



ENVRI
FAIR

Environmental Research Infrastructures building FAIR services accessible for society, innovation and research

Damien Boulanger, SEDOO, AERIS

École thématique 2019 – e-Envir – Gif-sur-Yvette



ENVRI-FAIR has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824068



EOSC – Déclaration de la Commission Européenne

Culture de la donnée et données FAIR



**EUROPEAN OPEN
SCIENCE CLOUD**

Findable 
Accessible 
Interoperable 
Reusable 

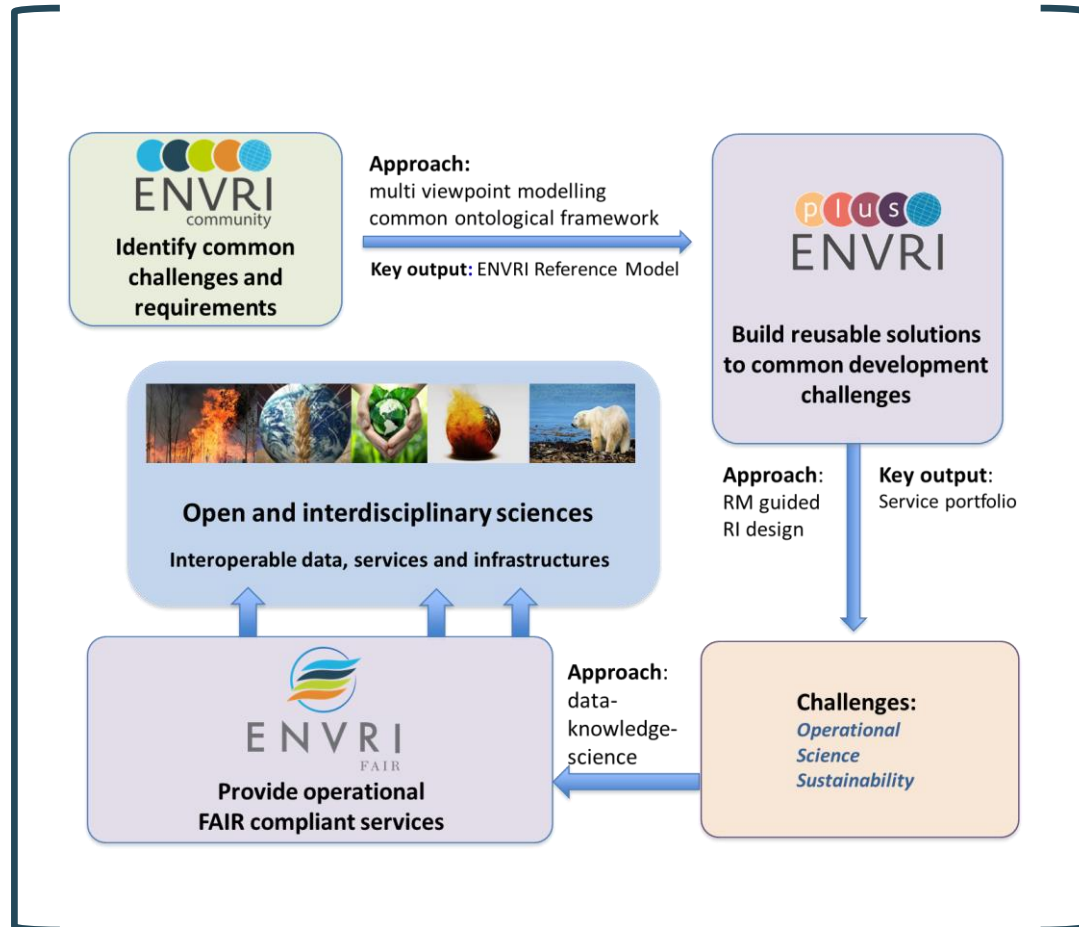
+

Reproductible

- Tous les chercheurs en Europe doivent avoir accès à un environnement de données de recherche ouvert, efficace et interdisciplinaire, basé sur les principes FAIR
- La science européenne doit être ancrée dans une culture commune de gestion des données, de sorte que les données de recherche soient reconnues comme un produit significatif de la recherche



Infrastructures de Recherche Européennes en Environnement (RIs)



ENVRI

- Construction de la communauté
- Cadre commun
- ENVRI Reference Model

ENVRIplus

- Développement communs
- Solutions partagées
- Service portfolio

ENVRI-FAIR

- Appliquer les principes FAIR dans les RIs
- Services conformes à FAIR
- Lier ENVRI à EOSC



Infrastructures de Recherche Européennes en Environnement (RIs)



ENVRI

- Construction de la communauté
- Cadre commun
- ENVRI Reference Model

ENVRIplus

- Développement communs
- Solutions partagées
- Service portfolio

ENVRI-FAIR

- Appliquer les principes FAIR dans les RIs
- Services conformes à FAIR
- Lier ENVRI à EOSC



Infrastructures de Recherche Européennes en Environnement (RIs)



Mission d'ENVRI-FAIR

Partager des expériences et trouver des solutions communes pour

- Le partage de données
- L'accessibilité
- L'interopérabilité

Développer des services communs

Exposer les services dans EOSC

Préparer ENVRI-hub



**EUROPEAN OPEN
SCIENCE CLOUD**



Le rôle d'ENVRI dans EOSC

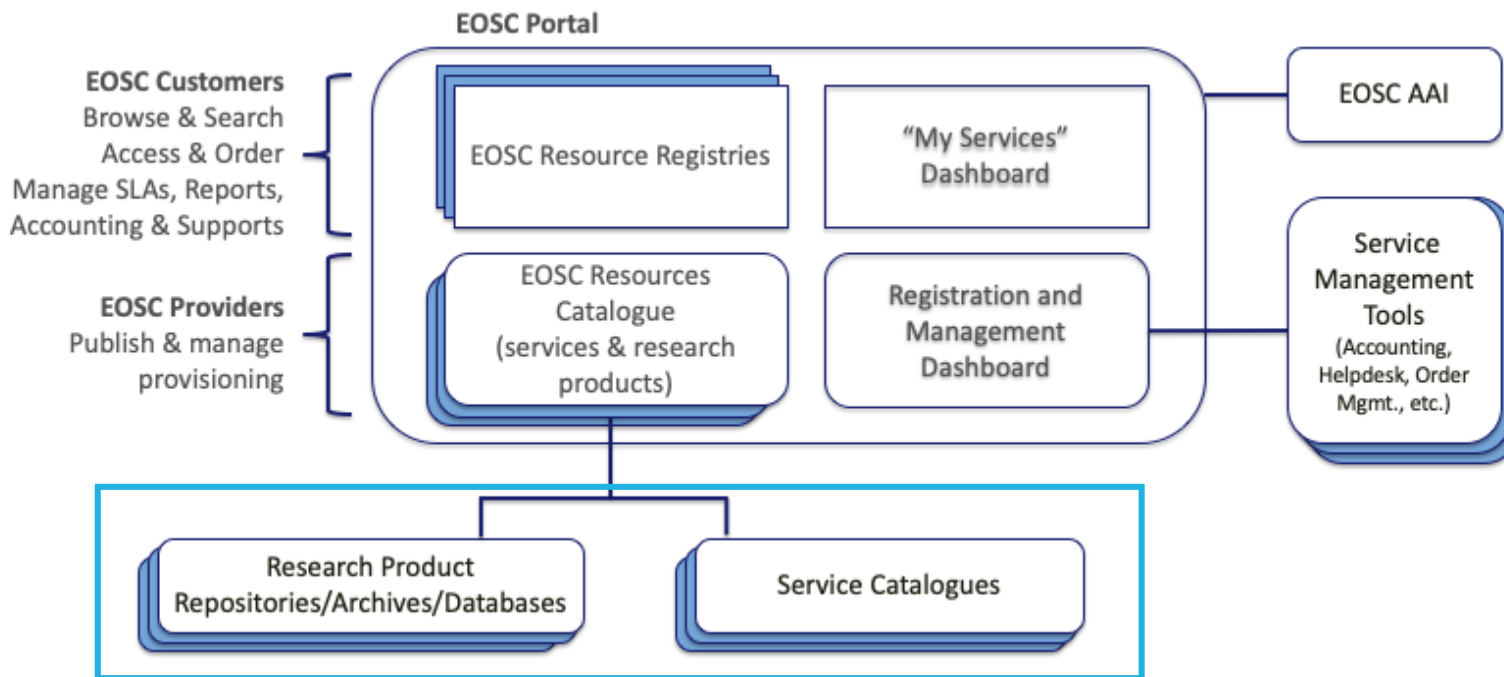


Illustration taken from EOSC Portal Concept 2.0, Version 2.2, July 2019

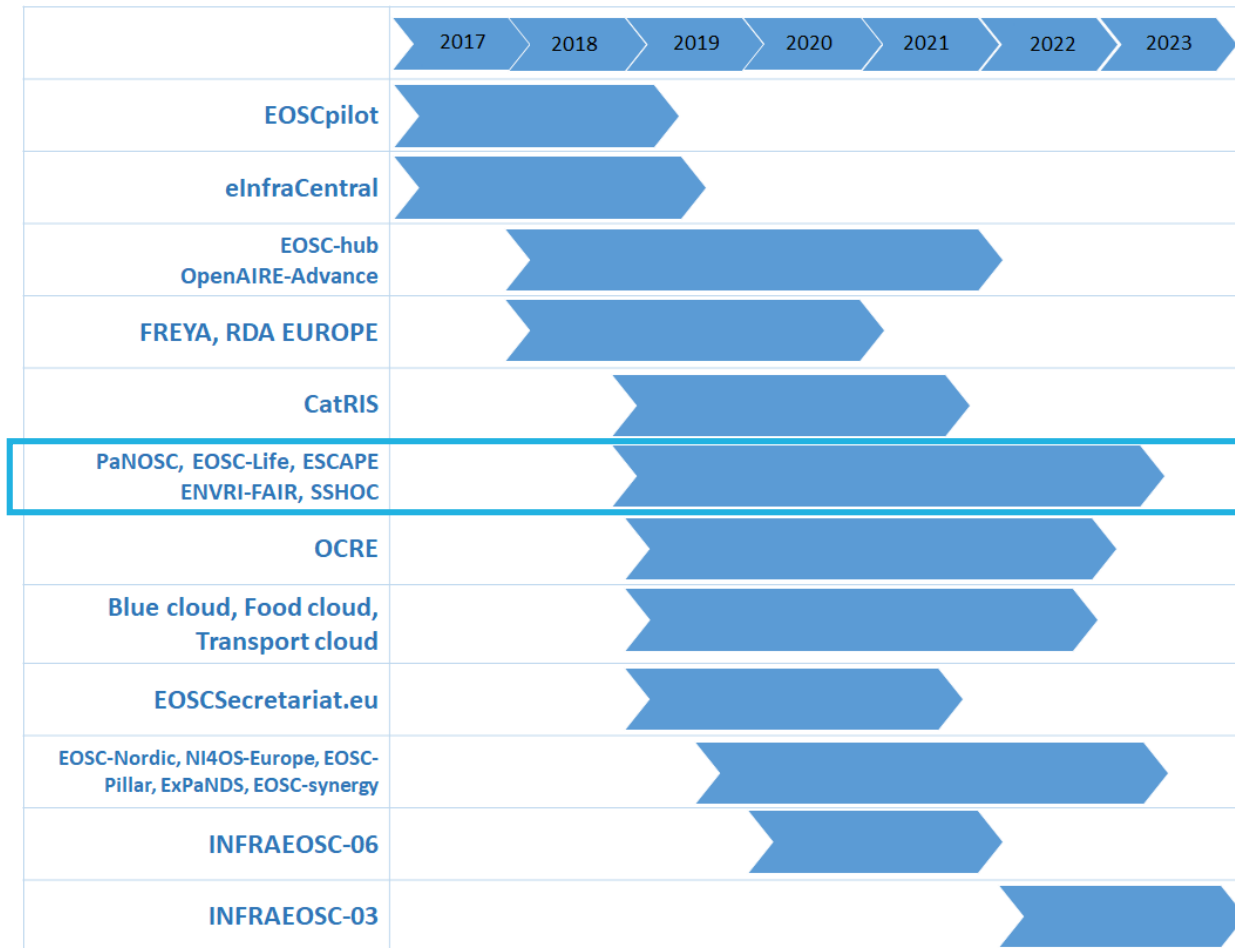
Le cluster ENVRI est un des projets principaux de la famille EOSC

ENVRI est considéré comme un des clusters les plus matures

La CE attend une forte contribution d'ENVRI aux services exposés dans EOSC



Le rôle d'ENVRI dans EOSC



Le cluster ENVRI est un des projets principaux de la famille EOSC

ENVRI est considéré comme un des clusters les plus matures

La CE attend une forte contribution d'ENVRI aux services exposés dans EOSC



Le projet

- Titre: ENVironmental Research Infrastructures building Fair services Accessible for society, Innovation and Research - ENVRI-FAIR
- Call: INFRAEOSC-04-2018
- Coordination: Andreas Petzold, FZ Jülich, Germany et Ari Asmi, University of Helsinki, Finland
- Consortium: 37 partenaires
- 13 Infrastructures : ESFRI Roadmap ou Landmark
- Budget: 19 M€
- Durée: 4 ans (Kick-off meeting en janvier 2019)



FAIR principles

<https://www.force11.org/group/fairgroup/fairprinciples>

TO BE **F**INDABLE:

- F1. (meta)data are assigned a globally unique and eternally persistent identifier.
- F2. data are described with rich metadata.
- F3. (meta)data are registered or indexed in a searchable resource
- F4. metadata specify the data identifier.

TO BE **A**CCESIBLE:

- A1 (meta)data are retrievable by their identifier using a standardized communications protocol.
 - A1.1 the protocol is open, free, and universally implementable.
 - A1.2 the protocol allows for an authentication and authorization procedure, where necessary.
- A2 metadata are accessible, even when the data are no longer available.

TO BE **I**NTEROPERABLE:

- I1. (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.
- I2. (meta)data use vocabularies that follow FAIR principles.
- I3. (meta)data include qualified references to other (meta)data.

TO BE **R**E-USABLE:

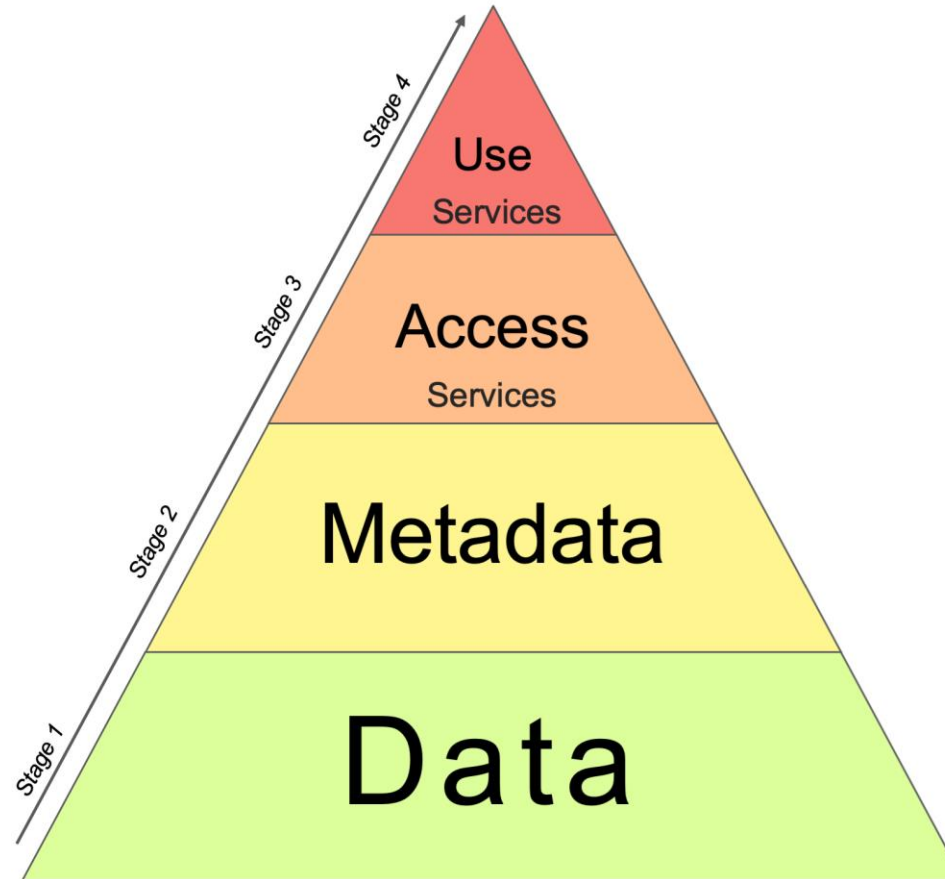
- R1. meta(data) have a plurality of accurate and relevant attributes.
 - R1.1. (meta)data are released with a clear and accessible data usage license.
 - R1.2. (meta)data are associated with their provenance.
 - R1.3. (meta)data meet domain-relevant community standards.

+ Computational Reproducibility



L'essentiel – Fourniture de données et de services FAIR

Findable 
Accessible 
Interoperable 
Reusable 
+
Reproducible



FAIR PRINCIPLES

A1. (meta)data are retrievable by their identifier using a standardized communications protocol.

A1.1. the protocol is open, free, and universally implementable.

A1.2. the protocol allows for an authentication and authorization procedure, where necessary.

F4. (meta)data are registered or indexed in a searchable resource.

F1. Metadata are assigned a globally unique and eternally persistent identifier.

F2. data are described with rich metadata.

F3. metadata specify the data identifier.

F4. metadata are registered or indexed in a searchable resource.

A2. metadata are accessible, even when the data are no longer available.

I1. metadata use a formal, accessible, shared, and broadly applicable.

I2. metadata use vocabularies that follow FAIR principles.

I3. metadata include qualified references to other metadata.

R1 (R1.1 - R1.2 - R1.3) Metadata are richly described with a plurality of accurate attributes.

F1. Data are assigned a globally unique and eternally persistent identifier.

F4. Data are registered or indexed in a searchable resource.

I1. Data use a formal, accessible, shared, and broadly applicable language for knowledge representation.

I2. Data use vocabularies that follow FAIR principles.

I3. Data include qualified references to other (meta)data.

R1.1. Data are released with a clear and accessible data usage license.

R1.3. Data meet domain-relevant community standards.

<https://zenodo.org/record/3299353#.XSXmcPwzbUJ>



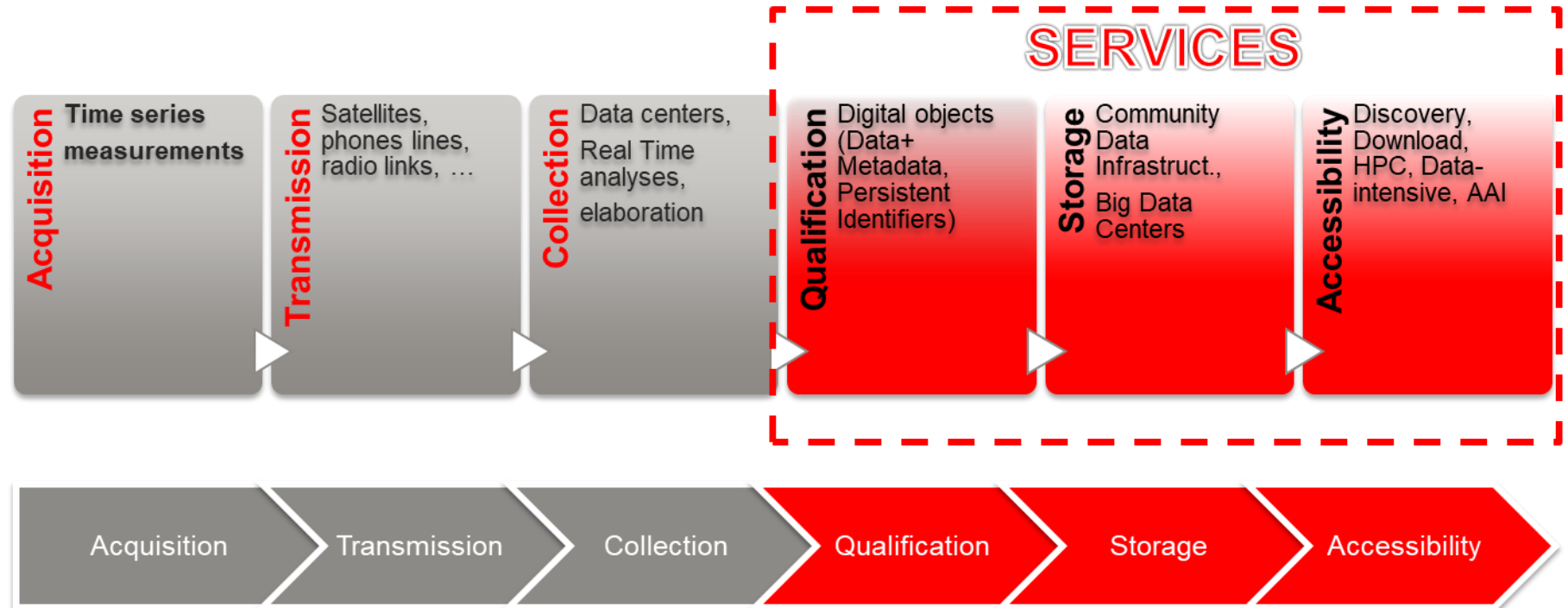
ENVRI
FAIR



L'essentiel – Fourniture de données et de services FAIR

Findable 
Accessible 
Interoperable 
Reusable 

+
Reproducible



Data acquisition, validation & standardization

Data collection, preservation, publication (PID, DOI)

Accessibility, integration, computation

by courtesy of M. Cocco



L'essentiel – Fourniture de données et de services FAIR

Les données produites par les RIs sont **standardisées** et soumises à un **contrôle qualité**, ce qui implique un **engagement durable des communautés scientifiques** et des efforts particuliers pour **harmoniser les données et les métadonnées**.

Hétérogénéité des données: données **brutes, produits, logiciels** et **services**.

Valeur des données = Utilité + Garantie

Utilité: les données sont utiles, harmonisées, normalisées

Garantie: les données sont accessibles et traçables



Principes de l'implémentation

Mise en œuvre par les infrastructures de recherche elles-mêmes

L'utilisation de services externes est recommandée en cas de besoin mais toutes les solutions doivent être transférables

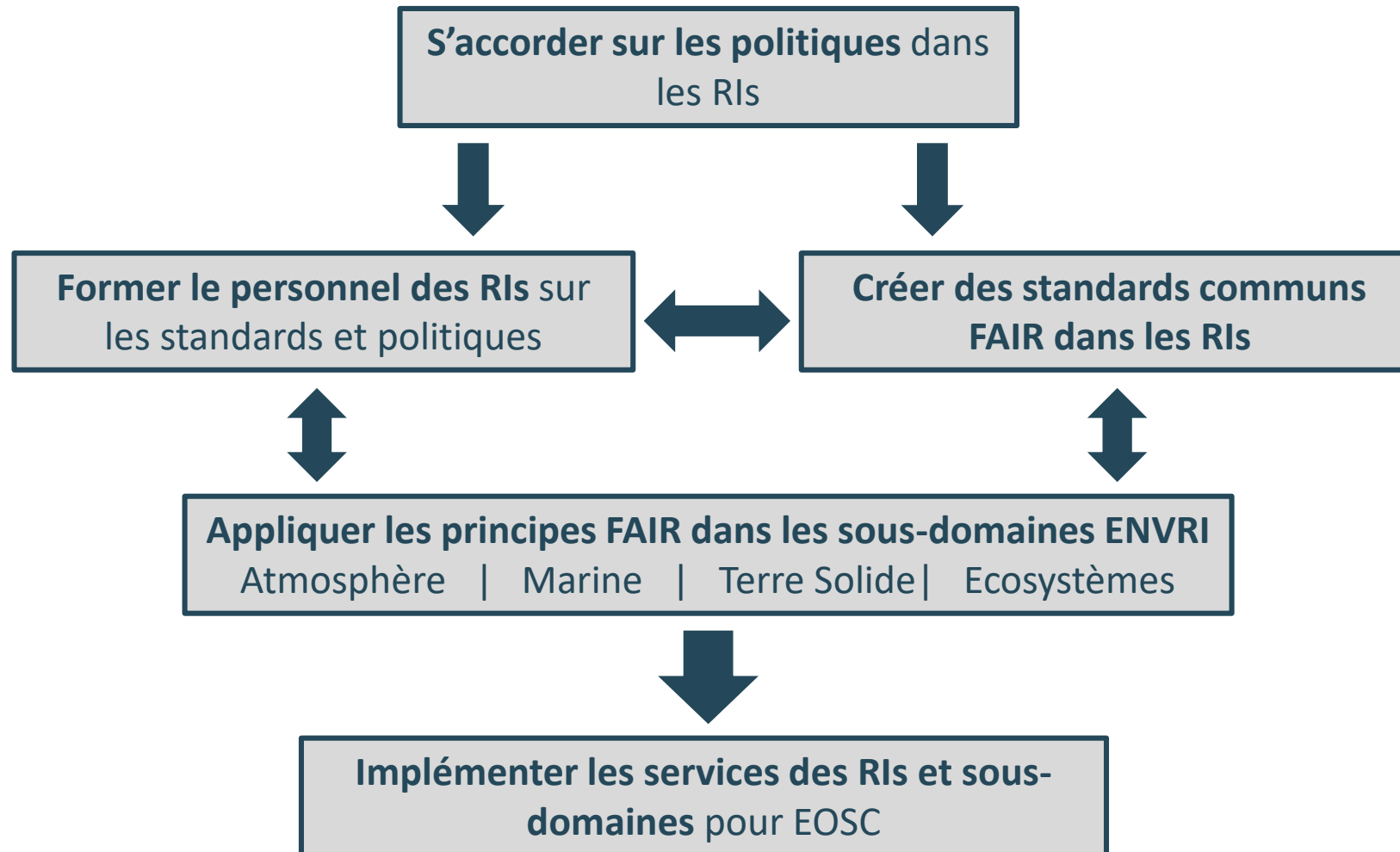
Les standards et l'implémentation sont hiérarchisés

- Au niveau cluster
- Au niveau des sous-domaines
- Au niveau des RIs

Toutes les **politiques et normes** sont soigneusement examinées avec les parties prenantes externes

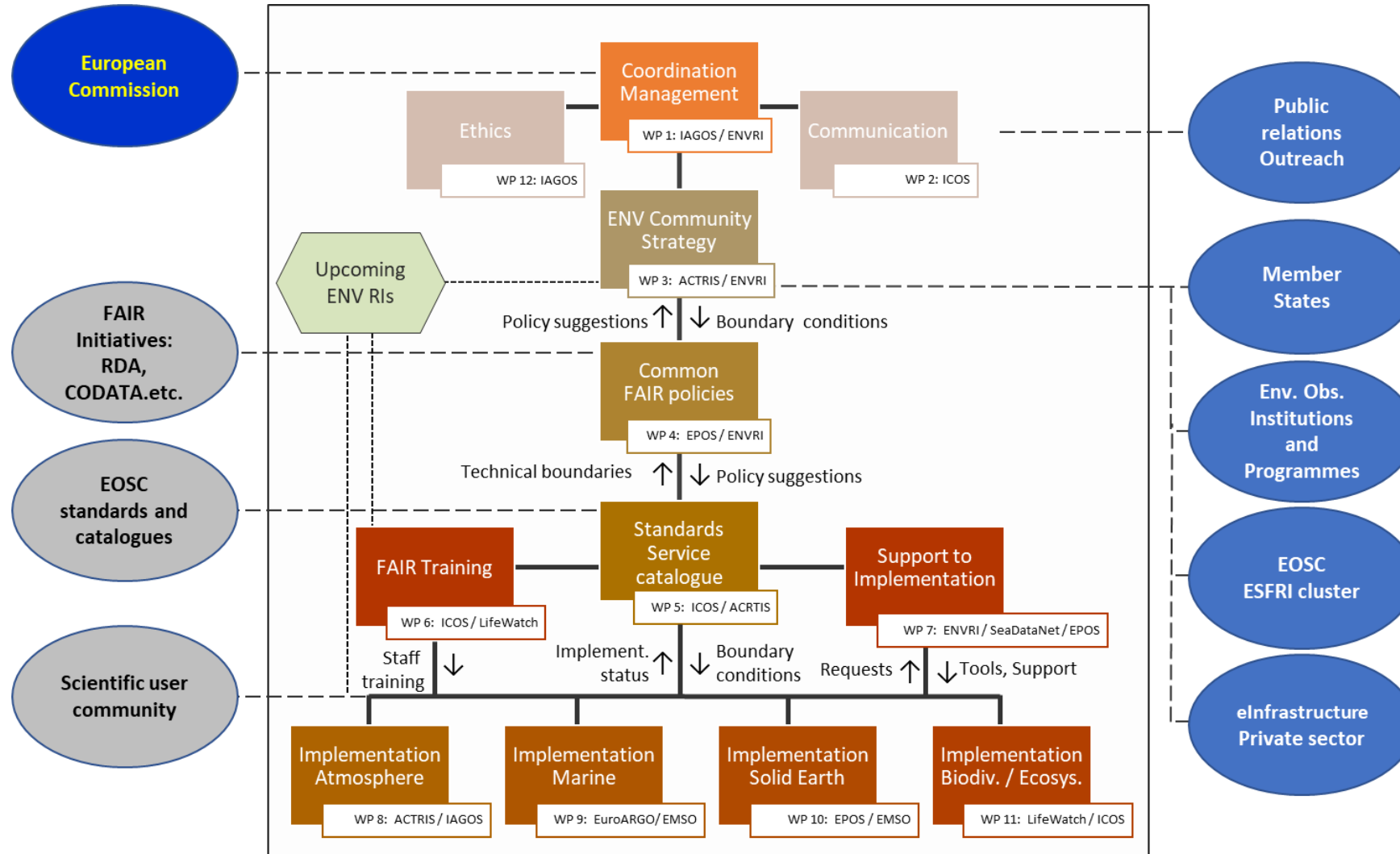


Workflow





Workflow





Exemple du WP8 : Atmosphere subdomain implementation

- Tâche 8.1: Coordination
- Tâche 8.2: Analyser le degré de “FAIRness” des RIs
- Tâche 8.3: Définir le plan d’implémentation au niveau du sous-domaine et des RIs
- Tâche 8.4: Implémentation
- Tâche 8.5: Démonstrateurs et expositions des services dans EOSC



1^{ère} étape : Evaluation du niveau de FAIRness

- Actions prioritaires pour début 2019
 - Créer une vue d'ensemble de la maturité FAIR des RIs
 - Elaborer un plan d'implémentation pour chaque RI
- Création de deux questionnaires
 - Questionnaire #1 : informations qualitatives axées sur la mise en œuvre et les solutions en place dans les RIs
 - Questionnaire #2 : basé sur FAIR metrics / GO FAIR, réponses sous forme d'URI (<https://github.com/FAIRMetrics/Metrics/blob/master/MaturityIndicators/Gen1/ALL.pdf> , Wilkinson)
- Plan d'implémentation de chaque RI → septembre



Questionnaire #1

- Repository
- Data
- Metadata
- Access mechanisms
- DMP
- Data processing
- Semantics
- FAIRness

Access mechanisms			
33	How is authentication done?	<i>What are the methods of authentication which are supported by your RI? Examples are Shibboleth paired with eduGain, OAuth, but many other methods are in use.</i>	Homemade system with local user database (mongodb) SSO implementation planned in 2019 in the frame of AERIS (ORCID, shibboleth, etc.)
34	Do you maintain an own user database?	<i>In many cases local user databases are used by repositories to store identities and to pair them with authorization information.</i>	yes
35	Do you use ORCID in your AAI?	<i>ORCIDs are widely used for publication purposes, do you make use of ORCID for AAI purposes?</i>	In progress and should be done in 2019
36	What is the major access technology	N/a	http get
37	How is authorization done?	N/a	Stored in local user database: - login/password for the web portal - tokens for the rest services (in progress)
38	Which specific licenses do you use for your data?	<i>Do you use for example Creative common licenses or similar?</i>	IAGOS license: http://www.iagos-data.fr/#CMSConsultPlace:DATA_POLICY change for a CC will be discussed
39	Are metadata openly available?	<i>Are your metadata openly accessible via some access mechanism?</i>	yes via CSW through AERIS metadata repository (geonetwork) also accessible throught the AERIS metadata catalogue (https://en.aeris-data.fr/catalogue)
Data Management Plans			
40	Do you use or provide specific DMP tools? If so, which DMP tool are you using or advocating in your community?		No DMP yet
41	Do you apply special data publishing steps?	<i>Often specific data curation steps are taken before publishing data. Provide specific metadata as required for example by DataCite and create DOIs.</i>	yes, automatic metadata conversion from IAGOS/AERIS pivot format to datacite metadata schema
Data processing			
42	Do you apply special data [processing] steps?	<i>Duplicate question?</i>	yes. Manual and automatic data qualification (from raw to final data). Automatic Level 2 data merging and formatting before curation.



Questionnaire #2

- Ressource ID
- Metadata
 - Identifier persistence
 - Metadata for discoverability
 - Identifier in Metadata
 - Indexed in Searchable Resource
- Access
 - Access Protocol
 - Access authorization
 - Metadata Longevity
- Ontologies, semantics
 - Use [of] a Knowledge Representation Language
 - Use [of] FAIR vocabularies
 - Use [of] Qualified References
- Re-usability
 - Accessible usage license
 - Detailed Provenance
 - Meets Community Standards

RESOURCE ID		The uniqueness of an identifier is a necessary condition to unambiguously refer that resource, and that resource alone. Examples of identifier schemes include, but are not limited to URN, IRI, DOI, Handle, trusty URI, LSID, etc. For an in-depth understanding of the issues around identifiers, please see	
4	Please provide the IRI of the resource to be evaluated *	What must be provided? Unique URL to a dataset that preferably uses a registered identifier scheme.	http://doi.org/10.25326/06
Identifier persistence		When using an identifier scheme this will have widespread implications for resource lookup, linking, and data sharing. Providers of digital resources must ensure that they have a policy to manage changes in their identifier scheme, with a specific emphasis on maintaining/redirecting previously	
5	F1A: Please provide the IRI for a registered identifier schema for your resource's IRI (e.g. DOI, HTTP): *	What must be provided? URL to a registered identifier scheme.	https://www.doi.org/hb.html
6	F1B: Please provide the IRI to the document describing the persistence policy for the identifier of this(meta)data (this may be a document from your service provider, e.g. Zenodo, UniProt, etc.):	In simple words: What should be provided? A URL that resolves to a document containing the relevant policy.	https://www.doi.org/doi_handbook/6_Policies.html
Metadata for discoverability		This metric is intended to test the format of the metadata: Machine readability of metadata makes it possible to optimize discovery. For instance, Web search engines suggest the use of particular structured metadata elements to optimize search. Thus, the machine-readability aspect can help people and	
7	F2A: Please provide the IRI to a document that contains machine-readable metadata for the digital resource	What must be provided? A URL to a document that contains machine-readable metadata for the digital resource.	http://catalogue2.sedoo.fr/geonetwork/srv/api/records/575882c0-64ce-4648-bb19-00030d5d63af/formatters/xml
8	F2B: Please provide the IRI for the file format of this metadata:	The file format must be specified preferably in a URL to a vocabulary.	http://www.isotc211.org/2005/gmd http://www.isotc211.org/2005/gmd/gmd.xsd http://www.isotc211.org/2005/gmx http://www.isotc211.org/2005/gmx/gmx.xsd http://www.isotc211.org/2005/srv http://schemas.opengis.net/iso/19139/20060504/srv/srv.xsd
Identifier in Metadata		The discovery of a digital object should be possible from its metadata. For this to happen, the metadata must explicitly contain the identifier for the digital resource it describes, and this should be present in the form of a qualified reference, indicating some manner of "about" relationship.	
9	F3: Please provide the IRI of the metadata	What must be provided? The GUID (Global identifier - URL) of the metadata of the digital resource (dataset) it describes.	http://catalogue2.sedoo.fr/geonetwork/srv/api/records/575882c0-64ce-4648-bb19-00030d5d63af/formatters/xml
10	F3: Please provide the IRI of the data described by the metadata above	(note that this one, or the one above MUST be the same as the one above)	http://doi.org/10.25326/06



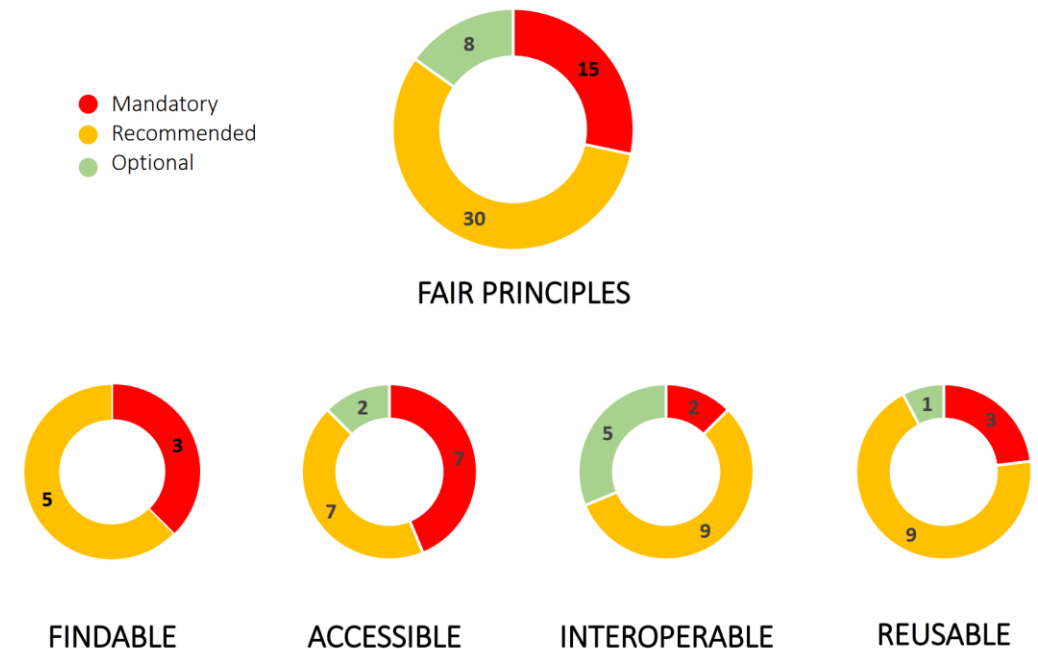
Liens avec RDA

- FAIR Data Maturity Model WG
 - Définition d'indicateurs communs correspondant aux principes FAIR
- Participation active de la communauté ENVRI aux groupes de travail RDA
 - Provenance
 - Citation
 - PID
 - Etc.

<https://www.rd-alliance.org/groups/fair-data-maturity-model-wg>

Development | Weighting Stats

Distribution of the weight of the indicators





En phase avec Data Terra



- Authentification
- Catalogue
- DOI
- Licences
- Gestion / Publication de vocabulaire
- Certification / DMP

- Authentification
- PID, identifiants internes en plus des DOI
- Licences (données et métadonnées)
- Vocabulaires
- Provenance (traçabilité des données)
- Recherche sémantique
- Evaluation de la « FAIRness »



Perspectives France

- Coordination des activités françaises au niveau Data Terra
- Retours aux ateliers Inter-pôles et Groupe Technique Data Terra
- Une fois les questionnaires rodés ils seront mis à disposition de la communauté



ENVRI
FAIR

THANKS



envri.eu/envri-fair



[@ENVRIcomm](https://twitter.com/ENVRIcomm)



[ENVRI community](#)



facebook.com/ENVRIcomm