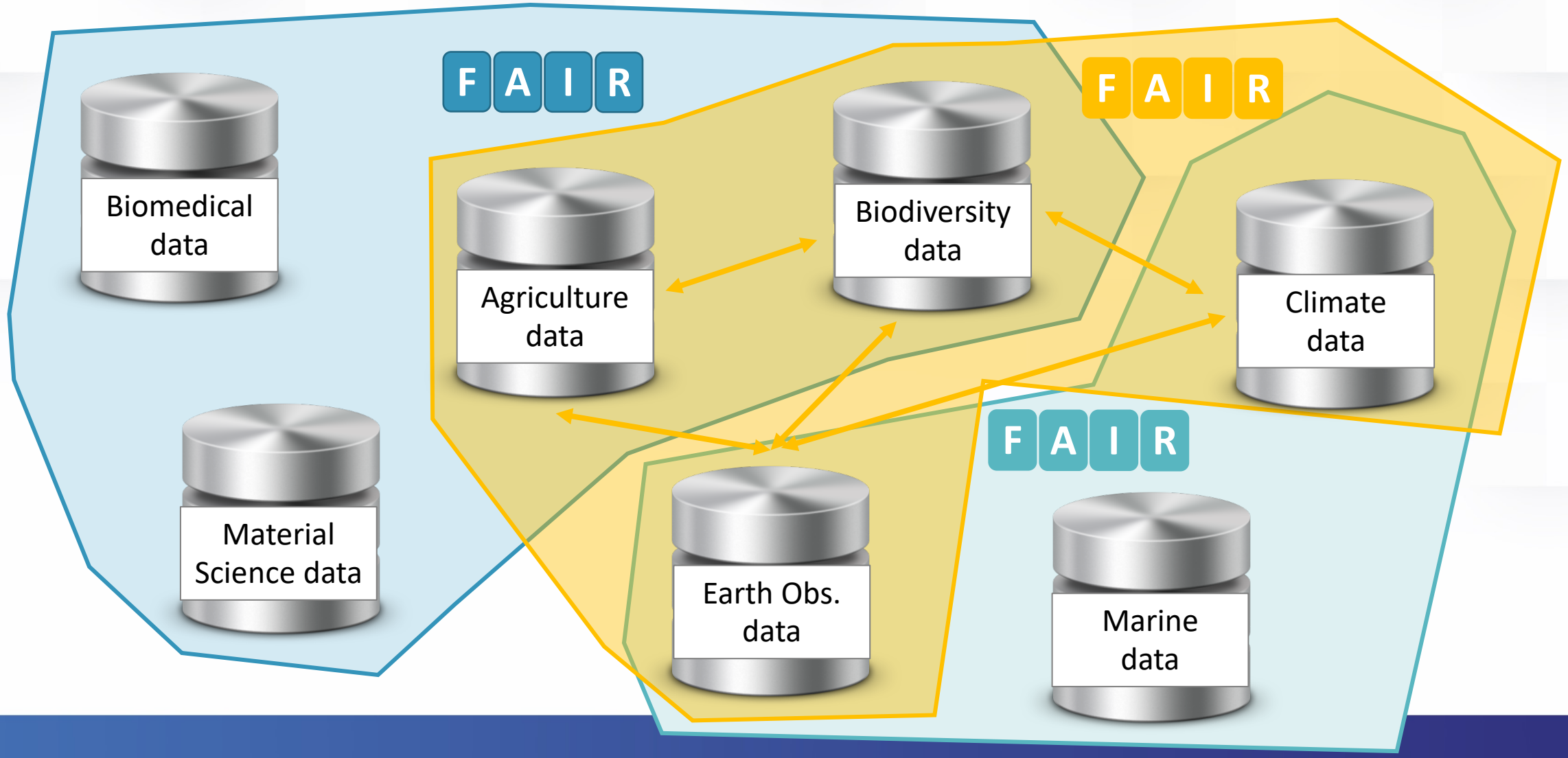
Pourquoi construire un espace fédéré de données FAIR ?
EOSC-Pillar: un projet pluridisciplinaire



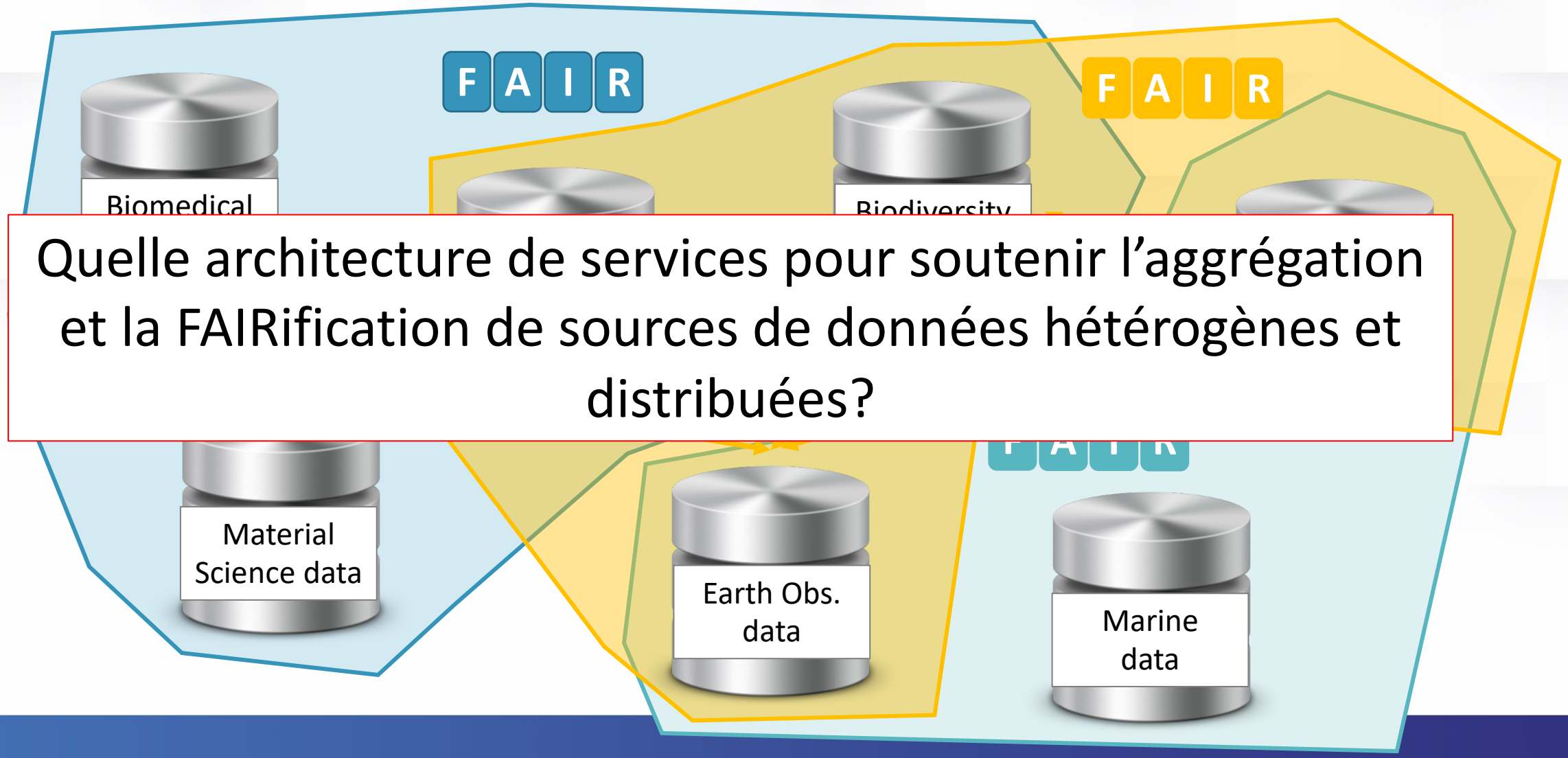
Pourquoi construire un espace fédéré de données FAIR ?
EOSC-Pillar: un projet pluridisciplinaire



Pourquoi construire un espace fédéré de données FAIR ?
EOSC-Pillar: un projet pluridisciplinaire



Pourquoi construire un espace fédéré de données FAIR ?
EOSC-Pillar: un projet pluridisciplinaire



Quelle architecture de services pour soutenir l'aggrégation et la FAIRification de sources de données hétérogènes et distribuées?

Les principes FAIR: un guide pour des informations utilisables par les machines

Considérations techniques

- Identifiant globalement unique persistent et résolvable
- Métadonnées
  - Provenance
  - Format commun
  - Vocabulaires/Ontologies FAIR
  - Intègre les standards de métadonnées spécifiques aux domaines
  - Persistentes
- Protocole d'accès gratuit, ouvert et universel
- Licenses (lisible par les humains et les machines)

Services existants

Publier des données FAIR

FAIR DATA POINT

Enrichir les (méta)données avec de la sémantique

FAIRIFIER

SEMAPHORA

Evaluer le niveau de FAIRification

F-UJI

FAIRshake

FAIRsharing.org
standards, databases, policies

Tâche 5.1

Fournisseur de données/
Opérateurs d’entrepôts

Tâche 5.2

Utilisateurs des données

EOSC-Pillar

Coordination and Harmonisation of National & Thematic Initiatives to support EOSC

INES Centre Informatique National de l'Enseignement Supérieur

Tâche 5.1

Fournisseur de données/
Opérateurs d'entrepôts

Tâche 5.2

Utilisateurs des données

Consiglio Nazionale delle Ricerche

Federated FAIR Data Space tool set

FAIR DATA POINT

FAIRIFIER

SEMAPHORA

F-UJI

FAIRshake

FAIRsharing.org
standards, databases, policies

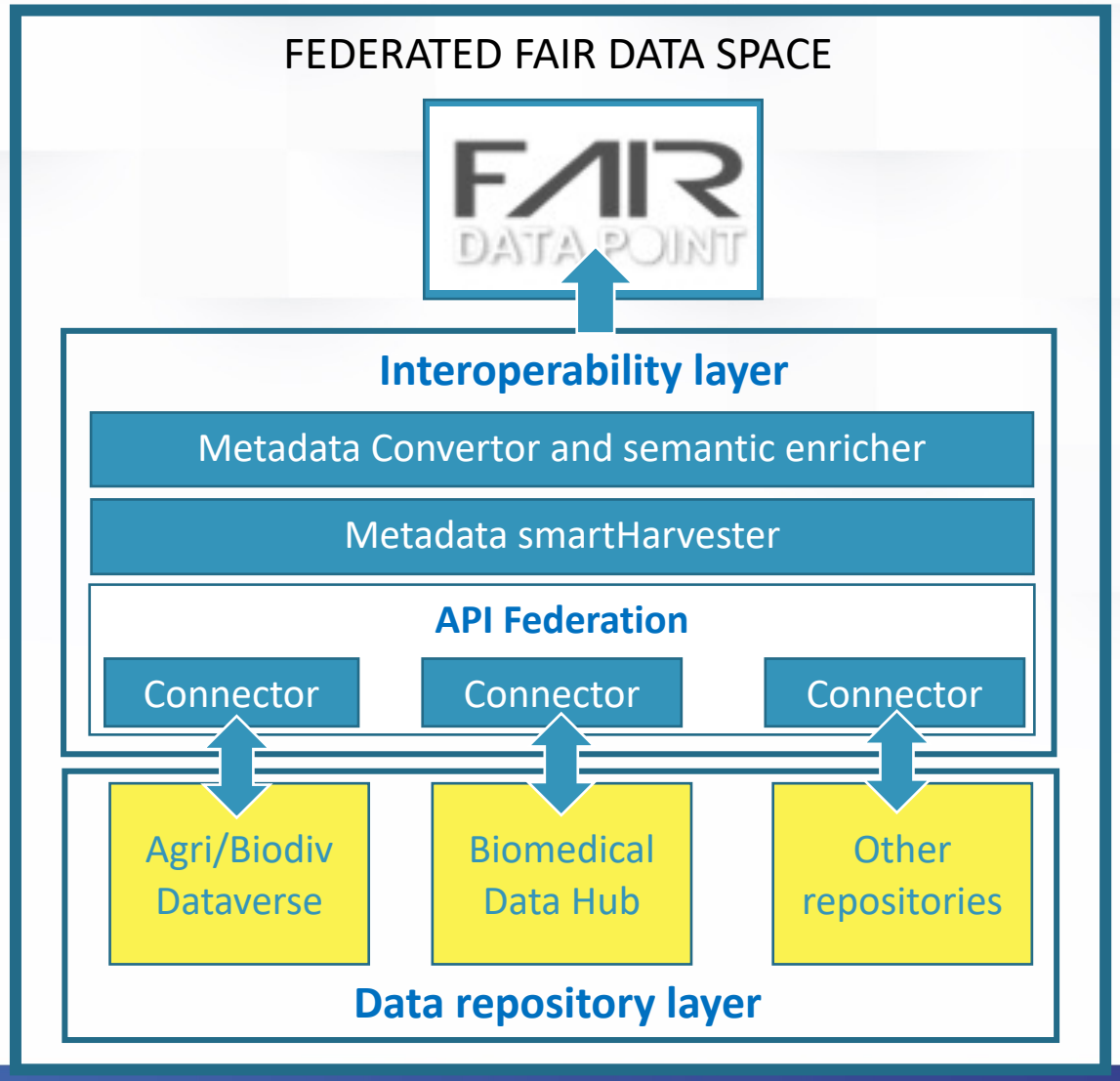
Aggrégation/Enrichissement/Evaluation

D4SCIENCE INFRASTRUCTURE

Visualisation/Recherche

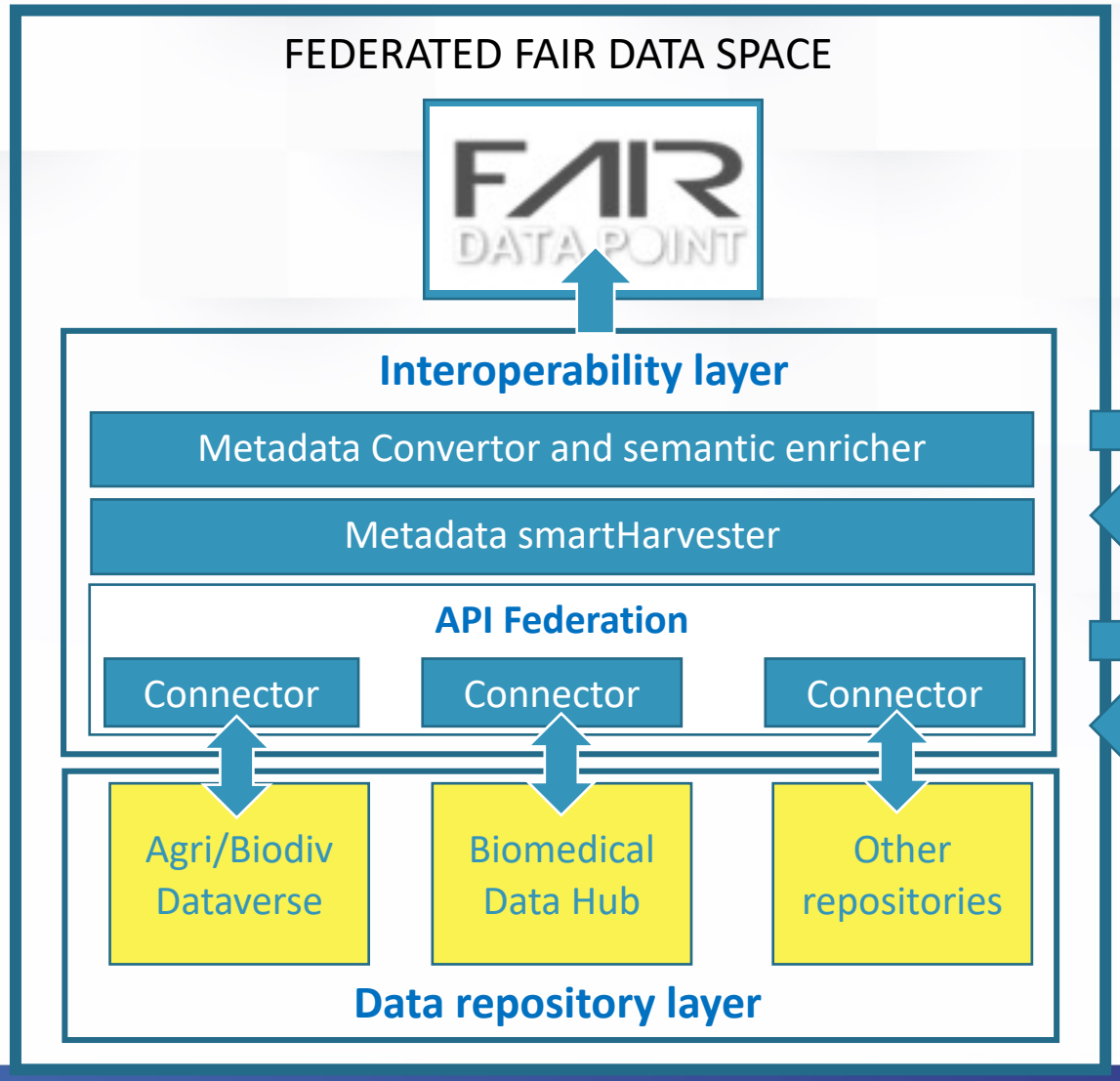
Une architecture simple

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



Une architecture simple

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



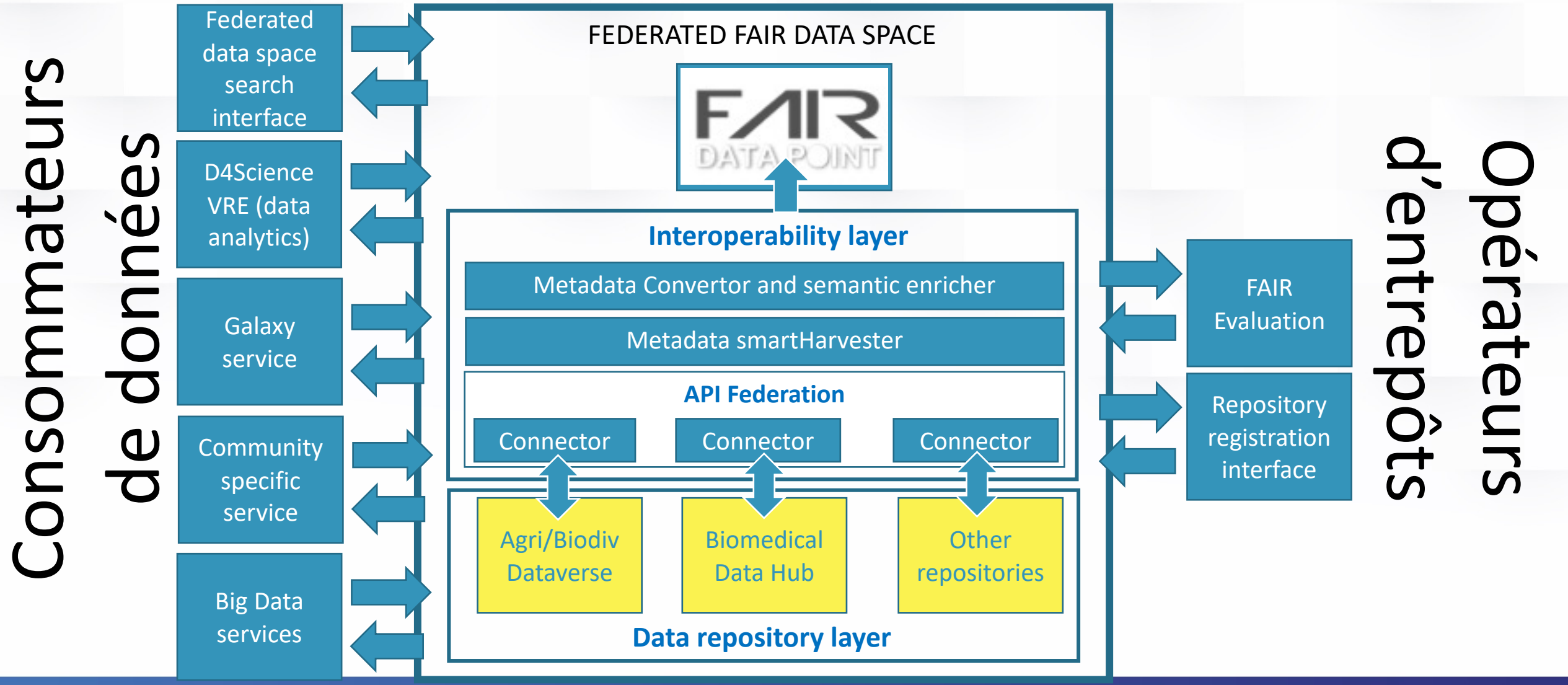
FAIR Evaluation

Repository registration interface

Opérateurs d'entrepôts

Une architecture simple

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



Un déploiement simple et évolutif

Un ensemble de services open-source facilement déployables et évolutifs sur n'importe quelle infrastructure cloud en utilisant Kubernetes

Créer des espaces fédérés de données à la volée pour des utilisations diverses

Enregistrer un entrepôt dans plusieurs espaces fédérés sans re-entrer les informations

Bénéfices

- ✓ **Crée rapidement des espaces fédérés avec des ressources distribuées et hétérogènes au sein de la même organization, entre organisations,…**
- ✓ **Pas ou peu de changements nécessaires pour les entrepôts**
- ✓ **Rend le contenu des entrepôts FAIR**
- ✓ **Des méthodes d'accès multiples pour utiliser les données (SparQL, REST API and web UI)**
- ✓ **Se connecte facilement avec n'importe quel service d'analyse (Galaxy, VREs, PANGEO, HPC/HTC, ...)**

Un workflow centré sur les opérateurs d'entrepôts

Enregistre l'entrepôt

Fourni des métadonnées décrivant l'entrepôt

Un workflow centré sur les opérateurs d'entrepôts

Enregistre l'entrepôt

Fourni des métadonnées décrivant l'entrepôt

Enregistre l'API

Fourni une description de l'API

Un workflow centré sur les opérateurs d'entrepôts

Enregistre l'entrepôt

Fourni des métadonnées décrivant l'entrepôt

Enregistre l'API

Fourni une description de l'API

Aligne les métadonnées

W3C

Data Catalog Vocabulary (DCAT) Version 2

W3C Recommendation 04 February 2020

Un workflow centré sur les opérateurs d'entrepôts

Enregistre l'entrepôt

Fourni des métadonnées décrivant l'entrepôt

Enregistre l'API

Fourni une description de l'API

Aligne les métadonnées

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

Enrichi les métadonnées

Semantic Web

Un workflow centré sur les opérateurs d'entrepôts

Enregistre l'entrepôt

Fourni des métadonnées décrivant l'entrepôt

Enregistre l'API

Fourni une description de l'API

Aligne les métadonnées

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

Enrichi les métadonnées

Semantic Web

Publie les métadonnées

FAIR DATA POINT

Un workflow centré sur les opérateurs d'entrepôts

Enregistre l'entrepôt

Fourni des métadonnées décrivant l'entrepôt

Enregistre l'API

Fourni une description de l'API

Aligne les métadonnées

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

Enrichi les métadonnées

Semantic Web

Publie les métadonnées

FAIR DATA POINT

Evalue FAIRness

Cherche et réutilise

Un workflow centré sur les opérateurs d'entrepôts

Enregistre l'entrepôt

Fourni des métadonnées décrivant l'entrepôt

Enregistre l'API

Fourni une description de l'API

Aligne les métadonnées

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

Enrichi les métadonnées

Semantic Web

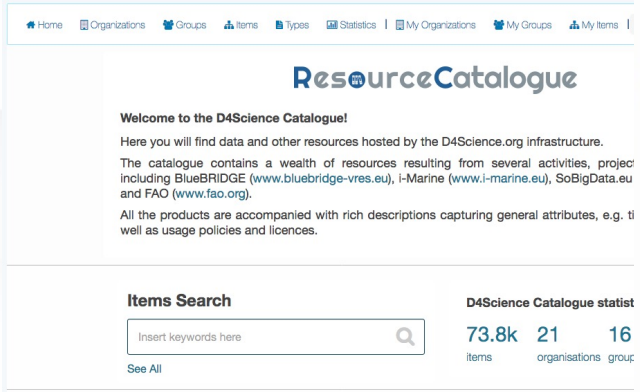
Publie les métadonnées

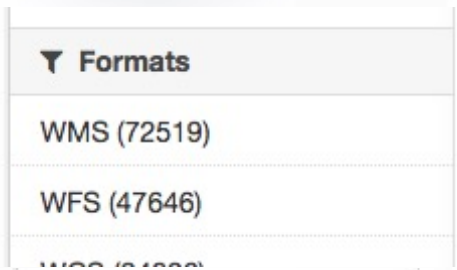
FAIR DATA POINT

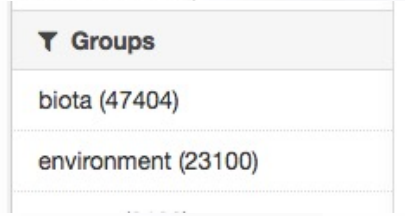
Evalue FAIRness

Cherche et réutilise

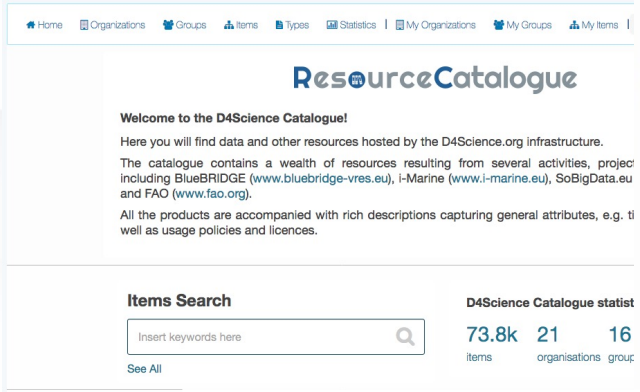
Accéder aux données via un catalogue D4Science

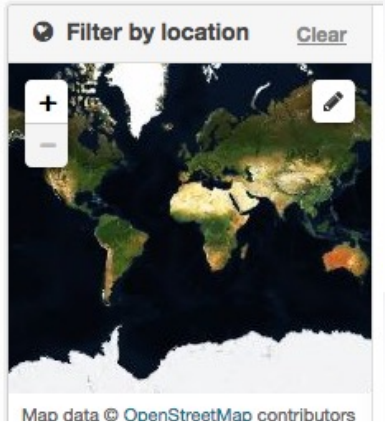


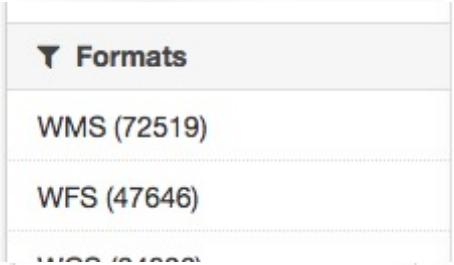


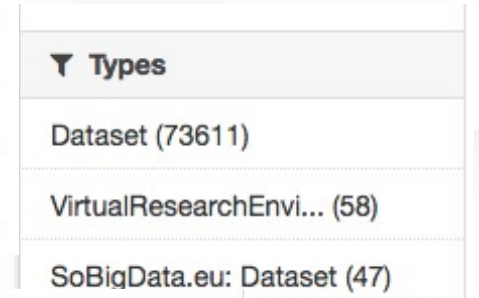


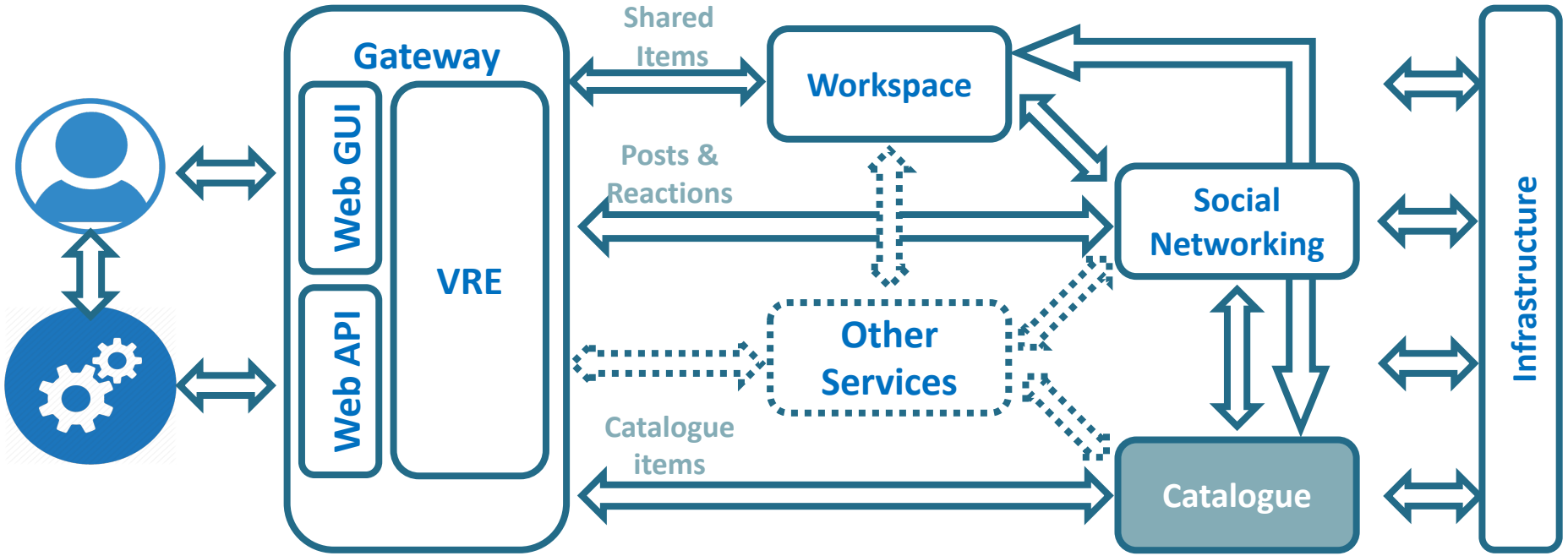
Accéder aux données via un catalogue D4Science

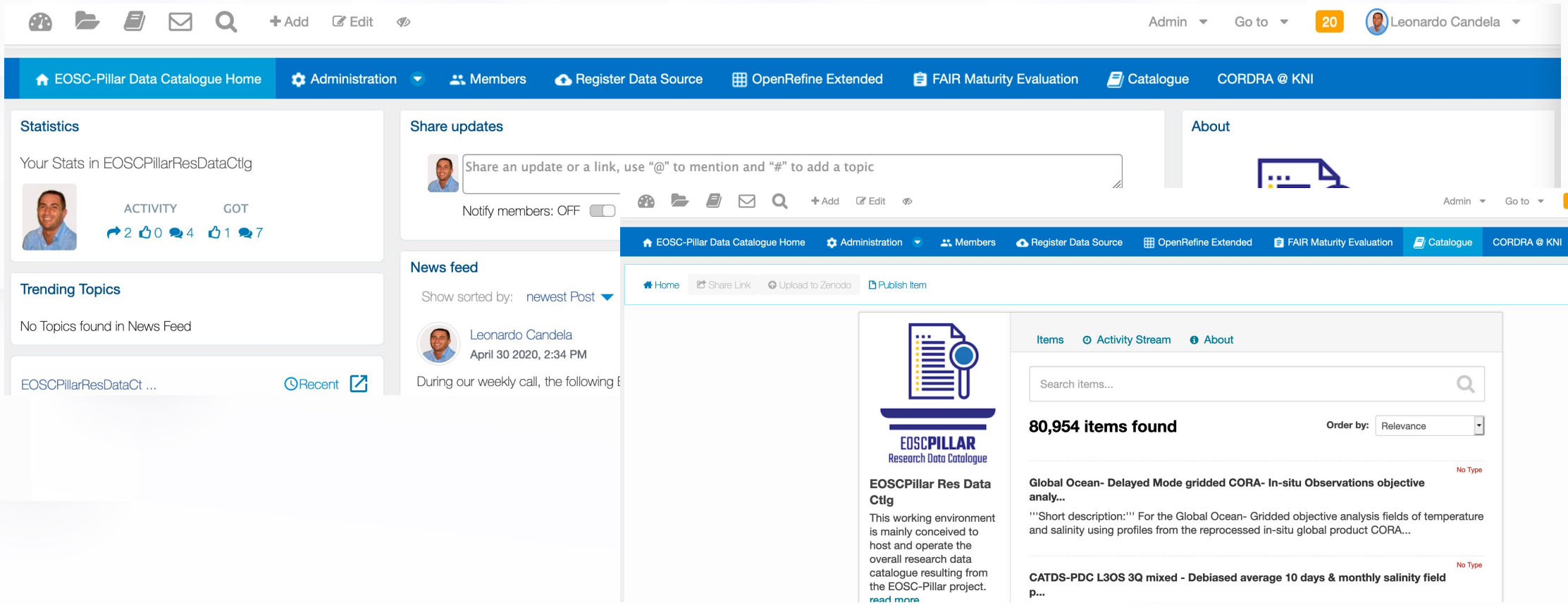











Add
Edit
Admin
Go to
20
Leonardo Candela
EOSC-Pillar Data Catalogue Home
Administration
Members
Register Data Source
OpenRefine Extended
FAIR Maturity Evaluation
Catalogue
CORDRA @ KNI
Statistics
Your Stats in EOSCPillarResDataCtlg
ACTIVITY
GOT
2 0 4 1 7
Trending Topics
No Topics found in News Feed
EOSCPillarResDataCt ...
Recent
Share updates
Share an update or a link, use "@" to mention and "#" to add a topic
Notify members: OFF
News feed
Show sorted by:
newest Post
Leonardo Candela
April 30 2020, 2:34 PM
During our weekly call, the following
About
Home
Share Link
Upload to Zenodo
Publish Item
Items
Activity Stream
About
Search items...
80,954 items found
Order by:
Relevance
EOSCPILLAR
Research Data Catalogue
EOSCPillar Res Data Ctlg
This working environment is mainly conceived to host and operate the overall research data catalogue resulting from the EOSC-Pillar project.
read more
No Type
Global Ocean- Delayed Mode gridded CORA- In-situ Observations objective analy...
'''Short description:''' For the Global Ocean- Gridded objective analysis fields of temperature and salinity using profiles from the reprocessed in-situ global product CORA...
No Type
CATDS-PDC L3OS 3Q mixed - Debiased average 10 days & monthly salinity field p...

Un workflow centré sur les opérateurs d'entrepôts

Enregistre l'entrepôt

Fourni des métadonnées décrivant l'entrepôt

Enregistre l'API

Fourni une description de l'API

Aligne les métadonnées

W3C

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

Enrichi les métadonnées

W3C Semantic Web

Publie les métadonnées

FAIR DATA POINT

Besoin d'un espace sémantique harmonisé permettant d'accéder aux concepts pour enrichir les métadonnées

Un workflow centré sur les opérateurs d'entrepôts

Register Data	Register API	Map	Enrich	Publish
			Web	

Comment pouvons nous créer un tel espace harmonisé permettant d'utiliser les artéfacts sémantiques existants pour les différentes communautés pour améliorer le niveau de FAIRness des données?

Besoin d'un espace sémantique harmonisé permettant d'accéder aux concepts pour enrichir les métadonnées

Analyse des ressources ontologiques utilisées par les communautés

	A	B	C	D	E	F	G
1	**Questions on semantic artefacts (SA) common in the discipline(s) of: Use Case 5**						
2	We are interested in semantic artefacts (i.e. glossaries, controlled vocabularies, thesauri, ontologies, ...) that are common in the discipline(s) covered by this use case. Please answer all questions below considering everything you know on SAs in this discipline/these disciplines.						
3	To resolve any potential misunderstanding arising around the concept of ontologies, we are using instead the concept of semantic artefacts coined in the FAIRsFAIR project and defined as such within the FAIR Semantics recommendations (https://doi.org/10.5281/zenodo.3707985): "semantic artefact is defined in this work as a machine-actionable and -readable formalisation of a conceptualisation enabling sharing and reuse by humans and machines. These artefacts may have a broad range of formalisation, from loose set of terms, taxonomies, thesauri to higher-order logics. Moreover, semantic artefacts are serialised using a variety of digital representation formats, e.g., RDF Turtle, OWL-RDF, XML, JSON-LD.". For more information regarding the concept, please read the introductory part of the FAIR Semantics document. If you have any questions, please reach out to lefranc@cines.fr.						
4							
5	Please think about SAs that are important or common in the discipline(s) covered by this use case: Are there any repositories in which these SAs are available? Please list the name of each repository and provide a link. In case you know whether the access to this repository is regulated, please choose the respective category in the row "access regulation". Please also fill in the information which "discipline(s)" the repository covers. Please add rows to the table below if the number of rows is not sufficient. If you would like to add any comment, please use the						
6		**Name of the repository**	**Link to the repository**	**Access regulation**	**Discipline(s)**	**Comment**	**Added by**
7	**Repository 1**						
8	**Repository 2**						
9	**Repository 3**						
10	**Repository 4**						
11	**Repository 5**						
12	**Repository 6**						
13	**Repository 7**						
14	**Repository 8**						
15	**Repository 9**						
16	**Repository 10**						
17	**Repository 11**						
18	**Repository 12**						
19	**Repository 13**						
20	**Repository 14**						
21	**Repository 15**						
22							
23							
24	If there are important/common semantic artefacts that are **not** stored in repositories, please provide the name, link and a short description of the ten most important SAs. Please take all discipline(s) into account that are covered by this use case. If you would like to add any comment, please use the respective column.						
25		**Name of the SA**	**Link to the SA**	**Description**	**Comment**		**Added by**
26	**Sem. Art. 1**	CIDOC-CRM	UC 6.9		Archeology UC; See Franco Niccolucci		Adeline Joffr
27	**Sem. Art. 2**	PACTOLS	https://pactols.frantiq.f	https://www.frantiq.fr/tag/	Archeology Vocab		Adeline Joffr
28	**Sem. Art. 3**				Archeology, LIDAR data, to be checked by Adeline with scientists;		Adeline Joffr
29	**Sem. Art. 4**				Conservatory for 3D data has specific data model		Adeline Joffr
30	**Sem. Art. 5**	JEL Classification System	https://www.aeaweb.or	Standard method of classifying scholarly literature in the field of economics			Bénédicte Ku
31	**Sem. Art. 6**	ACM Computing Classificatio	https://dl.acm.org/ccs				Bénédicte Ku
32	**Sem. Art. 7**	Thesaurus INRAE	https://vocabulaires.irs	Agriculture, Environment			Bénédicte Ku

- 1 questionnaire pour collecter les informations sur les ressources sémantiques existantes dans chaque communauté:
  - Les entrepôts sémantiques (AgroPortal, Bioportal, NVS,...)
  - Les ontologies/vocabulaires contrôlés

Analyse des ressources ontologiques utilisées par les communautés

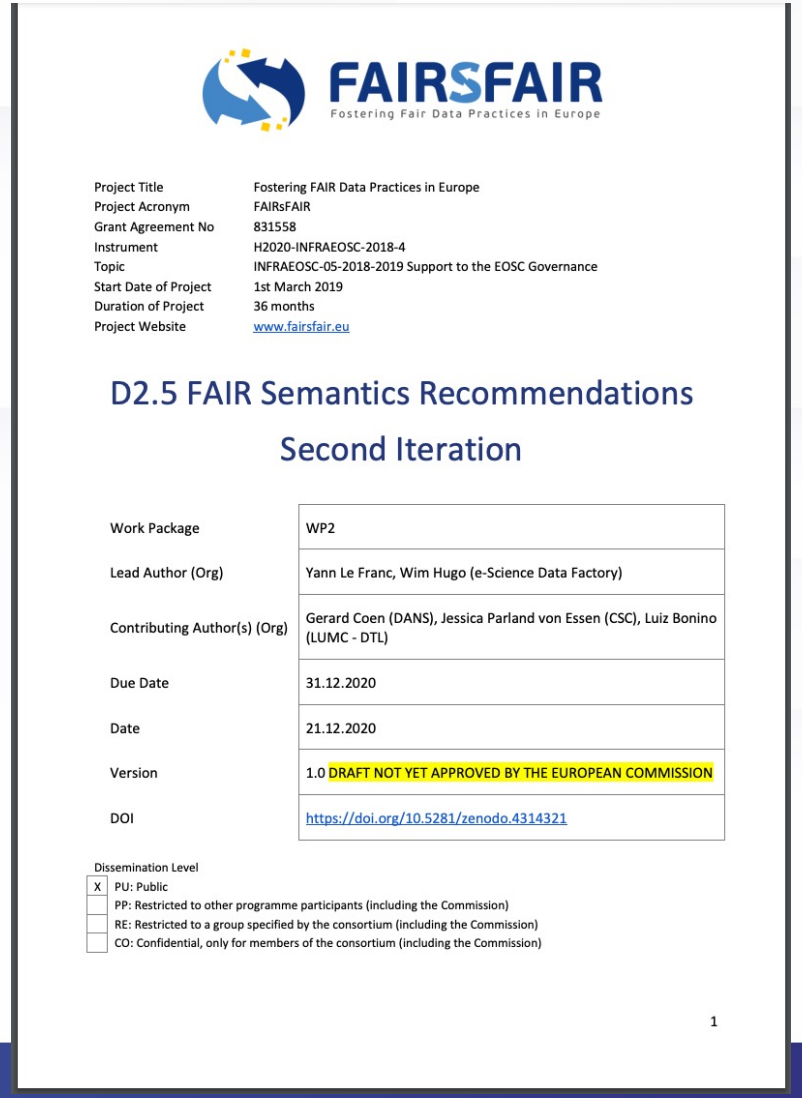
	Topic	What would we like to know?	How can we measure this?	What is the content of one specific variable measuring (one aspect of) what we want to know?	type of variable or categories of this variable on the left	Variabename
2	general info	What is the name of the SA?	enter the name of the SA in a textfield	the name of the SA	text field	gen_name
3	general info	in which UC is the SA used?	use information in the Excel file	the names of the UC	text field	gen_uc
4	general info	How weak or strong is the SA?	use the information in the Excel file to classify the SA based on the categories by Leo Obrst 2010 as cited in FAIRsFAIR 2020	one of the categories by Leo Obrst 2010 as cited in FAIRsFAIR 2020	list hierarchy thesaurus formal ontology don't know	gen_class
5	general_info	To what extent is the SA machine-readable?	Classify the ontology in how well machines can read the content	extent of machine-readability according to our categories	Machine visualisable Machine parseable Machine understandable	gen_machine
6	general info	In which format is the file saved/ available?	format of the file (e.g. xml, pdf…)	format of the SA	text field of format	
7	general info	Is the SA publicly available?	use information in Excel file	Is the SA publicly available?	yes no don't know	gen_publ
8	general info	Is the SA available free of charge?	use information in Excel file	Is the SA accessible at no cost?	yes no don't know	gen_cost
9	general info	Is the SA publicly available via a repository?	use information in Excel file	Is the SA publicly available via a repository?	yes no don't know	gen_repo
10	general info	In which repository is the SA publicly available?	use information in Excel file + desk research	URL to the repository in which the SA is available	text field	gen_repo_which
11	general info	Has the SA been developped by the UC?	look in the Excel sheet or ask UC whether they have developped a specific SA	Has the SA been developped by the UC?	yes no don't know	gen_dev
12	general info	If the UC has developped a SA, do they plan to publish the SA?	ask UC	Do you plan to publish the SA?	yes no don't know	gen_dev_pub
13	PID	Does the SA have a persistent identifier (PID) like a DOI?	By means of a variable that captures whether the SA currently is identifiable by a PID or not	Does the SA have a persistent identifier (PID) like a DOI?	yes no don't know	pid_yesno
14	PID	Which persistent identifier(s) (PID) is/are currently associated with the SA?	by means of a series of dummy variables, one per PID we are interested in	Which DOI is associated to the SA?	text field	pid_which_doi
15	PID			Is a URN associated to the SA?	yes + text field no don't know	pid_which_urn
16	PID			Is a Handle associated to the SA?	yes + text field no don't know	pid_which_handle
17	PID			Is a persistent URL associated to the SA?	yes + text field no don't know	pid_which_purl

	Name of the SA	Link to the SA	Description	Comment	Added by
Sem. Art. 1	CIDOC-CRM	UC 6.9		Archeology UC; See Franco Niccolucci	Adeline Joffr
Sem. Art. 2	PACTOLS	https://pactols.frantiq.f	https://www.frantiq.fr/tag/	Archeology Vocab	Adeline Joffr
Sem. Art. 3				Archeology, LIDAR data, to be checked by Adeline with scientists;	Adeline Joffr
Sem. Art. 4				Conservatory for 3D data has specific data model	Adeline Joffr
Sem. Art. 5	JEL Classification System	https://www.aeaweb.or	Standard method of classifying scholarly literature in the field of economics		Bénédicte Ku
Sem. Art. 6	ACM Computing Classificatio	https://dl.acm.org/ccs			Bénédicte Ku
Sem. Art. 7	Thesaurus INRAE	https://vocabulaires.irs	Agriculture, Environment		Bénédicte Ku

- 1 questionnaire pour collecter les informations sur les ressources sémantiques existantes dans chaque communauté:
  - Les entrepôts sémantiques (AgroPortal, Bioportal, NVS,…)
  - Les ontologies/vocabulaires contrôlés
- Evaluation des ressources selon divers critères et notamment les recommandations FAIR Semantics proposées par le projet FAIRsFAIR
- Etablir une liste de ressources accessibles par les machines pour aider à l'enrichissement sémantique dans l'espace fédéré de données FAIR

FAIR Semantics recommendations

- Développées avec la contribution d'une communauté d'experts de domaines différents (science de la terre, géospatial, biodiversité, écology, biomédecine,…)
- 17 recommandations générales/10 bonnes pratiques
- Alignées avec les recommandations existantes proposées par les différentes communautés (OBO Foundry, IOF principles,…)
- Seconde itération des recommandations FAIR Semantics alignées avec RFC 2119 livrée Janvier 2021 https://doi.org/10.5281/zenodo.4314321
- **La suite:** établir un cadre pratique pour l'implémentation des recommandations et les utiliser pour l'évaluation du niveau de FAIRness



FAIRSFAIR
Fostering Fair Data Practices in Europe

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

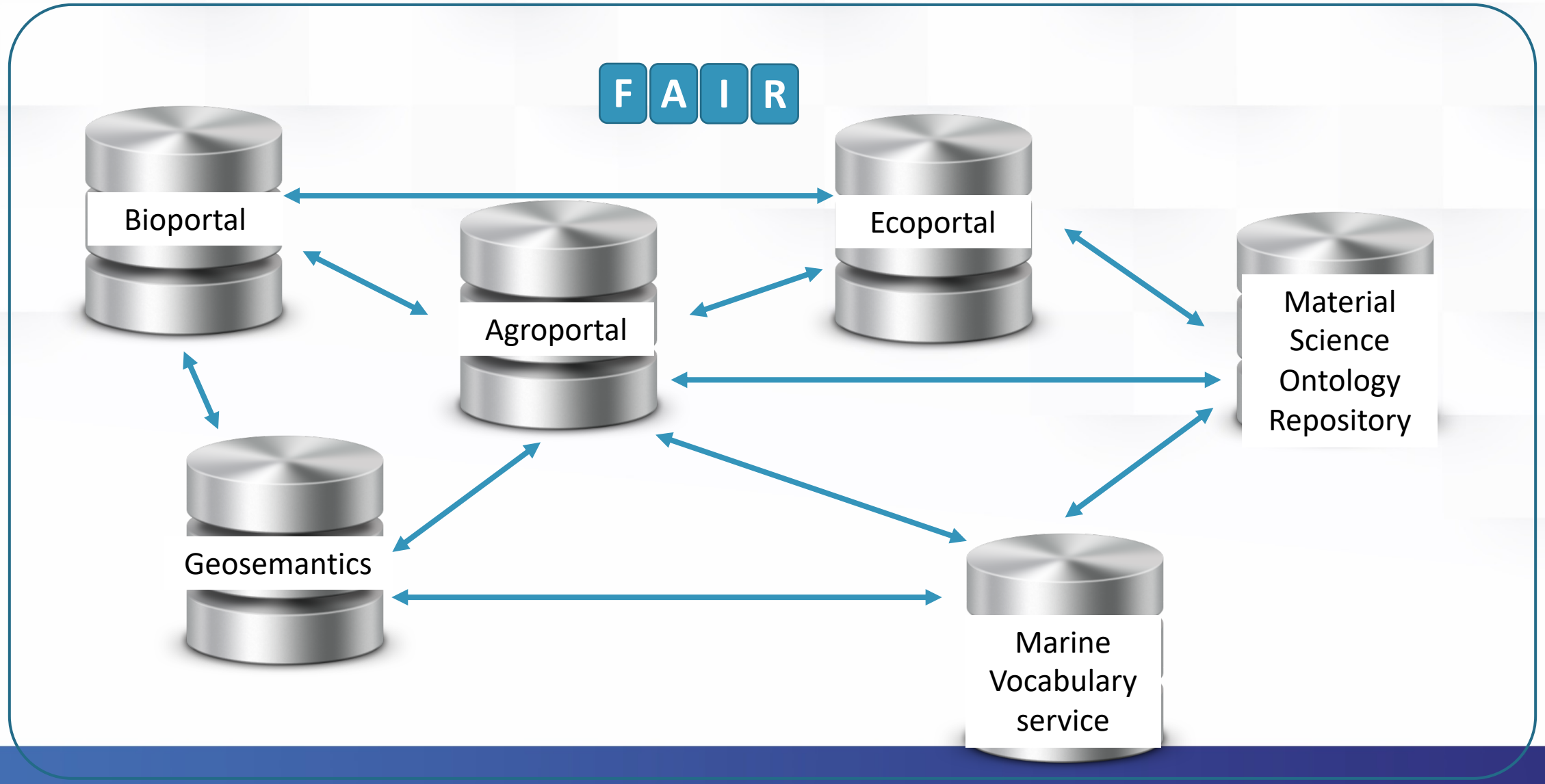
D2.5 FAIR Semantics Recommendations
Second Iteration

Work Package	WP2
Lead Author (Org)	Yann Le Franc, Wim Hugo (e-Science Data Factory)
Contributing Author(s) (Org)	Gerard Coen (DANS), Jessica Parland von Essen (CSC), Luiz Bonino (LUMC - DTL)
Due Date	31.12.2020
Date	21.12.2020
Version	1.0 DRAFT NOT YET APPROVED BY THE EUROPEAN COMMISSION
DOI	https://doi.org/10.5281/zenodo.4314321

Dissemination Level
[X] 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)

1

Créer un espace fédéré d'ontologies FAIR



Formation et Support

Promouvoir une culture de la donnée FAIR et l'intégration des bonnes pratiques et outils permettant de rendre les données FAIR

Consiglio Nazionale delle Ricerche

Tâche 5.3

Un catalogue de ressources sur les bonnes pratiques de gestion des données recherches FAIR

Tâche 5.4

UNIVERSITEIT GENT

Former les data stewards et les chercheurs d'aujourd'hui et de demain

Un pilote de formation

- 47 cours organisés sur la période 2019-2020

- 6 modules
  - Science Ouverte
  - Open Access
  - Gestion des données de recherche (plan de gestion des données)
  - Principes FAIR
  - EOSC basics
  - Data Stewardship

- Différentes typologies de cours:
  - Cours en ligne ouverts à tous
  - Cours en collaboration avec les infrastructures de recherche italienne (Science de la Terre, Santé)
  - Cours en collaboration avec universités (Université de Gent, Université de Pise, Scuola Normale Superiore)

- Prochaines étapes: extension à d'autres pays et à d'autres domaines avec notamment les Sciences Humaines et Sociales

Un catalogue de contenus pour la formation et le support

- Une collection de ressources accessibles en ligne pour le FAIR data stewardship et la gestion des données de recherche

Formation

- Objectif: acquérir des connaissances et **développer des compétences** en particulier sur des sujets autour de la gestion des données scientifiques
- Exemples: cours, slides, livres, tutoriaux, guidelines, jeux, vidéos...

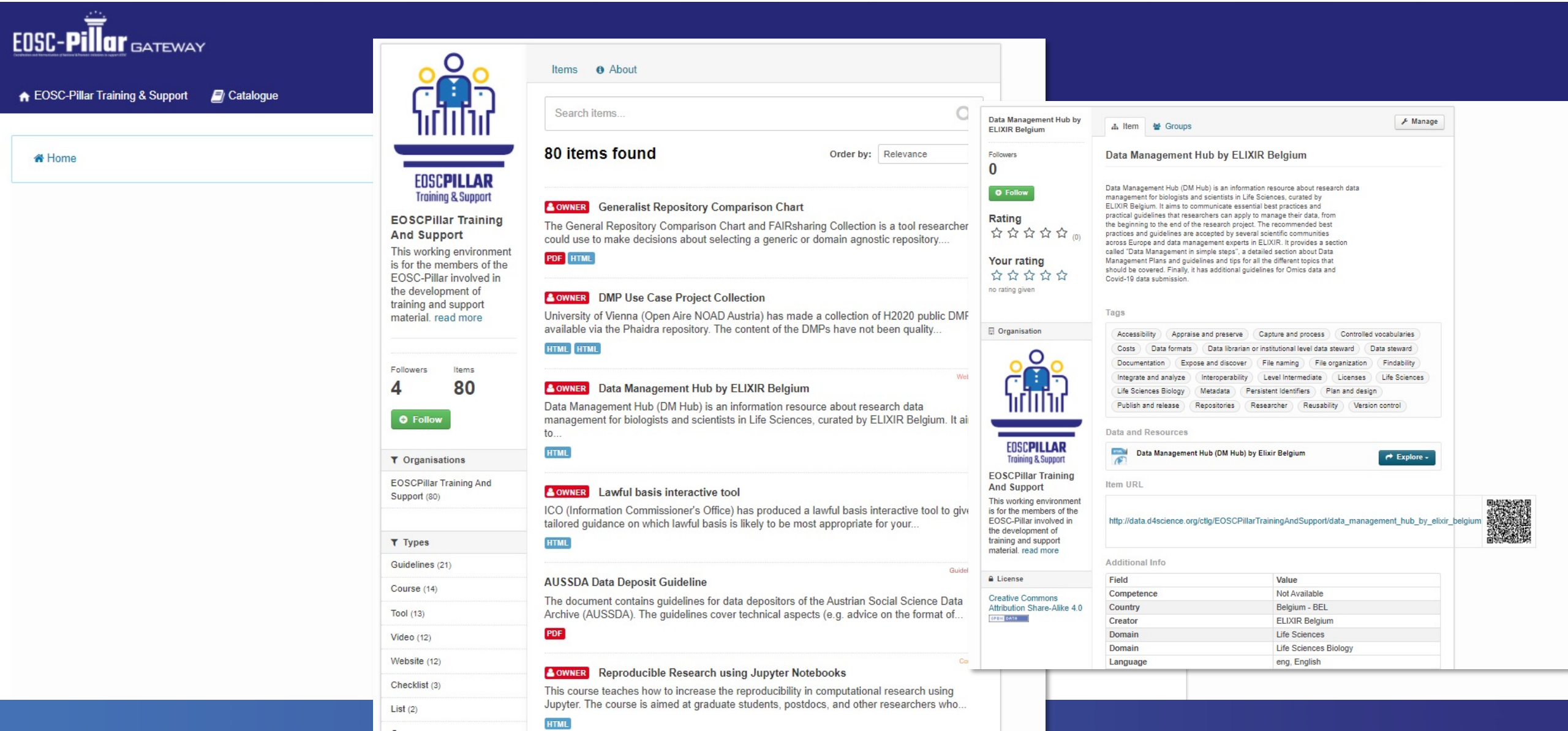
Support

- Objectif: **Outils et solutions prêtes à l'emploi**
- Exemples: checklists, arbres de décisions, guidelines, outils…

- Alignement avec d'autres projets et groups de travail autour de la formation et des compétences permettant une meilleure connectivité au sein d'EOSC

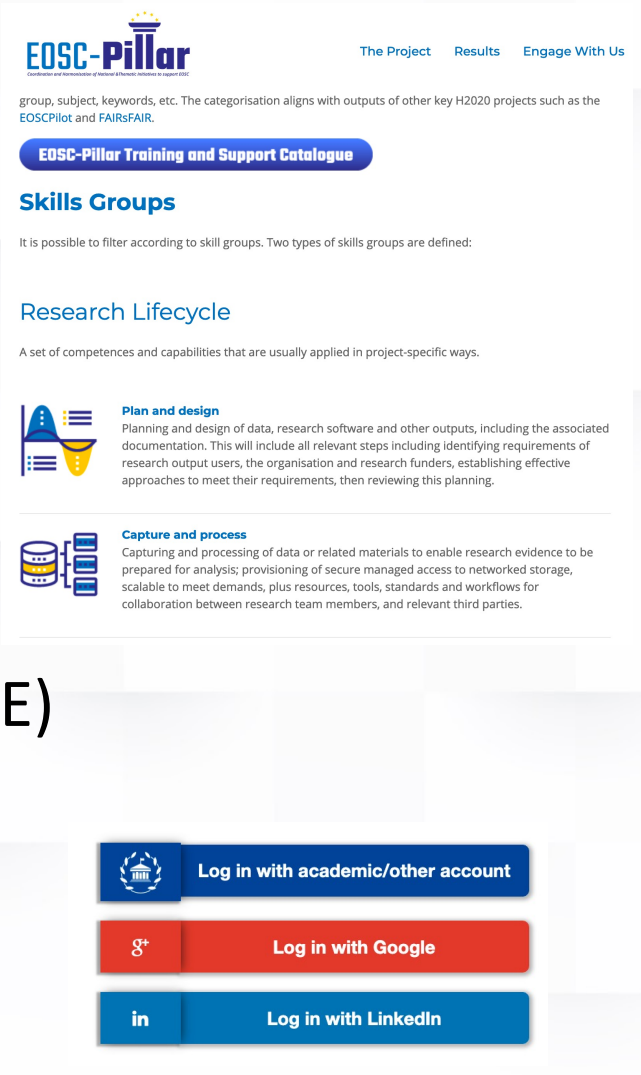
EOSC-Pillar has received funding from the European Union's Horizon 2020 research and innovation Programme under Grant Agreement No. 857650.

Un catalogue basé sur la technologie D4Science



Accèder au catalogue

- Via le site du projet (comme invité):
  - https://www.eosc-pillar.eu/training-and-support-catalogue
- Un catalogue pour les invités
  - https://eosc-pillar.d4science.org/web/eoscpillartrainingandsupport
- Un catalogue pour les “curateurs” et “utilisateurs connus” (via VRE)
  - https://eosc-pillar.d4science.org/group/eoscpillartrainingandsupport
- Autres matériel
  - Webinar 12-11-2020 recording and slides
  - FAQs



EOSC-Pillar

Coordination and Harmonisation of National &Thematic Initiatives to support EOSC

Thank you!

Get in touch with us!

www.eosc-pillar.eu

@EoscPillar

/company/eosc-pillar

EOSC-Pillar has received funding from the European Union's Horizon 2020 research and innovation Programme under Grant Agreement No. 857650.