

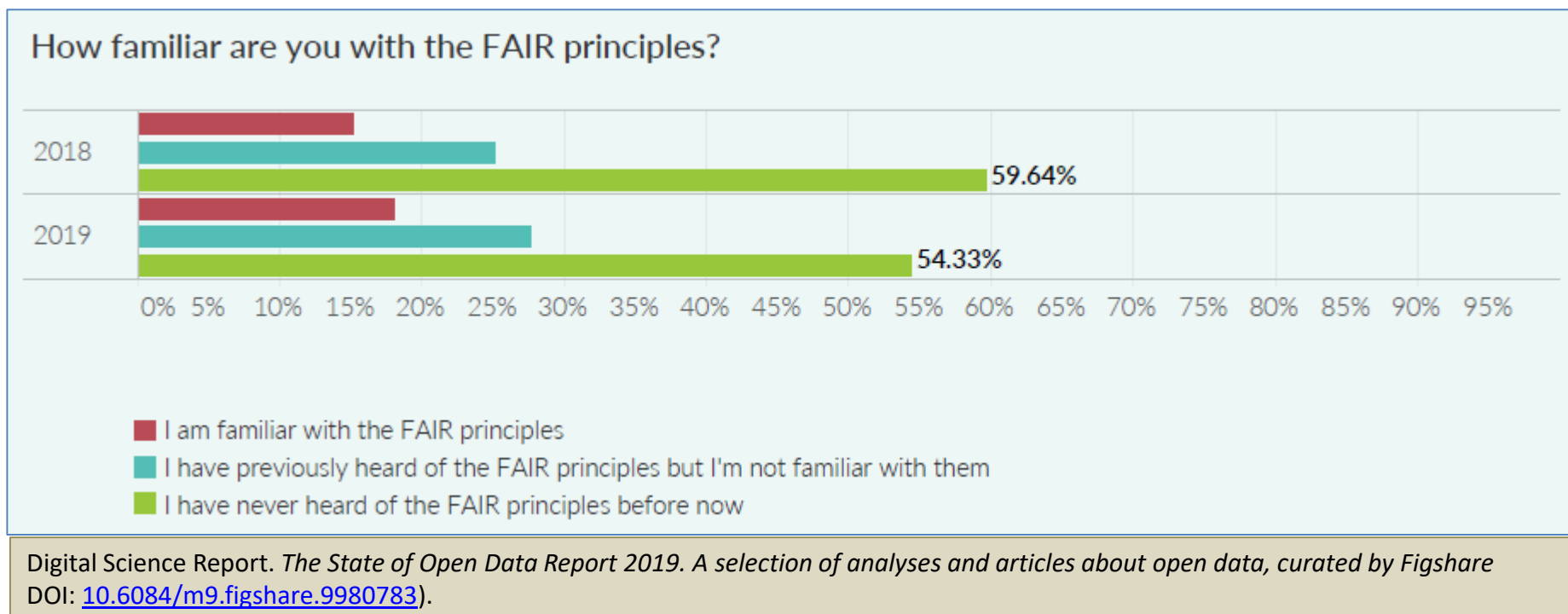
Produire des données FAIR

Sylvie Cocaud. Partager les données de la recherche à l'Inra : pourquoi, comment ?
Centre Antilles-Guyane, 25-27 novembre 2019



Sondage...

Enquête Figshare 2019 : + de 8000 participants de 190 pays



Qui connaît les principes FAIR ?

Oui	41%
Je ne suis pas sûr	27%
Non	32%

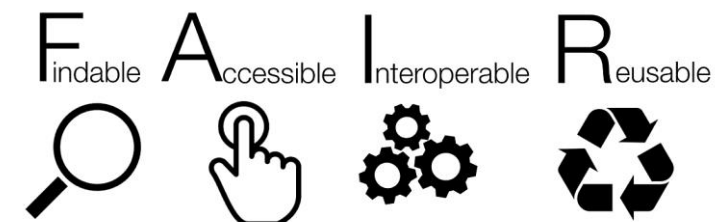
Qui a mis en pratique les principes FAIR ?



Origine des principes FAIR



ist@inra



SangyaPundir [[CC BY-SA 4.0](https://creativecommons.org/licenses/by-sa/4.0/)]

“all research objects should be Findable, Accessible, Interoperable and Reusable (FAIR) both for machines and for people”

In Web of Science Core Collection

836



Times Cited

Au 28/10/2019



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

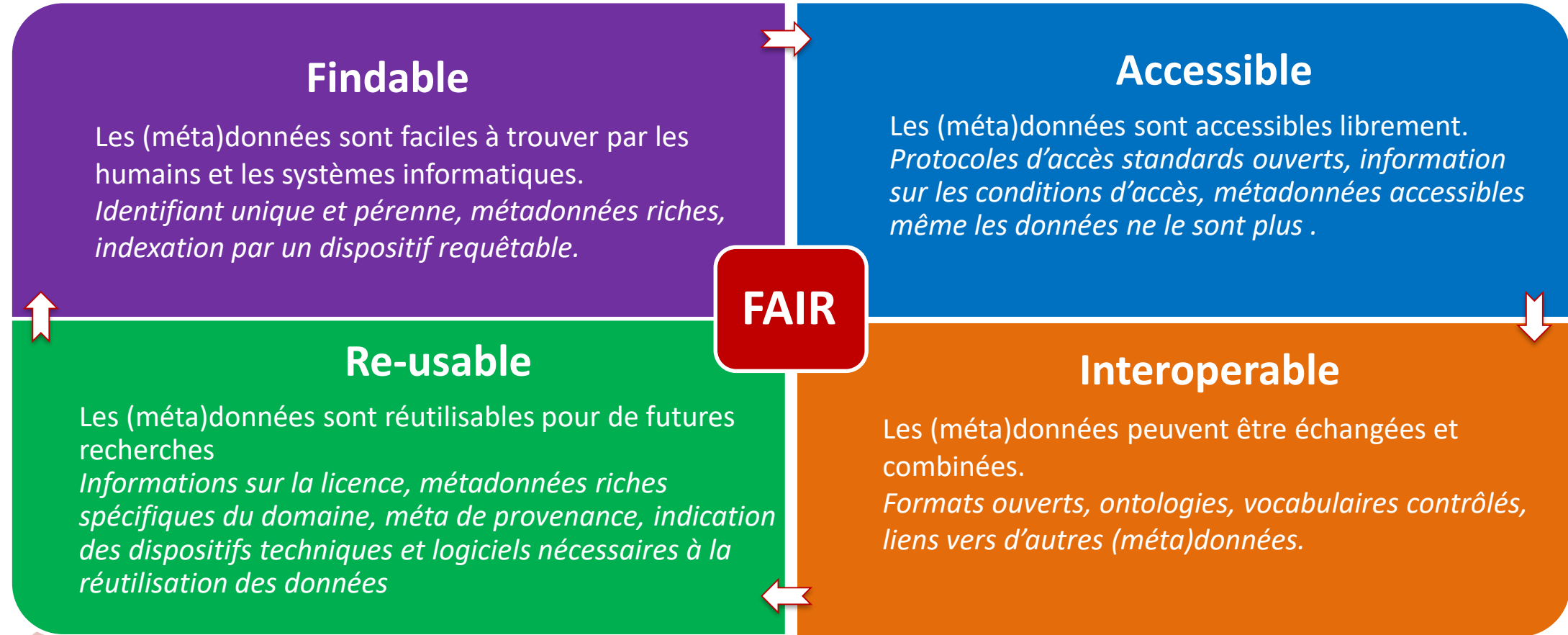
Wilkinson MD, Dumontier M, Aalbersberg IJ, et al. **The FAIR Guiding Principles for scientific data management and stewardship.** *Sci Data*. 2016;3:160018. Published 2016 Mar 15. doi:[10.1038/sdata.2016.18](https://doi.org/10.1038/sdata.2016.18)

Objectifs des principes FAIR

Volume
Variété, complexité
Vitesse de création

} des données

Les principes FAIR mettent l'accent sur l'exploitabilité des données par les machines (sans intervention humaine ou avec une intervention humaine minimale)



ist@inra

Le principe Findable

- F1. les (méta)données sont associées à un identifiant universellement unique, et pérenne**
- F2. les données sont décrites avec des métadonnées riches (voir aussi R1)**
- F3. les métadonnées incluent clairement et explicitement l'identifiant des données qu'elles décrivent**
- F4. les (méta)données sont enregistrées ou indexées dans un dispositif qui permet de les retrouver**

Le principe Accessible

A1. les (méta)données peuvent être récupérées grâce à leur identifiant via un protocole de communication normalisé

A1.1 le protocole est ouvert, libre et universellement applicable

A1.2 le protocole prévoit une procédure d'authentification et d'autorisation, si nécessaire

A2. les métadonnées sont accessibles, même lorsque les données ne sont plus disponibles

Le principe Interoperable

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

Le principe Reusable

R1. les données sont décrites de manière riche avec des attributs précis et pertinents

R1.1. les (méta)données sont publiées avec une **licence** d'utilisation explicite et accessible

R1.2. les (méta)données sont associées à une **provenance** détaillée

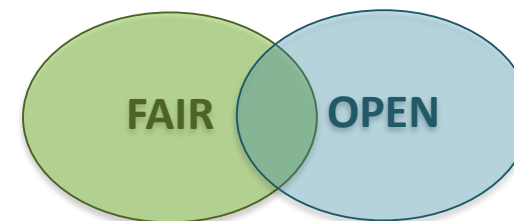
R1.3. les (méta)données respectent les **standards de domaine** pertinents

<https://fairsharing.org/standards/>

Les principes FAIR

- Principes directeurs,
- Applicables à tous les objets de la recherche
- Interprétables, implémentables de différentes façons
- Processus continu

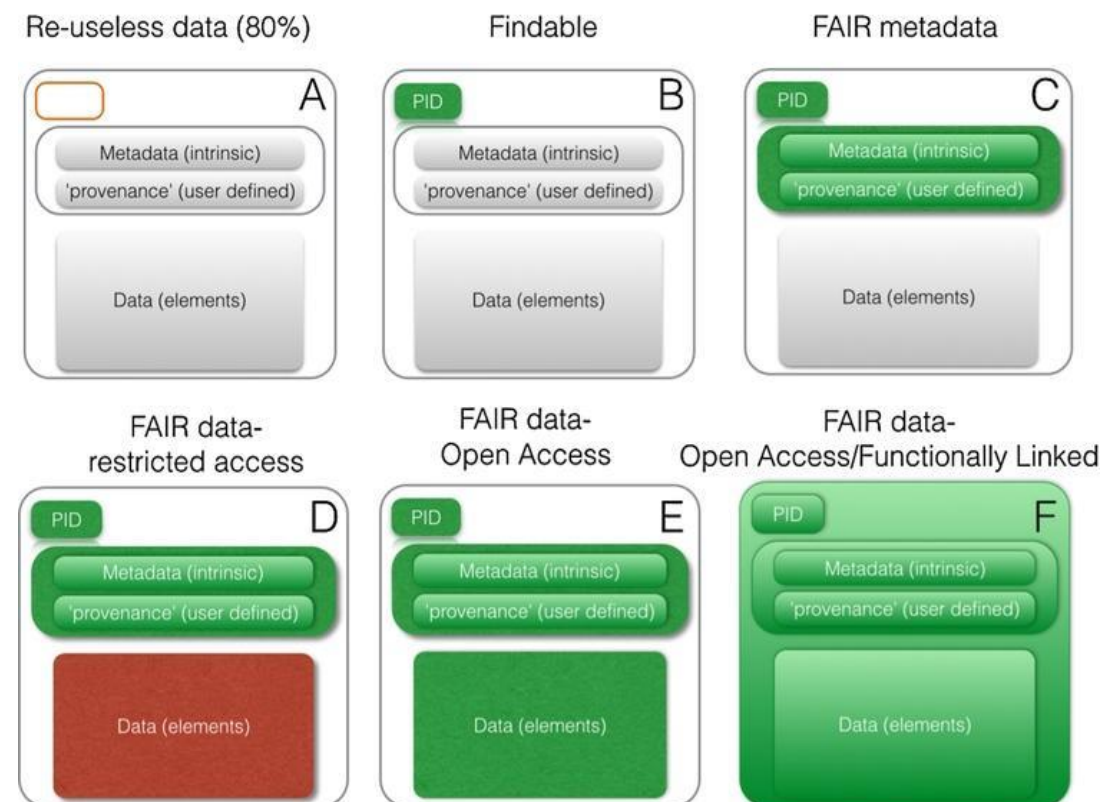
The more FAIR data is, the easier it is for computers to integrate it and derive additional value from it. <http://m.euretos.com/newspage.php?id=22>



Data as increasingly FAIR Digital Objects

open

close



Mons B, et al. Cloudy, increasingly FAIR; revisiting the FAIR Data guiding principles for the European Open Science Cloud. Information Services & Use. 1 janv 2017;37(1):49-56. DOI: [10.3233/ISU-170824](https://doi.org/10.3233/ISU-170824)

Mes données sont-elles FAIR ?

Outils d'(auto)évaluation

5 ★ DATA RATING TOOL du CSIRO

<http://oznome.csiro.au/5star/>



FAIR self-assessment tool

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

The screenshot shows the FAIR self-assessment tool interface. At the top, there is a search bar and navigation links for home, news, events, programs, and about. The main heading is "FAIR self-assessment tool". Below this, the "Findable" section is expanded, showing four questions with dropdown menus for answers. The questions and their selected answers are:

- 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? Yes
- How is the data described with metadata? Comprehensively (see suggestion) using a recognised formal
- What type of repository or registry is the metadata record in? Data is in one place but discoverable through several registrie

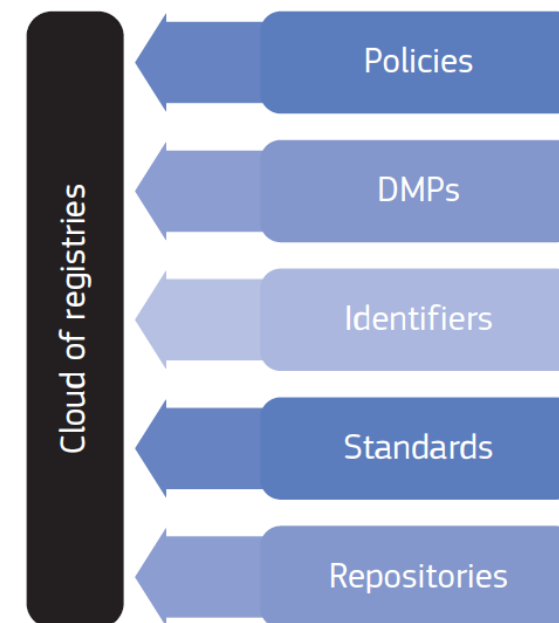
Below the "Findable" section, there are sections for "Accessible", "Interoperable", and "Reusable", each with an information icon. At the bottom, there is a "Total across F.A.I.R." section with a green progress bar.

Comment faire FAIR ?

FAIR Rédiger un PGD

F	Attribuer des identifiants pérennes et non ambigus
F	Citer les données réutilisées
FI	Choisir des schémas de métadonnées standard pour décrire/citer les données
FAiR	Déposer les données dans un entrepôt de confiance, dans des formats de fichiers ouverts et standards
A	Préciser les conditions d'accès, accès libre quand c'est possible
I	Utiliser des vocabulaires largement partagés dans le domaine
I	Utiliser (ou développer(et documenter) des ontologies de domaine pour définir les données sans ambiguïté
AI	Utiliser des systèmes offrant des protocoles d'échange standard (OAI-PMH, Sword, SPARQL...) et des APIs
R	Indiquer la licence d'utilisation de manière explicite
IR	Mettre en place des procédures de nettoyage des données, d'enrichissement des métadonnées, pour améliorer la qualité
R	Fournir des métadonnées les plus riches possible, des informations sur la provenance des données, tout autre document nécessaire

S'appuyer sur les composants d'un « écosystème FAIR »



European Commission. **Turning FAIR into reality. Final Report and Action Plan from the European Commission Expert Group on FAIR Data.** 2018.
https://ec.europa.eu/info/publications/turning-fair-reality_en

Data Inra : un entrepôt FAIR ?



Estimation réalisée avec l'ARDC
[FAIR self-assessment tool](#)



■ Présent : DOI ; Métadonnées indiquant l'identifiant ; entrepôt permettant la recherche
▨ Utilisateur : Métadonnées riches en utilisant des vocabulaires

F1. les métadonnées et les données sont associées à un identifiant global, unique et pérenne	😊
F2. les données sont décrites avec des métadonnées riches	😐 Démarche utilisateur
F3. les métadonnées et les données sont enregistrées ou indexées dans un dispositif qui permet de les retrouver	😊
F4. les métadonnées incluent clairement et explicitement l'identifiant des données qu'elles décrivent	😊

Data Inra : un entrepôt FAIR ?



Estimation réalisée avec l'ARDC
[FAIR self-assessment tool](#)



■ Présent : Données accessible avec le DOI, via protocole HTTP, métadonnées toujours disponibles

A1. les métadonnées et les données peuvent être récupérées grâce à leur identifiant via un protocole de communication normalisé	😊
A1.1 le protocole est ouvert, libre et universellement applicable	😊
A1.2 le protocole prévoit une procédure d'authentification et d'autorisation, si nécessaire	😊
A2. les métadonnées sont accessibles, même lorsque les données ne sont plus disponibles	😊



Data Inra : un entrepôt FAIR ?







Estimation réalisée avec l'ARDC
[FAIR self-assessment tool](#)



■ Présent : Métadonnées standardisées et machine readable

▨ Utilisateur : Format de fichiers ouverts et machine-readable, Vocabulaires avec URI

11. Les métadonnées et les données utilisent un langage formel, accessible, partagé et largement adopté pour la représentation des connaissances	 métadonnées  Données Démarche utilisateur
12. les métadonnées et les données utilisent des vocabulaires qui respectent les principes FAIR	 Démarche utilisateur
13. les métadonnées et les données incluent des liens « qualifiés » vers d'autres métadonnées et données	 Démarche utilisateur

Data Inra : un entrepôt FAIR ?



Estimation réalisée avec l'ARDC
[FAIR self-assessment tool](#)



■ Présent : Licences disponibles

▨ Utilisateur : Provenance des données, Standards correspondant à la communauté, ...

R1. les données sont décrites de manière riche avec des attributs précis et pertinents	
R1.1. les métadonnées et les données sont publiées avec une licence d'utilisation explicite et accessible	😊
R1.2. les métadonnées et les données sont associées à une provenance détaillée	😐 Démarche utilisateur
R1.3. les métadonnées et les données respectent les standards de domaine pertinents	😐 Démarche utilisateur



Pour en savoir plus...



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

A bottom-up international approach for the practical implementation of the European Open Science Cloud (EOSC) as a part of a global Internet of FAIR data & services

<https://www.go-fair.org/fair-principles/>



- [FAIR Data Maturity Model WG](#)
- [FAIRSharing Registry: connecting data policies, standards & databases WG](#)
- [Sharing Rewards and Credit \(SHARC\) IG](#)

Esther Dzalé Yeumo. 2018. **Les principes FAIR**. Cahier des Techniques de l'Inra, n. sp. « Données de la recherche ».

https://www6.inra.fr/cahier_des_techniques/content/download/5106/51886/version/4/file/CTh2018bis_Art1_DZA-4.pdf

