

# Une méthodologie et un outil d'évaluation du niveau de « FAIRness » pour les ressources sémantiques : le cas d'AgroPortal

IC 2021, Bordeaux



Fo@SIN



Emna Amdouni \* et Clément Jonquet \* \*\*

\* LIRMM, Université de Montpellier, CNRS, France

\*\* MISTEA, Université de Montpellier, INRAE, Institut Agro, France

[emna.amdouni@lirmm.fr](mailto:emna.amdouni@lirmm.fr) et [jonquet@lirmm.fr](mailto:jonquet@lirmm.fr)

ORCID : 0000-0002-2930-5938

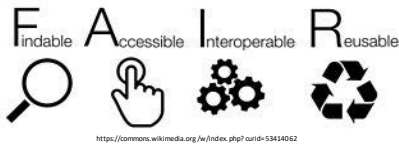
ORCID : 0000-0002-2404-1582



## Contexte

# Définition- Historique

- Les principes FAIR définissent les fondements d'un partage de données **F**aciles à trouver (F), **A**ccessibles (A), **I**nteropérables (I), **R**éutilisables (R) pour l'humain et la machine.
- Enoncés par le groupe FORCE 11



## SCIENTIFIC DATA

**OPEN** Comment: The FAIR Guiding Principles for scientific data management and stewardship

SUBJECT CATEGORIES  
 • Research data  
 • Publication characteristics

Mark D. Wilkinson et al.\*

Received: 30 December 2015  
 Accepted: 12 February 2016  
 Published: 15 March 2016

There is an urgent need to improve the infrastructure supporting the reuse of scholarly data. A diverse set of stakeholders—representing academia, industry, funding agencies, and scholarly publishers—have come together to design and jointly endorse a concise and measurable set of principles that we refer to as the FAIR Data Principles. The intent is that these may act as a guideline for those wishing to enhance the reusability of their data holdings. Distinct from peer initiatives that focus on the human scholar, the FAIR Principles put specific emphasis on enhancing the ability of machines to automatically find and use the data, in addition to supporting its reuse by individuals. This Comment is the first formal publication of the FAIR Principles, and includes the rationale behind them, and some exemplar implementations in the community.

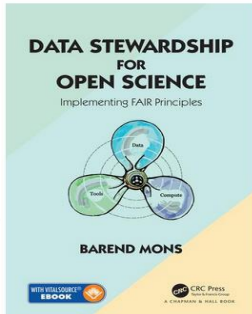


# Comment les principes se traduisent-ils ?

## 15 principes FAIR

- **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 they describe.
  - F4. (Meta)data are registered or indexed in a searchable resource.
- **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 where necessary.
  - A2. Metadata are accessible, even when the data are no longer available.
- **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.
- **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.

## Attention, FAIR n'est pas ...



- **Standard/ norme** : les principes FAIR ont pour but de guider les gestionnaires de données vers des choix d'implémentation qui facilitent la découverte et l'utilisation de ces données.
- **Technologie** : les principes FAIR sont indépendants de toute technologie.
- **Science ouverte** : Le choix du niveau de disponibilité de la donnée FAIR est laissée à la discrétion du publiant. Le principe R implique de fournir suffisamment d'informations (conditions d'accès, de provenance et d'usage).

30/06/2021

Journées Ingénierie des Connaissances 2021, Bordeaux

5

## La vision FAIR

- Adoptés par plusieurs :
  - Organismes de financement
    - Commission européenne
  - Organisations politiques
    - G20
  - Communautés scientifiques
    - EOSC



[https://www.go-fair.org/wp-content/uploads/2020/03/Data-Together\\_March-2020.pdf](https://www.go-fair.org/wp-content/uploads/2020/03/Data-Together_March-2020.pdf)

30/06/2021

Journées Ingénierie des Connaissances 2021, Bordeaux

6



The MIT Press Journals

Books Journals Digital Resources About Contact

Home | Data Intelligence | List of Issues | Volume 2, No. 1-2 | Ontology, Ontologies and the "I" of FAIR



Quarterly  
Founded: 2018  
E-ISSN: 2641-435X

More About *Data Intelligence* ▾

#### Journal Resources

Editorial Info  
Peer Review  
Abstracting and Indexing  
Release Schedule  
Advertising Info

## Ontology, Ontologies and the "I" of FAIR

Giancarlo Guizzardi 

Posted Online January 31, 2020  
[https://doi.org/10.1162/dint\\_e\\_00040](https://doi.org/10.1162/dint_e_00040)

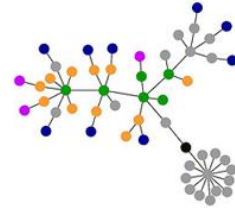
© 2019 Chinese Academy of Sciences Published under a Creative Commons Attribution 4.0 International (CC BY 4.0) license

#### Data Intelligence

Volume 2 | Issue 1-2 | Winter-Spring 2020  
p.181-191

This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. For a full description of the license, please visit <https://creativecommons.org/licenses/by/4.0/legalcode>.

**Keywords:** FAIR, Semantic Interoperability, Ontology, ontologies



**Les ontologies sont  
considérées comme  
un élément clé pour  
l'implémentation des  
principes FAIR**

# Problèmes liés aux principes FAIR

# Problèmes liés aux principes FAIR

- Les principes FAIR sont décrits à un niveau très générique et leur mise en œuvre a été mise de côté au début.



- **Variété des principes FAIR** : objectives, mesurables, subjectives, techniques, communautaires ...
- **Variété des interprétations** : 50 nuances de FAIR !

## SCIENTIFIC DATA

OPEN Comment: A design framework and exemplar metrics for FAIRness

Mark D. Wilkinson<sup>1</sup>, Susanna-Assunta Sansone<sup>2</sup>, Erik Schultze<sup>3</sup>, Peter Dumas<sup>4</sup>, Luis Olivera-Bonini de Silva Santos<sup>5,6</sup> & Michel Dumortier<sup>7</sup>

Received 28 November 2020  
Accepted 9 May 2021  
Published 28 Jan 2021

The FAIR Principles<sup>1</sup> (<https://doi.org/10.21961/43846v2g>) provide guidelines for the publication of digital research such as datasets, code, workflows and research objects, in a manner that enables their Findability, Accessibility, Interoperability and Reusability (FAIR). The Principles were rapidly being adopted by publishers, funders, and pre-decisionary infrastructure programmes and societies. The Principles are expected, in that they do not merely define how to achieve a state of 'FAIRness', but rather they describe a continuum of factors, activities and behaviors that will move a digital resource closer to that goal. This article has led to a wide range of interpretations of FAIRness, with some researchers even claiming to already 'be FAIR'. The increasing number of such statements, the emergence of alternative and self-assessments of FAIRness<sup>2,3</sup>, and the need of data and service providers, journals, funding agencies, and regulatory bodies to qualitatively or quantitatively evaluate such claims, led us to self-organize and establish a FAIR Maturity group (<http://fairmaturity.org>) to pursue the goal of defining ways to measure FAIRness.

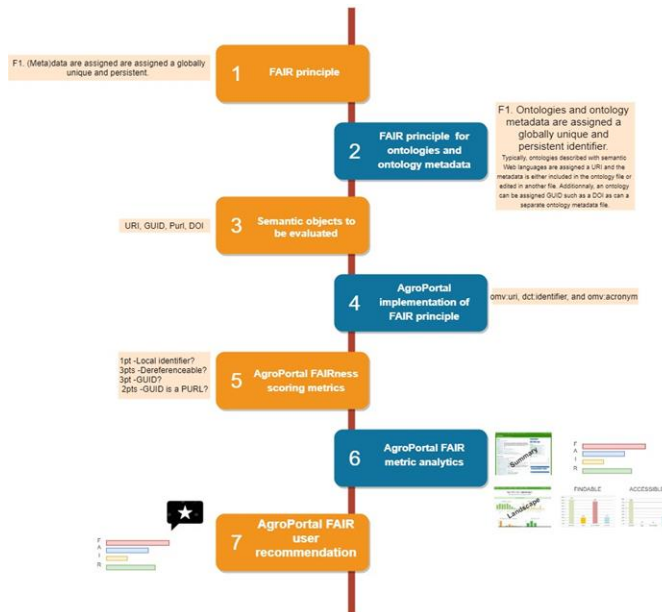


Besoin de fournir des bonnes pratiques et des outils à la communauté.

# Objectif

# Objectif

## Une approche et un outil d'évaluation automatique du niveau de «FAIRness» d'une ontologie



# Travaux autour du «FAIRness assessment»

30/06/2021

Journées Ingénierie des Connaissances 2021, Bordeaux

13

## Etat de l'art

Actuellement, plusieurs **approches et outils génériques** pour l'auto-évaluation d'une ressource :

- Manuel
  - **FDMM** et SHARC
- Semi-automatique
  - FAIR self-assessment tool
- Automatique
  - FAIR Metrics et FAIR-Aware



Matrice FDMM V0.04

PRINCIPLE	INDICATEUR	INDICATORS	PRIORITY	
F	F1	F1-01M	Metadata is identified by a persistent identifier	Mandatory
	F1	F1-01D	Data is identified by a persistent identifier	Mandatory
	F1	F1-02M	Metadata is identified by a universally unique identifier	Mandatory
	F1	F1-02D	Data is identified by a universally unique identifier	Mandatory
	F2	F2-01M	Sufficient metadata is provided to allow discovery, following domain/discipline-specific metadata standard	Recommended
	F3	F3-01M	Metadata includes the identifier for the data	Recommended
	F4	F4-01M	Metadata is offered/published/expressed in such a way that it can be harvested and indexed	Mandatory
	A1	A1-01M	Metadata includes information about access conditions	Mandatory
	A1	A1-01D	Data can be accessed manually (i.e. with human intervention)	Mandatory
	A1	A1-02D	Data can be accessed automatically (i.e. by a computer program)	Recommended
A	A1	A1-02M	Metadata identifier resolves to a metadata record	Mandatory
	A1	A1-02D	Data identifier resolves to a digital object	Mandatory
	A1	A1-03M	Metadata is accessed through standardised protocol	Recommended
	A1	A1-03D	Data is accessible through standardised protocol	Recommended
	A1.1	A1.1-01M	Metadata is accessible through a free access protocol	Mandatory
	A1.1	A1.1-01D	Data is accessible through a free access protocol	Recommended
	A1.1	A1.1-02M	Metadata is accessible through an open-source access protocol	Recommended
	A1.1	A1.1-02D	Data is accessible through an open-source access protocol	Recommended
	A1.1	A1.1-03D	Conditions to be taken by a user to get access to the data are self documented	Recommended
	A1.2	A1.2-01M	Metadata includes information relevant for access control	Mandatory
A1.2	A1.2-01D	Data is accessible through an access protocol that supports authentication	Optional	
A1.2	A1.2-02D	Data is accessible through an access protocol that supports authentication	Optional	
A2	A2-01M	Metadata is guaranteed to remain available after data is no longer available	Mandatory	

<https://docs.google.com/document/d/1pDGL3-BbJu18KifZUI3AizKLHXGdImPtpEWmeg/edit#heading=h.gjdxs>

30/06/2021

Journées Ingénierie des Connaissances 2021, Bordeaux

14

## Etat de l'art

Actuellement, plusieurs **approches et outils génériques** pour l'auto-évaluation d'une ressource :

- Manuel
  - SHARC et FDMM
- Semi-automatique
  - FAIR self-assessment tool
- Automatique
  - FAIR Metrics et FAIR-Aware

FAIR self-assessment tool A R D C  
Australian Research Data Commons

Findable ?

Accessible ?

How accessible is the data? Fully accessible to persons who meet explicitly stated condition

Is the data available online without requiring specialised protocols or tools once access has been approved? Non-standard web service (e.g. OpenAPI/Swagger/Informal AP)

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

Interoperable ?

Reusable ?

Total across FAIR [Progress bar]

<https://www.ands-nectar-rds.org.au/fair-tool>

30/06/2021

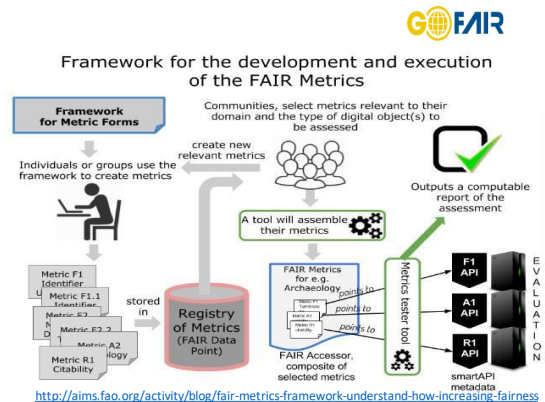
Journées Ingénierie des Connaissances 2021, Bordeaux

15

## Etat de l'art

Actuellement, plusieurs **approches et outils génériques** pour l'auto-évaluation d'une ressource :

- Manuel
  - SHARC et FDMM
- Semi-automatique
  - FAIR self-assessment tool
- Automatique
  - FAIR Metrics et FAIR-Aware



30/06/2021

Journées Ingénierie des Connaissances 2021, Bordeaux

16



## Etat de l'art

- Ces deux dernières années, **des approches plus spécifiques** autour de l'évaluation des ressources sémantiques selon les principes FAIR, ont été publiées :
  - Recommandations projet EU FAIRs FAIR
  - Recommandations de Poveda et al.\*
  - Outil DBPedia Archivio

## Notre méthodologie d'évaluation du niveau de «FAIRness» pour les ressources sémantiques

## Approche «Bottom-Up»

### • Etape 1 : Sélection

Category	Generic (A)		Specific (B, C)			
Format	grid		principles	scheme	recommendations	
Approach	SHARC	FDMM	5-stars V	MIRO	FAIRsFAIR	Poveda et al.
Year	2017	2018	2011	2017	2020	2020
FAIR movement	after	after	before	before	after	after
FAIR priorities	yes	yes	n/a	n/a	no	no

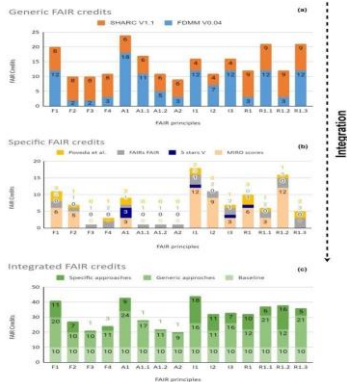
## Approche «Bottom-Up»

### • Etape 2 : Alignement

	SHARC	FDMM	MIRO	5*V	FsF	Poveda et al.
F1	X	X	X	-	X	X
F2	X	X	X	-	X	X
F3	X	X	-	-	X	-
F4	X	X	-	-	X	X
A1	X	X	X	X	X	X
A1.1	X	X	-	-	X	-
A1.2	X	X	-	-	X	-
A2	X	X	-	-	X	-
I1	X	X	X	X	X	X
I2	X	X	X	-	X	-
I3	X	X	X	X	X	X
R1	X	X	X	X	-	X
R1.1	X	X	X	-	X	X
R1.2	X	X	X	-	X	X
R1.3	X	X	-	-	X	X

# Approche «Bottom-Up»

## • Etape 3 : Mesurage



Principe	Base line	SHA RC	FDMM	5-stars V	MIRO	FAIRs FAIR	Poveda et al.	Credits		
F	F1	10	8	12	0	6	2	3	41	113
	F2	10	8	2	0	5	1	1	27	
	F3	10	8	2	0	0	1	0	21	
	F4	10	8	3	0	0	2	1	24	
A	A1	10	6	18	3	3	1	2	43	113
	A1.1	10	6	11	0	0	1	0	28	
	A1.2	10	6	5	0	0	1	0	22	
I	A2	10	6	3	0	0	1	0	20	109
	I1	10	4	12	1	12	3	2	44	
	I2	10	4	7	0	9	2	0	32	
R	I3	10	4	12	1	3	2	1	33	143
	R1	10	9	3	1	6	0	3	32	
	R1.1	10	9	12	0	3	2	1	37	
	R1.2	10	9	3	0	12	3	1	38	
Total credits									478	

30/06/2021

Journées Ingénierie des Connaissances 2021, Bordeaux

21

# Approche «Bottom-Up»

## • Etape 4 : Formulation

68 FAIR questions

<https://github.com/agroportal/fairness>

### Liste 1. Questions d'évaluation de F1 (41 crédits)

**Q1.** Une ontologie a-t-elle un identifiant "local", c'est-à-dire un identifiant globalement unique et potentiellement persistant mais attribué par le développeur (ou l'organisation de qui développe l'ontologie) ? **9 cts**

**Q2.** Une ontologie fournit-elle un identifiant "externe" supplémentaire, c'est-à-dire un identifiant globalement unique et persistant attribué par un organisme accrédité ? **6 cts**

**Q3.** Si oui, cet identifiant externe est-il un DOI ? **5 cts**

**Q4.** Les métadonnées de l'ontologie sont-elles incluses dans le fichier d'ontologie et partagent-elles par conséquent les mêmes identifiants ? **6 cts**

**Q5.** Sinon, l'ensemble des métadonnées est-il clairement identifié par son propre GUPRI ? **6 cts**

**Q6.** Une ontologie fournit-elle un URI spécifique à la version ? **4 cts**

**Q7.** Si oui, cet URI est-il résolvable/ déréférencable ? **5 cts**

30/06/2021

Journées Ingénierie des Connaissances 2021, Bordeaux

22

# Approche «Bottom-Up»



## Etape 5 : Représentation (1/2)

### « Findable »



- PID  
owl:ontologyIRI, dct:identifiant, owl:versionIRI



- Métadonnées  
72 métadonnées optionnelles



- Métadonnées avec PID  
N/A



- Référentiel de données  
schema:includedInDataCatalog

### « Accessible »



- Protocole standard  
owl:ontologyIRI, dct: identifiant, sd:endpoint



- Protocole libre et ouvert  
N/A



- Authentification  
Schema:includedInDataCatalog



- Accès pérenne aux métadonnées  
dct:hasVersion, omv:status

23

# Approche «Bottom-Up»



## Etape 5 : Représentation (2/2)

### « Interoperable »



- Vocabulaire  
Omv:hasOntologyLanguage, omv:hasFormalityLevel, omv:hasOntologySyntax, dct:hasFormat, dct:isFormatOf, voaf:similar, voaf:metadataVoc, dct:relation, dct:isPartOf, voaf:specializes, schema:translationOfWork, voaf:generalizes

F-A-I-R



- Vocabulaire FAIR  
owl:imports, voaf:hasEquivalenceWith, owl:priorVersion



- Métadonnées liées  
Mod:ontologyInUse, omv:endorsedBy

### « Reusable »



- Métadonnées avec attributs  
mod:prefLabelProperty, mod:synonymProperty, mod:definitionProperty, mod:authorProperty, bpm:obsoleteProperty, mod:hierarchyProperty, mod:obsoleteParent, mod:maxDepth, mod:maxChildCount, mod:averageChildCount, mod:classesWithOneChild, mod:classesWithNoDefinition



- License  
dct:license, dct:rightsHolder, dct:accessRights, cc:morePermissions, cc:useGuidelines



- Provenance  
dct:source, prov:wasGeneratedBy, prov:wasInvalidatedBy, dct:accrualMethod, dct:accrualPeriodicity, dct:accrualPolicy, omv:versionInfo, vann:changes, dct:hasVersion, omv:usedOntologyEngineeringTool, omv:usedOntologyEngineeringMethodology, omv:conformsToKnowledgeRepresentationParadigm, omv:designedForOntologyTask, mod:competencyQuestion, dct:fundedBy



- Standards de la communauté  
mod:group

24

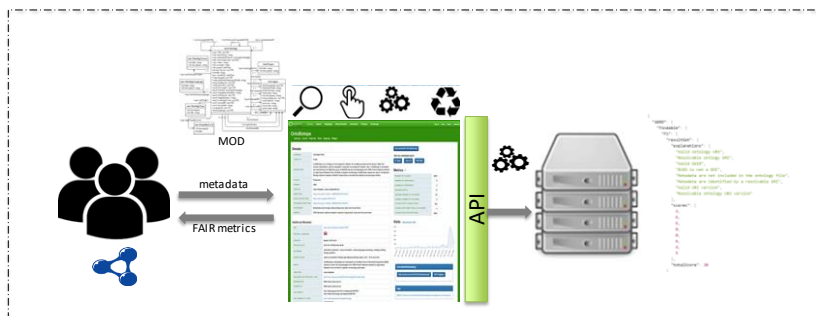
# Notre outil d'évaluation du niveau de «FAIRness» pour les ressources sémantiques

30/06/2021

Journées Ingénierie des Connaissances 2021, Bordeaux

25

## Service Web «FAIRness assessment»



[http://services.agroportal.lirmm.fr/fairness\\_assessment/?portal=agroportal&ontology=ANAEETHES](http://services.agroportal.lirmm.fr/fairness_assessment/?portal=agroportal&ontology=ANAEETHES)

 <https://github.com/agroportal/fairness>

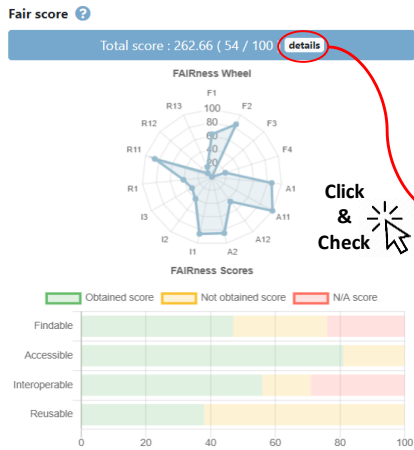
30/06/2021

Journées Ingénierie des Connaissances 2021, Bordeaux

26

# Prototype : Interface & Rapport d'évaluation

Bientôt une nouvelle rubrique !



Evaluation détaillée en JSON-LD :

```

{
  "@context": "https://www.w3.org/ns/dx/v2/context",
  "@type": "dx:EvaluationReport",
  "id": "https://www.ebi.ac.uk/ontology/evaluation/2021/01/01/1234567890",
  "title": "Evaluation report for a FAIR identifier (i.e., a globally unique and persistent identifier assigned to the metadata (or including organization))",
  "questions": [
    {
      "question": "P101: Does an ontology have a 'local' identifier (i.e., a globally unique and potentially persistent identifier assigned by the developer (or developing organization))?",
      "testResult": "Success",
      "explanation": "Valid ontology URI"
    },
    {
      "question": "P102: If yes, is the identifier a resolvable/referencable URI?",
      "testResult": "Success",
      "explanation": "Resolvable ontology URI"
    },
    {
      "question": "P103: Does an ontology provide an additional 'external' identifier (i.e., a globally unique and persistent identifier assigned by an accredited body)?",
      "testResult": "Success",
      "explanation": "Valid DOI"
    },
    {
      "question": "P104: If yes, is this external identifier a DOI?",
      "testResult": "Success",
      "explanation": "Valid DOI for the GUID"
    },
    {
      "question": "P105: Are the ontology metadata included in the ontology file 'held' (consequently share the same identifier)?",
      "testResult": "Fail",
      "explanation": "Metadata are not included in the ontology file"
    },
    {
      "question": "P106: If not, is the metadata record clearly identified by its",
      "testResult": "Success",
      "explanation": "Metadata are"
    }
  ]
}

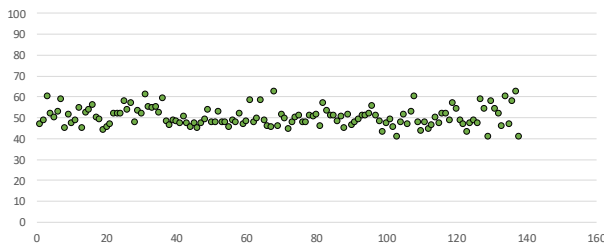
```

<http://services.stageportal.lirmm.fr/fairness/v2/?portal=stageportal&ontology=ANAETHES>

27

## Expérimentation

FAIRness values in AgroPortal



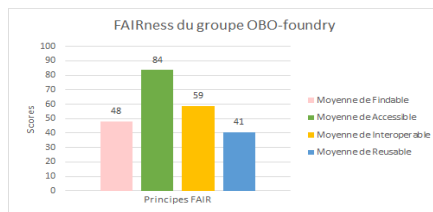
Mars 2021

### AgroPortal

- N=134
- Moyenne= 50 et médiane= 49

### OBO- Foundry

- N=24
- Moyenne= 58
- Meilleure= Phenotype And Trait Ontology PATO (score de 65)
- BFO score = 55



28

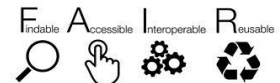
# Synthèse et mise en perspective

30/06/2021

Journées Ingénierie des Connaissances 2021, Bordeaux

29

## Les apports de notre travail



- Contrairement aux méthodes existantes, notre travail :
  - Considère des approches historiques, génériques et spécifiques : **alignement, intégration et mesurage**
  - Définit une **métrique** : **pas FAIR, FAIR ou FAIRer ?**
  - Définit un ensemble de **questions** pour tous les principes FAIR
  - Propose des **métadonnées** couvrant 70% des principes FAIR
- Notre travail est implémenté dans un **portail d'ontologies**

30/06/2021

Journées Ingénierie des Connaissances 2021, Bordeaux

30

## Enjeux scientifiques ...



- Plusieurs enjeux scientifiques liés à l'évaluation du niveau de «FAIRness» des ressources sémantiques nécessitent d'être traités par la communauté nous citons à titre d'exemples :
  - Etablir un consensus pour garantir **la persistance** des URIs (ex. un service d'enregistrement d'identifiants)
  - Proposer **un ensemble de métadonnées** à évaluer pour chaque principe
  - Fournir des mécanismes de **standardisation** et d'échange des métadonnées afin de faciliter leur récupération par les **moteurs de recherche**
  - Désigner **un organisme de validation** des métriques FAIR
  - ...

## Nos perspectives

- Rendre AgroPortal **un portail d'ontologies et de vocabulaires FAIR** :
  - Supporter la vision FAIR
  - Produire des ontologies et vocabulaires interopérables et persistants
  - Guider les utilisateurs dans le choix de la « meilleure » ontologie et le développement d'une ontologie FAIR
- Continuer nos efforts pour **la standardisation et l'interopérabilité** des métadonnées des ressources sémantiques au sein des initiatives internationales de la communauté FAIR.





## Remerciements



**LIRMM**

Jérôme Lamarque

Syphax Bouazzouni



**FooSIN**



**D2KAB**



RESEARCH DATA ALLIANCE



**FAIRSFAR**  
Fostering Fair Data Practices in Europe

# Merci de votre attention !

[emna.amdouni@lirmm.fr](mailto:emna.amdouni@lirmm.fr)