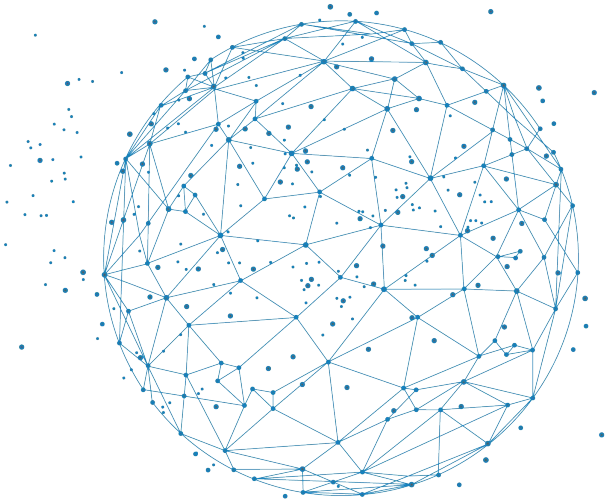
Gestion des données en Sciences de la Vie



FAIR DATA by IFB

ARTbio IBPS

migale

INSTITUT PASTEUR

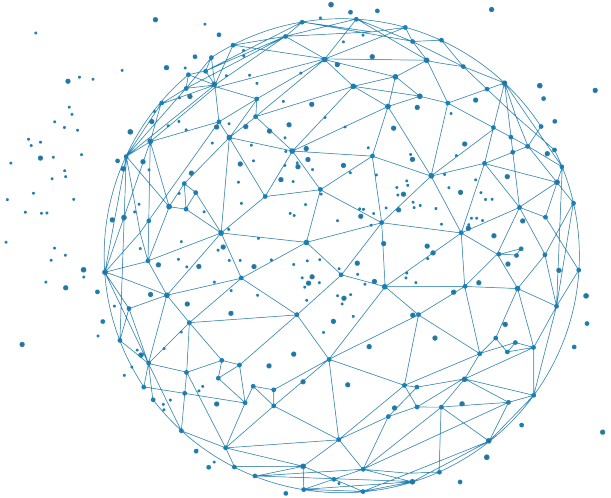
PLANT BIOINFORMATICS FACILITY

IFB INSTITUT FRANÇAIS DE BIOINFORMATIQUE

elixir FRANCE

cnrs INRAE Inserm cea

Introduction



- Montage de la formation par un Gdt IFB pendant le confinement en 2020

H. Chiapello T. Denecker J.-F. Dufayard P. Lieby L. Maurel G. Sarah J. Seiler F. de Lamotte

- 2 éditions “nationales” en distanciel synchrone en 2021 : mars et octobre

- 2 éditions “régionales” en présentiel en 2022 : édition “francilienne” - Sorbonne Univ - 22 et 23 mars + édition “Grand Est” - Strasbourg les 23 et 24 mars

FAIR data - Organisation par les plateformes franciliennes de l'IFB, 22 et 23 mars 2022.

- ARTbio


Christophe Antoniewski


Naïra Naouar


Léa Bellenger (helper)

- IFB


Hélène Chiapello


Thomas Denecker


Frédéric de Lamotte

- Institut Pasteur Paris


Anne-Caroline Delétoille


Fanny Sébire


Lucie Lamothe (helper)

- Migale


Valentin Loux


Cédric Midoux

- PlantBioinfoPF


Célia Michotey


Cyril Pommier

- Horaires : 9h30 -12h30 & 13h30 - 17h30 les mardi et mercredi, 1 pause vers 10h15 et vers 15h15
- En présentiel
- … donc interactif ! Nous avons planifié
  - Beaucoup de (courtes) séances pratiques avec des outils dédiés comme :
    - Scrumblr http://scrumblr.ca
    - Mentimeter https://www.menti.com
  - Des séquences d'échanges : n'hésitez pas à partager vos expériences sur le sujet
- Espace Slack
  - Inscrivez vous à slack.com
  - Tutoriel de connexion
  - Rejoignez le canal **#fairdata-paris-mars2022**
-

Objectifs pédagogiques

Après cette formation, vous connaîtrez

- les différents points fondamentaux (théoriques, pratiques, juridiques) en lien avec la politique nationale d'ouverture des données de la recherche
- les ressources nationales et internationales accessibles à la communauté scientifique ainsi que les solutions proposées par l'IFB pour gérer les données d'un projet de recherche
- les principales notions et ressources concernant les métadonnées en biologie
- les outils et principes permettant de rédiger un Plan de Gestion de Données (PGD)

Contenu de la formation : 1 module par ½ journée

- **Module 1 : Les données de la Recherche et leur centralité dans le processus de recherche**

Crise de reproductibilité - Vers FAIR - Cycle de vie des données - Le plan de gestion des données

- **Module 2 : La vie des données pendant le projet : guide de bonnes pratiques**

Introduction - Le nommage des fichiers - Format de fichier - Stockage et accès -Outils et solutions

- **Module 3 : Les Métadonnées : les standards en sciences de la Vie et retour d'expérience sur la soumission dans des entrepôts internationaux**

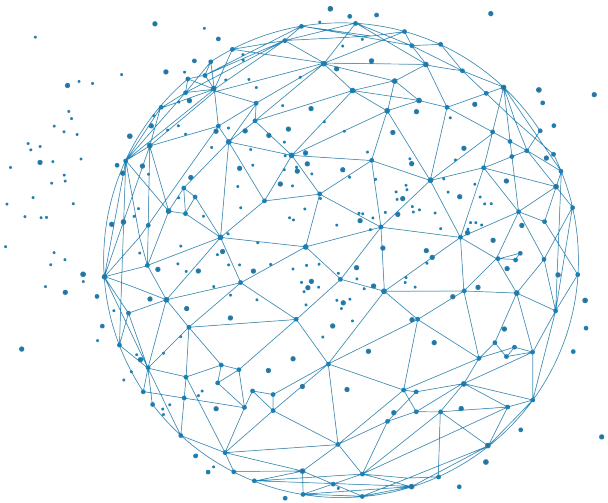
Introduction aux métadonnées - Choix des métadonnées en sciences de la Vie - Retour d'expérience sur la soumission dans les dépôts internationaux

- **Module 4 : Partage et valorisation des données**

Cadre Juridique - Les licences sur les données - Quel entrepôt pour quelles données ?

 Supports ici : https://moodle.france-bioinformatique.fr/course/view.php?id=8

- Chacun se présente rapidement :
  - Nom, prénom, affiliation professionnelle
  - Vos activités professionnelles, votre lien aux données de la Recherche
  - Vos attentes vis à vis de la formation
  - Décide de la proposition qu'on lui soumet



Module 1 : Les données de la Recherche

ARTbio IBPS

migale

INSTITUT PASTEUR

PLANT BIOINFORMATICS FACILITY

IFB INSTITUT FRANÇAIS DE BIOINFORMATIQUE

elixir FRANCE

cnrs

INRAE

Inserm

cea

INVESTIR L'AVENIR

Les intervenants du module 1

Christophe Antoniewski – https://orcid.org/0000-0001-7709-2116

Célia Michotey – https://orcid.org/0000-0003-1877-1703

Naïra Naouar – https://orcid.org/0000-0003-2161-8603

Fanny Sébire – https://orcid.org/0000-0002-6301-7147

Merci à :

Frédéric de Lamotte - https://orcid.org/0000-0003-4234-1172

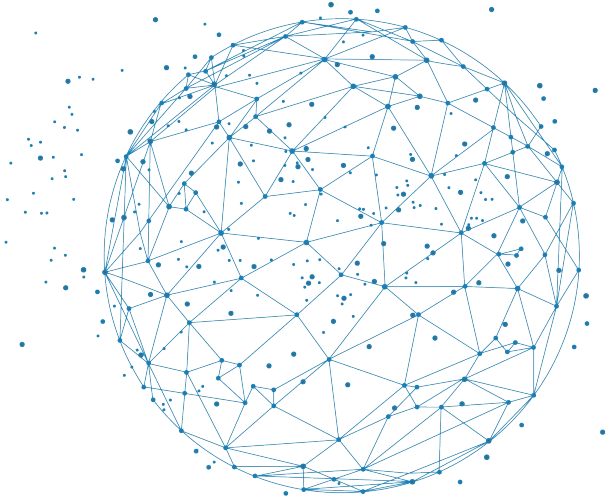
Jean-François Dufayard - https://orcid.org/0000-0002-7427-6822

Paulette Lieby - https://orcid.org/0000-0002-9289-9652

https://www.youtube.com/watch?v=YcuE54E9coI

FAIR pourquoi faire ?

Une réponse aux crises des sciences biomédicales

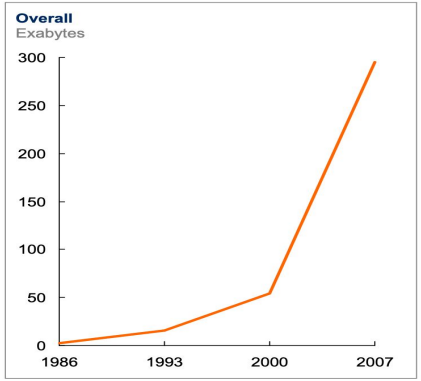


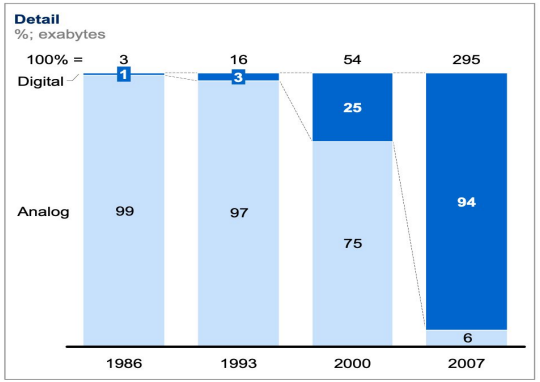
La disruption numérique, une bascule brutale

1 000 000 000 000 000 000 000 bytes (octets) 1 exaoctet = 10^{18} octets, 1 zetaoctet = 10^{21} octets

eo po to go mo ko

250 millions de millions **160** millions de millions **20** millions de milliards **1.4** millions de millions **294** milliards



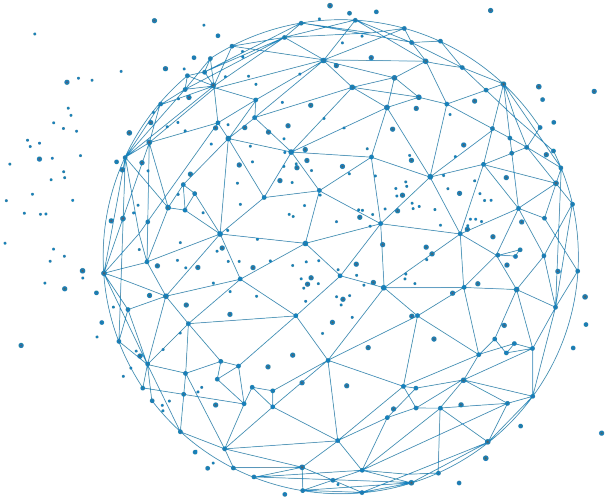


The World's Technological Capacity to Store, Communicate, and Compute Information. Science. 2011;332.

… Qui bouscule l'existant

- La première compagnie de taxi n'en possède aucun (Uber)
- Le premier fournisseur de logement n'en possède pas (AirBnB)
- La première compagnie de téléphonie ne possède pas de standard (Skype)
- Le premier fournisseur d'info ne crée pas de contenu (Facebook)
- Le premier diffuseur de film ne possède pas de salle de cinéma (Netflix)

Les technologies numériques ont transformé la biologie et médecine



Dites nous tout sur http://scrumblr.ca/disruption !

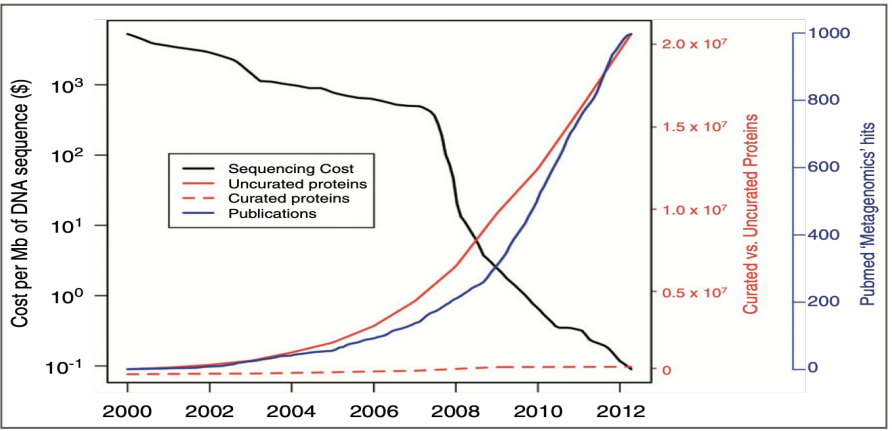
Le déluge des données en Science

Les techniques à haut débit induisent une chute des coûts et une explosion de la production de données

Génome humain :

en 1990 = **13 ans** et **3 Milliards** $

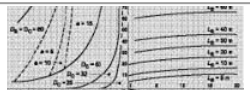
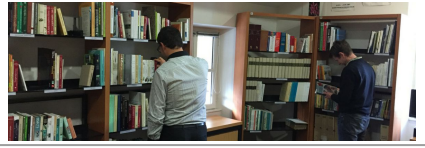
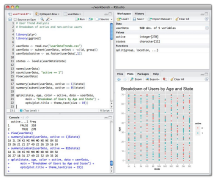
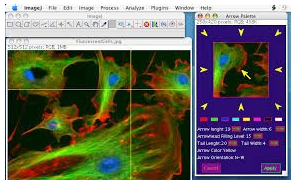
en 2015 = **quelques heures** et **1000** $



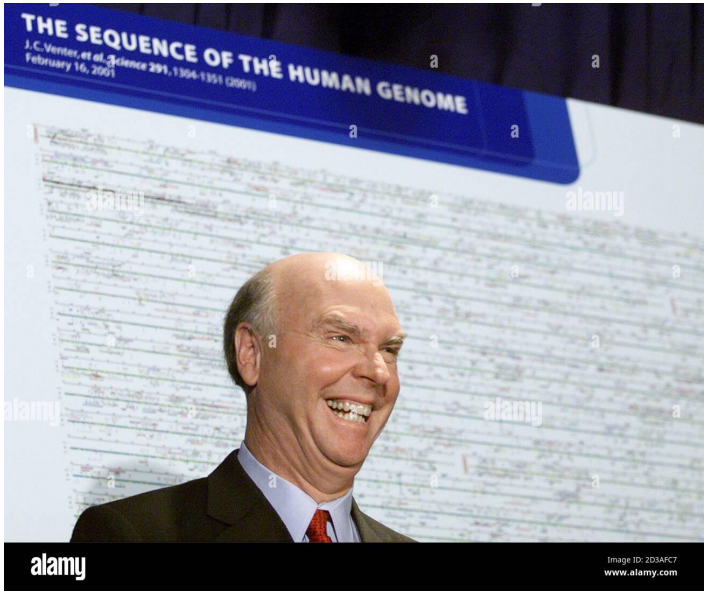
→ La quantité de données à **stocker** et **analyser** explose

→ Le ***rendement*** d'analyse chute

Un changement d'échelle qui impacte profondément la science

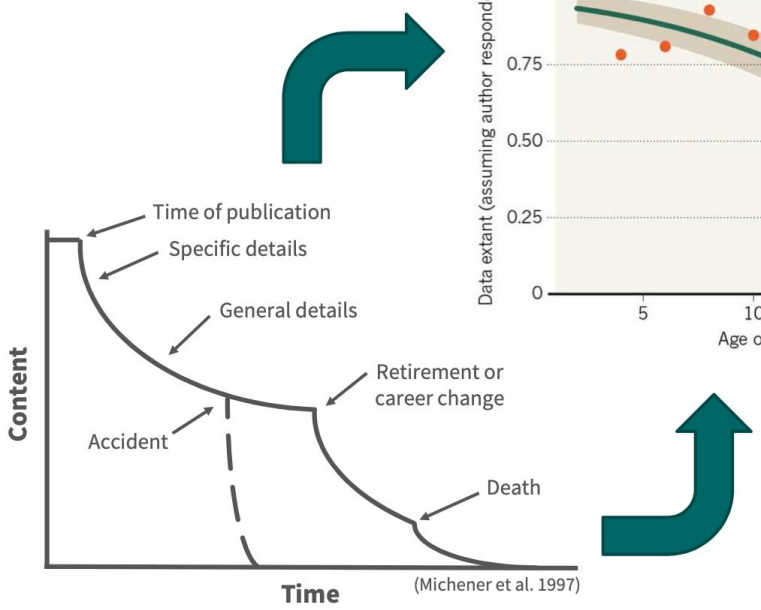
Les étapes	AVANT	MAINTENANT
Concevoir l'expérimentation	Une connaissance des dispositifs expérimentaux existants accessible à un où quelques individus. Un volume d'observations attendues à taille humaine	Une matrice de technologies qui échappe à un expérimentateur individuel Possibilité d'utiliser comme de générer une quantité massive de données expérimentales
Collecter des résultats		
Analyser des résultats	   	 
Diffuser le savoir		

Un changement d’échelle qui modifie la place de la science dans la société



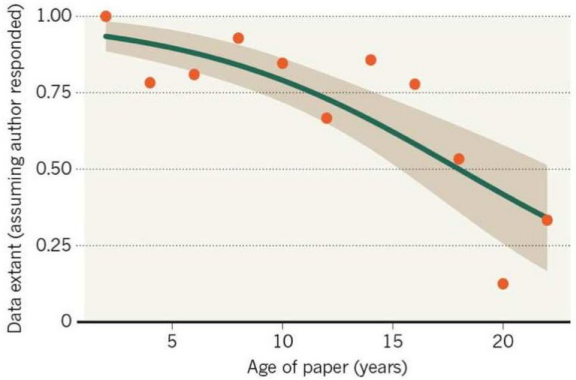
Les ravages du temps

Data Entropy



MISSING DATA

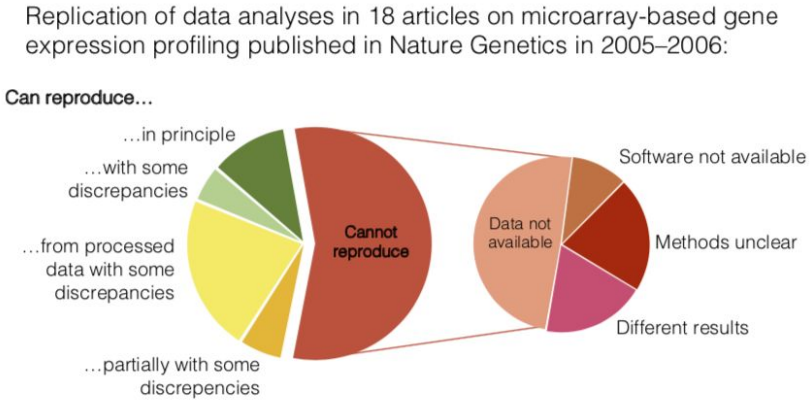
As research articles age, the odds of their raw data being extant drop dramatically.



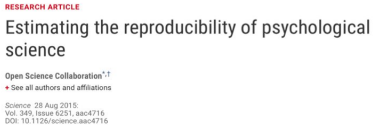
Vines, T. H. *et al.* Curr. Biol. http://dx.doi.org/10.1016/j.cub.2013.11.014 (2013).

3

La crise de reproductibilité

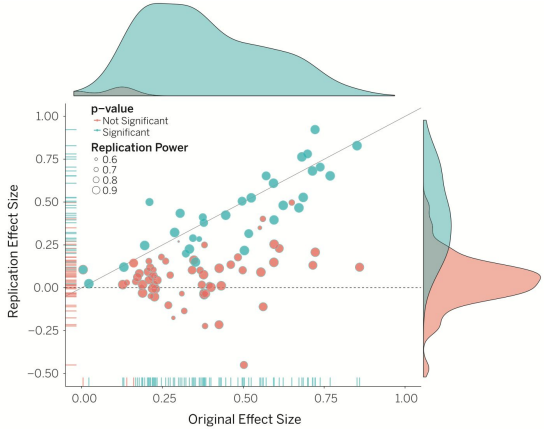


Repeatability of published microarray gene expression analyses. Nat Genet. **2009**;41: 149–155. doi:10.1038/ng.295



The *Reproducibility project* set out to replicate 100 experiments published in high-impact psychology journals.

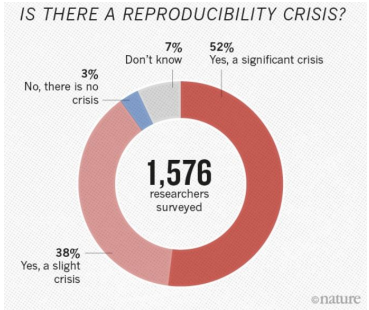
About one-half to two-thirds of the original findings could not be observed in the replication study.



Estimating the reproducibility of psychological science. Science. **2015**;349: aac4716. doi:10.1126/science.aac4716

Step	Reference																	
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
Mapping	■	■	■	■	■	■	■	■		■	■	■	■	■	■	■	■	■	■
Realignment		■								■	■					■	■		■
Recalibration				■						■	■						■		■
Initial variant detection	■	■	■	■		■	■	■		■	■			■		■	■	■	■

Next-generation sequencing data interpretation: enhancing reproducibility and accessibility. Nat Rev Genet. **2012**;13: 667–672.



Is there a reproducibility crisis in science? Nature. **2016**. doi:10.1038/d41586-019-00067-3

Les premières formalisations du FAIR

- Accept that the computational component is becoming an integral component of biomedical research
- Always provide access to primary data
- Record versions of all auxiliary data sets used during the analysis
- Note the exact versions of software used
- Record all parameters even if defaults are used
- Provide all custom scripts
- Do not reinvent the wheel

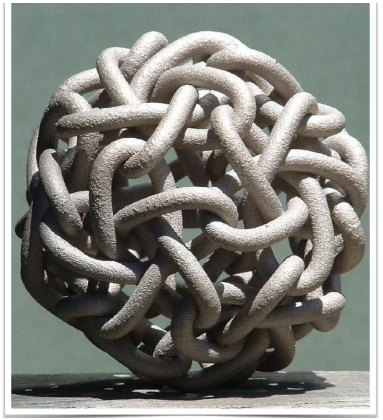
Anton Nekrutenko and James Taylor
Nat Rev Genet. **2012**;13: 667–672.

Box 3 | **Guidelines for reproducibility**

- Accept that the computational component is becoming an integral component of biomedical research. As the life sciences are becoming increasingly data-driven, there will be no escape from computation and data handling. Familiarize yourself with best practices of scientific computing using existing educational resources, such as the Software Carpentry project[41]. Implementing good computational practices in your group will automatically take care of many of the points listed below.
- Always provide access to primary data. It is obvious that without access to the original data sets, any claims made in a publication cannot be verified. In situations in which the data cannot be made public (for example, clinical data sets under Institutional Review Board protection), they should be deposited in controlled access repositories (such as dbGaP[42]), where they can be retrieved by authorized users. One potential issue with this point is the fact that there is currently a debate on what constitutes primary data. Storing images generated by some next-generation sequencing (NGS) machines on a large scale has long been unfeasible. Public sequencing archives, such as those at the National Center for Biotechnology Information (NCBI) and the European Bioinformatics Institute (EBI), are still accepting sequencing reads as submissions and should be used. Going forward, other formats, such as aligned data in BAM format, are likely to be used (as is already done by the 1000 Genomes Project).
- Record versions of all auxiliary data sets used during the analysis. For example, in most NGS analyses, such as variant discovery detailed here, sequencing reads are compared against a reference genome. It is crucial to record which reference genome was used because, just as software has versions and cars have model years, genomes have build identifiers. For example, the latest human genome build distributed by the UCSC Genome Browser is called hg19 (it is derived from the GRC37 build prepared by the Genome Reference Consortium) and has the highest number of functional annotations (7,330 annotation types) and should be the preferred version to be used. Note that the latest version may not always be the best choice. The latest mouse genome build (mm10) has only a fraction of annotations (258 tracks) compared with its predecessor (mm9, which has 2,096 tracks). Thus, it would be easier to interpret results of an NGS experiment mapped to the mm9 build even though mm10 has an additional 48 megabases of actual sequence.
- Note the exact versions of software used. Different versions of the same software often produce different results, and important bug fixes may have implications to results produced with a particular version of a tool.
- Record all parameters, even if defaults are used. Although the reason to record all parameters requires no explanation, we emphasize the importance of explaining default settings for reproducibility. A clause 'software was used with default settings' is found in many publications. However, the meaning of default settings often changes between versions of software and can be quite difficult to track down when a substantial amount of time has passed since publication. Thus, record what the default settings actually are.
- Provide all custom scripts. With the complexity of NGS analysis, it is often unavoidable to create simple scripts that carry out such straightforward tasks as, for example, changing data formats. Such scripts must be made accessible as any other part of the analysis.
- Do not reinvent the wheel. It pays to reuse existing software. Integrative frameworks and associated application stores already house hundreds of tools (for example, as of May 2012, Galaxy ToolShed contains ~1,700 tools). It is likely that a script for a particular problem has been already written. Ask around through existing resources such as SEQanswers[43] and BioStar[44].

Les racines du mouvement “#OpenScience”

- La crise de reproductibilité
- La crise de déontologie scientifique (P-Hacking)
  P-hacking in clinical trials and how incentives shape the distribution of results across phases.
  Proc Natl Acad Sci U S A. **2020**;117: 13386–13392.)
- La crise politique (éthique/démocratique)
  (indicateurs-ego metrics, producteurs de savoir biomedical privé, marché de l’édition scientifique)
- La crise de la “bioinformatique”

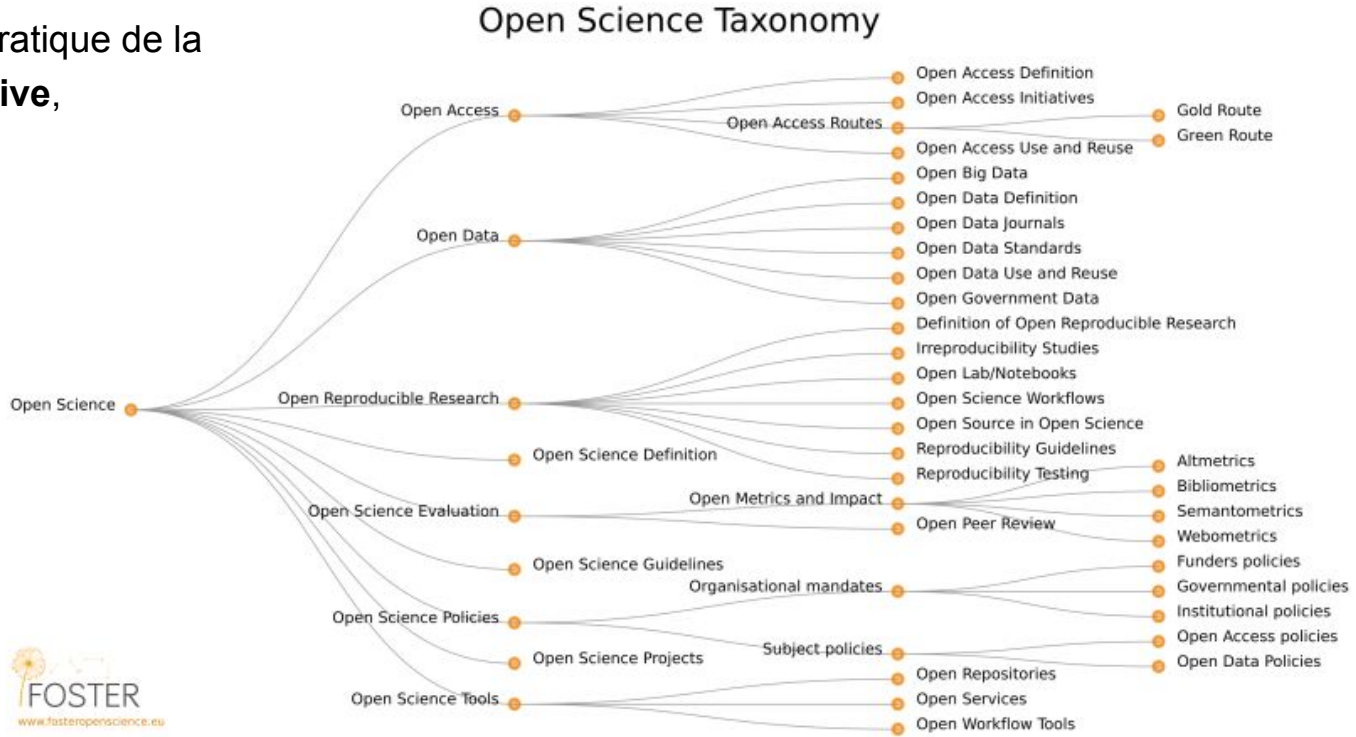


→Améliorer la science et l’innovation

- **démocratiser l’accès aux savoirs**
- **rendre la science plus cumulative**, plus fortement étayée par les données, plus transparente
- **augmenter l’efficacité de la recherche** en évitant de dupliquer les efforts, en réutilisant des données ou du matériel scientifique
- favoriser la **confiance des citoyens dans la science**

La Science Ouverte

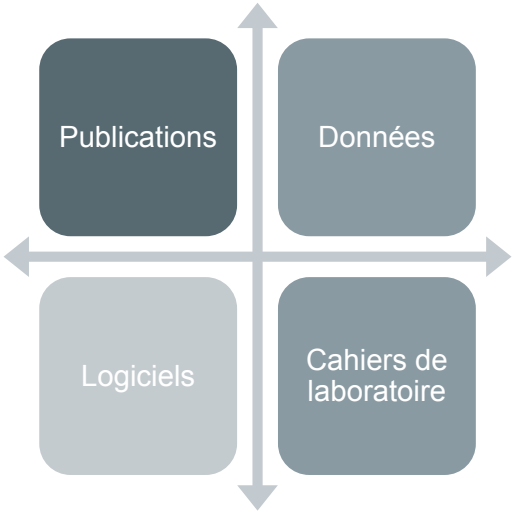
Nouvelle forme de pratique de la science : **collaborative**, **participative et interdisciplinaire**



Source : Knoth, Petr; Pontika, Nancy (2015): Open Science Taxonomy. figshare. Figure. (https://doi.org/10.6084/m9.figshare.1508606.v3)

FAIR et Science Ouverte

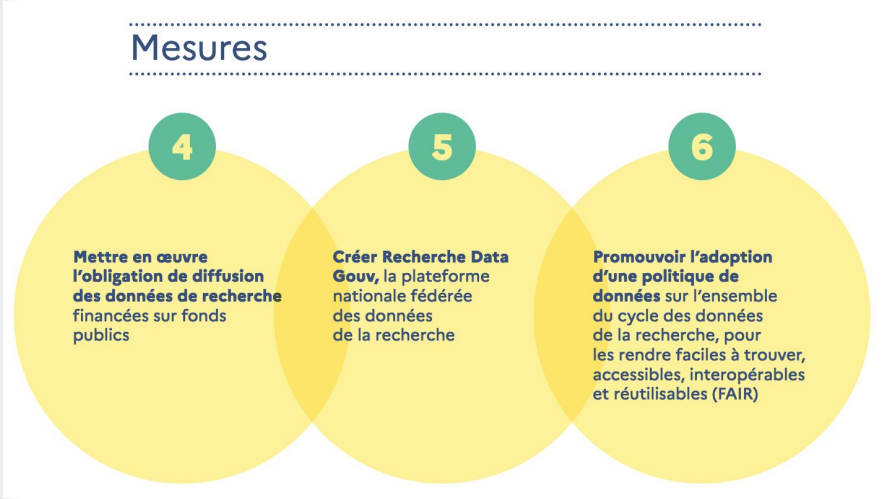
- Rendre les **résultats de la recherche scientifique** accessibles à tous (un des aspects de la Science Ouverte)
- En permettant une diffusion et une réutilisation sans entrave des données et résultats de la recherche



Contexte national

2018

2021



Citez les types de données de la recherche que vous connaissez

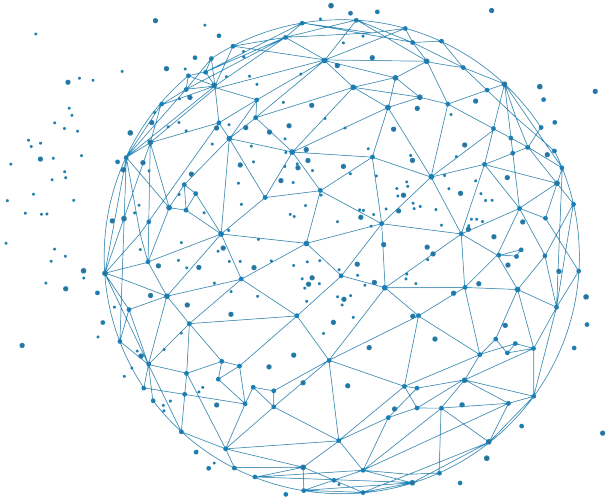
En utilisant le tableau à Post-It disponible ici http://scrumblr.ca/typesdedonnees

Définition OCDE

L'Organisation de coopération et de développement économiques (OCDE) est une organisation internationale qui œuvre pour la mise en place de politiques meilleures pour **une vie meilleure**. Notre objectif est de promouvoir des politiques publiques qui favorisent la prospérité, l'égalité des chances et le bien-être pour tous.

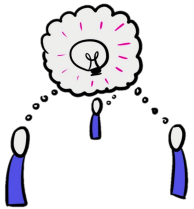
- Les données de recherche sont les **preuves** qui sous-tendent la réponse à la question de recherche et peuvent être utilisées pour **valider** les **résultats**, quelle que soit leur forme (i.e. imprimée, numérique ou physique).
- Il peut s'agir de **renseignements quantitatifs** ou d'énoncés **qualitatifs** recueillis par les chercheurs dans le cadre de leurs travaux par **expérimentation, observation, modélisation, entrevue** ou autres méthodes, ou de renseignements tirés de preuves existantes.
- Les données peuvent être **brutes** ou **primaires** (par exemple, directement issues de mesures ou de collectes) ou **dérivées** de données primaires par analyse ou interprétation (e.g. nettoyées ou extraites d'un ensemble de données plus vaste), ou encore dérivées de sources existantes dont les droits peuvent être détenus par d'autres.

Vers le FAIR



Réutiliser les données ? Où est le problème ?

https://youtu.be/66oNv_DJuPc



Notez les points marquants (bon ou mauvais) en gestion des données

Comment rendre des données réutilisables ?

https://www.menti.com/xtqhkq1kf4

- Notez 5 conditions nécessaires
  - 1 seul mot à la fois
  - en français
  - sans majuscule

- A partir du nuage de mots créé collectivement quels regroupements pouvons-nous faire ?

Les principes FAIR

Les principes FAIR sont un ensemble de lignes directrices visant à rendre les données **Facilement trouvables, Accessibles, Interopérables et Réutilisables**.

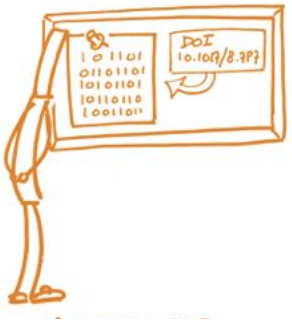
- Ils fournissent des orientations pour la gestion des données scientifiques et sont pertinents pour toutes les parties prenantes de l'écosystème numérique
- Ils s'adressent directement aux producteurs et aux éditeurs de données afin de promouvoir une utilisation maximale des données de recherche
- Ils mettent l'accent sur la capacité des machines à gérer des données de façon automatique, avec le minimum d'interventions humaines

<u>Références</u> :

- Wilkinson 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
- Les principes FAIR. DORANum. http://doi.org/10.13143/z7s6-ed26

Faciliter la découverte des données (et de leurs métadonnées)
tant pour les humains que pour les machines

- Les données ont un **PID** (Persistent Identifier, identifiant unique et pérenne)
- Les données sont décrites par des **métadonnées**
- Ces métadonnées incluent le PID des données qu'elles décrivent
- Les données sont déposées dans un **entrepôt de données**



Permettre l'accès aux données et leur téléchargement,

ce qui peut inclure l'authentification et l'autorisation.

- Les données sont accessibles à travers un **protocole de communication standard**
- Ce protocole est **libre et ouvert**
- Ce protocole permet un accès par **authentification** si besoin
- Les **métadonnées restent accessibles** même si les données ne le sont pas (disparues ou inaccessibles)

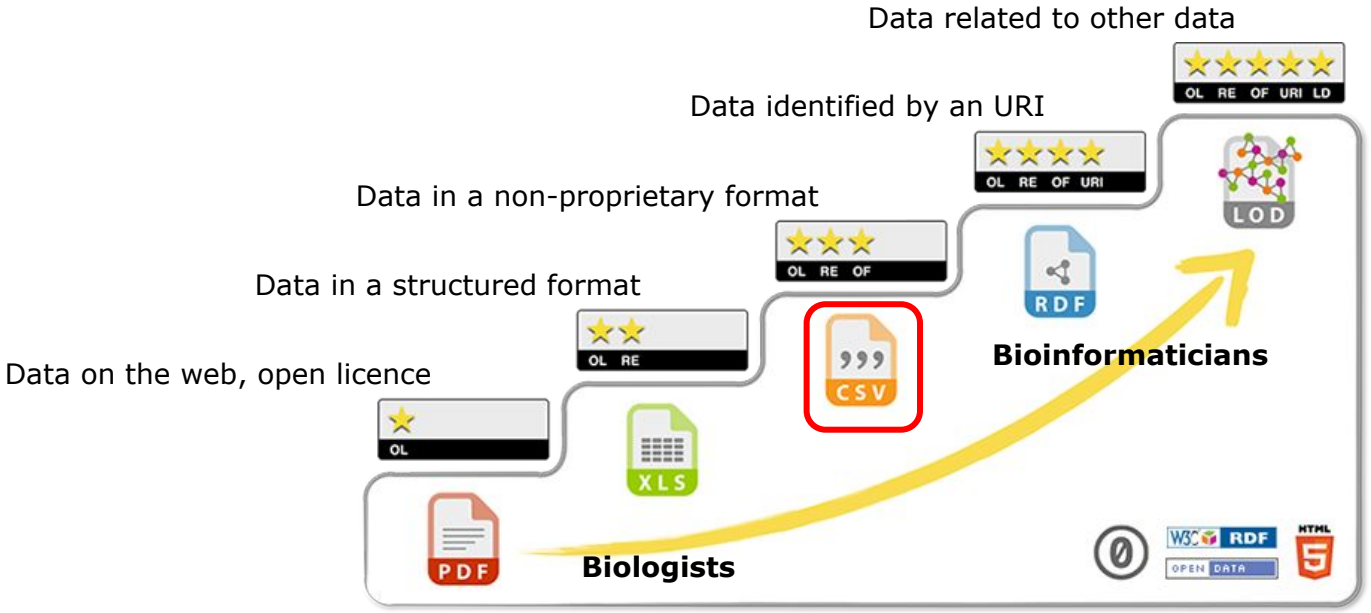
Permettre l'exploitation et l’intégration des données
quel que soit l'environnement informatique utilisé

- Les données sont **décrites avec un vocabulaire contrôlé**
- Le vocabulaire utilisé **respecte les principes FAIR**
- Les **métadonnées sont contextualisées** avec des liens vers d'autres données

Permettre la réutilisation des données pour de futures recherches

- Les métadonnées contiennent toutes les informations qui peuvent être utiles (**pluralité d'attributs**)
- Une **licence de réutilisation** est attribuée aux données
- La description des données indique leur **provenance**
- Le partage des données suit les **standards de la communauté scientifique**

5 * Open Data

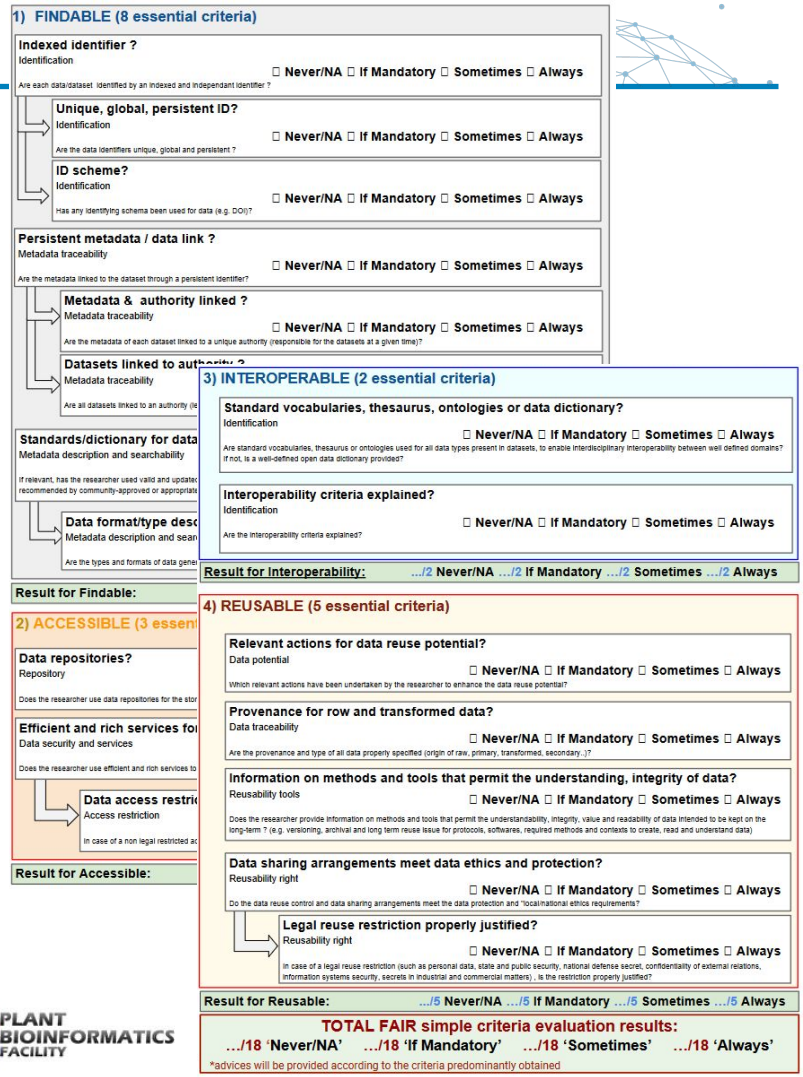


La progression vers FAIR et l'Open Data nécessite une coopération multidisciplinaire :

- Biologistes
- Bioinformaticiens
- Spécialistes des ontologies/sémantiques

Evaluation de la FAIRness des données

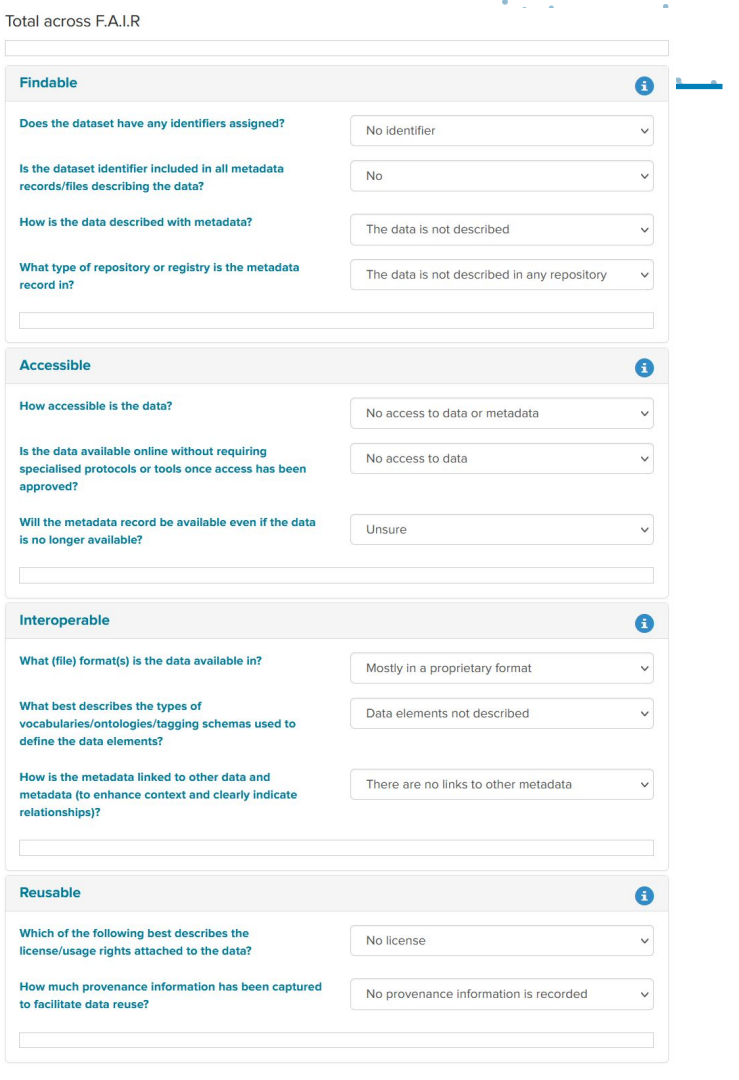
- Grille SHARC (SHAring Reward & Credit)
  - RDA (Research Data Alliance)
  - Grille simplifiée : https://zenodo.org/record/2551500#.X4hC_-2re70
  - Grille complète : https://zenodo.org/record/3922069



Evaluation de la FAIRness des données

- Grille SHARC (SHAring Reward & Credit)
  - RDA (Research Data Alliance)
  - https://zenodo.org/record/2551500#.X4hC_-2re70

- FAIR self assessment tool
  - ARDC (Australian Research Data Commons)
  - https://ardc.edu.au/resources/aboutdata/fair-data/fair-self-assessment-tool/

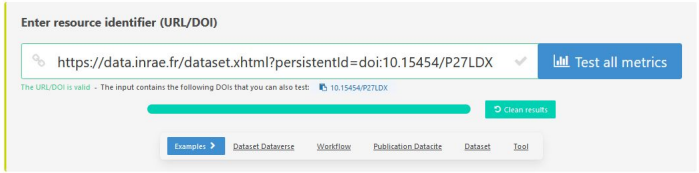


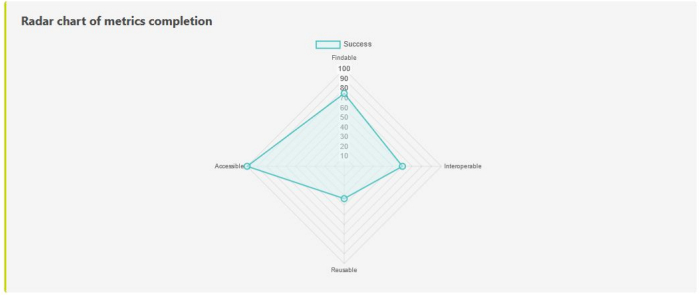
Evaluation de la FAIRness des données

- Grille SHARC (SHAring Reward & Credit)
  - RDA (Research Data Alliance)
  - https://zenodo.org/record/2551500#.X4hC_-2re70

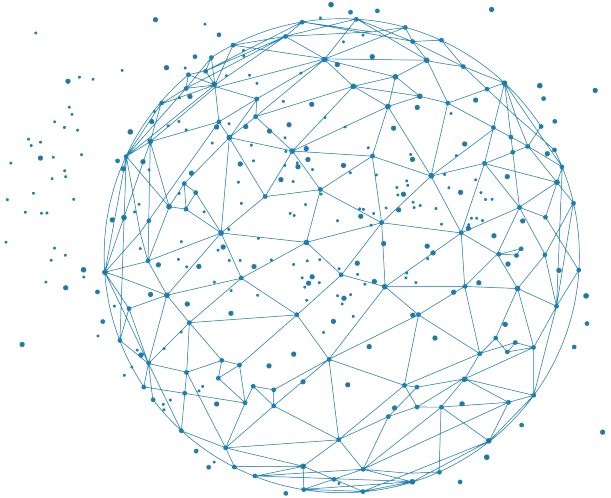
- FAIR self assessment tool
  - ARDC (Australian Research Data Commons)
  - https://ardc.edu.au/resources/aboutdata/fair-data/fair-self-assessment-tool/

- FAIR checker
  - IFB
  - https://fair-checker.france-bioinformatique.fr/base_metrics



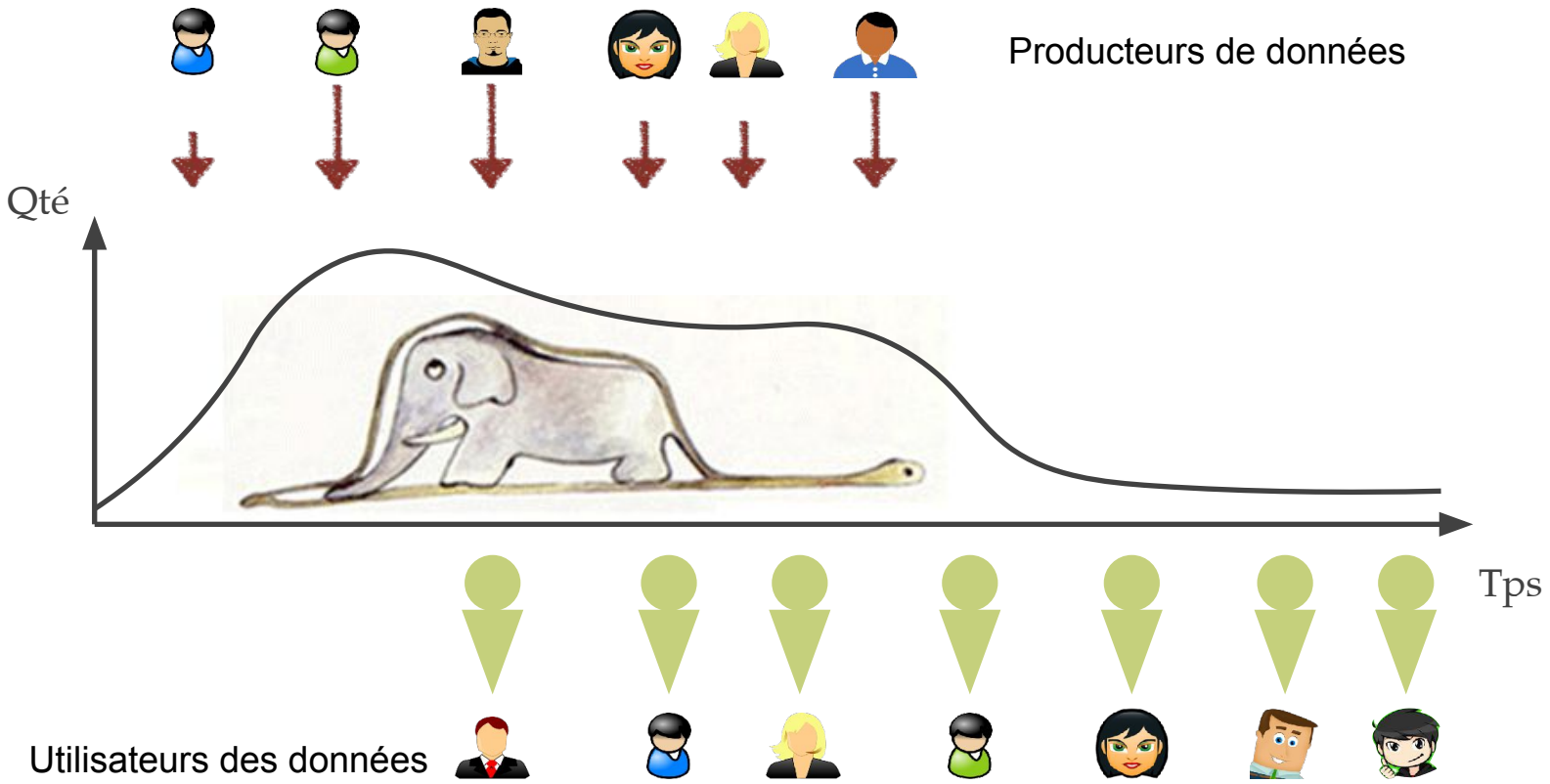


Cycle de vie des données

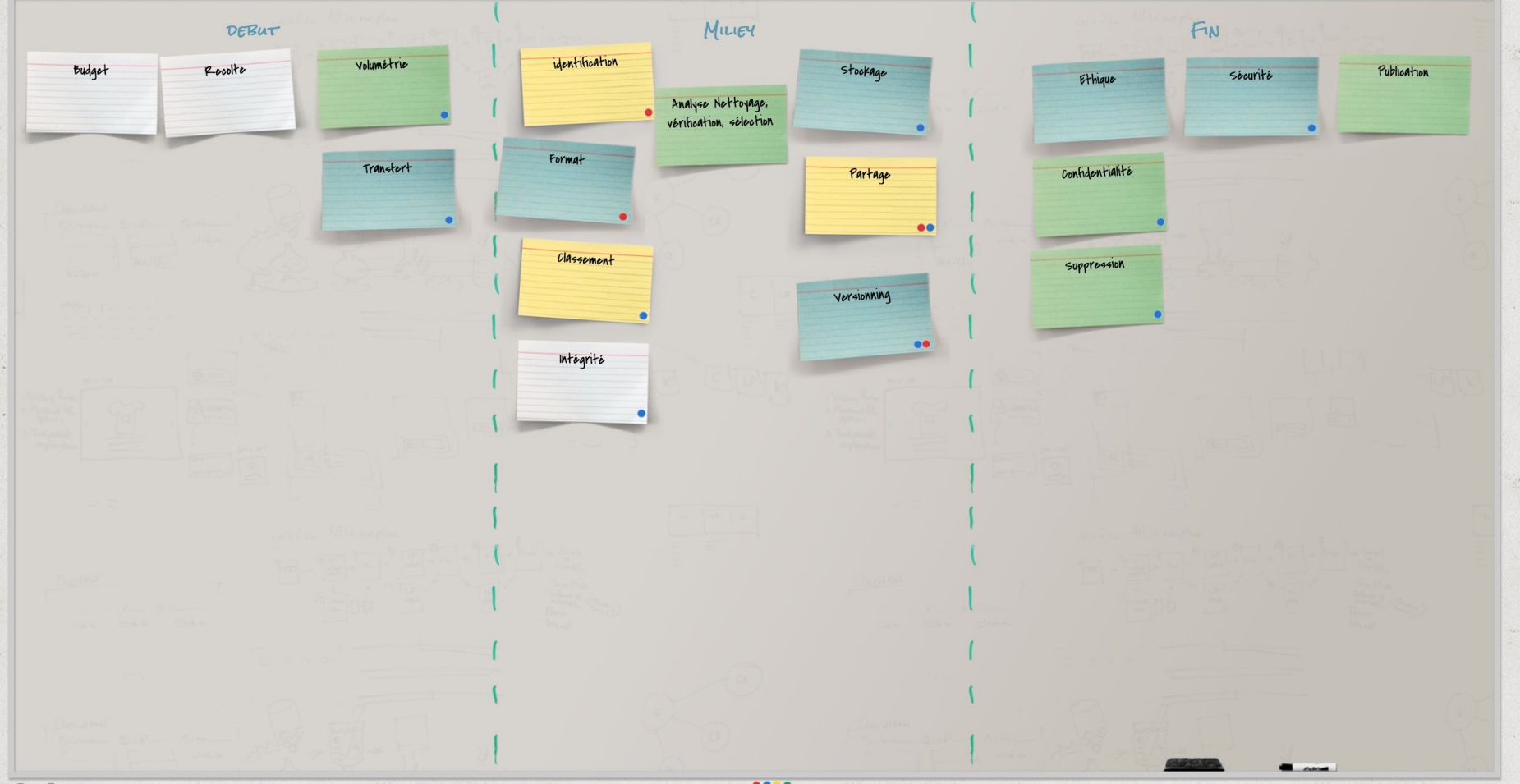


- Plusieurs temporalités
    - Le temps d'une thèse
    - Le temps d'un projet de recherche
    - Le temps de vie de la thématique dans le labo
    - Le temps de vie de la thématique dans l'institution
    - Le temps de vie de la thématique …

Un projet sur la durée



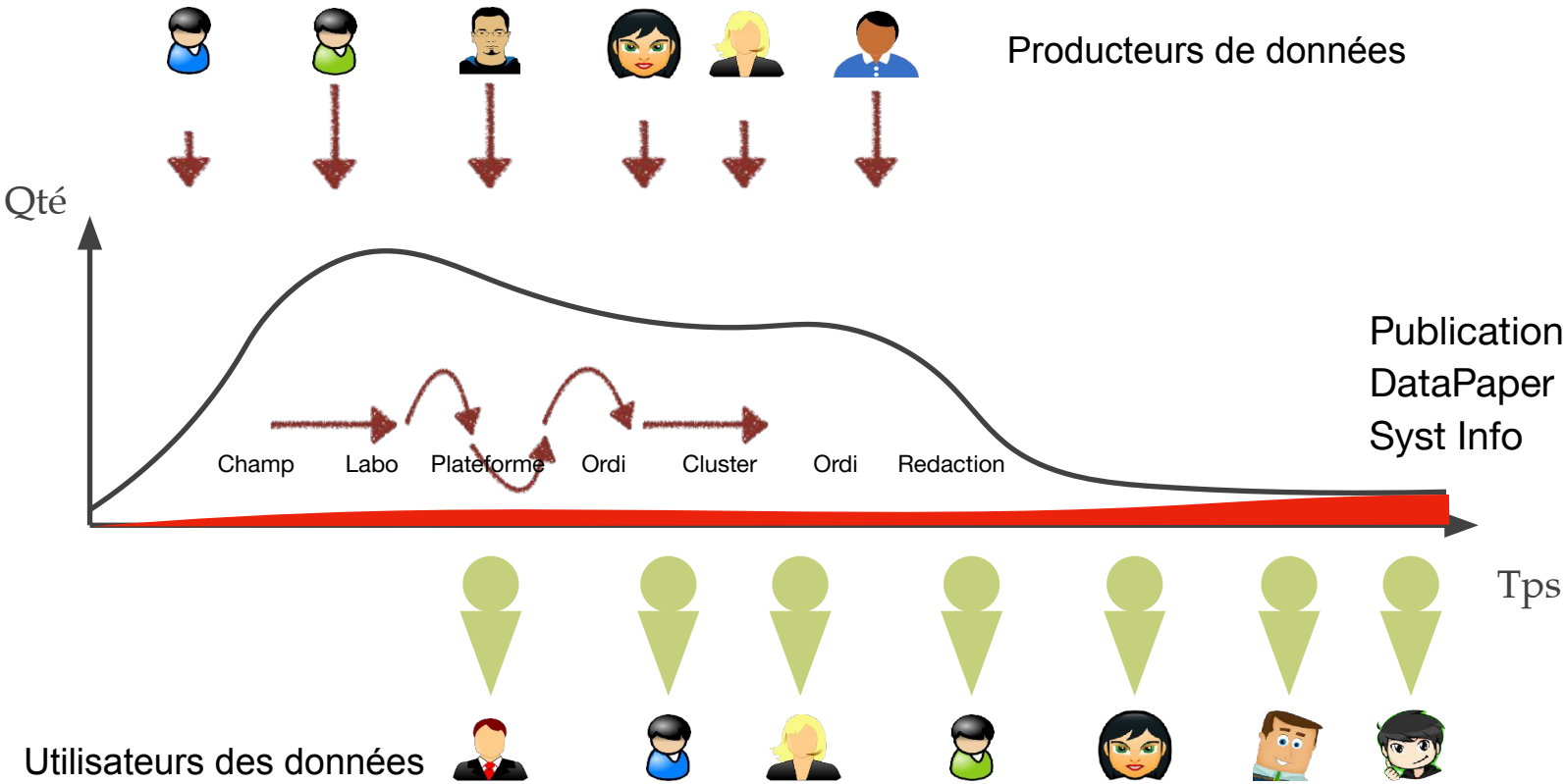
Connectez vous sur le scrumblr : http://scrumblr.ca/cycledata
Rédigez et positionnez des post-it concernant tous les points d'attention à avoir le long d'un projet, de sa conception jusqu'à sa valorisation



connected:

Fred (you)

Un projet sur la durée

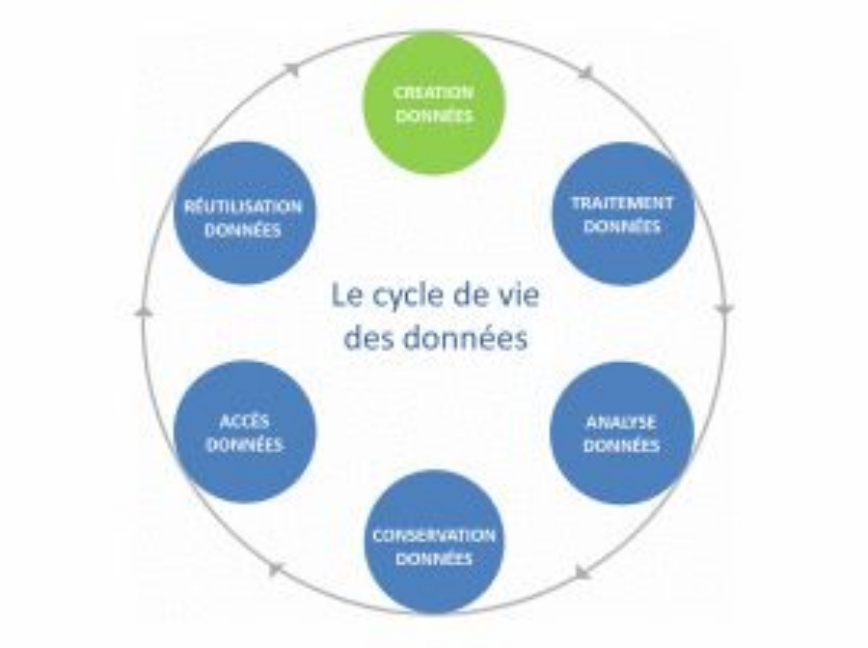


Le modèle de UK Data Archive définit les six étapes suivantes :

- **Création ou collecte** des données (creating data) ;
- **Traitement** des données (processing data) ;
- **Analyse** des données (analysing data) ;
- **Conservation** des données (preserving data) ;
- **Accès** aux données (giving access to data / data discovery) ;
- **Réutilisation** des données (reusing data).

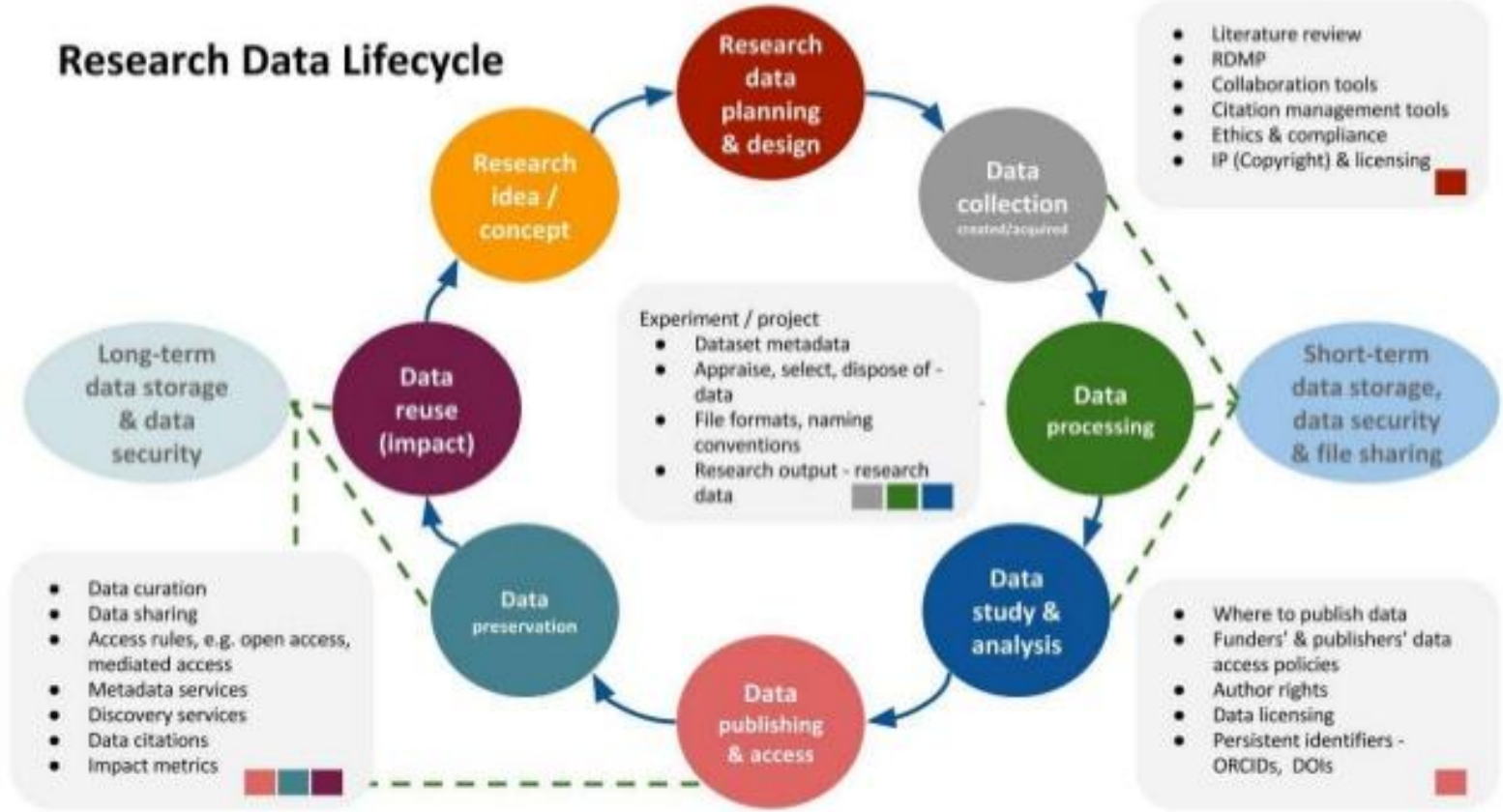
Une introduction à la gestion et au partage des données de la recherche - Le cycle de vie des données

Les étapes



[Une introduction à la gestion et au partage des données de la recherche - Le cycle de vie des données]

Une vue plus détaillée

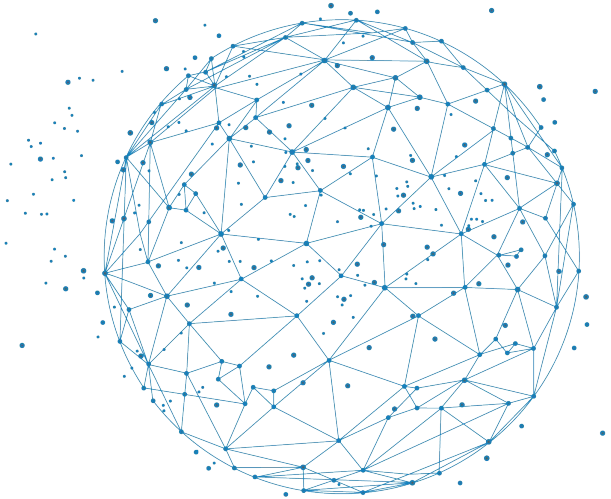


Donc, dans la “vraie vie“, gérer quoi ?

- **Le passé**
  - Le leg (du doctorant précédent …)
  - La biblio à T0
  - Les méthodes pré existantes
- **Le présent**
  - Les manipes
  - La création de connaissance (méthodes, posters …)
- **Le futur**
  - Le manuscrit
  - Les publications

- **Des échantillons**
  - dans les frigos
  - dans les tiroirs
- **Des fichiers**
  - des petits, des gros
  - un peu partout (PC, cloud, cluster)
  - des données brutes, du code, des résultats
- **De la connaissance**
  - des méthodes, du code
  - des systèmes d’information
  - des publications

Plan de Gestion de Données



PGD de projet

Document qui définit comment seront gérées les données d'un projet **pendant et après le projet**

- La difficulté : penser à traiter toutes les étapes du cycle de vie des données
- L'avantage : les modèles vous aident à penser à tout, en vous posant une série de questions

QUAND

Document évolutif = au moins 3 versions :

- Première version au début du projet de recherche
- Mises à jour régulières au cours du projet : versions intermédiaires
- Version finale à la fin du projet

QUI

Rédaction par l'équipe de recherche coordinatrice du projet

Modèle de PGD = une liste de questions à remplir pour rédiger un PGD

Modèles existants :

- Modèles des financeurs : ANR, Commission Européenne…
- Modèle Science Europe - https://doi.org/10.5281/zenodo.4915862
- Modèles institutionnels, des centres de calcul/stockage… (INRAe, CEA, Institut Pasteur, IN2P3, Unviversités, Grandes écoles…)

Peu / pas de modèles spécifiques à des types de données

Pour quoi faire ?

- **Plan** : on planifie (donc on anticipe)
- **Gestion** : on gère, on fait fructifier (on commence déjà par ne plus perdre)
- **Données** : à bien définir au préalable

→ Se poser les bonnes questions, le plus tôt possible

→ Penser à toutes les démarches à effectuer à chaque étape du cycle de vie des données

Les objectifs du PGD

- Assurer la reproductibilité des expériences
  - Décrire comment les données sont obtenues
- Faciliter la réutilisation des données
  - Garantir la compréhension des données
- Respecter le droit et les personnes
  - Clarifier le cadre juridique et éthique
- Éviter les pertes de données
  - Assurer un stockage adapté
- Clarifier les droits de réutilisation
  - Spécifier les modalités de partage
- Établir le rôle de chacun
  - Définir les responsabilités

Pour l'équipe de recherche

Se référer au PGD pour :

- Retrouver les données
- Comprendre les données
- Savoir où sont conservées les données …

Pour la communauté scientifique

Publier le PGD pour indiquer :

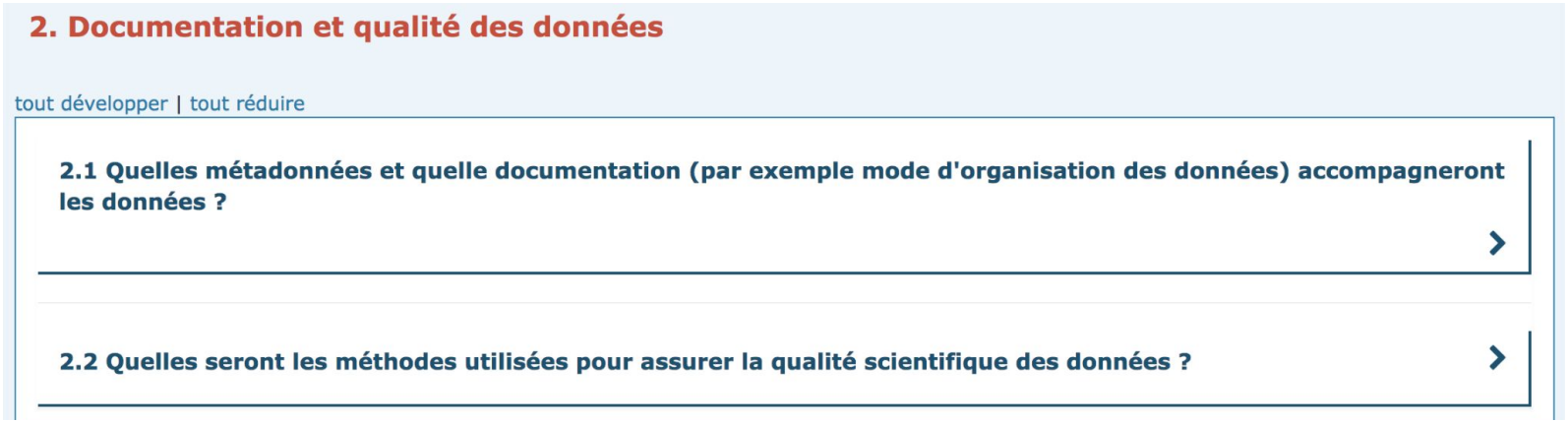
- Quelles données existent
- Où elles sont conservées
- Qui peut y accéder, sous quelles conditions…

Les thématiques abordées dans le PGD (modèle Science Europe)

→ Objectif : Assurer la reproductibilité des expériences

- Décrire les données

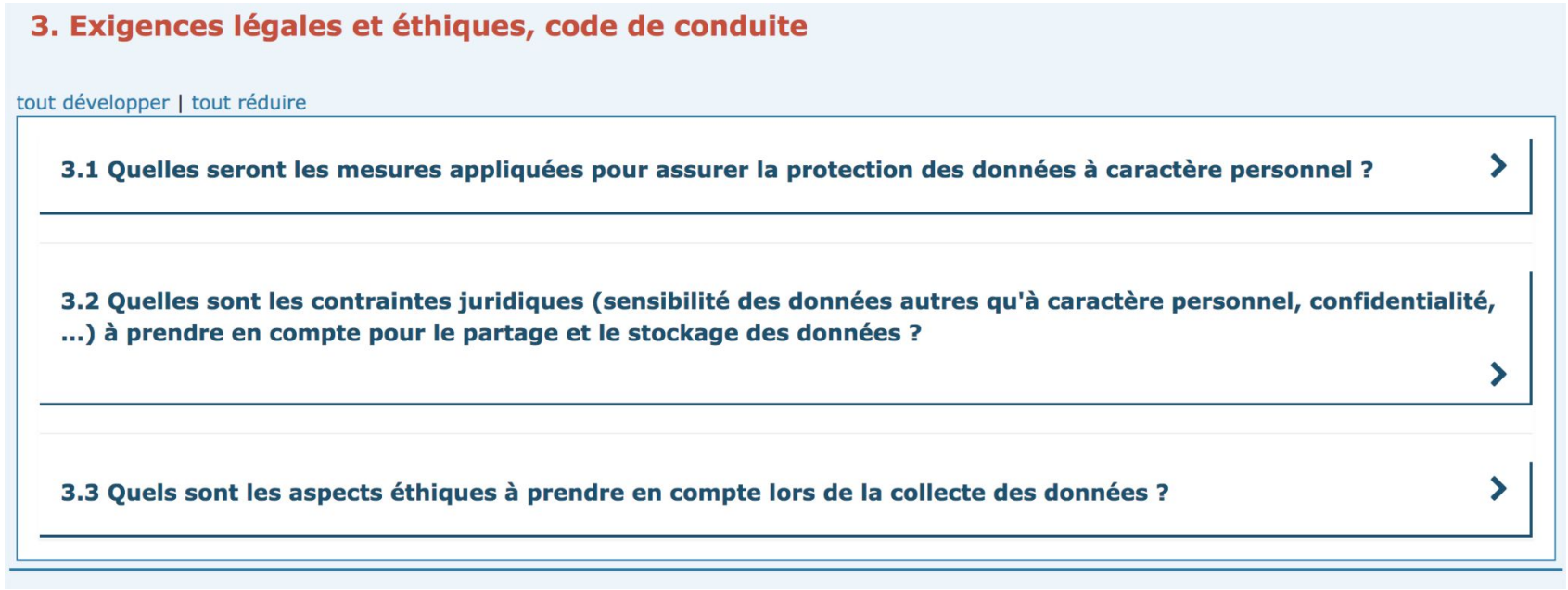
Les thématiques abordées dans le PGD (modèle Science Europe)



⟶ Objectif : Faciliter la réutilisation des données

- Garantir la compréhension des données

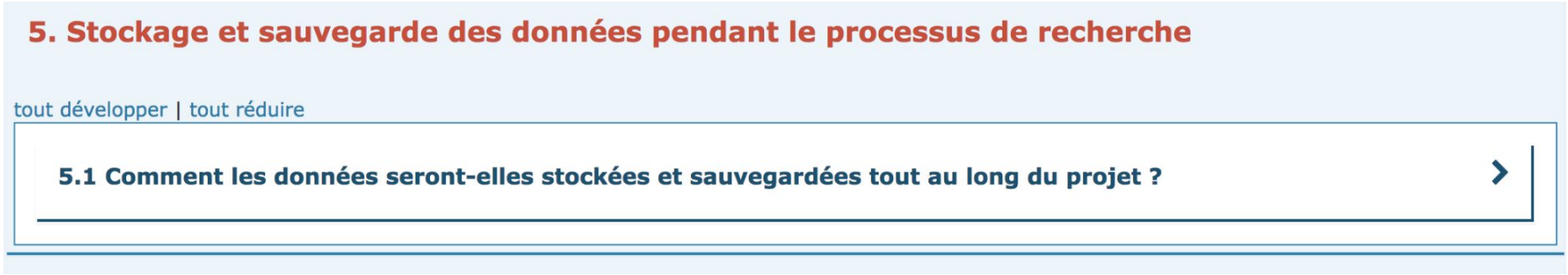
Les thématiques abordées dans le PGD (modèle Science Europe)



→ Objectif : Respecter le droit et les personnes

- Clarifier le cadre juridique et éthique

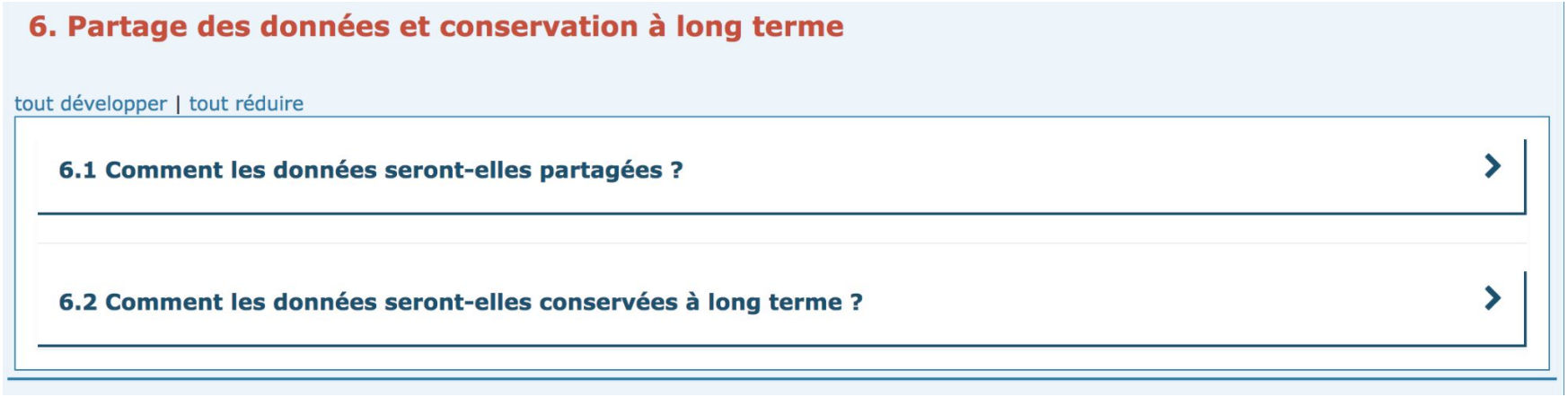
Les thématiques abordées dans le PGD (modèle Science Europe)



→ Objectif : Éviter les pertes de données

- Assurer un stockage adapté

Les thématiques abordées dans le PGD (modèle Science Europe)



→ Objectif : Clarifier les droits de réutilisation

- Spécifier les modalités de partage

Les thématiques abordées dans le PGD (modèle Science Europe)

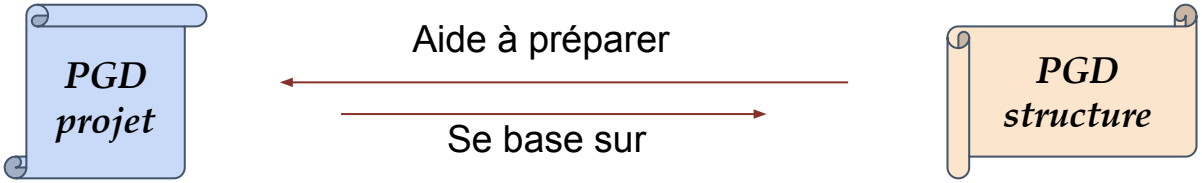
Liste des personnes contribuant à la gestion des produits de recherche au cours d'un projet et leurs rôles. L'attribution d'un rôle à une personne s'effectue dans l'onglet "Rédiger".

Nom	Affiliation	Rôles attribués (Produits de recherche associés)	
Anne-Caroline Delétoille	Institut Pasteur	• Personne contact pour les données (JD2 - PCR)	
Fanny SEBIRE	Institut Pasteur	• Coordinateur du projet • Personne contact pour les données (JD1 - Images) • Responsable du plan	

→ Objectif : Établir le rôle de chacun

- Définir les responsabilités

PGD de projet / PGD de structure



Projet de recherche

PGD à **durée déterminée**
PGD **plus spécifique**

Plateforme, Infrastructure, unité de recherche...

PGD à **durée indéterminée**
PGD **plus générique**

→ Fixer des règles sur la façon de gérer les données produites ou collectées par la structure

→ Se mettre d'accord avec le producteur des données

→ Harmoniser les pratiques

→ Guider les nouveaux arrivants

PGD publics sur DMP OPIDoR (projet et structure) : https://dmp.opidor.fr/public_plans

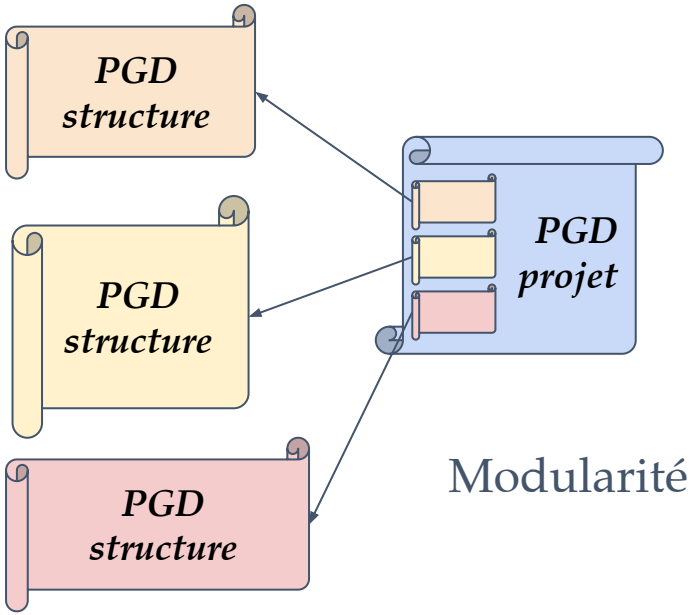
Exemples de PGD de projet :

- PGD du projet INFRAVEC2
- Collection de 841 PGD H2020
- 11 PGD primés

Exemples de PGD de structure :

- Plan de Gestion des Données du CIRM-BIA :
- Data management plan of the Plant Bioinformatics Facility

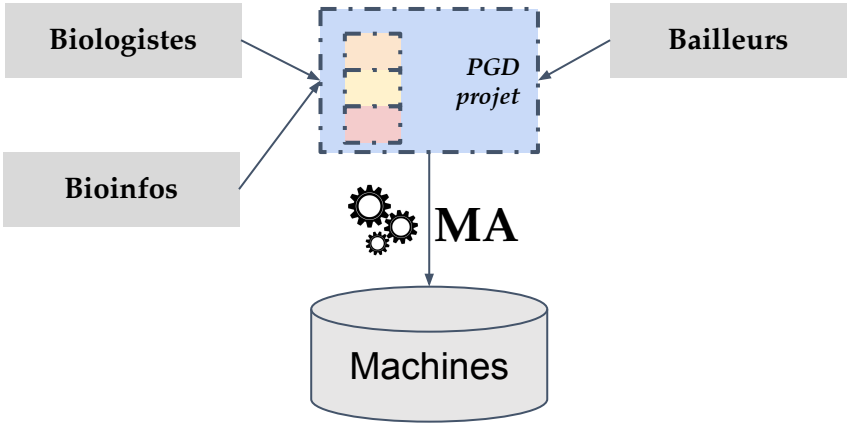
Machine-actionable DMPs



Modularité !

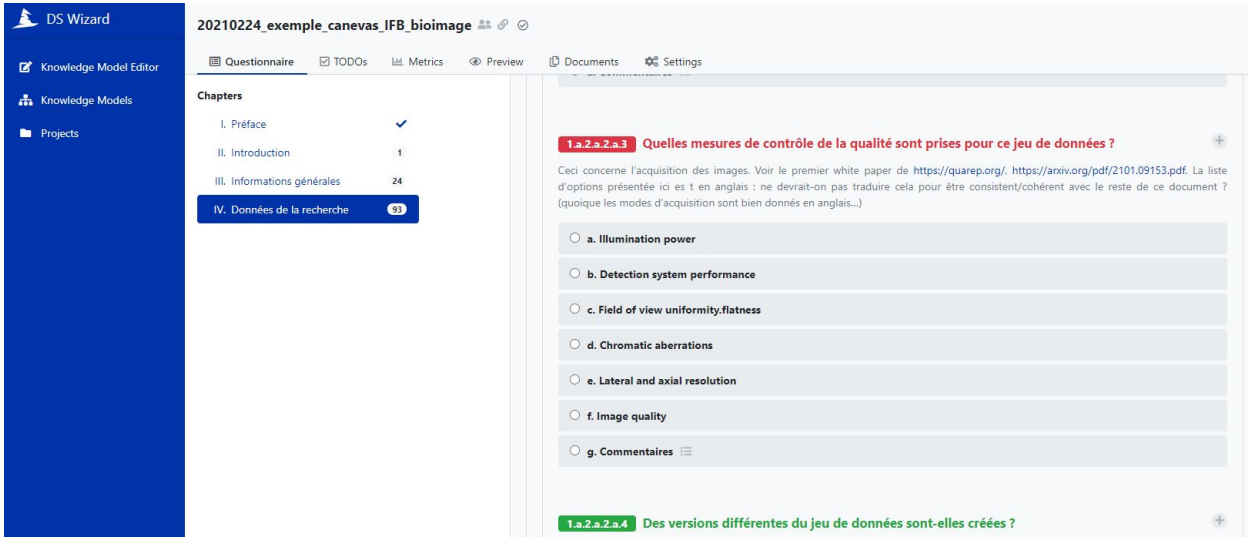
Machine actionable DMP : un plan de gestion lisible par les machines

Objectif : faire du Plan de gestion des Données un outil de configuration des environnements des infrastructures



Différents outils pour rédiger des PGD

- DMP OPIDoR - solution nationale
- DSW - Data Stewardship Wizard - solution européenne (ELIXIR)
- ARGOS - solution de la Commission Européenne



- Get familiar with the **DMP Lifecycle** and excel in **Open and FAIR RDM planning**
- **Co-create DMPs** and manage workload
- **Publish and cite** DMPs as living documents
- **Configure** DMPs to tailored community needs
- **Link** DMPs to research outputs, EOSC services and the **OpenAIRE Research Graph**