

Ecole Thématique DATA SDUE

**Guide de Survie dans la jungle des données
en Sciences de l'Univers et de l'Environnement (SDUE) :
Comment gérer les données pour les valoriser?**

« Paysage de la Science Ouverte »

Françoise Genova

La Science Ouverte

Le partage des données scientifiques

Françoise Genova

Observatoire Astronomique de Strasbourg



The top of the slide features a cosmic background with a dark blue and black sky filled with numerous stars and nebulae. Four blue squares are overlaid on this background: a small square in the top left, two medium squares in the top center, and a large square on the right side that extends below the black background into the white area of the slide.

AU NIVEAU POLITIQUE: DES PRECONISATIONS « TOP-DOWN »

□ L'OCDE, dès 2004...

- Les ministres de la Recherche et de la Technologie des pays de l'OCDE, plus l'Afrique du Sud, la Chine, Israël et la Russie
 - Declaration on Access to Research Data from Public Funding
 - Demandent que l'OCDE formule des principes et des directives pour l'accès aux données de la recherche obtenues sur financement publics
- OECD Principles and Guidelines for Access to Research Data from Public Funding, 2007
 - Openness, flexibility, transparency, legal conformity, protection of intellectual property, formal responsibility, professionalism, interoperability, quality, security, efficiency, accountability, sustainability
- Updated by OECD in 2021
 - Initiatives internationales, politiques actuelles dans les pays membres et partenaires, statut actuel, les lacunes et les défis

□ Les ministres de la recherche du G8/G7

2013

- i. To the greatest extent and with the fewest constraints possible **publicly funded scientific research data should be open**, while at the same time respecting concerns in relation to privacy, safety, security and commercial interests, whilst acknowledging the legitimate concerns of private partners.
- ii. Open scientific research data should be easily **discoverable, accessible, assessable, intelligible, useable, and wherever possible interoperable to specific quality standards**.
- iii. To maximise the value that can be realised from data, the mechanisms for delivering open scientific research data should be efficient and cost effective, and consistent with the potential benefits.
- iv. To ensure successful adoption by scientific communities, open scientific research data principles will need to be underpinned by an appropriate policy environment, including **recognition of researchers** fulfilling these principles, and **appropriate digital infrastructure**.

2016

Promote international coordination and collaboration to develop the appropriate technology, infrastructure, including digital networks, and human resources for the effective utilization of open science for the benefit of all.

□ Au niveau européen

- C. Moedas (Research, Science and Innovation): *Open Science, Open Innovation, Open to the World* (2015)
- Plans de Gestion de Données pour les projets
- European Open Science Cloud (EOSC) en développement
 - Web de données ouvertes et de services
- European Strategy Forum on Research Infrastructures (ESFRI)
 - Des infrastructures de données
 - Des questions sur les données pour toutes les infrastructures

□ En France



2016: Loi pour la République Numérique

- Promeut l'ouverture et la circulation des données et des connaissances

Plan National pour la Science Ouverte

2018

- Généraliser l'accès ouvert aux publications
- Structurer et ouvrir les données de la recherche
- S'inscrire dans une dynamique durable, européenne et internationale

2021

- Généraliser l'accès ouvert aux publications
- Structurer, partager et ouvrir les données de la recherche
- Ouvrir et promouvoir les codes sources produits par la recherche
- Transformer les pratiques pour faire de la science ouverte le principe par défaut

Le Comité pour la Science Ouverte (CoSO)

Le Comité pour les Services et Infrastructures Numériques (CoSIN)

□ Les mesures pour les données de la recherche

Deuxième axe

Structurer, partager et ouvrir les données de la recherche

12

Mesures

14

4

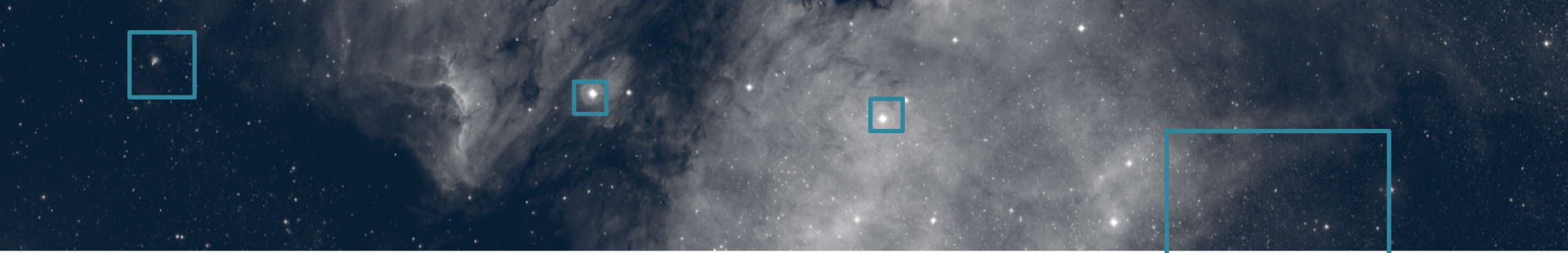
Mettre en œuvre l'obligation de diffusion des données de recherche financées sur fonds publics

5

Créer Recherche Data Gouv, la plateforme nationale fédérée des données de la recherche

6

Promouvoir l'adoption d'une politique de données sur l'ensemble du cycle des données de la recherche, pour les rendre faciles à trouver, accessibles, interopérables et réutilisables (FAIR)



UN OUTIL ESSENTIEL: LES PRINCIPES FAIR

□ Les propriétés à développer

- OECD « Guiding principles », 2007
Openness, flexibility, transparency, legal conformity, protection of intellectual property, formal responsibility, professionalism, interoperability, quality, security, efficiency, accountability, sustainability
- Ministres de la Recherche du G8, 2013
Open scientific research data should be easily discoverable, accessible, assessable, intelligible, useable, and wherever possible interoperable to specific quality standards
- FAIR – Wilkinson et al., 2016
Findable, Accessible, Interoperable, Reusable

□ 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

PID
Metadonnées
Standards
Licences

To be Accessible:

- A1. (meta)data are retrievable by their identifier using a standardized communications protocol
 - A1.1. the protocol is free, open 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 uses 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 data provenance
 - R1.3. (meta)data meet domain relevant community standards

Rôle des
communautés
thématiques

Figure 1: The FAIR guiding principles from Wilkinson et al. (2016).²

□ FAIR \neq Ouvert

- Les principes FAIR n'incluent pas le fait que les données sont ouvertes
- Il y a des exceptions légitimes - « aussi ouvert que possible »

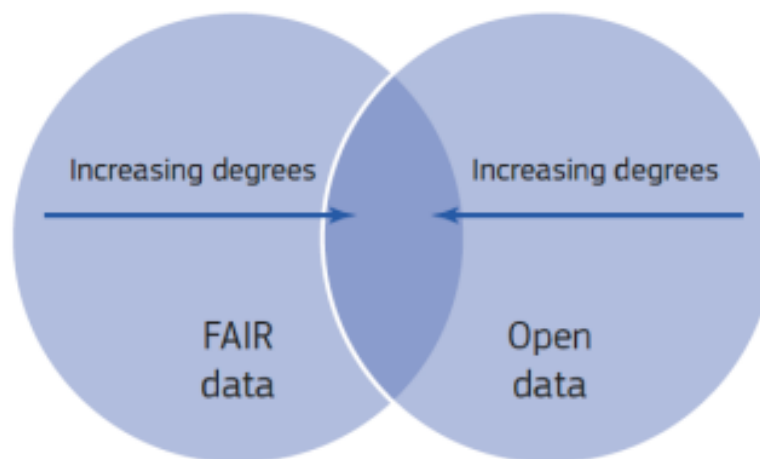


Figure 4. The relationship between FAIR and Open

From TFiR

□ Turning FAIR into Reality (2018)

Les différents éléments de l'écosystème à mettre en place

- Concepts – pourquoi FAIR?
- Créer une culture des données FAIR
- Créer un écosystème technique pour les données FAIR
- Création et construction des compétences
- Mesurer les évolutions
- Financement et pérennisation des données FAIR
- Plan d'action pour FAIR

Six Recommendations for Implementation of FAIR Practices



2020

1. Fund awareness-raising, training, education and community-specific support.
2. Fund development, adoption and maintenance of community standards, tools and infrastructure.
3. Incentivise development of community governance.
4. Translate FAIR guidelines for other digital objects.
5. Reward and recognise improvements of FAIR practice.
6. Develop and monitor adequate policies for FAIR data and research objects.



LA SCIENCE OUVERTE: UN INSTRUMENT POUR LA SCIENCE!

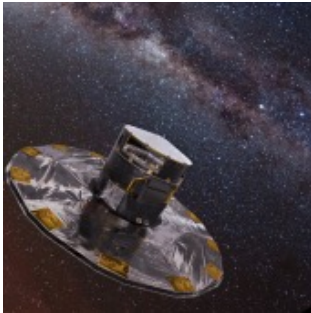
□ La science ouverte n'est pas seulement un objet politique!

- Une révolution dans la manière dont on fait de la science
Découvrir les données, y accéder, les réutiliser, les combiner
- Toutes les disciplines sont concernées
- Mais elles ne sont pas toutes au même niveau sur le sujet
- La plupart travaillent maintenant sur le sujet – l'astronomie a été pionnière

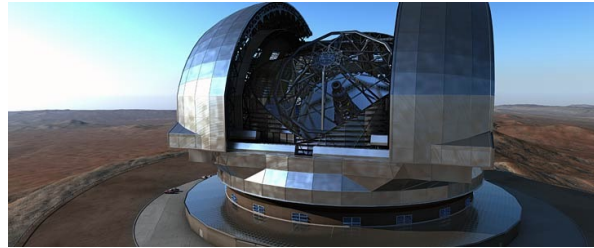
□ On met souvent en avant les utilisations interdisciplinaires des données...

- ... en particulier les Objectifs du Développement Durable
- Mais un usage interdisciplinaire des données repose sur des piliers disciplinaires solides
- C'est au niveau disciplinaire qu'on connaît les données et qu'on peut les rendre FAIR

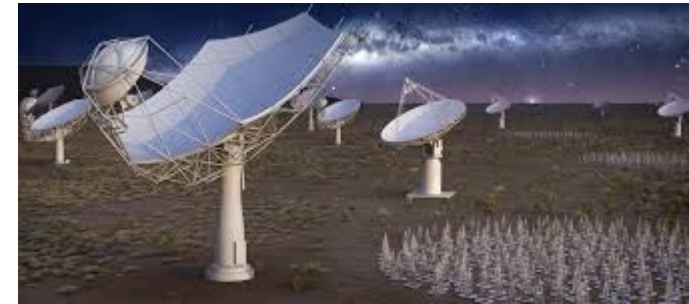
□ Les Infrastructures de Recherche en astronomie



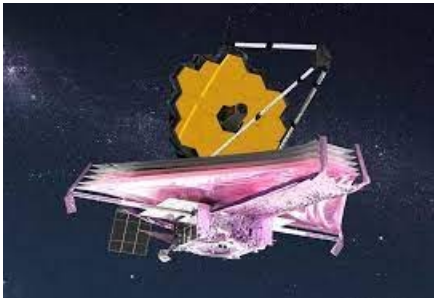
Gaia



ELT



SKA



JWST



Planck



CTA

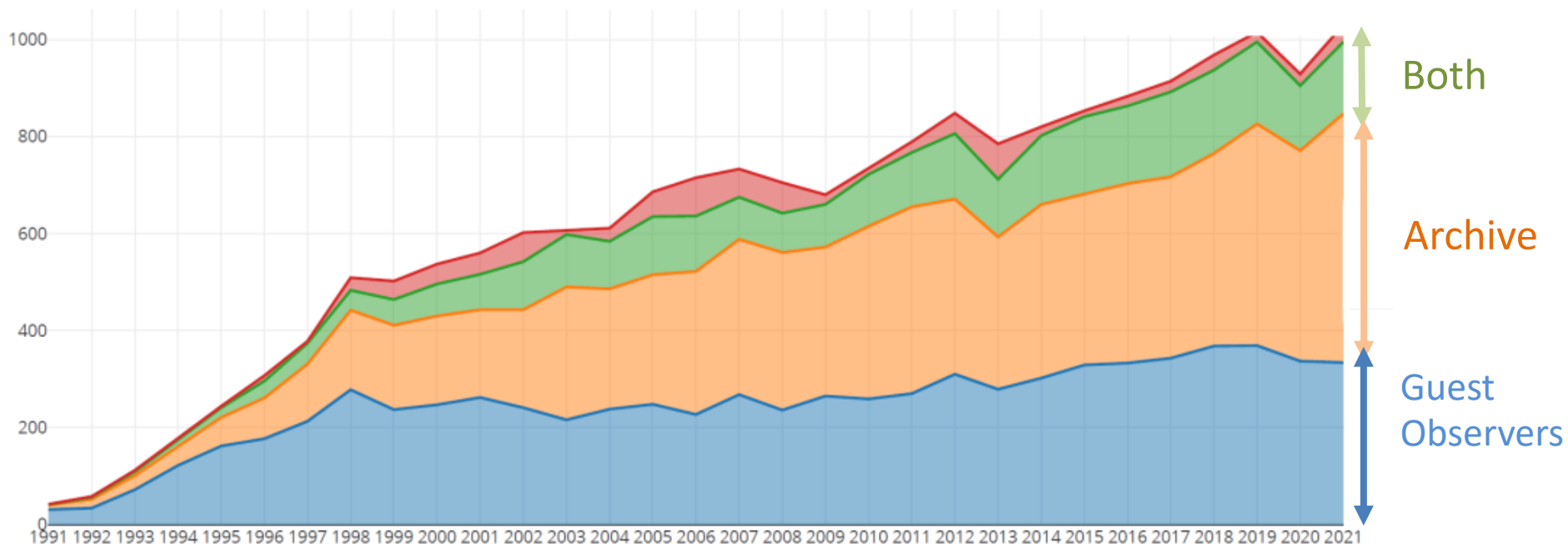
+ données

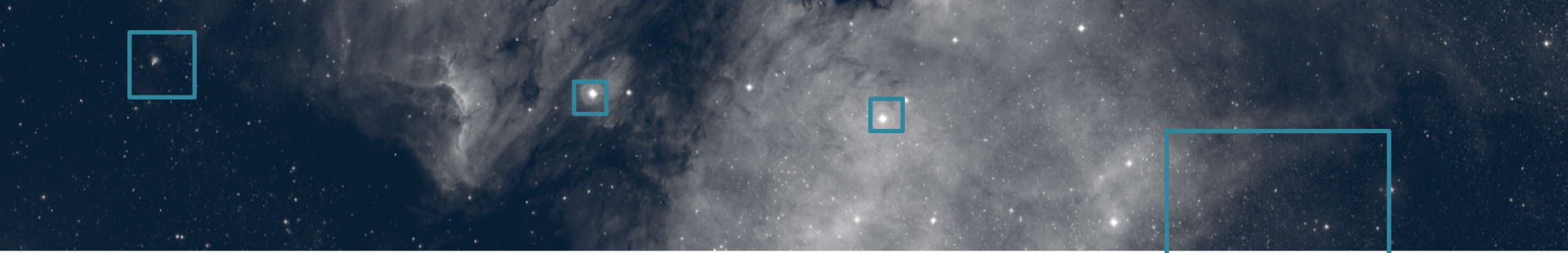
□ Les standards jouent un rôle essentiel

- Standards définis au niveau international
- Point de départ: le format FITS (1981)
 - Inclut les données et les métadonnées: les données sont **Réutilisables**
- Les standards pour Trouver (**F**) les données, y **Accéder** et les rendre **Interopérables** sont développés par l'International Virtual Observatory Alliance (IVOA) depuis 2002
 - Une couche fine au-dessus des données pour les rendre F, A et I
 - Les données et les outils pour les utiliser sont interopérables
- Vu par les utilisateurs: une infrastructure de données unique, mondiale et multipolaire
- Une infrastructure ouverte: tout le monde peut publier un service ou développer un outil
- L'infrastructure permet un accès transparent aux données et est invisible des utilisateurs
- Utilisation massive des données en ligne par la communauté

Impact of open data for the infrastructures: Hubble Space Telescope, published papers

<https://archive.stsci.edu/hst/bibliography/pubstat.html>





CONCLUSIONS

□ Conclusions (1)

- La Science Ouverte est un sujet chaud, en évolution rapide
- Nous sommes tous concernés!
- Une pression politique forte
- Une révolution dans la manière de faire de la science
- Partage des données, des logiciels, des techniques, etc
- Un écosystème complet

□ Conclusions (2)

- Parmi les éléments essentiels
 - La confiance (qualité des données, confiance dans la chaîne de production et de partage)
 - Pertinence vis-à-vis des besoins des utilisateurs
- BEAUCOUP de travail sous-jacent, invisible des utilisateurs
 - Chercheurs, curateurs des données, informaticiens
 - Accord international sur les standards
 - Essentiel, mais nécessite du temps et des efforts

□ Liens utiles

Documents « politiques »

OECD documents

OECD Principles and Guidelines for Access to Research Data from Public Funding, 2007, <https://doi.org/10.1787/9789264034020-en-fr>
Enhanced Access to Publicly Funded Data for Science, Technology and Innovation, 2020, link to read the document on line:
<https://read.oecd.org/10.1787/947717bc-en?format=read#page1>

G7/G8 Science Ministers

2013: <https://www.gov.uk/government/publications/g8-science-ministers-statement-london-12-june-2013>

2016: <http://www.international.gc.ca/g7/assets/pdfs/g7-2016-tsukuba-communique.pdf>

European Commission

Open Science, Open Innovation, Open to the World, A vision for Europe, 2016, <https://data.europa.eu/doi/10.2777/061652>

ESFRI

ESFRI web site: <https://www.esfri.eu/>

ESFRI Roadmap 2021: <https://roadmap2021.esfri.eu/>

France

Plan National pour la Science Ouverte/National Plan for Open Science, 2018 [Version française](#) [English version](#)

Deuxième Plan National pour la Science Ouverte/Second National Plan for Open Science [Version française](#) [English version](#)

FAIR

Wilkinson, M., Dumontier, M., Aalbersberg, I. et al. The FAIR Guiding Principles for scientific data management and stewardship. Sci Data 3, 160018 (2016). <https://doi.org/10.1038/sdata.2016.18>

Turning FAIR into Reality, Final report and action plan from the European Commission expert group on FAIR data, 2018, <https://data.europa.eu/doi/10.2777/1524>

Six Recommendations for implementation of FAIR practice by the FAIR in practice task force of the European open science cloud FAIR working group, 2020, <https://data.europa.eu/doi/10.2777/986252>