

FAIRISATION des Entrepôts

Introduction

Atelier JNSO 2019
19 Novembre 2019

Pascal Aventurier, Jean-Christophe Desconnets

IRD – Institut de Recherche pour le Développement
(France)

Pascal.aventurier@ird.fr Jean-christophe.desconnets@ird.fr



Programme de l'atelier

15h - 15h20 : Introduction

Les principes FAIR : définitions et enjeux. par Pascal Aventurier et Jean-Christophe Desconnets - IRD

15h20 - 16h : Comment évaluer le degré FAIR des entrepôts ? Les initiatives existantes

- Dataverse INRA et Fairisation par Dimitri Szabo - INRA
- L'expérience d'entrepôts Nesstar au CDSP par Alexandre Mairot et/ou Alina Danciu Science Po

16- 16h20 : Interopérabilité et principes FAIR

Enjeu des vocabulaires et de l'interopérabilité par Jessie Galasso Carbonel - IRD

16h20 - 17h Discussion

Animation de l'atelier : Pascal AVENTURIER (IRD), Jean-Christophe DESCONNETS (IRD) Jessie Galasso-Carbonel, Dimitri SZABO (Inra); Alexandre MAIROT (Sciences po)

FAIR Enjeux

- Augmenter la découverte et réutilisation (citabilité) des données pour les humains et les machines
- Décrire les jeux de données avec la granularité suffisante et avec des pratiques normalisées pour les données et les métadonnées
- Décrire les conditions de réutilisation et la provenance des données
- Exploiter les données dans différents contextes
- Définir des critères de qualité pour la gestion des données

Données accessibles et dans la mesure du possible « ouvertes »

FAIR enjeux (2)

Données FAIR dans un écosystème FAIR (DMP, PID, spécifications sur les métadonnées)

Dans des modèles interoperables

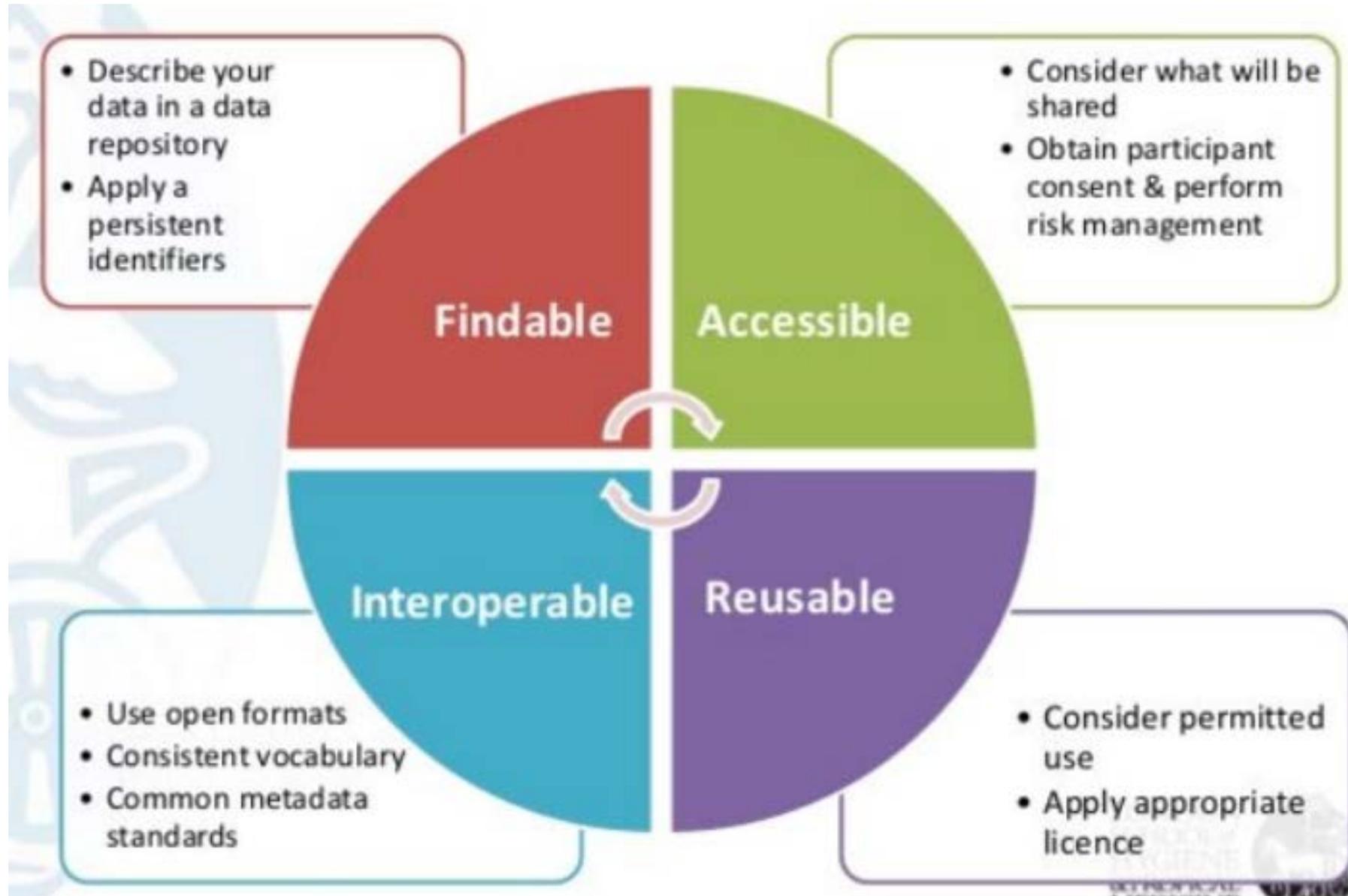
FAIR pour les humains et les machines

Augmentation des compétences data science et data stewardship

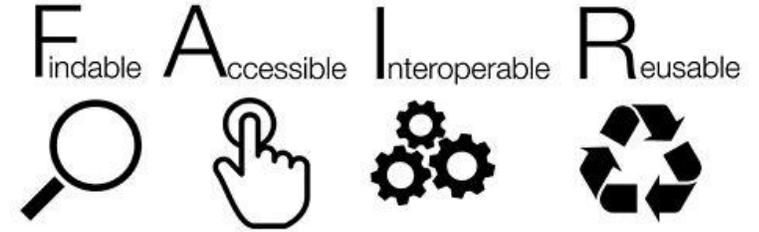
Financement pour FAIR apporte un fort retour sur investissement
Besoin de métriques pour évaluer le niveau FAIR des données

FAIR et données ouvertes sont complémentaires

Fair Principles



FAIR Data



- **To be Findable:**

- 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 ACCESSIBLE:**

- 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 INTEROPERABLE:**

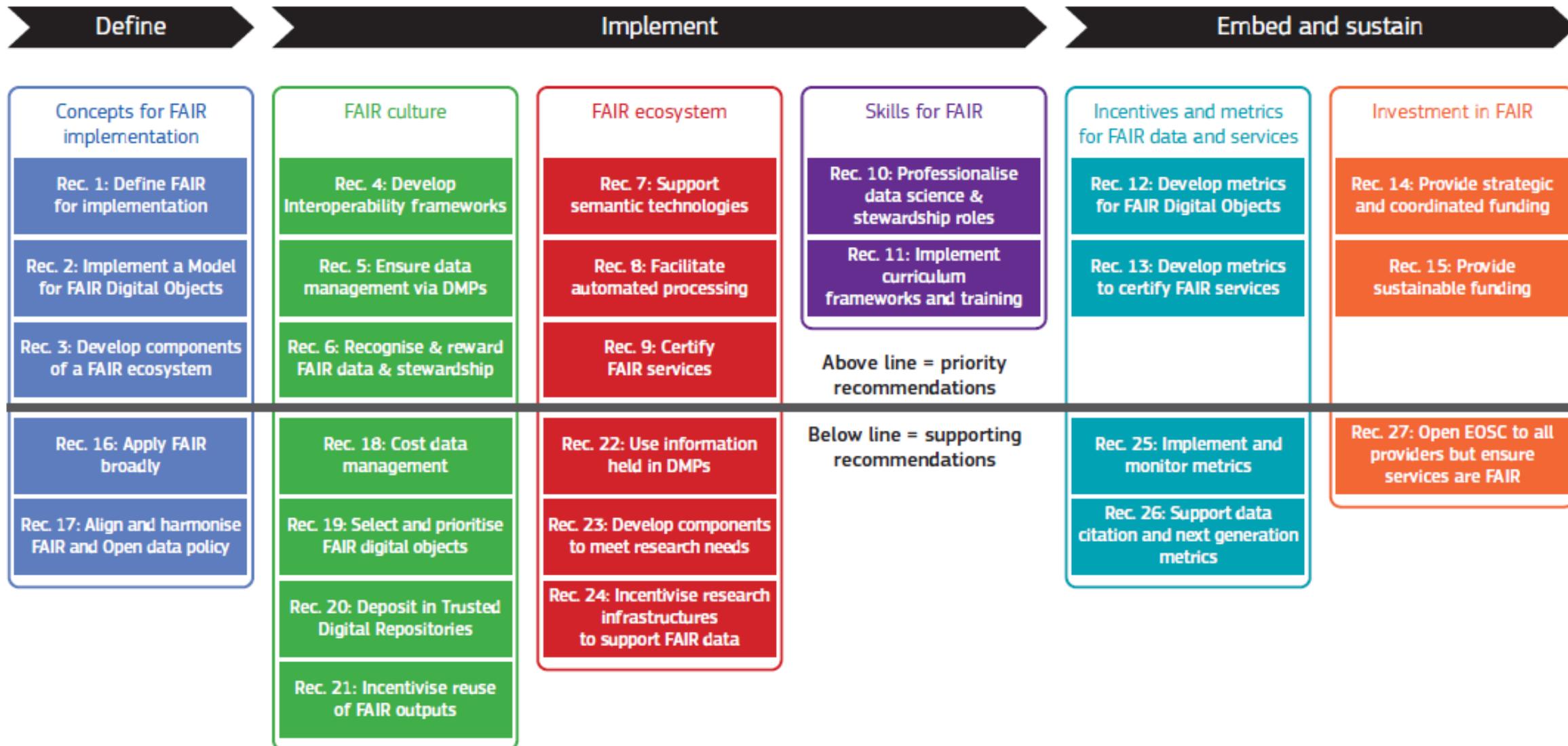
- I1. (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.
- I2. (meta)data use vocabularies that follow FAIR principles.
- I3. (meta)data include qualified references to other (meta)data.

- **TO BE RE-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.

Rapport Turning FAIR into Reality 2018

27 recommendations <https://op.europa.eu:443/en/publication-detail/-/publication/7769a148-f1f6-11e8-9982-01aa75ed71a1/language-en/format-PDF>



FAIR et évaluation des entrepôts

Interopérabilité et réusabilité

S'assurer que les données sont FAIR à travers les disciplines et les zones géographiques.

S'aligner sur les infrastructures internationales :

EOSC, Australian Research Data, NIH Data Commons, African Open Science Platform

REPOSITORIES SUPPORTING USERS TO ACHIEVE FAIRNESS

Findable:

- ✓ F1. (meta)data are assigned a globally unique and persistent identifier;
- ✓ F2. data are described with rich metadata;
- ✓ F3. metadata clearly and explicitly include the identifier of the data it describes;
- ✓ F4. (meta)data are registered or indexed in a searchable resource;

Interoperable:

- ✓ 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;

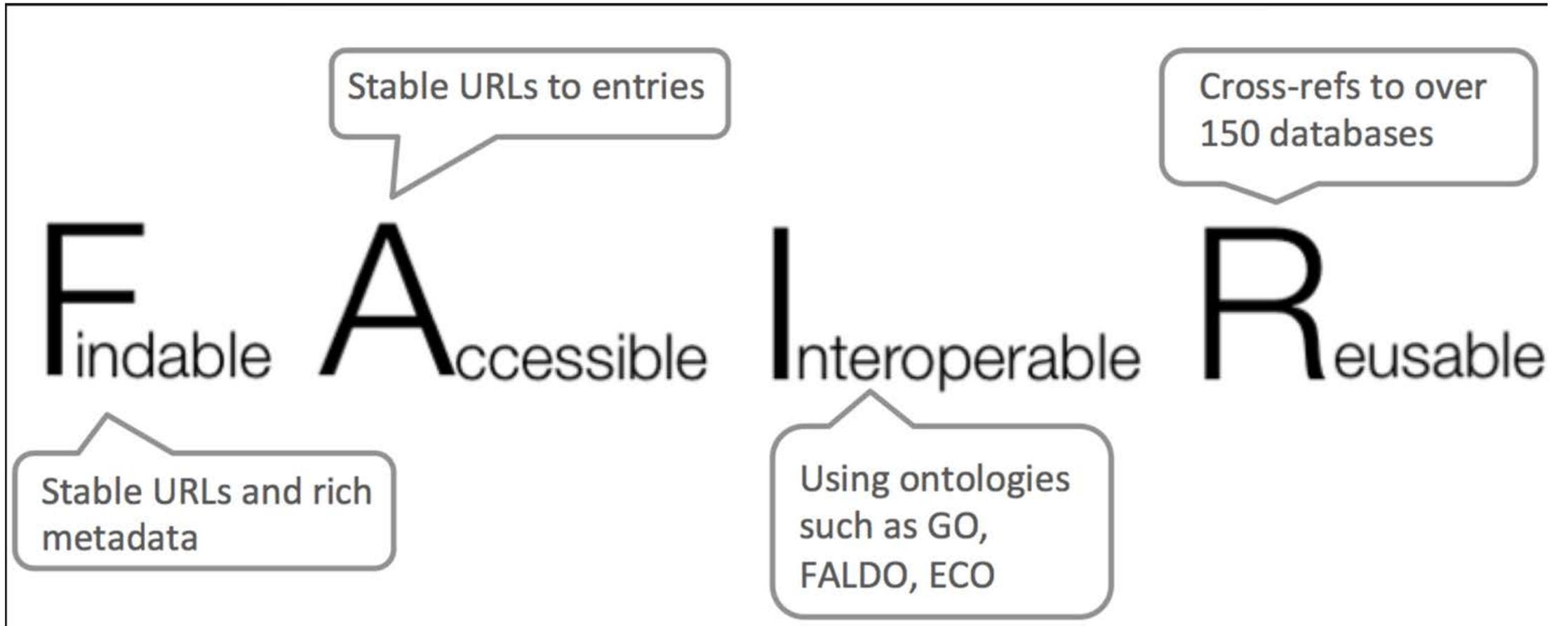
Accessible:

- ✓ 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;

Reusable:

- ✓ R1. (meta)data are richly described with 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 detailed provenance;
 - ✓ R1.3. (meta)data meet domain-relevant community standards;

FAIR UNIPROT



UniProt est une base de données de séquences de protéines

FAIR Data Maturity Model WG

[Home](#) » [Working And Interest Groups](#) » [Working Group](#) » [FAIR Data Maturity Model WG](#)

WG

Group details

Status: Recognised & Endorsed

Chair (s): Edit Herczog, Keith Russell

Secretariat Liaison: Lynn Yarmey

TAB Liaison: Jane Wyngaard

 WGs Producing deliverables (~6-12 months after RDA endorsement)

FAIR Data Maturity Model: core criteria to assess the implementation level of the FAIR data principles

Case Statement

FAIR Data Maturity Model WG Case Statement
23 September 2018

Comments 4

Outputs & Recommendations

Results of an Analysis of Existing FAIR Assessment Tools

Comments: 2



FAIR self-assessment tool

Findable



Does the dataset have any identifiers assigned?

Globally Unique, citable and persistent (e.g. DOI, PURL, ARK c

Is the dataset identifier included in all metadata records/files describing the data?

No

How is the data described with metadata?

Comprehensively (see suggestion) using a recognised formal r

What type of repository or registry is the metadata record in?

Data is in one place but discoverable through several registries



Accessible



How accessible is the data?

Publicly accessible

Is the data available online without requiring specialised protocols or tools once access has been approved?

File download from online location

Will the metadata record be available even if the data is no longer available?

Yes



Interoperable



What (file) format(s) is the data available in?

In a structured, open standard, non-machine-readable format

What best describes the types of vocabularies/ontologies/tagging schemas used to define the data elements?

No standards have been applied in the description of data elem

How is the metadata linked to other data and metadata (to enhance context and clearly indicate relationships)?

There are no links to other metadata



Reusable



Which of the following best describes the license/usage rights attached to the data?

Standard machine-readable license (e.g. Creative Commons)

How much provenance information has been captured to facilitate data reuse?

Partially recorded



Total across F.A.I.R



Toute une constellation !

« The layered cake of FAIR coordination: how many is too many? : Scientific Data ». s. d. Consulté le 29 octobre 2019.

<http://blogs.nature.com/scientificdata/2019/10/22/the-layered-cake/>.



NIH Data commons



EOSC pilot

The European Open Science
Cloud for Research Pilot Project



EUDAT & THE FAIR DATA PRINCIPLES

"Everybody wants to play FAIR,
but how do we put the principles into practice?"

FAIR des principes pour partager utile

Faciliter la découverte, l'accès, l'interopérabilité et la réutilisation des données.

--> Tout en étant réaliste (Un fichier CSV bien documenté / RDF plus difficile à produire)

Il faut en tenir compte tout au long du cycle de vie de la donnée

—>

Rédiger un PGD qui décrit la manière dont les données sont produites / obtenues, décrites, représentées, stockées, partagées

Objectifs de l'atelier

Mieux comprendre les principes FAIR dans les entrepôts et les actions à mener.

Illustration avec deux projets : INRA et NESTAR PROGEDO

Comprendre les enjeux autour de l'interopérabilité