

Les principes FAIR

Damien Boulanger, AERIS-SEDOO

Atelier ODATIS, 15 décembre 2020



Introduction

- Publication en 2016 de l'article “ The FAIR Guiding Principles for scientific data management and stewardship” par Wilkinson et al.

<https://doi.org/10.1038/sdata.2016.18>

- Principes FAIR rapidement acceptés par la communauté internationale
- Des pratiques déjà existantes → une formulation
- Les principes FAIR c'est rendre les données :

Findable 

- Facile à trouver

Accessible 

- Accessible

Interoperable 

- Interopérable

Reusable 

- Réutilisable

Findable

- Les données sont faciles à trouver par des humains ou des machines
- F1. Les données et les métadonnées ont un identificateur unique et persistant
e.g. DOI, handles, etc.
- F2. Les données sont décrites par des métadonnées riches (voir R1)
Formats de métadonnées standards: ISO 19115, Data cite, NetCDF, etc.
- F3. Les données et métadonnées sont enregistrées ou indexées dans un dispositif permettant de les rechercher
e.g. Portail de données projet/Pôles, GEOSS, Google Dataset search, etc.
- F4. Les métadonnées contiennent l'identificateur unique et persistant des données

Accessible

- Une fois les données trouvées, l'utilisateur doit savoir comment y accéder et s'il a besoin d'être authentifié et/ou autorisé
- A1. Les données et métadonnées sont accessibles par leur identifiant via un protocole standard de communication
 - A1.1 Ce protocole est ouvert, libre et peut être implémenté de manière universelle
e.g. REST, CSW, OAI-PMH, OPeNDAP, etc.
 - A1.2 Ce protocole peut gérer des procédures d'authentification et d'autorisation si nécessaire
e.g. ORCID, EduGAIN, Oauth, etc.
- A2 Les métadonnées restent accessibles même si les données ne sont plus disponibles
e.g. DMP, etc.

Interoperable

- Les données sont interoperables: intégration avec d'autres données, interoperabilité avec des applications ou des workflows

- I1. Les données et métadonnées utilisent un langage formel, accessible, partagé et largement applicable pour la représentation des connaissances

e.g. RDF, JSON-LD

- I2. Les données et métadonnées utilisent des vocabulaires qui eux-mêmes suivent les principes FAIR

e.g. GCMD, CF standard names, SeaDataNet, etc.

- I3. Les données et métadonnées contiennent des liens vers d'autres (méta)données

e.g. liens vers d'autres versions, articles, etc.

Reusable

- Les données doivent être réutilisables pour de futures recherches ou autres activités
- R1. Les (méta)données doivent être richement décrites, par une pluralité d'attributs précis et pertinents
 - R1.1. Elles sont fournies avec une licence d'utilisation claire et accessible
e.g. Creative Commons, etc.
 - R1.2. Elles sont associées à leur provenance
e.g. standard, W3C PROV, etc.
 - R1.3. Elles suivent les standards des communautés concernées

+ Reproductibilité (machine actionable)

Projets et communautés liés aux principes FAIR

- Research Data Alliance (RDA)

- Groupes d'intérêt et Groupes de travaux
- Beaucoup de recommandations FAIR: PIDs, citation, provenance, certification, etc.
- <https://www.rd-alliance.org/groups/rda-france>



- GO FAIR

- Initiative de plusieurs pays européens
- <https://www.go-fair.org>



- Comité de données pour la science et la technologie (CODATA)

- <https://codata.org/>



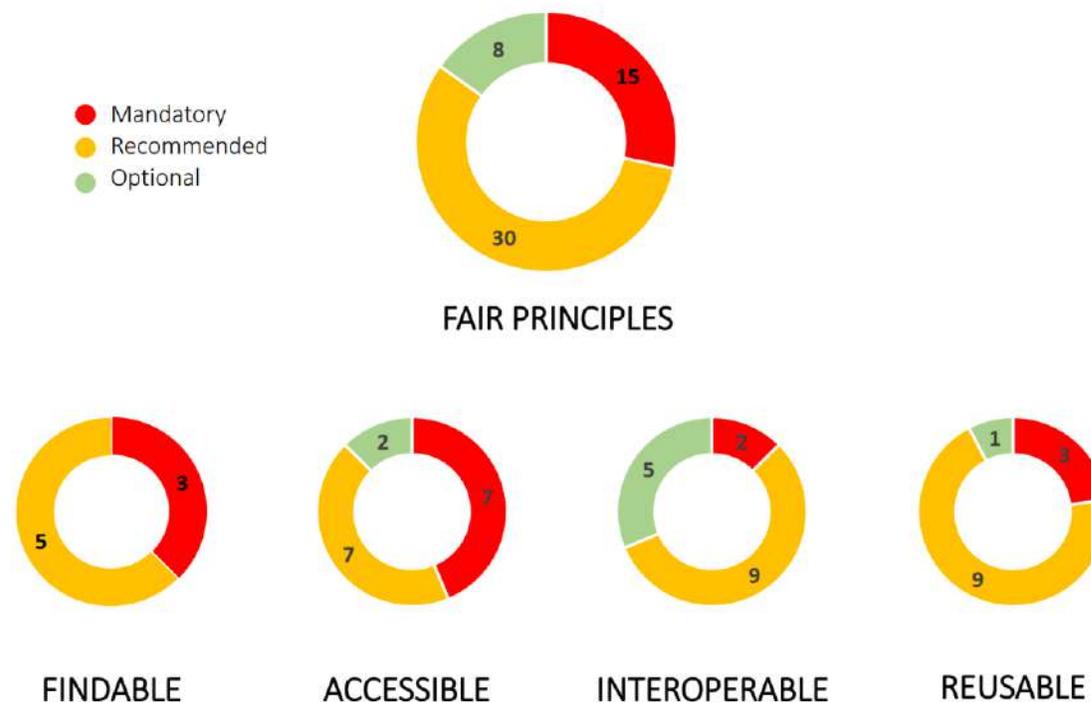
FAIR Data Maturity Model

- Définition d'indicateurs communs correspondant aux principes FAIR
- Utilisable pour l'évaluation du degré de « FAIRness »
- <https://www.rd-alliance.org/groups/fair-data-maturity-model-wg>



Development | Weighting Stats

Distribution of the weight of the indicators



FAIR Data Maturity Model

ID	PRINCIPLE	INDICATOR_ID	INDICATORS	PRIORITY	METRIC	VIZ	5	
1	F	F1	RDA-F1-01M	Metadata is identified by a persistent identifier	Essential	4 – fully implemented	4	1
2		F1	RDA-F1-01D	Data is identified by a persistent identifier	Essential	4 – fully implemented	4	1
3		F1	RDA-F1-02M	Metadata is identified by a globally unique identifier	Essential	0 – not applicable	0	0
4		F1	RDA-F1-02D	Data is identified by a globally unique identifier	Essential	4 – fully implemented	4	1
5		F2	RDA-F2-01M	Rich metadata is provided to allow discovery	Essential	4 – fully implemented	4	1
6		F3	RDA-F3-01M	Metadata includes the identifier for the data	Essential	4 – fully implemented	4	1
7		F4	RDA-F4-01M	Metadata is offered in such a way that it can be harvested and indexed	Essential	4 – fully implemented	4	1
8	A	A1	RDA-A1-01M	Metadata contains information to enable the user to get access to the data	Important	4 – fully implemented	4	1
9		A1	RDA-A1-02M	Metadata can be accessed manually (i.e. with human intervention)	Essential	4 – fully implemented	4	1
10		A1	RDA-A1-02D	Data can be accessed manually (i.e. with human intervention)	Essential	3 – in implementation phase	3	0
11		A1	RDA-A1-03M	Metadata identifier resolves to a metadata record	Essential	4 – fully implemented	4	1
12		A1	RDA-A1-03D	Data identifier resolves to a digital object	Essential	4 – fully implemented	4	1
13		A1	RDA-A1-04M	Metadata is accessed through standardised protocol	Essential	4 – fully implemented	4	1
14		A1	RDA-A1-04D	Data is accessible through standardised protocol	Essential	4 – fully implemented	4	1
15		A1	RDA-A1-05D	Data can be accessed automatically (i.e. by a computer program)	Important	4 – fully implemented	4	1
16		A1.1	RDA-A1.1-01M	Metadata is accessible through a free access protocol	Essential	4 – fully implemented	4	1
17		A1.1	RDA-A1.1-01D	Data is accessible through a free access protocol	Important	4 – fully implemented	4	1
18		A1.2	RDA-A1.2-01D	Data is accessible through an access protocol that supports authentication and authorisation	Useful	4 – fully implemented	4	1
19		A2	RDA-A2-01M	Metadata is guaranteed to remain available after data is no longer available	Essential	4 – fully implemented	4	1
20	I	I1	RDA-I1-01M	Metadata uses knowledge representation expressed in standardised format	Important	4 – fully implemented	4	1
21		I1	RDA-I1-01D	Data uses knowledge representation expressed in standardised format	Important	4 – fully implemented	4	1
22		I1	RDA-I1-02M	Metadata uses machine-understandable knowledge representation	Important	4 – fully implemented	4	1
23		I1	RDA-I1-02D	Data uses machine-understandable knowledge representation	Important	4 – fully implemented	4	1
24		I2	RDA-I2-01M	Metadata uses FAIR-compliant vocabularies	Important	4 – fully implemented	4	1
25		I2	RDA-I2-01D	Data uses FAIR-compliant vocabularies	Useful	4 – fully implemented	4	1

FAIR Implementation Profiles (FIP)

- Machine actionnable
- Plusieurs communautés
- Convergence matrix

FAIR Principles

Box 2 | The FAIR Guiding Principles

To be Findable:
F1. (metadata) are assigned a globally unique and persistent identifier
F2. data are described with rich metadata (defined by R1 below)
F3. metadata clearly and explicitly include the identifier of the data it describes
F4. (metadata) are registered or indexed in a searchable resource

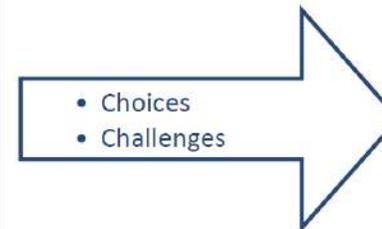
To be Accessible:
A1. (metadata) are retrievable by their identifier using a standardized communications protocol
A1.1 the protocol is open, free, and universally implementable
A1.2 the protocol allows for an authentication and authorization procedure, where necessary
A2. metadata are accessible, even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

To be Interoperable:
I1. (metadata) use a formal, accessible, shared, and broadly applicable language for knowledge representation.
I2. (metadata) use vocabularies that follow FAIR principles
I3. (metadata) include qualified references to other (metadata)

To be Reusable:
R1. (metadata) are richly described with a plurality of accurate and relevant attributes
R1.1. (metadata) are released with a clear and accessible data usage license
R1.2. (metadata) are associated with detailed provenance
R1.3. (metadata) meet domain-relevant community standards



FAIR Implementations

- 
- Choices
 - Challenges



Evaluate FAIRness



<https://www.go-fair.org/today/fair-matrix/>

FAIRsFAIR

- Initiative CoreTrustSeal+FAIR
 - Lier les principes FAIR aux exigences CTS
 - Proposer une nouvelle version des exigences CTS
- CoreTrustSeal plus FAIR Overview:
<https://zenodo.org/record/4003630#.X9DZ17PjJPZ>
- Supporte dix entrepôts de données pour la certification qui fournissent leur retours sur la démarche



FAIRsFAIR
Fostering Fair Data Practices in Europe

CoreTrustSeal+FAIR

