



# Implémenter les principes FAIR, Restitution, sensibilisation

Annie Hofstetter

► **To cite this version:**

Annie Hofstetter. Implémenter les principes FAIR, Restitution, sensibilisation. CEE-M Environmental seminar, Apr 2020, Montpellier, France. hal-02871121

**HAL Id: hal-02871121**

**<https://hal.inrae.fr/hal-02871121>**

Submitted on 17 Jun 2020

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Sem/web inar

F<sub>indable</sub> A<sub>ccessible</sub> I<sub>nteroperable</sub> R<sub>eusable</sub>



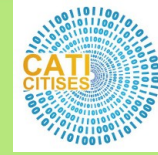
© SangyaPundir - under CC By-SA license

Implémenter les principes FAIR  
Restitution, sensibilisation  
Annie Hofstetter

**INRAE**



# Atelier implémenter les principes FAIR à INRAE (Paris, 4-6 03 2020)



- Organisé par le CATI PROSODIe, avec le soutien de la DIPSO, dans le cadre de l'appel à projet inter Cati
- Porté par Sophie Aubin, Tovo Rabemanantsoa, Esther Dzalé-Yeumo
- Wilfried Heintz, Michaël Alaux, Annie Hofstetter, Patrice Buche, Eric Latrille, Sylvain Jason, Bruno Perret, Pascal Neveu, Delphine Steinbach, Jean-Marc Rousselle, Sylvie Cocaud, Daniel Jacob, Romain David
- avec la collaboration des Cati CITISES, CODEX, DIISCICO, GEDEOP, GREP, Plantbreed et PEPI GD
- 50 participants



reproductibilité  
traçabilité  
évaluation  
dictionnaire  
métavocabulaire  
Wilkinson  
forge DMP  
ontologie  
protocole  
FAIR  
données  
accessible  
Agrilogue  
entrepôt  
description  
réutilisable  
authentification  
interopérabilité  
pérennité

# Définition - Historique

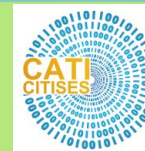


*Tout ce que vous avez toujours voulu savoir sans jamais oser le faire*

- Article de référence Wilkinson et al 2016  
FAIR acronyme et principes : 2014
- Ni un standard ni une norme
- Indépendance de toute technologie
- Différent de l'open data

**Objectif : tendre vers la qualité des données, leur accès, leur description, leur réutilisation**

# Principes



## Box 2 | The FAIR Guiding Principles

### To be Findable:

- F1. (meta)data 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. (meta)data are registered or indexed in a searchable resource

### 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 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



## Findable

- *F1* Les données et les métadonnées sont identifiées par un identifiant global unique et pérenne
- *F2* Les métadonnées décrivant les données sont riches
- *F3* Les données et les métadonnées sont enregistrées et indexées dans un dispositif permettant de les rechercher
- *F4* Les métadonnées spécifient l'identifiant de la donnée

## Accessible

- *A1* Les données et les métadonnées sont accessibles par leur identifiant via un protocole de communication standardisé
  - *A1.1* Le protocole utilisé est ouvert, libre et peut être implémenté de manière universelle
  - *A1.2* Le protocole utilisé permet l'authentification et l'autorisation si besoin
- *A2* Les métadonnées sont accessibles même quand les données ne le sont plus

## Interoperable

- *I1* Les données et les métadonnées utilisent un langage formel, accessible, partagé et largement applicable pour la représentation des connaissances
- *I2* Les données et les métadonnées utilisent des vocabulaires qui respectent les principes FAIR
- *I3* Les données et les métadonnées incluent des liens spécifiés vers d'autres (méta)données

## Reusable

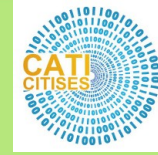
- *R1* Les données et les métadonnées ont des attributs multiples et pertinents
  - *R1.1* Les données et les métadonnées sont mises à disposition selon une licence explicite et accessible
  - *R1.2* Les données et les métadonnées sont associées à leur provenance
  - *R1.3* Les données et les métadonnées correspondent aux standards des communautés indiquées



Le clou de la présentation...

ou comment FAIR ?





- Grille Sharc simplifiée  
SHARC = SHARing Reward and credit  
GIS interdisciplinaire dans le cadre de la Research Data Alliance
- [Poster de présentation](#)
- Grille d'évaluation  
récupérée [ici](#)
- Grille simplifiée :  
[site prosodie](#)  
ou séquence de travail [RDA](#)
- Autre : data maturity model

Templates for FAIRness evaluation criteria - RDA-SHARC ig

[Extensive Template](#) [Minimal Template](#) [List of criteria](#) [Synthetic criteria tree](#) [Glossary](#) [Read me](#)

	A	B	C	D	E	F	G
1	<b>Scientist Name / Project Name, Date</b>		Tick the proper case using '1' (to enable the automatic count)	Never/NA	Mandatory	Sometimes	Always
3	<b>Essential / Recommended / Desirable</b>						
4	<b>1) FINDABLE (8 essential criteria / total of 12 criteria)</b>						
5	<b>Indexed identifier? (F01_E)</b>						
6	Identification						
7	Are each data/dataset identified by an indexed and independant identifier?						
8	<b>Unique, global, persistent identifier? (F02_E)</b>						
9	Identification						
10	Are the data identifiers unique, global and persistent?						
11	<b>Identifier scheme? (F03_E)</b>						
12	Identification						
13	Has any identifying schema been used for data (e.g., DOI)?						
14	<b>Persistent metadata &amp; data link? (F04_E)</b>						
15	Metadata traceability						
16	Are the metadata linked to the dataset through a persistent identifier?						
17	<b>Persistent metadata &amp; authority link? (F05_E)</b>						
18	Metadata traceability						
19	Are the metadata of each dataset linked to a unique authority (responsible for the datasets at a given time)?						
20	<b>Persistent datasets &amp; authority link? (F06_E)</b>						
21	Metadata traceability						
22	Are all datasets linked to an authority over time (legal entity, e.g., institution, association or established body) through a unique and persistent identifier?						
23	<b>Standards / dictionary for data description? (F07_E)</b>						
24	Metadata description and searchability						
25	Choose the question corresponding to your case: If standards exist, has the researcher used valid and updated standards for data description AND are the data standards approved by community- or appropriate authorities? If no standards exist, has the researcher created a well described community-approved data dictionary?						
26	<b>Search keywords? (F08_R)</b>						
27	Metadata description and searchability						
28	Does the researcher use community-approved well-described search keywords from a controlled vocabulary / thesaurus / ontology?						
29	<b>Metadata semantic validator? (F09_R)</b>						
30	Metadata description and searchability						
31	Does the researcher use a semantic validator for metadata (recommended by standardisation authorities or community-approved)?						
32	<b>Data format / type description? (F10_E)</b>						
33	Metadata description and searchability						

Scientist Name /  
Project Name, Date

To enable the automatic counting of the data entered boxes:

"1" corresponding to Yes if the indicator is SATISFIED  
 "0" corresponding to No if the indicator is NOT SATISFIED  
 "NA" if the indicator is considered as NON APPLICABLE  
 "?" if the indicator is NOT UNDERSTANDABLE or if you have any clue to respond

Applied

is the indicator  
reachable ?

Essential / Recommended / Desirable

**1) FINDABLE (8 essential criteria / total of 12 criteria)**

**Indexed identifier? (F01\_E)**

Identification

Are each data/dataset identified by an indexed and independent identifier?

E

**Unique, global, persistent identifier? (F02\_E)**

Identification

Are the data identifiers unique, global and persistent?

E

**Identifier scheme? (F03\_E)**

Identification

Has any identifying schema been used for data (e.g., DOI)?

E

**Persistent metadata & data link? (F04\_E)**

Metadata traceability

Are the metadata linked to the dataset through a persistent identifier?

E

**Persistent metadata & authority link? (F05\_E)**

Metadata traceability

Are the metadata of each dataset linked to a unique authority (responsible for the datasets at a given time)?

E

**Persistent datasets & authority link? (F06\_E)**

Metadata traceability

Are all datasets linked to an authority over time (legal entity, e.g., institution, association or established body) through a unique and persistent identifier?

E

**Standards / dictionary for data description? (F07\_E)**

Metadata description and searchability

Choose the question corresponding to your case:

If standards exist, has the researcher used valid and updated standards for data description AND are the data standards approved by community- or appropriate authorities?

If no standards exist, has the researcher created a well described community-approved data dictionary?

E

**Search keywords? (F08\_R)**

Metadata description and searchability

Does the researcher use community-approved well-described search keywords from a controlled vocabulary / thesaurus / ontology?

R

**Metadata semantic validator? (F09\_R)**

Metadata description and searchability

Does the researcher use a semantic validator for metadata (recommended by standardisation authorities or community-approved)?

R

**Data format / type description? (F10\_E)**

Metadata description and searchability

Are the types and formats of data generated / collected well described as recommended by the community or by standardisation authorities?

E

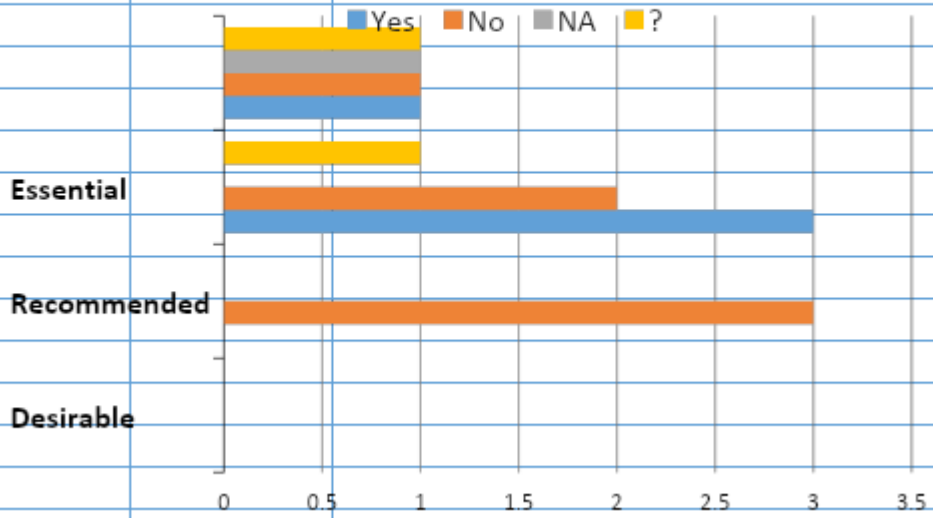
**Versioning system? (F11\_R)**

Metadata description and searchability

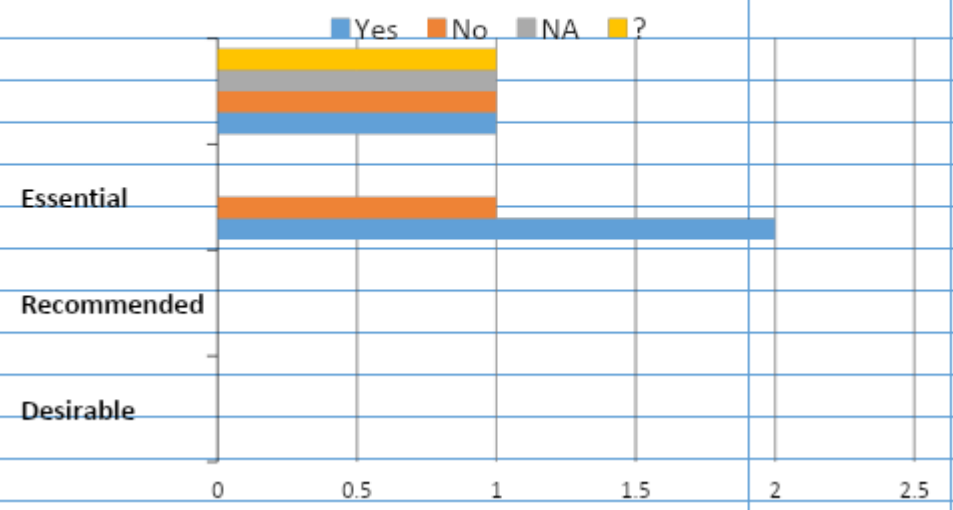
Does the researcher use a community-approved well described versioning system (for a proper tracking of versions)?

R

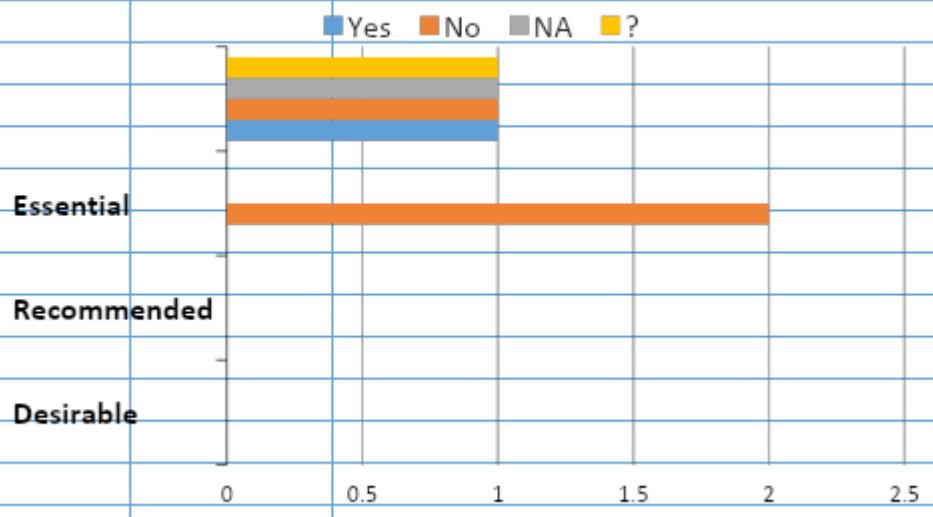
## Findable



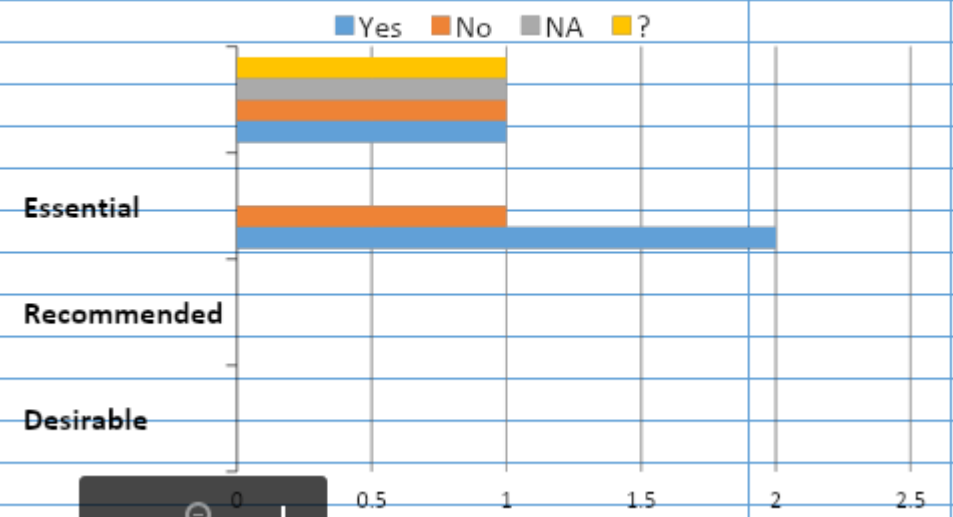
## Accessible



## Interoperable



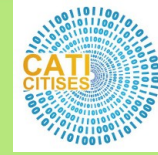
## Reusable





Des focus pour y voir plus clair...

# Focus DATAINRAE



- Portail institutionnel donc LDAP INRAE, basé sur logiciel **Dataverse**
  - Ensemble de services, plus qu'un entrepôt ;
    - gestion
    - Partage → Attribution d'un **DOI** → accès, génération de template pour les datapapers
    - recherche de données (attention données diffusables...)
  - **Objectif :**
    - **Politique open science (open access, open data)**
    - **Exigences des financeurs dont la Commission Européenne**
    - **Mise en conformité avec la loi pour une République numérique**
    - **Exigences des revues**
- <https://data.inrae.fr>
- Formation IST

**Principes  
FAIR !**

# Focus PGD ou DMP

## Plan de Gestion des Données



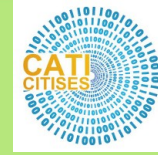
- Accompagner la qualité des données dans la recherche = bonne(s) pratique(s)
- **Objectif :**
  - **Exigences des financeurs**
  - **Transparence scientifique**
  - **Favoriser la réutilisation des données**
  - **Publications**

**Rendre  
les données  
FAIR !**

<https://opidor.fr>

- Document itératif
- Recommandations de Inrae outils DMP OPIDoR
- Actions en cours : DMP Moisa et autre outil... à suivre

# Focus ID, etc. Focus Agrilogue



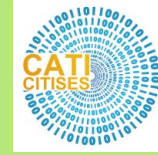
- IDENT l'intérêt d'un identifiant n'est plus à démontrer !
- DOI via Datainrae
- ORCID = open researcher and contributor ID
- ID HAL...
  
- Agrilogue outil mis en place par la Mission Bases de données de ECOSOCIO

**<https://odr.inra.fr/agrilogue/>**

Pas de données !



# Focus éléments de réflexion



- Aspects juridiques de l'utilisation des données !...
- Outils d'aide à la réflexion, se poser les bonnes questions (se poser des questions)
- Grille SHARC à méditer
- Article en cours de révision Journal of economic surveys  
Recherche reproductible pour l'économie empirique et  
l'économétrie  
collectif d'ITA EcoSocio