Science Ouverte et PGD

Contexte et Opportunités

Frédéric de Lamotte - ORCID 0000-0003-4234-1172

Oana Vigy - ORCID 0000-0002-5000-8735

- Début d’un processus
  - Vous avez déjà des compétences
    - Spécialistes
      - Tout le monde va progresser grâce à nos échanges

- Interactif !
  - N’hésitez pas à partager vos expériences sur le sujet

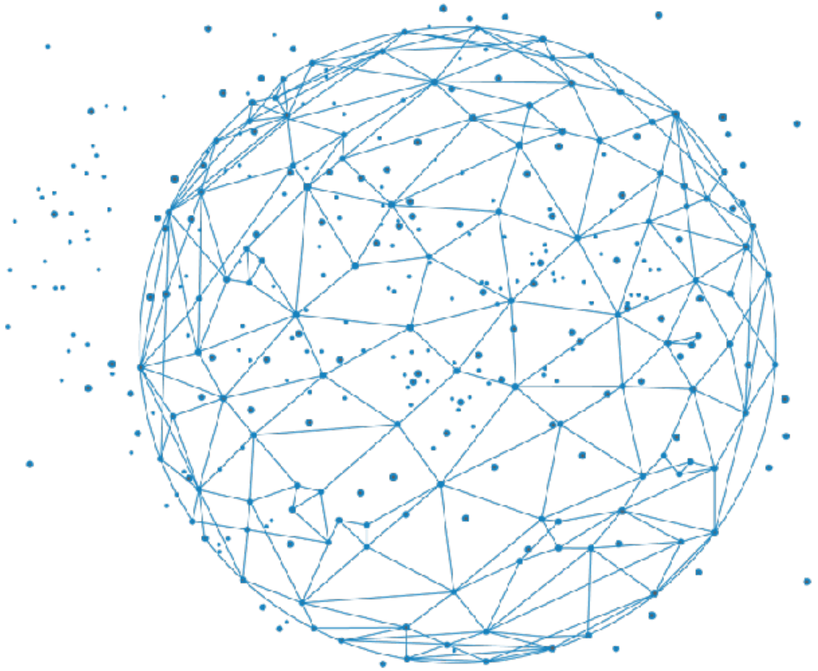
- Dans une ambiance détenue et bienveillante

frederic.de-lamotte@inrae.fr

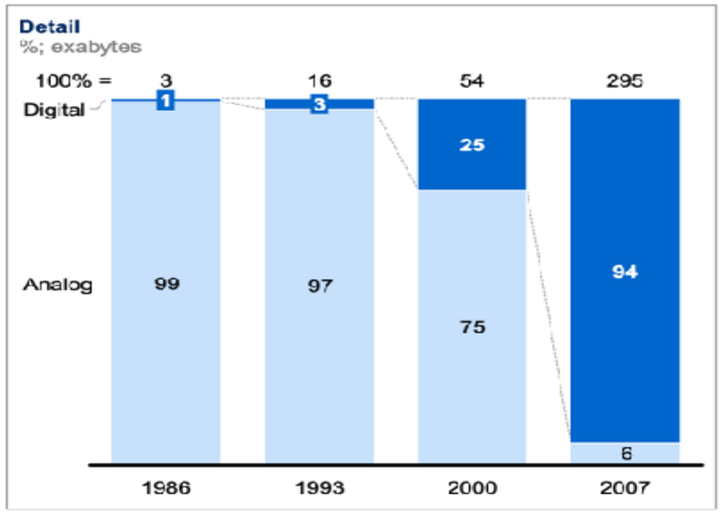
oana.vigy@fpp.cnrs.fr

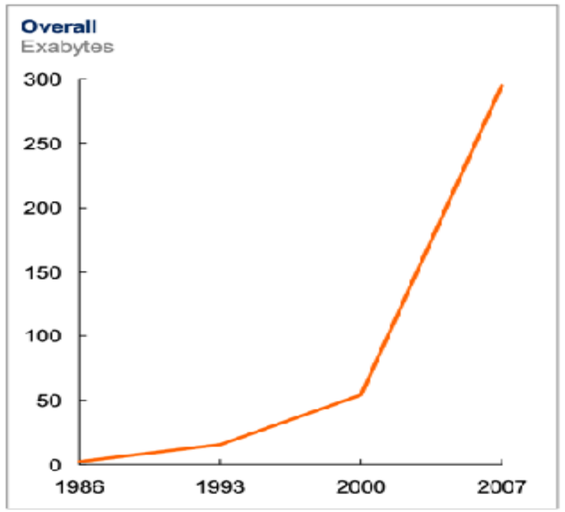
FAIR pourquoi faire ?

Pourquoi maintenant ?



La disruption numérique, une bascule brutale



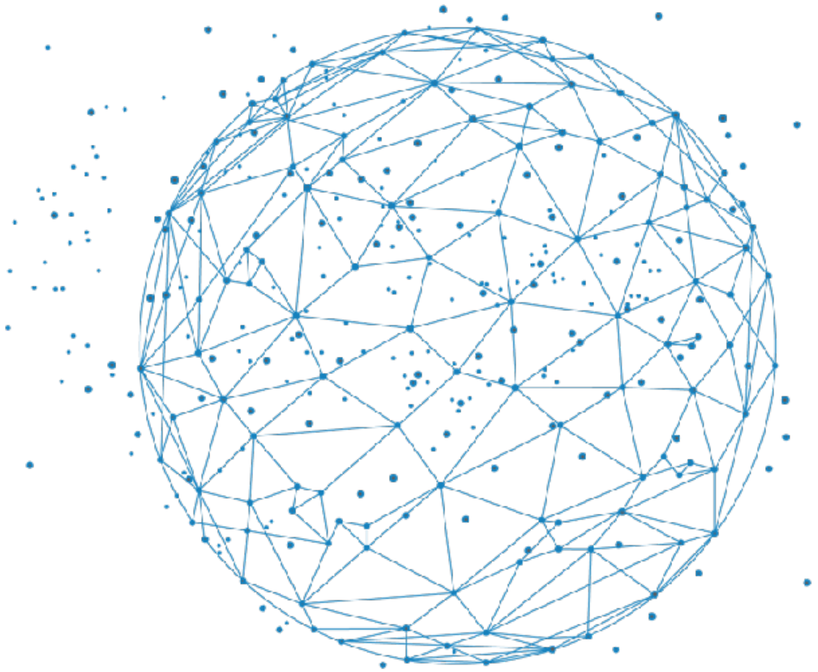


The World's Technological Capacity to Store, Communicate, and Compute Information. Science. 2011;332.

… 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 santé



Et pour vous, d'où est venue la disruption numérique ?

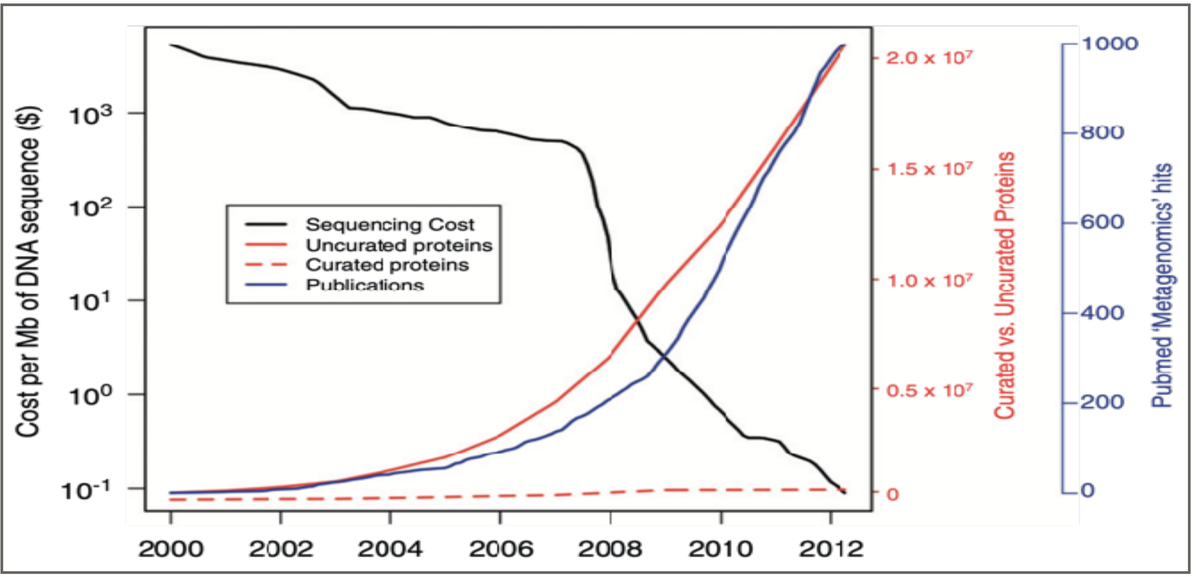
Le déluge des données en Science

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** $

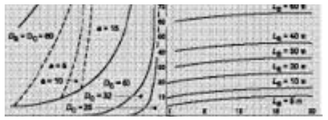
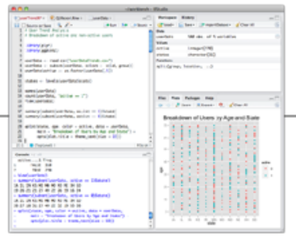
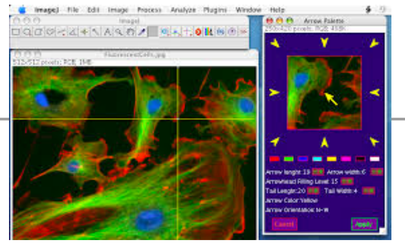
en 2015 = **quelques heures** et **1000** $



→ 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 et de générer une quantité massive de données expérimentales
Collecter des résultats		
Analyser des résultats	  	 
Diffuser le savoir	 	 

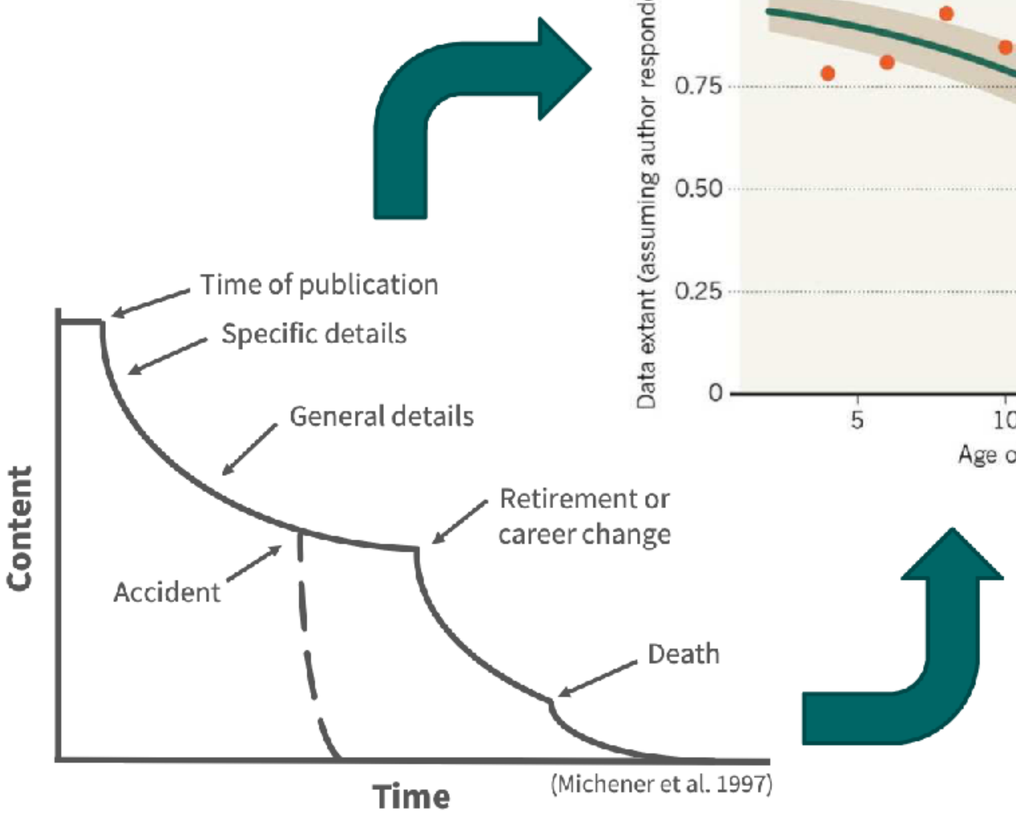
Un changement d'échelle qui modifie la place de la science dans la société

The Sequence of the Human Genome

J. Craig Venter,[1]* Mark D. Adams,[1] Eugene W. Myers,[1] Peter W. Li,[1] Richard J. Mural,[1] Granger G. Sutton,[1] Hamilton O. Smith,[1] Mark Yandell,[1] Cheryl A. Evans,[1] Robert A. Holt,[1] Jeannine D. Gocayne,[1] Peter Amanatides,[1] Richard M. Ballew,[1] Daniel H. Huson,[1] Jennifer Russo Wortman,[1] Qing Zhang,[1] Chinnappa D. Kodira,[1] Xiangqun H. Zheng,[1] Lin Chen,[1] Marian Skupski,[1] Gangadharan Subramanian,[1] Paul D. Thomas,[1] Jinghui Zhang,[1] George L. Gabor Miklos,[2] Catherine Nelson,[3] Samuel Broder,[1] Andrew G. Clark,[4] Joe Nadeau,[5] Victor A. McKusick,[6] Norton Zinder,[7] Arnold J. Levine,[7] Richard J. Roberts,[8] Mel Simon,[9] Carolyn Slayman,[10] Michael Hunkapiller,[11] Randall Bolanos,[1] Arthur Delcher,[1] Ian Dew,[1] Daniel Fasulo,[1] Michael Flanigan,[1] Liliana Florea,[1] Aaron Halpern,[1] Sridhar Hannenhalli,[1] Saul Kravitz,[1] Samuel Levy,[1] Clark Mobarry,[1] Knut Reinert,[1] Karin Remington,[1] Jane Abu-Threideh,[1] Ellen Beasley,[1] Kendra Biddick,[1] Vivien Bonazzi,[1] Rhonda Brandon,[1] Michele Cargill,[1] Ishwar Chandramouliswaran,[1] Rosane Charlab,[1] Kabir Chaturvedi,[1] Zuoming Deng,[1] Valentina Di Francesco,[1] Patrick Dunn,[1] Karen Eilbeck,[1] Carlos Evangelista,[1] Andrei E. Gabrielian,[1] Weiniu Gan,[1] Wangmao Ge,[1] Fangcheng Gong,[1] Zhiping Gu,[1] Ping Guan,[1] Thomas J. Heiman,[1] Maureen E. Higgins,[1] Rui-Ru Ji,[1] Zhaoxi Ke,[1] Karen A. Ketchum,[1] Zhongwu Lai,[1] Yiding Lei,[1] Zhenya Li,[1] Jiayin Li,[1] Yong Liang,[1] Xiaoying Lin,[1] Fu Lu,[1] Gennady V. Merkulov,[1] Natalia Milshina,[1] Helen M. Moore,[1] Ashwinikumar K Naik,[1] Vaibhav A. Narayan,[1] Beena Neelam,[1] Deborah Nusskern,[1] Douglas B. Rusch,[1] Steven Salzberg,[12] Wei Shao,[1] Bixiong Shue,[1] Jingtao Sun,[1] Zhen Yuan Wang,[1] Aihui Wang,[1] Xin Wang,[1] Jian Wang,[1] Ming-Hui Wei,[1] Ron Wides,[13] Chunlin Xiao,[1] Chunhua Yan,[1] Alison Yao,[1] Jane Ye,[1] Ming Zhan,[1] Weiqing Zhang,[1] Hongyu Zhang,[1] Qi Zhao,[1] Liansheng Zheng,[1] Fei Zhong,[1] Wenyan Zhong,[1] Shiaoping C. Zhu,[1] Shaying Zhao,[12] Dennis Gilbert,[1] Suzanna Baumhueter,[1] Gene Spier,[1] Christine Carter,[1] Anibal Cravchik,[1] Trevor Woodage,[1] Feroze Ali,[1] Huijin An,[1] Aderonke Awe,[1] Danita Baldwin,[1] Holly Baden,[1] Mary Barnstead,[1] Ian Barrow,[1] Karen Beeson,[1] Dana Busam,[1] Amy Carver,[1] Angela Center,[1] Ming Lai Cheng,[1] Liz Curry,[1] Steve Danaher,[1] Lionel Davenport,[1] Raymond Desilets,[1] Susanne Dietz,[1] Kristina Dodson,[1] Lisa Doup,[1] Steven Ferriera,[1] Neha Garg,[1] Andres Gluecksmann,[1] Brit Hart,[1] Jason Haynes,[1] Charles Haynes,[1] Cheryl Heiner,[1] Suzanne Hladun,[1] Damon Hostin,[1] Jarrett Houck,[1] Timothy Howland,[1] Chinyere Ibegwam,[1] Jeffery Johnson,[1] Francis Kalush,[1] Lesley Kline,[1] Shashi Koduru,[1] Amy Love,[1] Felecia Mann,[1] David May,[1] Steven McCawley,[1] Tina McIntosh,[1] Ivy McMullen,[1] Mee Moy,[1] Linda Moy,[1] Brian Murphy,[1] Keith Nelson,[1] Cynthia Pfannkoch,[1] Eric Pratts,[1] Vinita Puri,[1] Hina Qureshi,[1] Matthew Reardon,[1] Robert Rodriguez,[1] Yu-Hui Rogers,[1] Deanna Romblad,[1] Bob Ruhfel,[1] Richard Scott,[1] Cynthia Sitter,[1] Michelle Smallwood,[1] Erin Stewart,[1] Renee Strong,[1] Ellen Suh,[1] Reginald Thomas,[1] Ni Ni Tint,[1] Sukyee Tse,[1] Claire Vech,[1] Gary Wang,[1] Jeremy Wetter,[1] Sherita Williams,[1] Monica Williams,[1] Sandra Windsor,[1] Emily Winn-Deen,[1] Keriellen Wolfe,[1] Jayshree Zaveri,[1] Karena Zaveri,[1] Josep F. Abril,[14] Roderic Guigó,[14] Michael J. Campbell,[1] Kimmen V. Sjolander,[1] Brian Karlak,[1] Anish Kejariwal,[1] Huaiyu Mi,[1] Betty Lazareva,[1] Thomas Hatton,[1] Apurva Narechania,[1] Karen Diemer,[1] Anushya Muruganujan,[1] Nan Guo,[1] Shinji Sato,[1] Vineet Bafna,[1] Sorin Istrail,[1] Ross Lippert,[1] Russell Schwartz,[1] Brian Walenz,[1] Shibu Yooseph,[1] David Allen,[1] Anand Basu,[1] James Baxendale,[1] Louis Blick,[1] Marcelo Caminha,[1] John Carnes-Stine,[1] Parris Caulk,[1] Yen-Hui Chiang,[1] My Coyne,[1] Carl Dahlke,[1] Anne Deslattes Mays,[1] Maria Dombroski,[1] Michael Donnelly,[1] Dale Ely,[1] Shiva Esparham,[1] Carl Fosler,[1] Harold Gire,[1] Stephen Glanowski,[1] Kenneth Glasser,[1] Anna Glodek,[1] Mark Gorokhov,[1] Ken Graham,[1] Barry Gropman,[1] Michael Harris,[1] Jeremy Heil,[1] Scott Henderson,[1] Jeffrey Hoover,[1] Donald Jennings,[1] Catherine Jordan,[1] James Jordan,[1] John Kasha,[1] Leonid Kagan,[1] Cheryl Kraft,[1] Alexander Levitsky,[1] Mark Lewis,[1] Xiangjun Liu,[1] John Lopez,[1] Daniel Ma,[1] William Majoros,[1] Joe McDaniel,[1] Sean Murphy,[1] Matthew Newman,[1] Trung Nguyen,[1] Ngoc Nguyen,[1] Marc Nodell,[1] Sue Pan,[1] Jim Peck,[1] Marshall Peterson,[1] William Rowe,[1] Robert Sanders,[1] John Scott,[1] Michael Simpson,[1] Thomas Smith,[1] Arlan Sprague,[1] Timothy Stockwell,[1] Russell Turner,[1] Eli Venter,[1] Mei Wang,[1] Meiyuan Wen,[1] David Wu,[1] Mitchell Wu,[1] Ashley Xia,[1] Ali Zandieh,[1] Xiaohong Zhu[1]

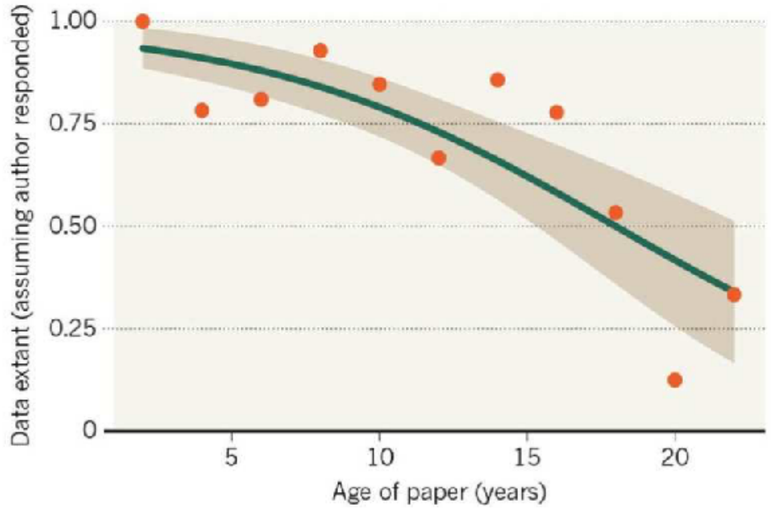
Les ravages du temps

Data Entropy



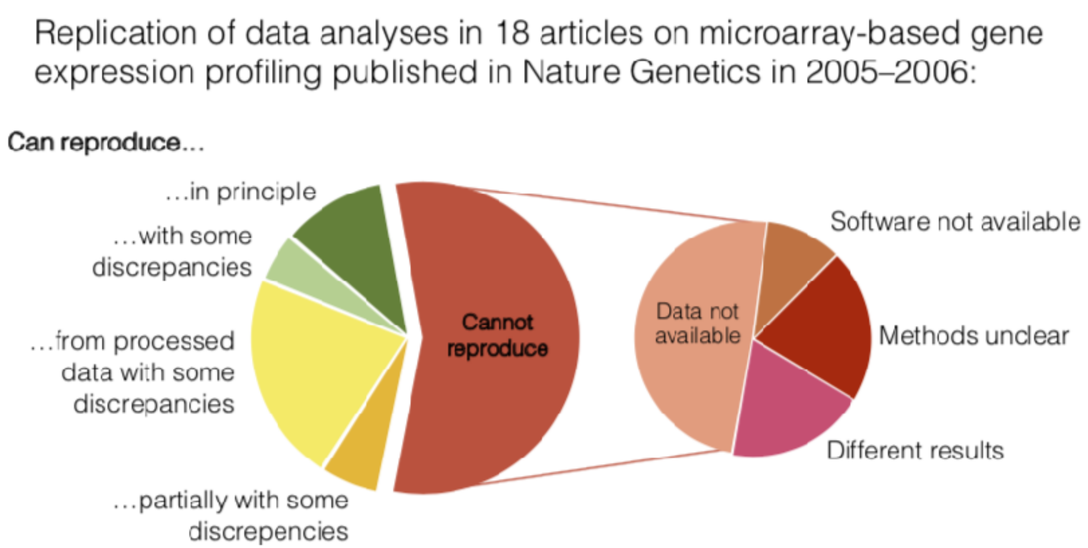
MISSING DATA

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

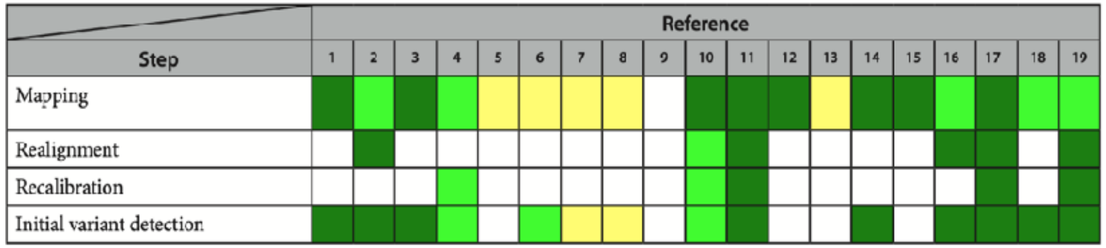


Vines, T. H. *et al. Curr. Biol.* http://dx.doi.org/10.1016/j.cub.2013.11.014 (2013).

La crise de reproductibilité

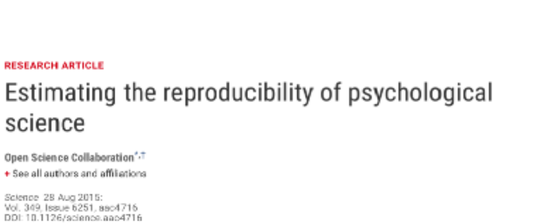


Repeatability of published microarray gene expression analyses. Nat Genet. **2009**;41: 149–155. doi:10.1038/ng.295



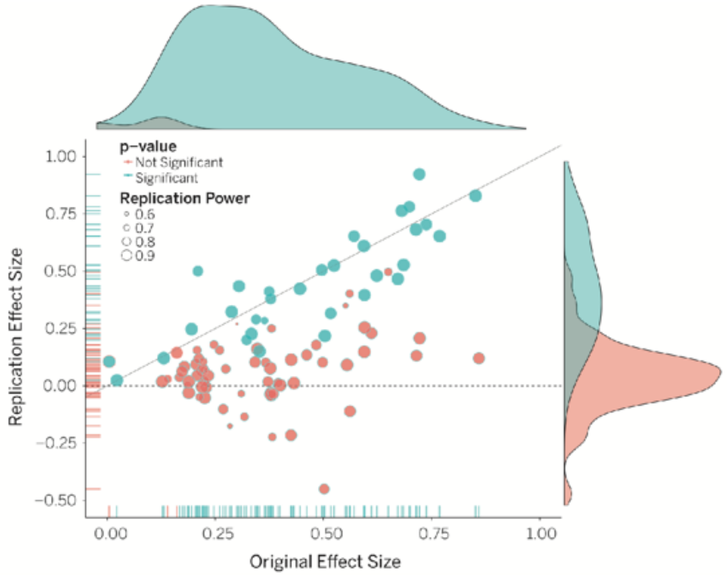
Step	Reference 1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
Mapping																			
Realignment																			
Recalibration																			
Initial variant detection																			

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

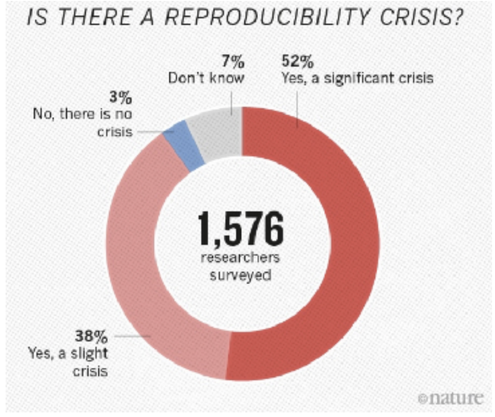


The *Reproducibility project* set out to replicate 100 experiments published in high-impact psychology journals.

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

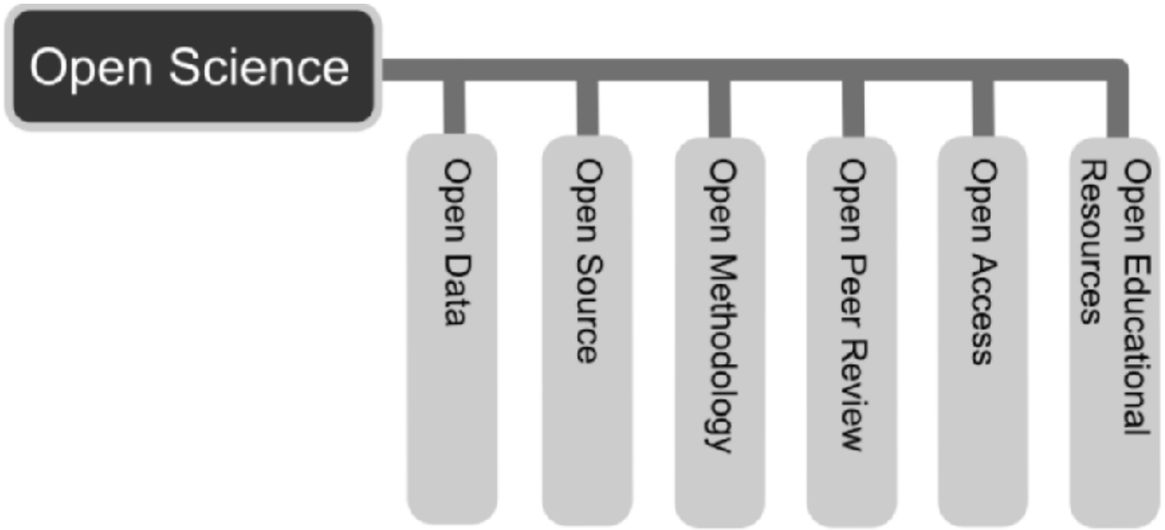


Estimating the reproducibility of psychological science. Science. **2015**;349: aac4716. doi:10.1126/science.aac4716



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

- Une solution ?



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)

  (Publish or Perish & indicateurs / ego metrics, producteurs de savoir biomédical privés, marché de l’édition scientifique)
- La crise de la “bioinformatique”

→Améliorer la science et l’innovation

démocratiser l’accès aux savoirs

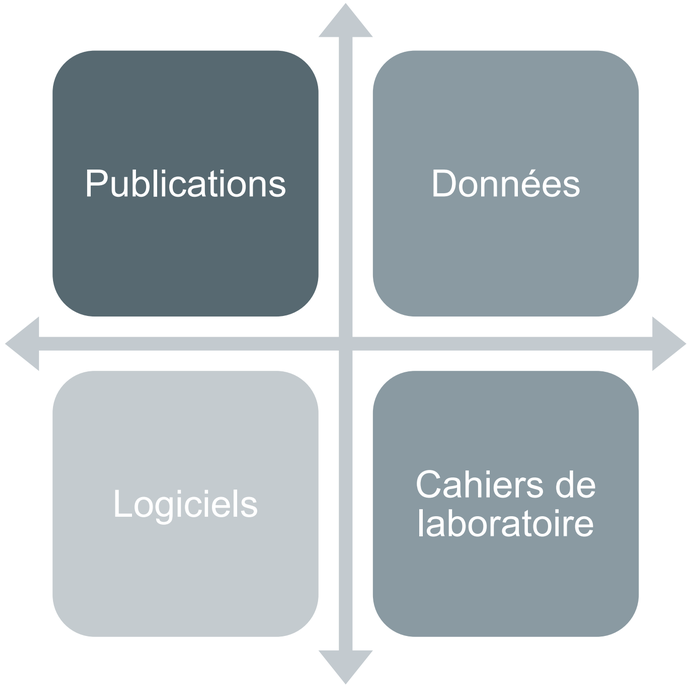
rendre la science plus cumulative, plus fortement étayée par les données, plus transparente

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**

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

Mesures

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)

D’accord, mais… Qu’est ce qu’une **donnée** de la recherche ?

Citez les types de données de la recherche que vous connaissez

En utilisant le tableau à Post-It
http://scrumblr.ca/Data_Type

Définition OCDE

- 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.

Réutiliser les données ? Où est le problème ?

https://youtu.be/66oNv_DJuPc

@picto-dico

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

Data Sharing and Management Snafu in 3 Short Acts

by Karen Hanson, Alisa Surkis & Karen Yacobucci
NYU Health Sciences Libraries

August 3, 2012 (Last Update: December 12, 2012)

Citez 5 conditions

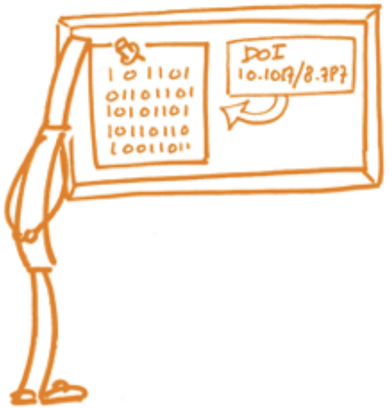
En utilisant le tableau à Post-It
http://scrumblr.ca/solutions

Les principes FAIR

Findable / Facilement trouvable

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)

Interoperable / Intéropérable

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

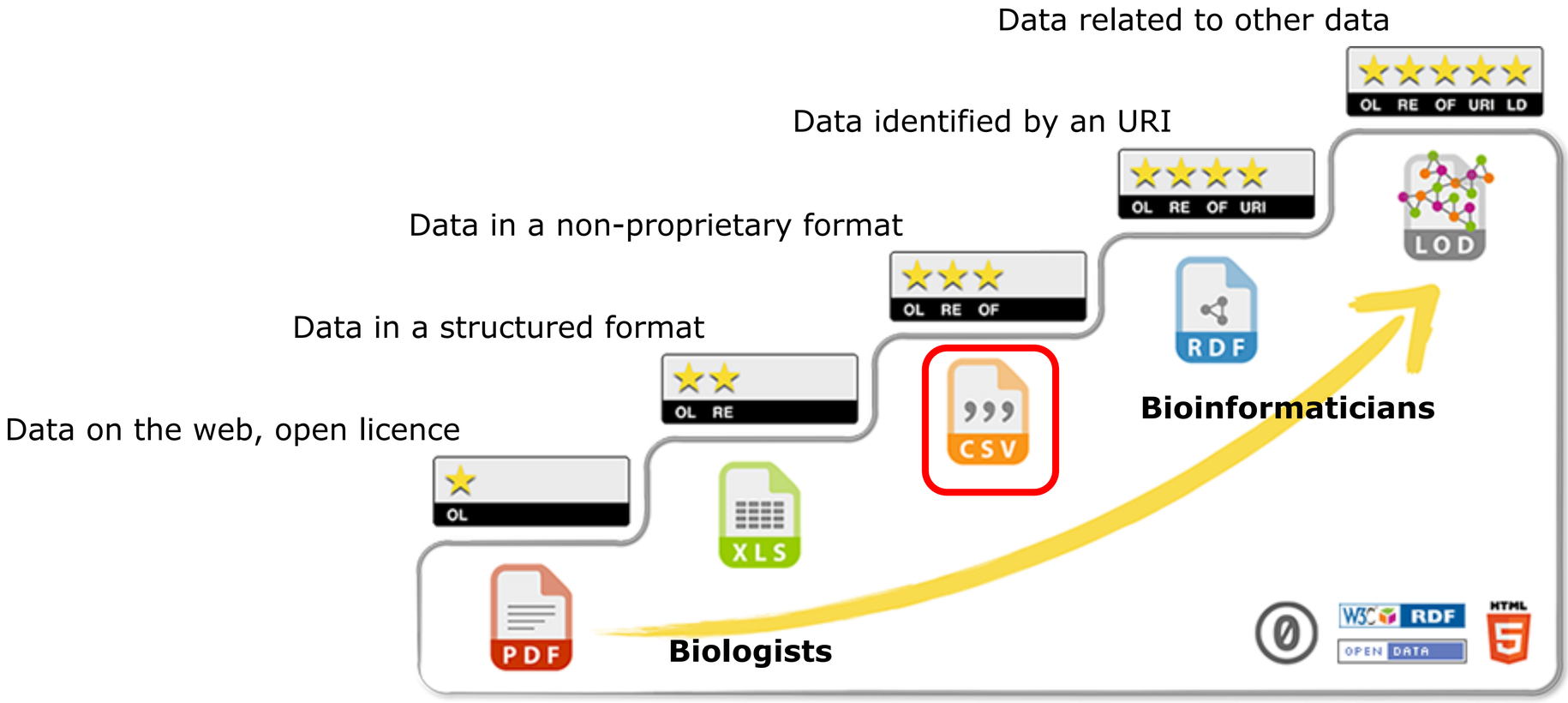
- Les données sont **décrites avec un vocabulaire contrôlé**
- Le vocabulaire utilisé **respecte les principes FAIR**
- Les **métadonnées sont contextualisées** avec des liens vers d'autres données

Reusable / Réutilisable

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 :

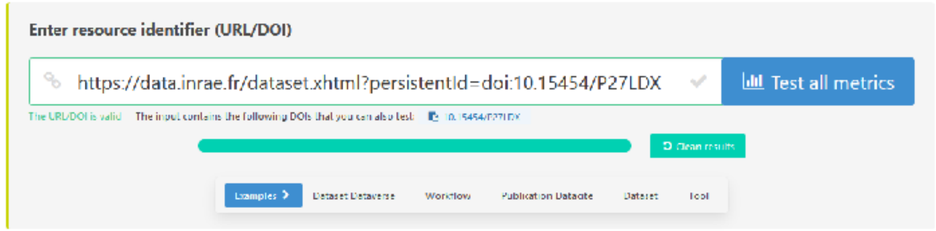
- Biologistes
- Bioinformaticiens
- Spécialistes des ontologies/sémantiques

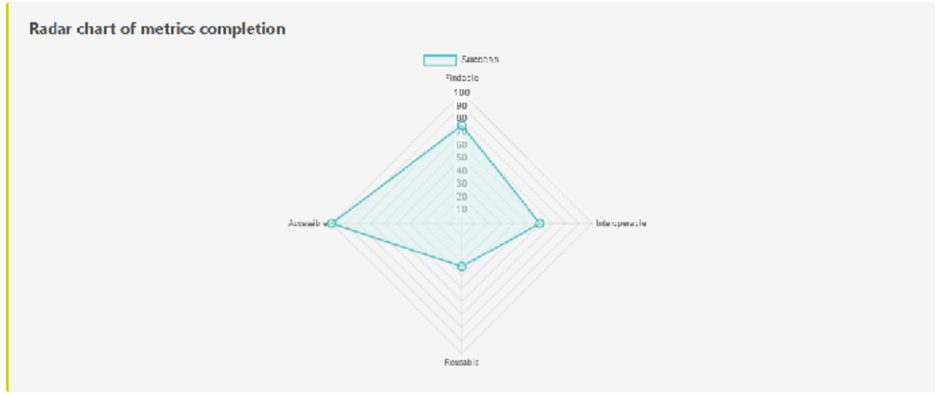
Evaluation de la FAIRness des données

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

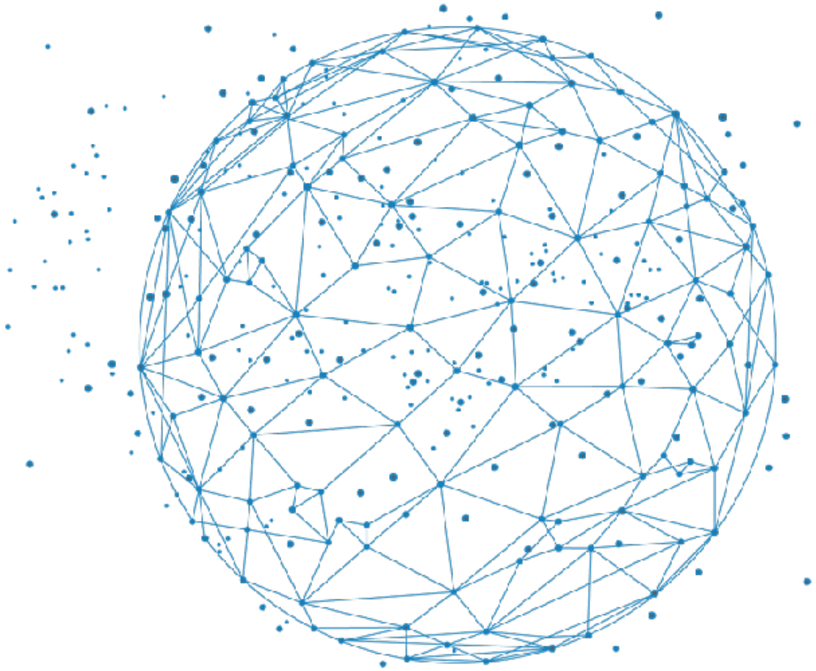
- FAIR self assessment tool
  - ARDC (Australian Research Data Commons)
  - https://ardc.edu.au/resources/aboutdata/fair-data/fair-self-assessment-tool/

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

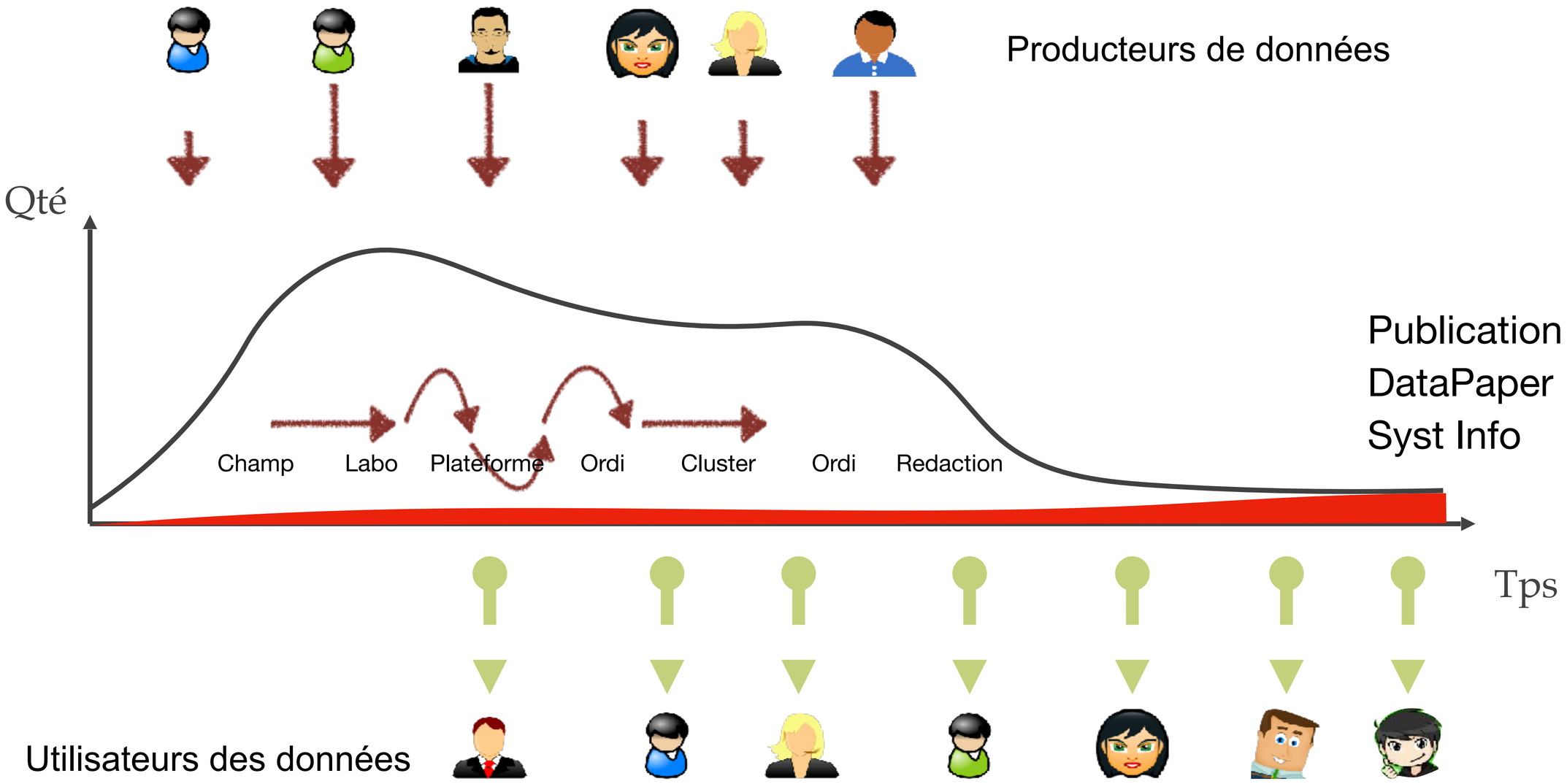




Cycle de vie des données



Un projet sur la durée



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

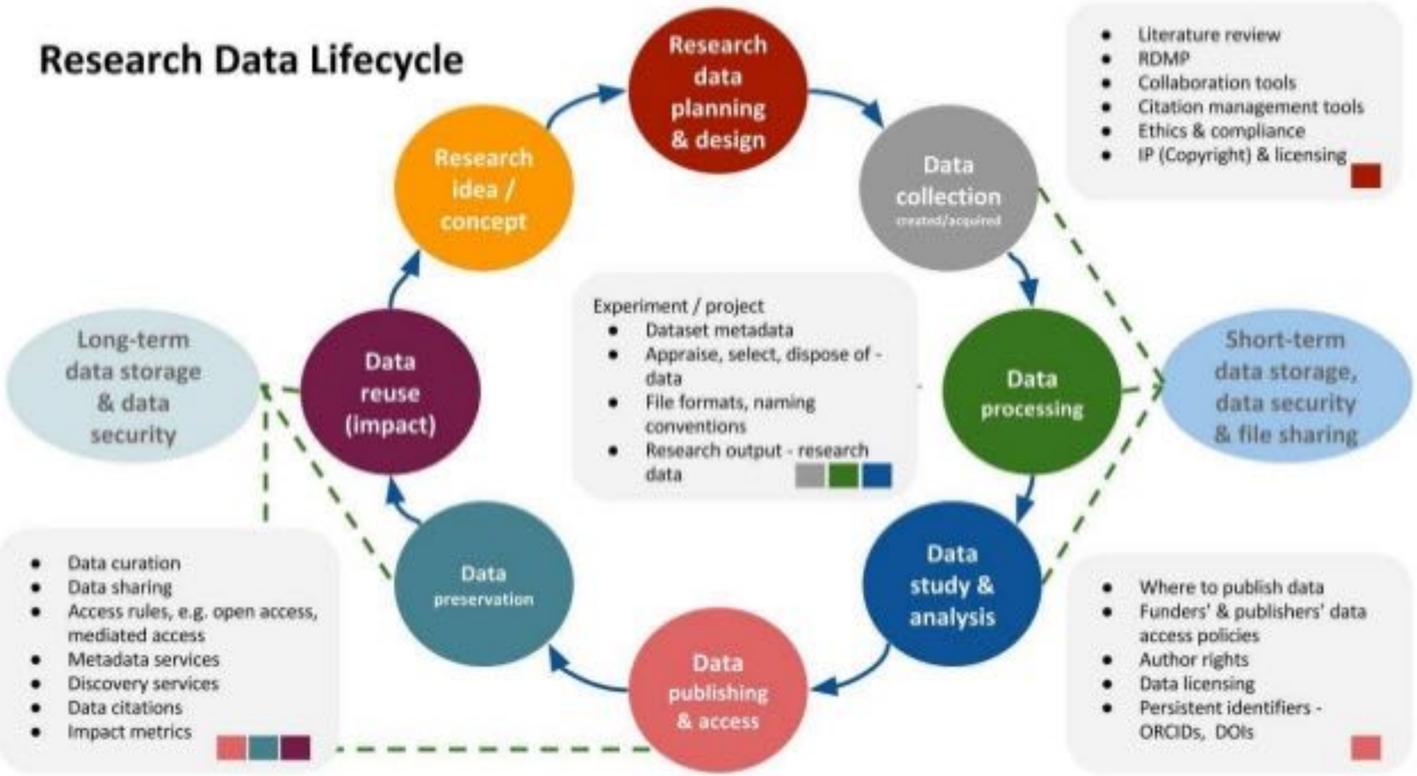
- **Création ou collecte** des données (creating data) ;
- **Traitement** des données (processing data) ;
- **Analyse** des données (analysing data) ;
- **Conservation** des données (preserving data) ;
- **Accès** aux données (giving access to data / data discovery) ;
- **Réutilisation** des données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

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

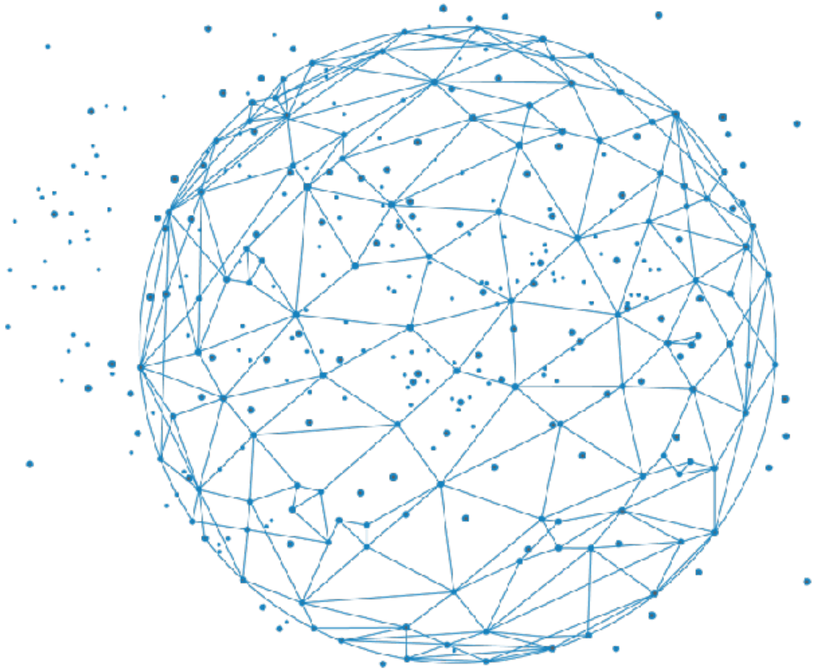
Les étapes

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

Une vue plus détaillée



Plan de Gestion de Données



Avez vous déjà pris l’avion ?

https://www.enac.fr/fr/les-formations-au-pilotage

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 :

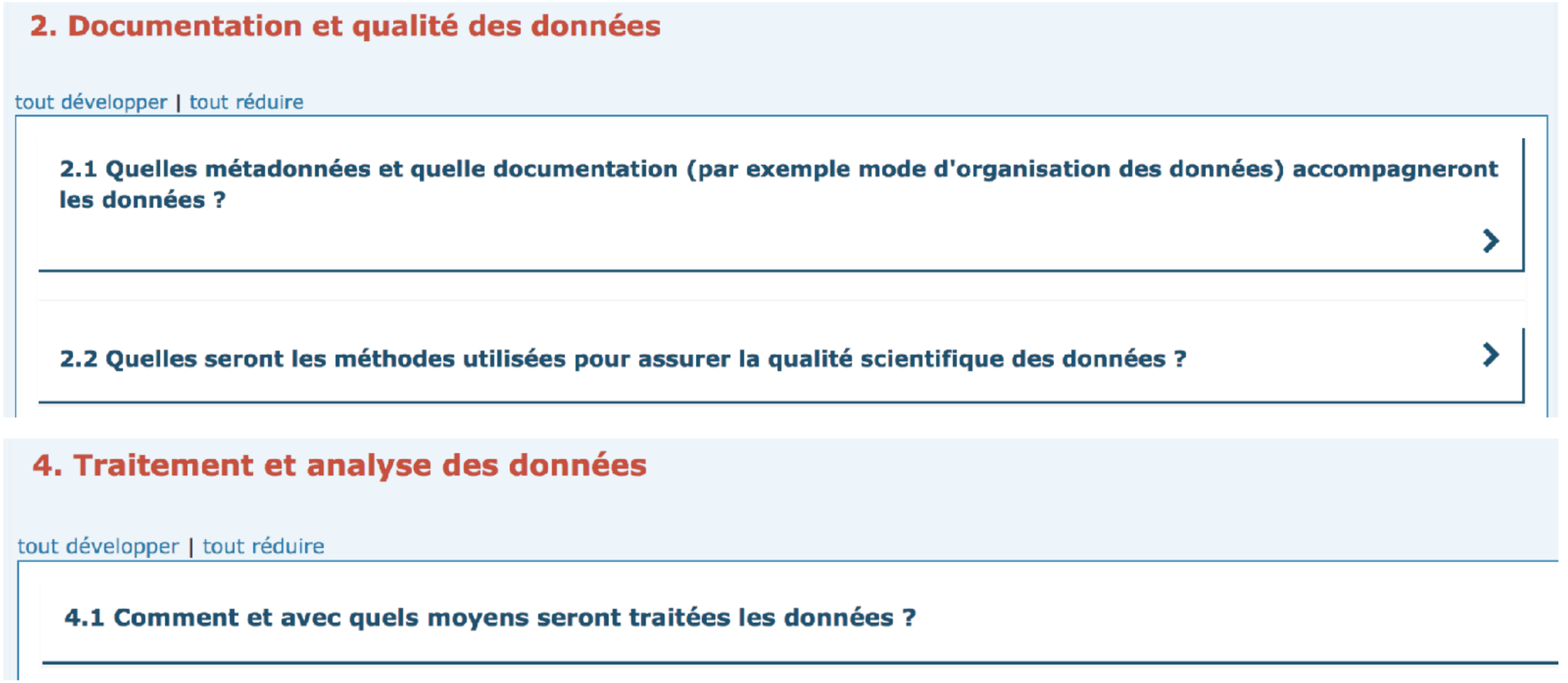
- Quelles données existent
- Où elles sont conservées
- Qui peut y accéder, sous quelles conditions…

Les thématiques abordées dans le PGD (modèle Science Europe)

→ Objectif : Assurer la reproductibilité des expériences

- Décrire les données

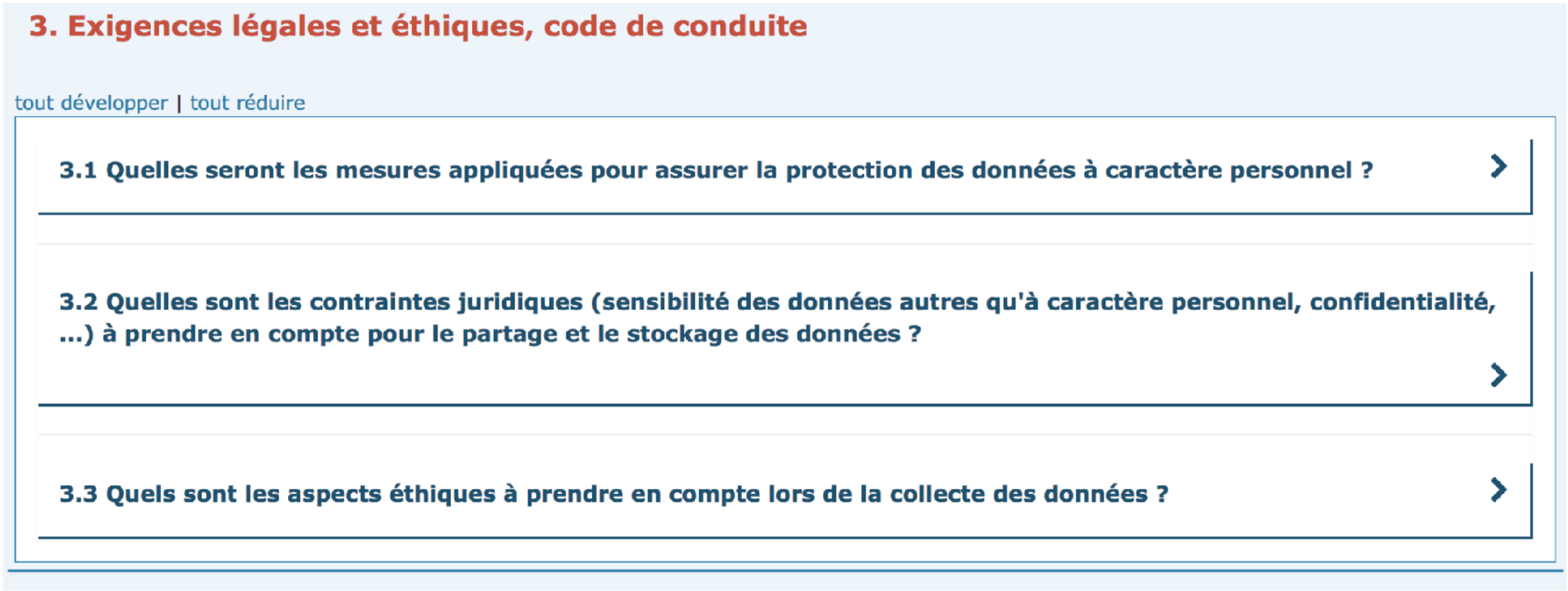
Les thématiques abordées dans le PGD (modèle Science Europe)



→ Objectif : Faciliter la réutilisation des données

- Garantir la compréhension des données

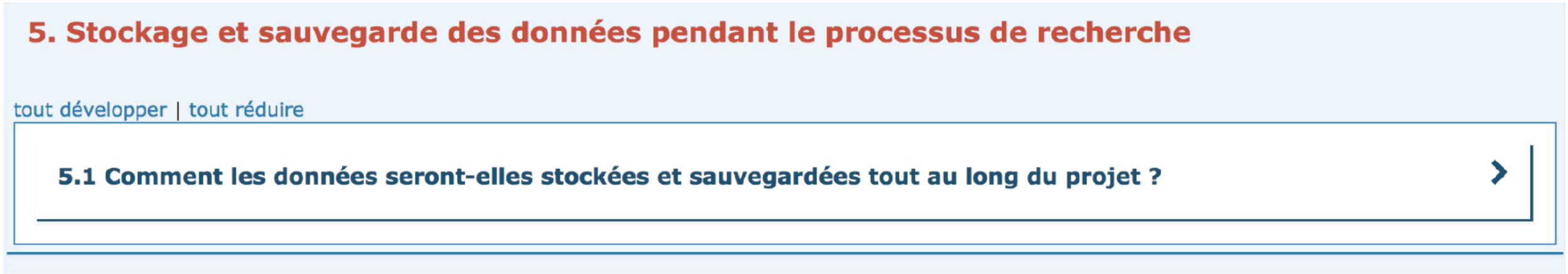
Les thématiques abordées dans le PGD (modèle Science Europe)



→ Objectif : Respecter le droit et les personnes

- Clarifier le cadre juridique et éthique

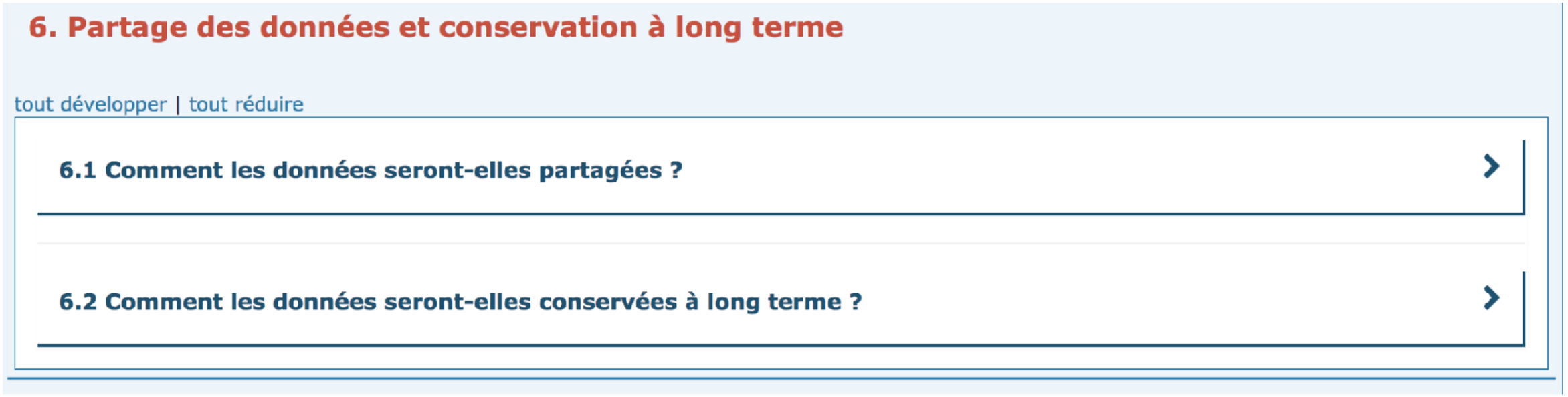
Les thématiques abordées dans le PGD (modèle Science Europe)



→ 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)

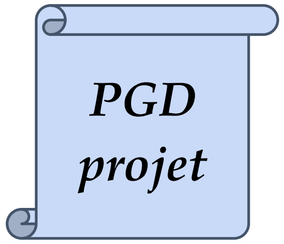
Liste des personnes contribuant à la gestion des produits de recherche au cours d'un projet et leurs rôles. L'attribution d'un rôle à une personne s'effectue dans l'onglet "Rédiger".

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

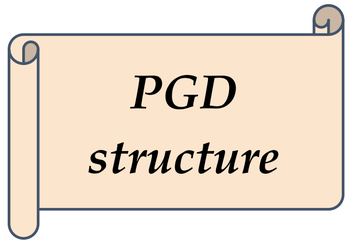
- Définir les responsabilités

PGD de projet / PGD de structure



Aide à préparer

Se base sur



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

Exemples de PGD

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

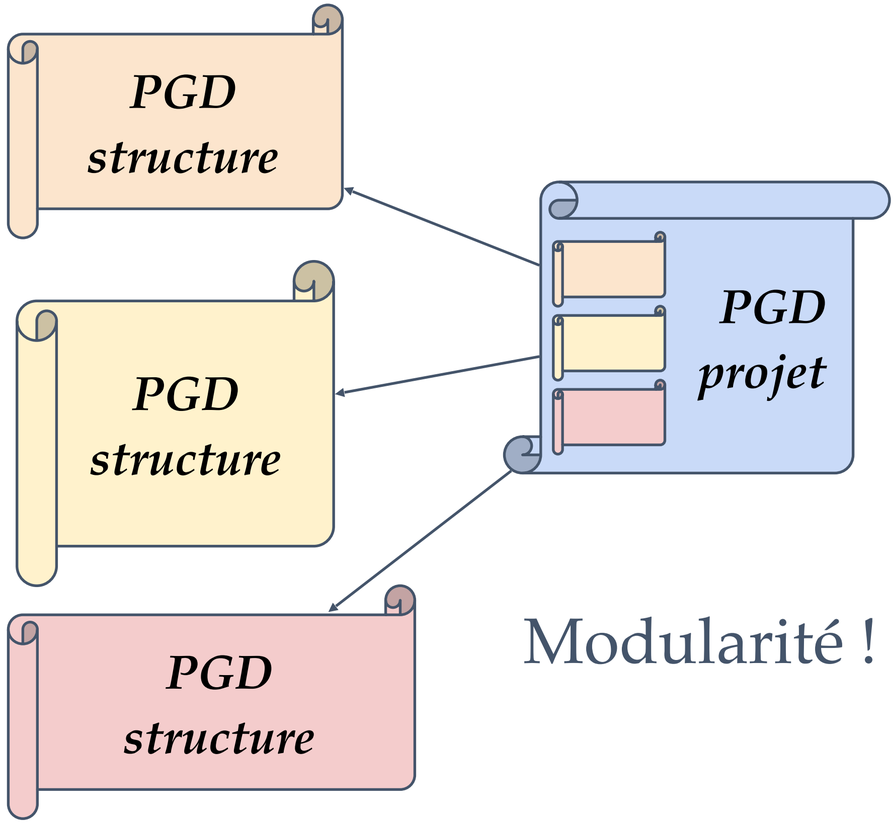
Exemples de PGD de projet :

- 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

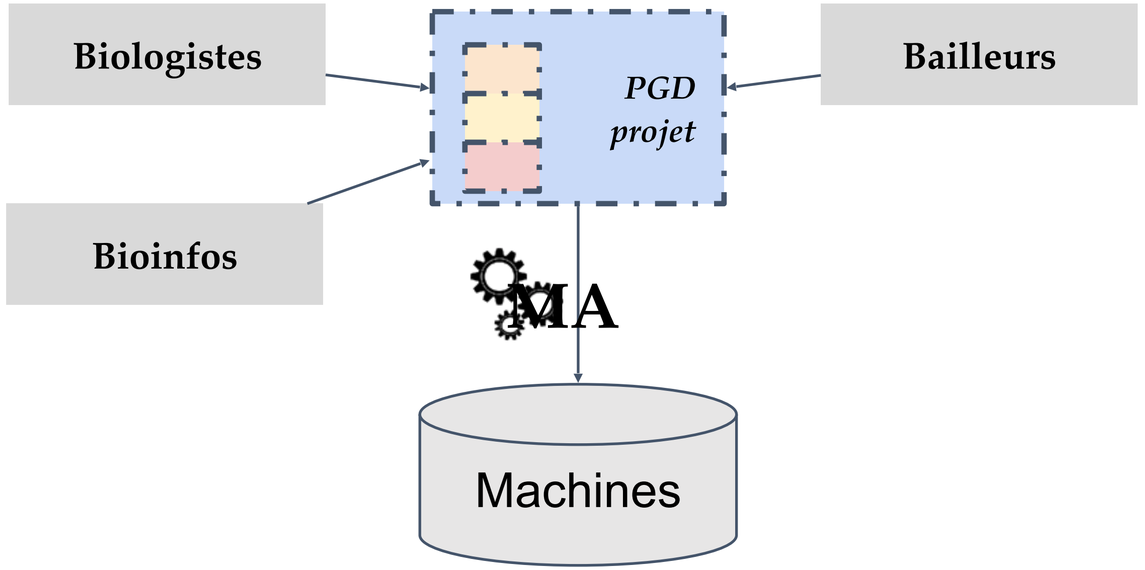
Machine-actionable DMPs



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**

Devoirs à faire à la maison

- Liste exhaustive de vos donnés
- Réfléchir à
  - Où mettre les données pendant le projet
    - Volume, besoins d'accès
  - Où mettre les données pour les partager
    - Comment les documenter
- Comment gérer « le reste »
  - CR, annexes, gestion …
- Stratégie de nommage de tous les fichiers

- Un grand merci à :

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

Christophe Antoniewski

Naïra Naouar

Anne-Caroline Delétoille

Fanny Sébire

Valentin Loux

Cédric Midoux

Célia Michotey

Cyril Pommier

Et merci à toutes celles et ceux qui nous ont stimulé par leurs questions et encouragé par leur enthousiasme