

Collecter, préserver, annoter, organiser, réutiliser nos données

Essayons !



C'est complexe !

- ❖ Contexte
- ❖ Éléments de réflexion (à partager autour de vous)
- ❖ Quelques pistes

Fred de Lamotte - Montpellier

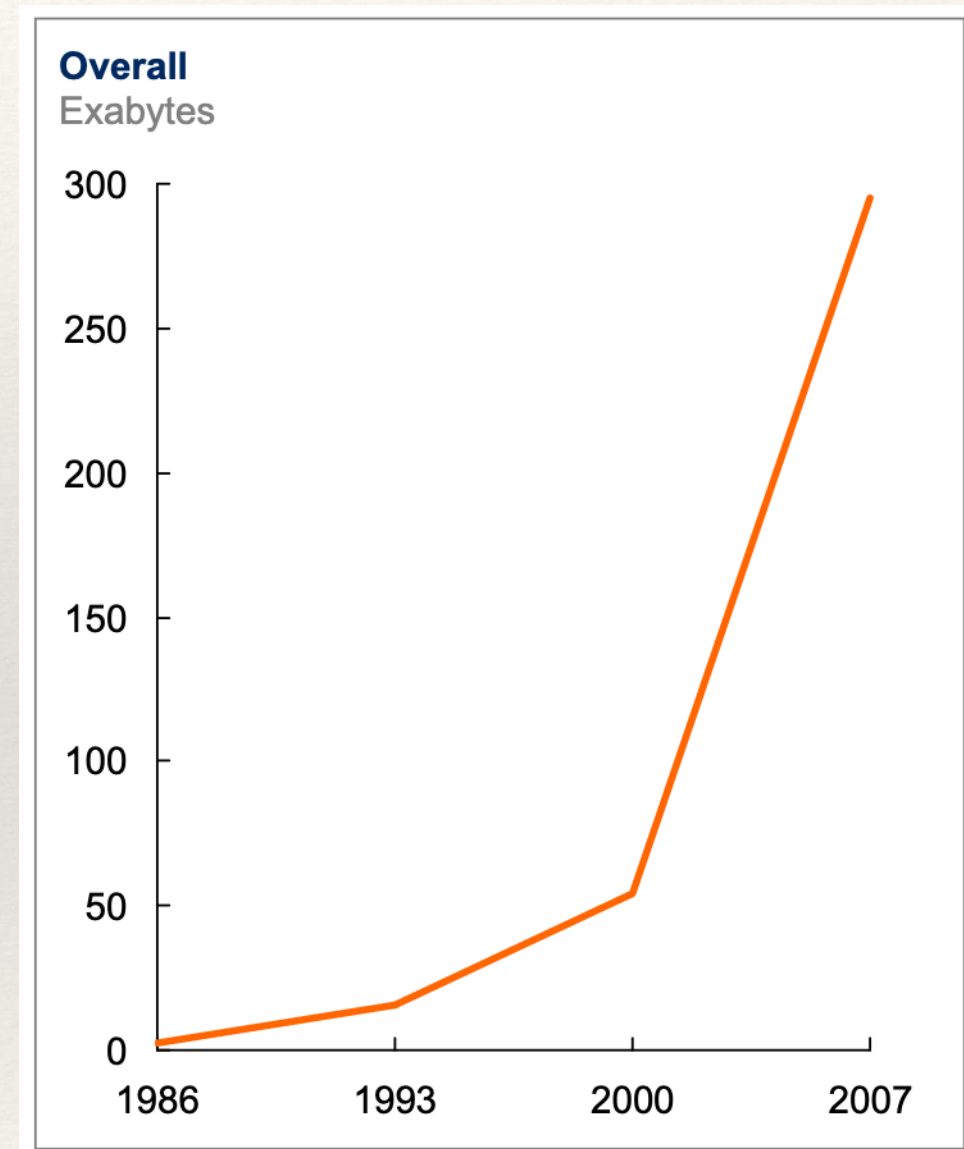
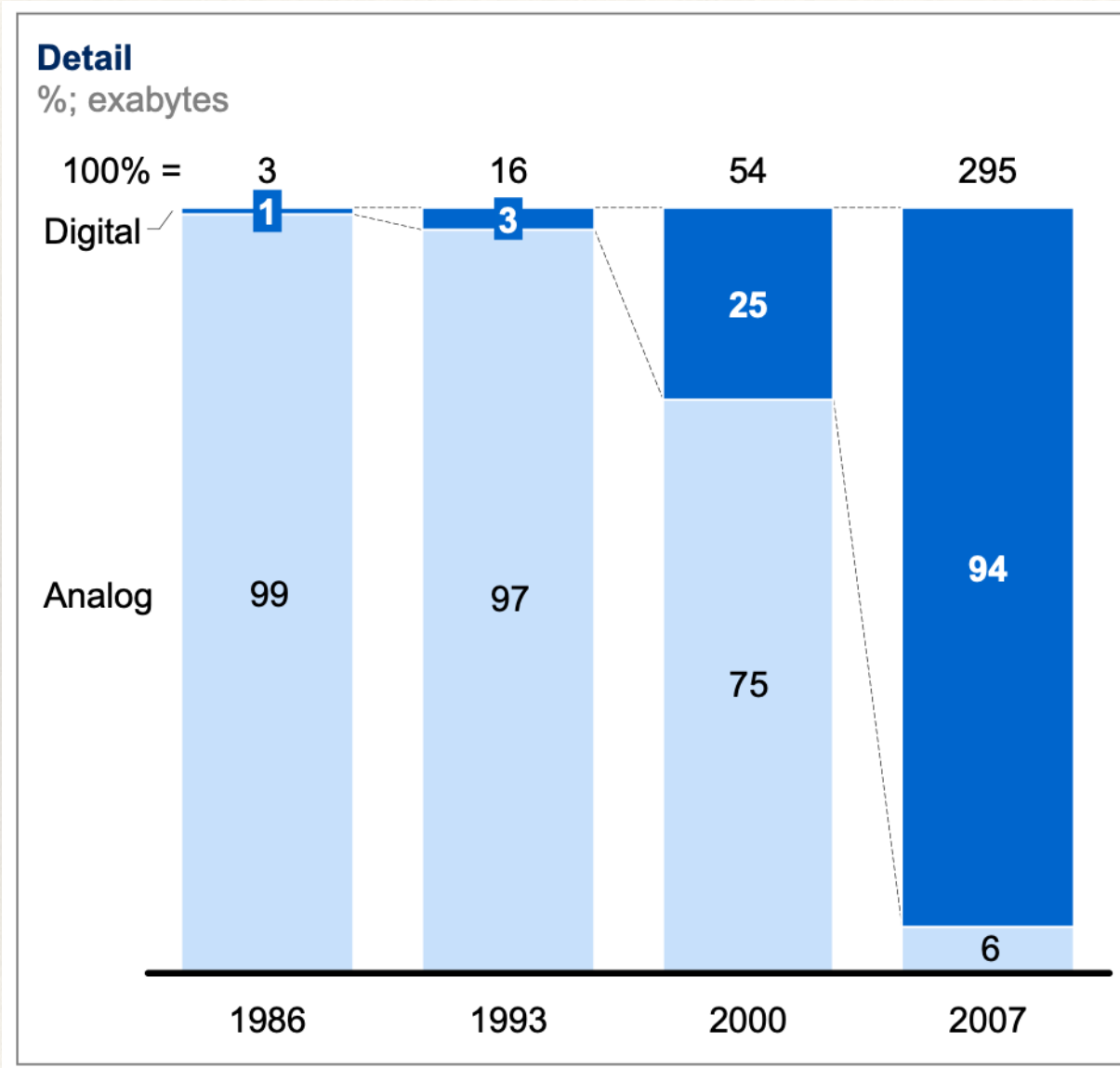
INRAE



Que se passe-t-il ?

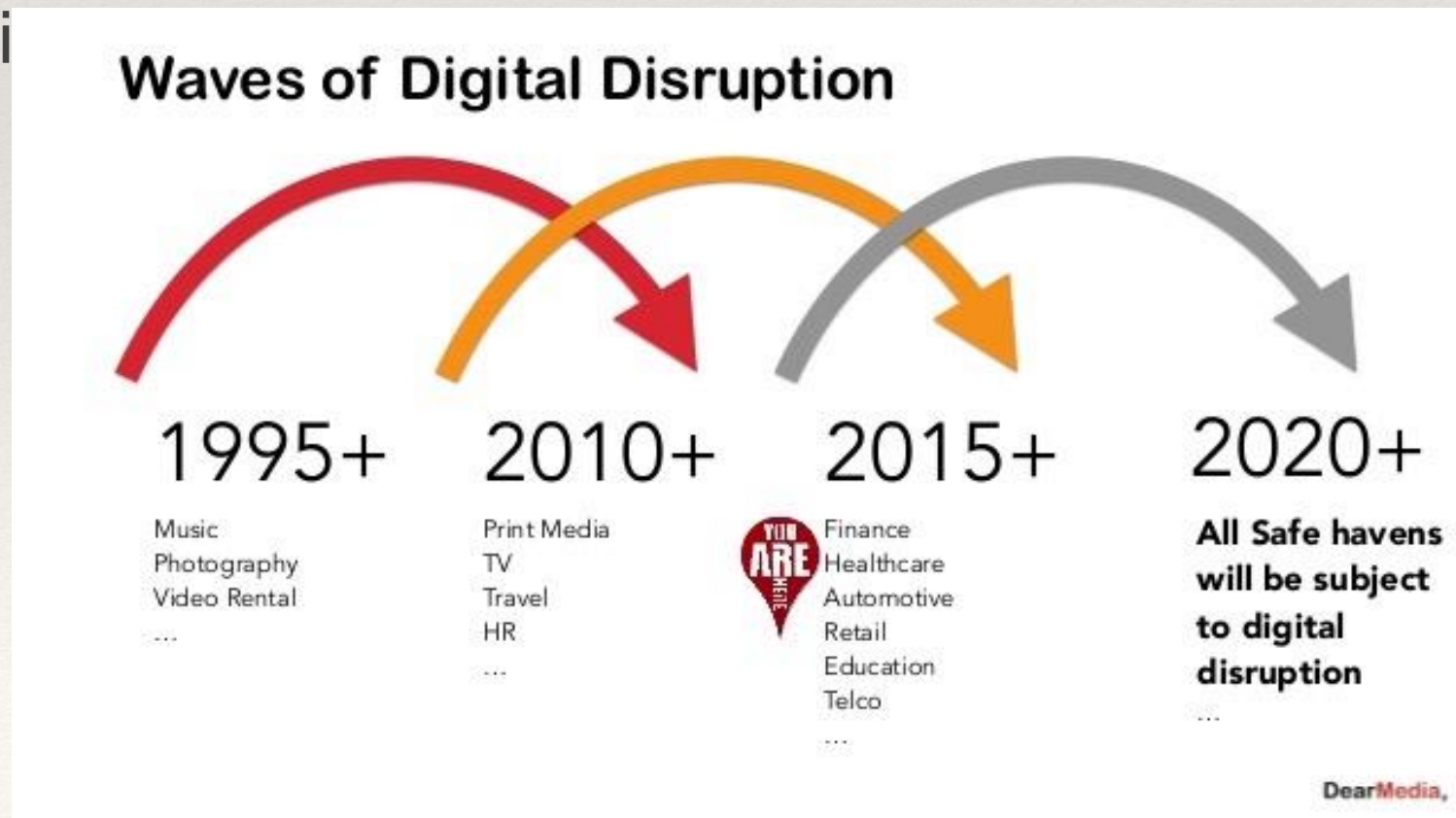


La disruption numérique : une bascule brutale



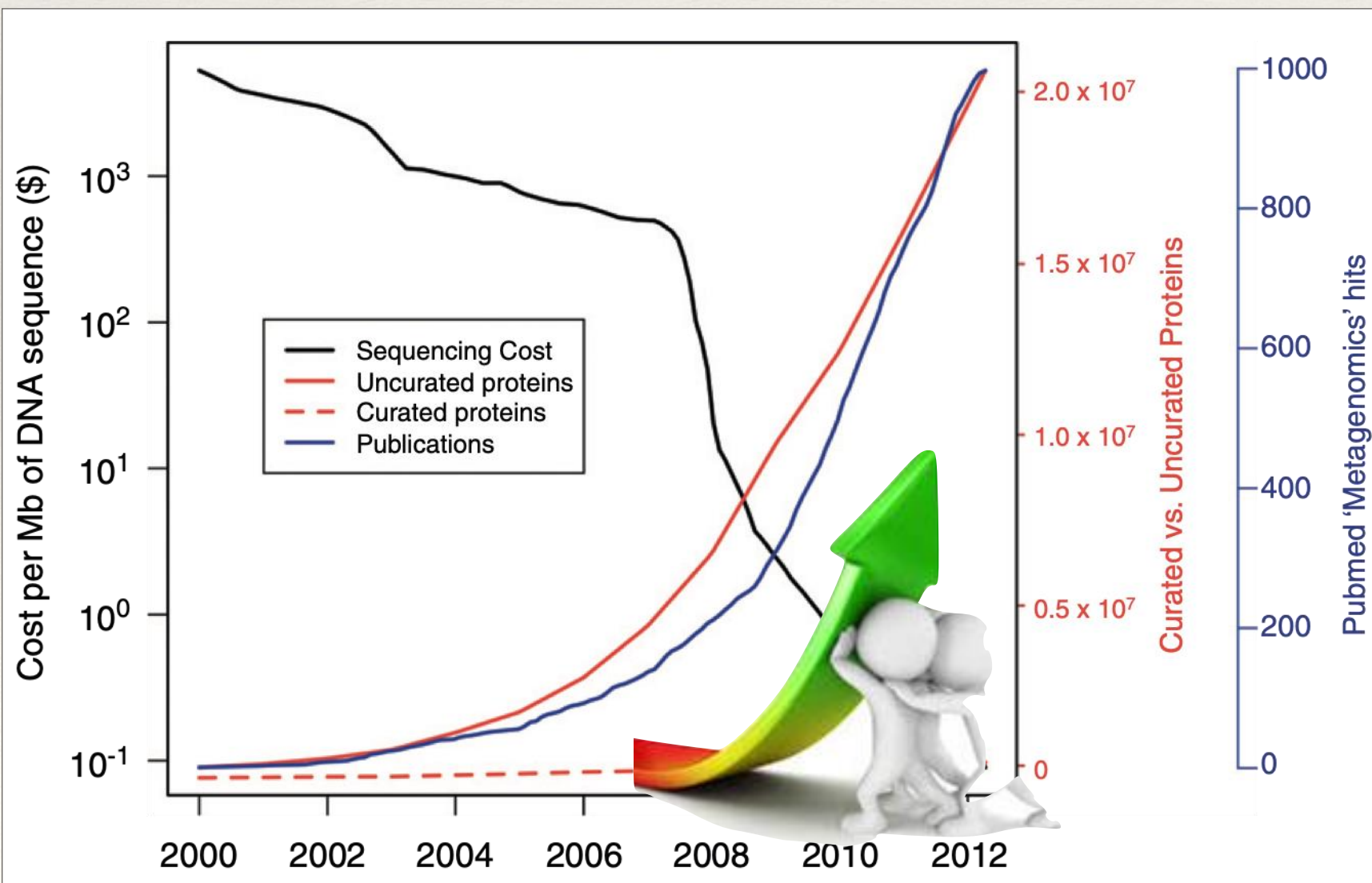
Qui change notre monde

- ❖ 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 distributeur de cinéma (Netflix)



Et en Science ?

- Les techniques à haut débit provoquent un déluge de données
eExemple du 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'approche

Avant

- 1 Concevoir l'expérimentation
- 2 Collecter des résultats
- 3 Analyser des résultats

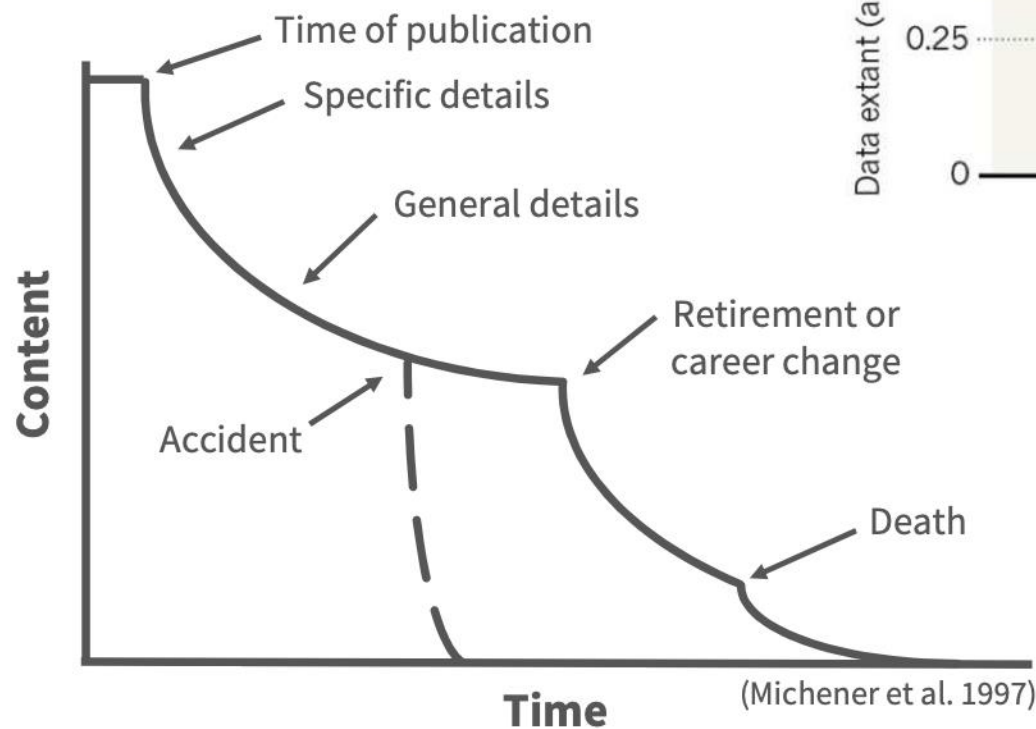
Un changement de paradigme

Maintenant

- 1 Générer massivement des données
- 2 Organiser (stocker, documenter, annoter)
- 3 Analyser (extraire de l'information)
- 4 Diffuser l'information

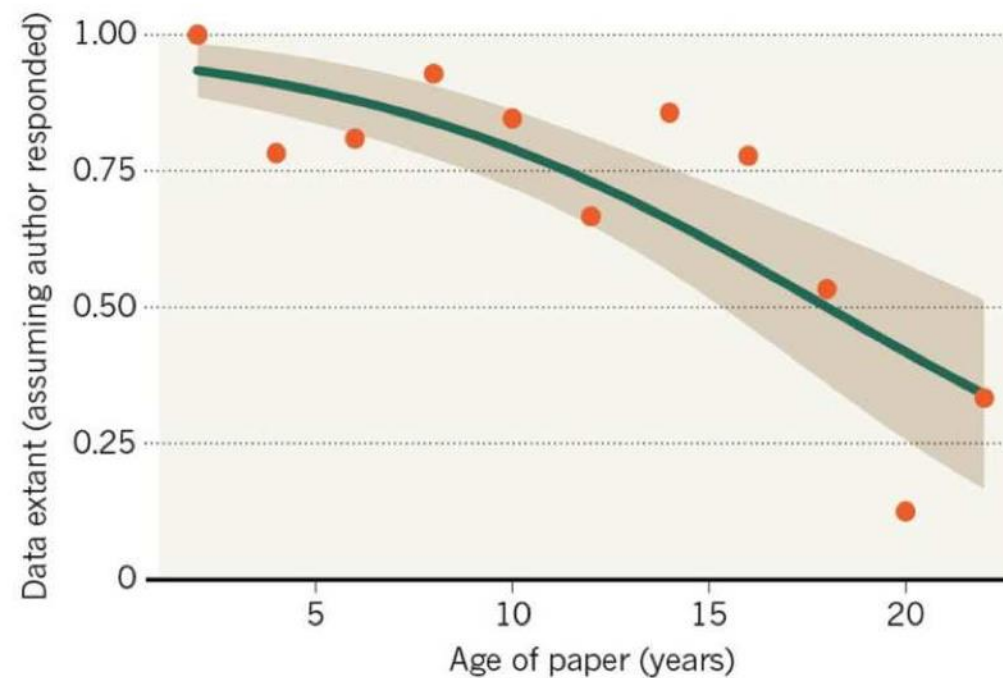
Les ravages du temps

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).

Les défis de la reproductibilité

RESEARCH ARTICLE

Estimating the reproducibility of psychological science

Open Science Collaboration^{*,†}

+ See all authors and affiliations

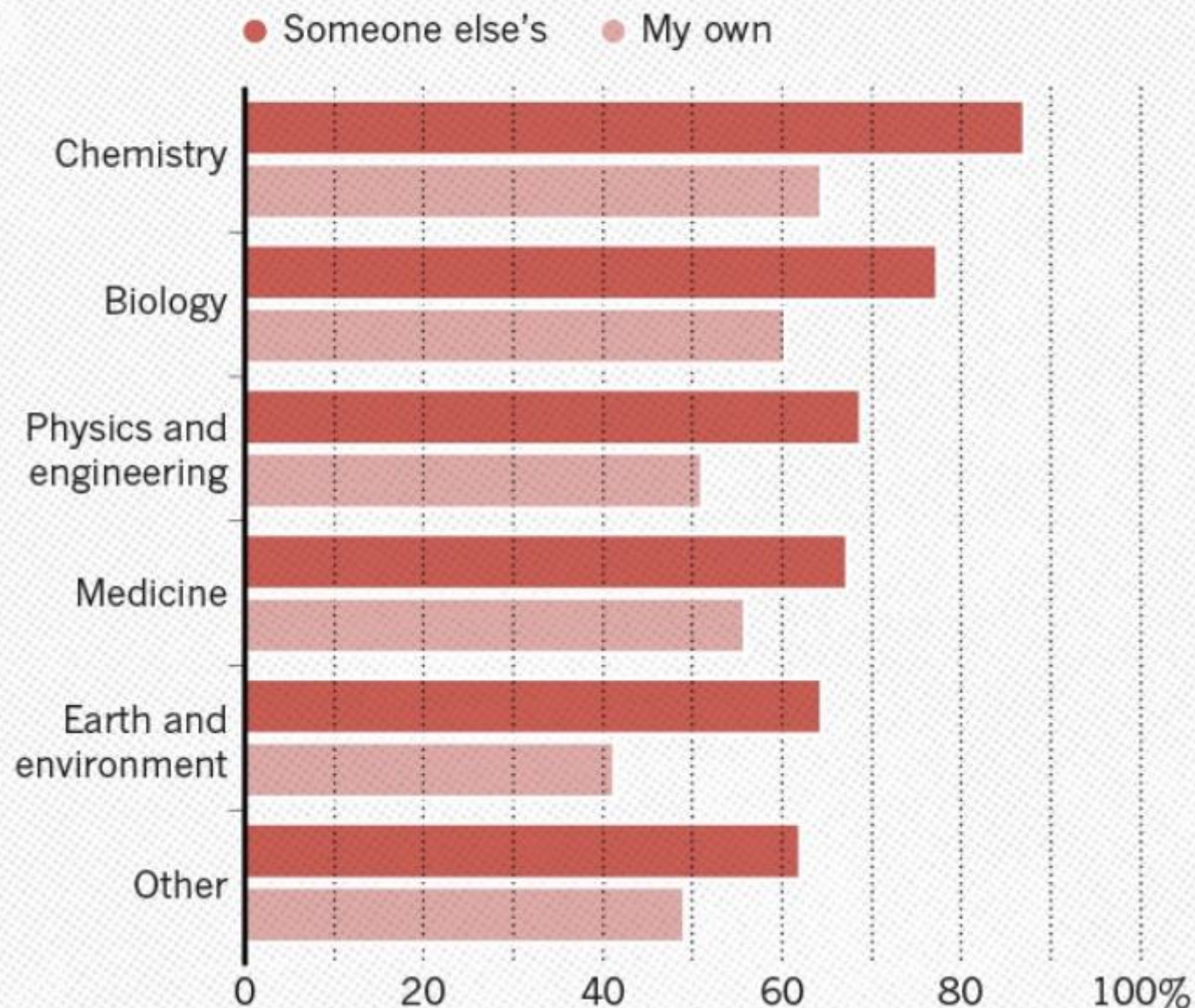
Science 28 Aug 2015:
Vol. 349, Issue 6251, aac4716
DOI: 10.1126/science.aac4716

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.

HAVE YOU FAILED TO REPRODUCE AN EXPERIMENT?

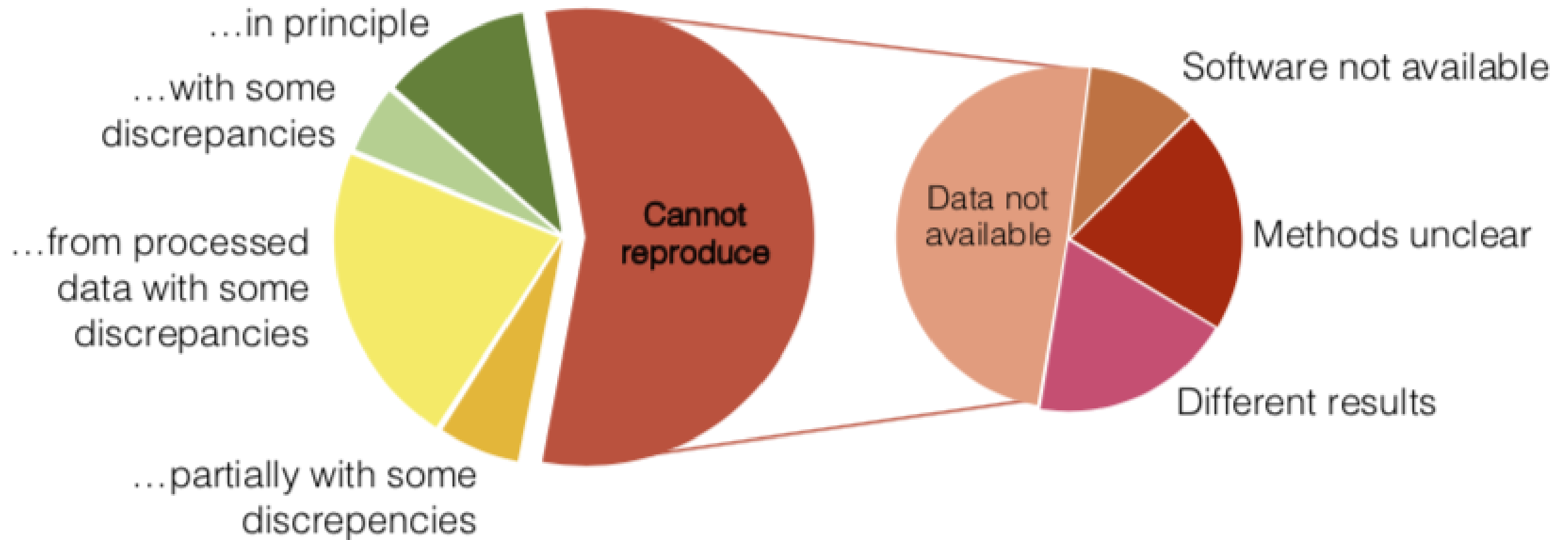
Most scientists have experienced failure to reproduce results.



1,500 scientists lift the lid on reproducibility". Nature. 533: 452–454 - 2016

Replication of data analyses in 18 articles on microarray-based gene expression profiling published in Nature Genetics in 2005–2006:

Can reproduce...



Summary of the efforts to replicate the published articles
Adopted from: Ioannidis et al. Repeatability of published research findings. *Nature Genetics* **41** (2009) doi:10.1038/ng.295

Open Science

Open Data

Open Source

Open Methodology

Open Peer Review

Open Access

Open Educational Resources

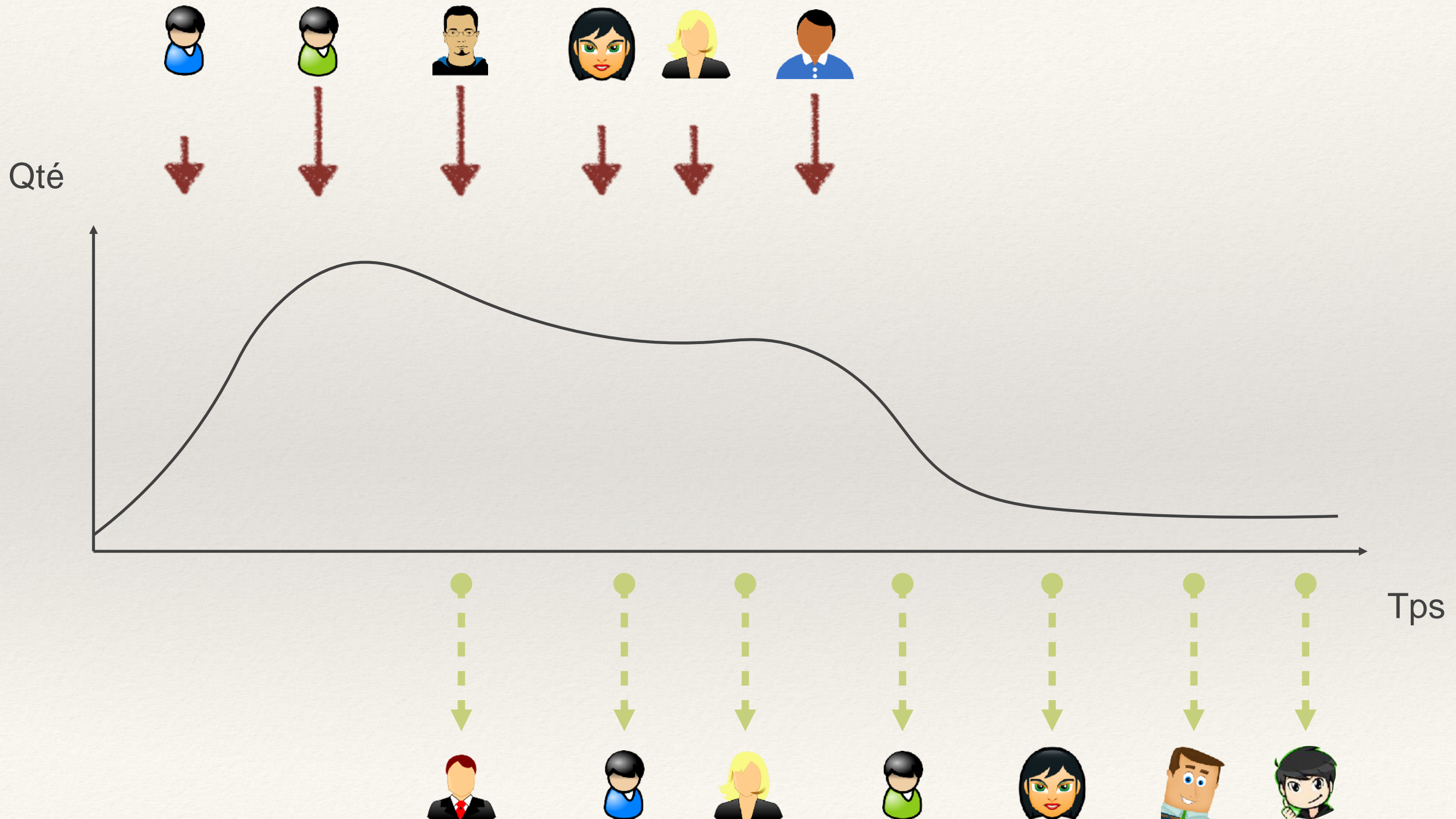
Qu'est-ce que la science ouverte ?

- **La science ouverte, c'est la diffusion sans entrave des publications et des données de la recherche.** Elle s'appuie sur l'opportunité que représente la mutation numérique pour développer l'accès ouvert aux publications et – autant que possible – aux données de la recherche.
- **Son objectif :** faire sortir la recherche financée sur fonds publics du cadre confiné des bases de données fermées. Elle réduit les efforts dupliqués dans la collecte, la création, le transfert et la réutilisation du matériel scientifique. Elle augmente ainsi l'efficacité de la recherche.
- **La science ouverte vise à construire un écosystème dans lequel la science est plus cumulative, plus fortement étayée par des données, plus transparente, plus rapide et d'accès plus universel.** Elle induit une démocratisation de l'accès aux savoirs, utile à la recherche, à la formation, à l'économie, à la société.
- **Elle favorise les avancées scientifiques ainsi que l'innovation,** les progrès économiques et sociaux, en France, dans les pays développés et dans les pays en développement.
- **Elle constitue un levier pour l'intégrité scientifique** et favorise la confiance des citoyens dans la science. Elle constitue un progrès scientifique et un progrès de société.

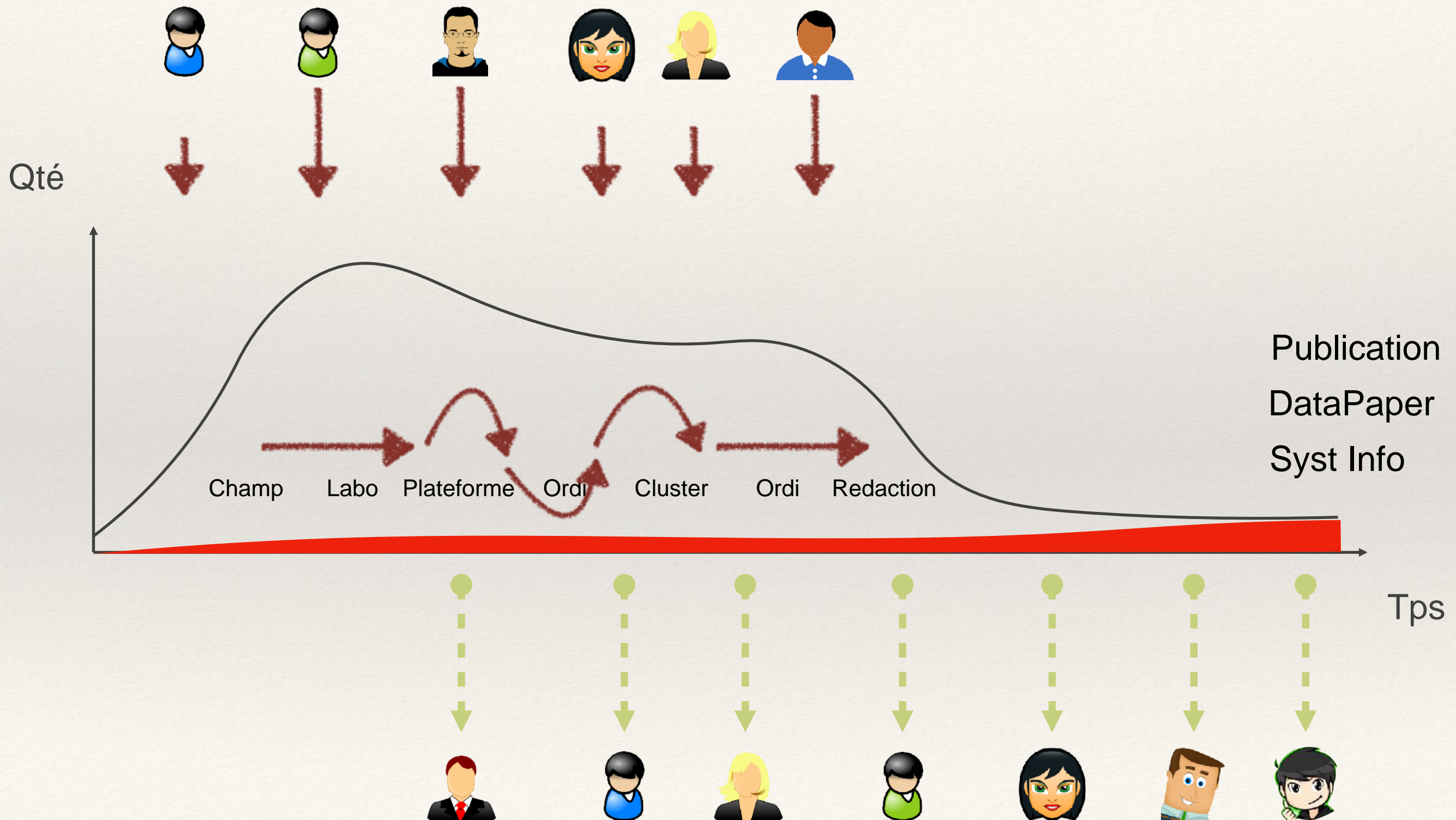
Bon d'accord ...

- ❖ Mais dans la « vraie vie », ça veut dire quoi ?

Un projet sur la durée



Un projet sur la durée



Les freins

- ❖ Volume (toujours croissant)
 - ❖ Vitesse (high throughput, omics ...)
 - ❖ Variété (biologie intégrative, approches méta ...)
 - ❖ Véracité (contrôle qualité)
 - ❖ Valeur (exploitation, ré-exploitation)
- Cela induit une énorme charge mentale pour les équipes

Un double enjeu

- ❖ Personnel
- ❖ Collectif

Des gains importants

- ❖ Confiance
- ❖ Efficience
- ❖ Ré-exploitation



Favoriser les changements de pratiques : carotte et baton

❖ Raisons externes :
Confiance des bailleurs et de la société civile

❖ Raisons internes :
Faciliter l'analyse, la ré-exploitation, le partage

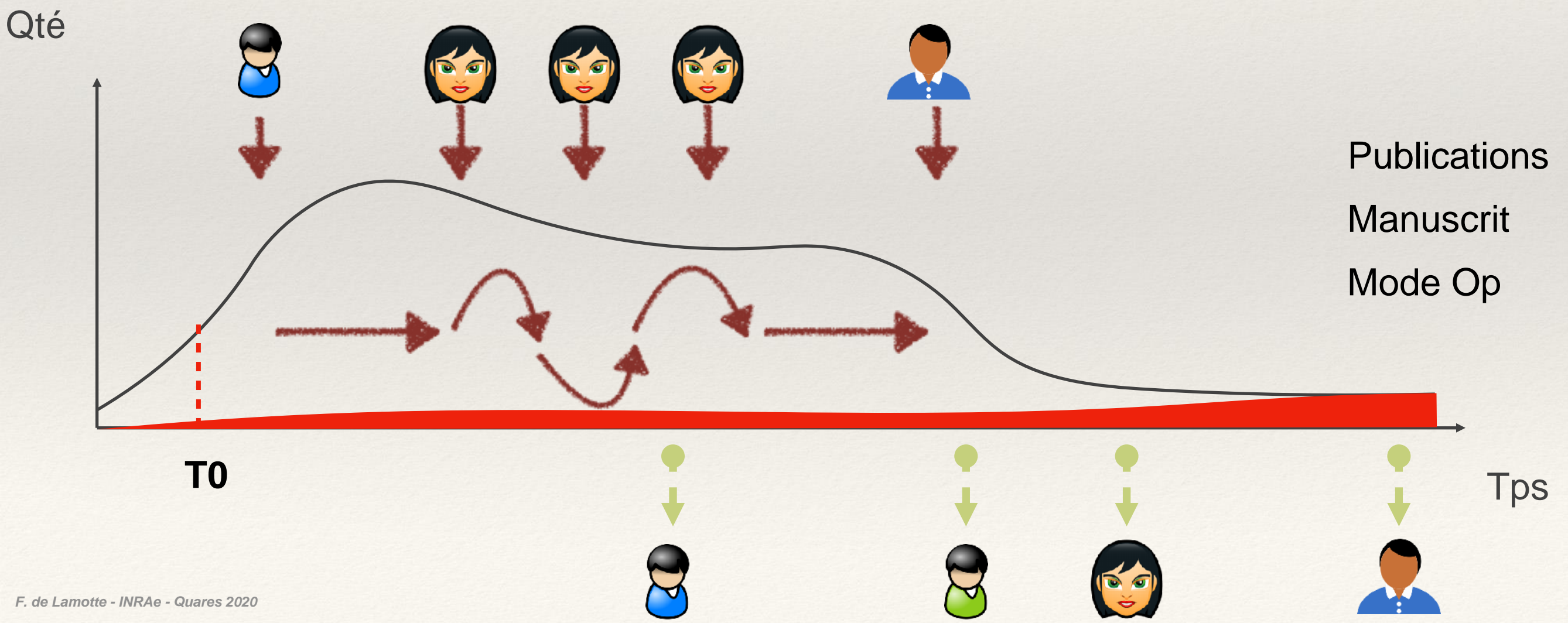
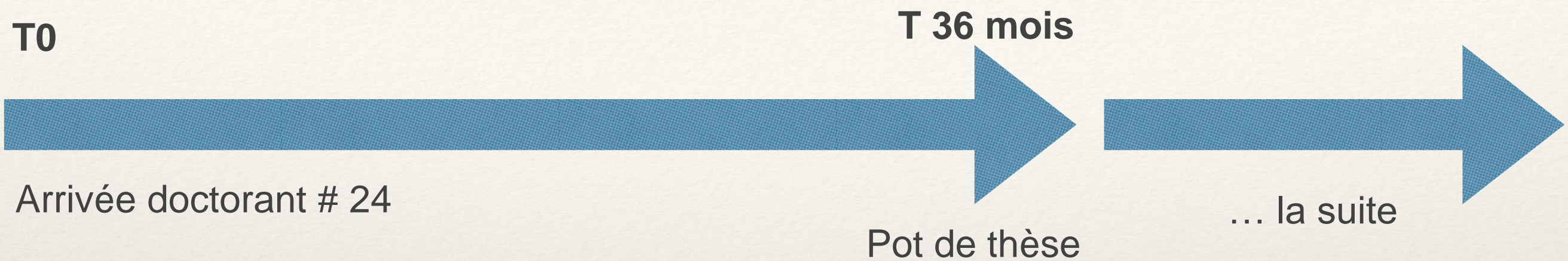


Construire une offre pour les différentes facettes de la Gestion des Données

- ❖ À la recherche d'un cahier des charges de la solution pour traiter tous ces problèmes :
 - de façon intégrée
 - et conviviale



Gérer quoi ? Comment ?



Et vous ...

- ❖ Comment faites (faisiez) vous ?

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**

- ❖ du code
- ❖ des publications

1) Un espace de préservation et de partage du savoir du groupe

The screenshot shows the OSFHOME interface for a project titled "Gérez vos données de la recherche - Formation". The page includes a navigation bar with links like "My Quick Files", "My Projects", "Search", "Support", "Donate", and the user profile "Frédéric de Lamotte". Below the navigation bar, there are tabs for "Files", "Wiki", "Analytics", "Registrations", "Contributors", "Add-ons", and "Settings".

The main content area is divided into several sections:

- Wiki:** A section titled "Page d'accueil" with instructions to write initials and a list of contributors: ADG (Alexandre Dehne Garcia), GVF (Germain Valentin Faity), JFM (Jean-François Martin), FZL (Frédéric de Lamotte), and CKM (Charlotte Kinowski-Moysan).
- Files:** A table listing files stored in OSF Storage (Germany - Frankfurt). The files include presentations (pptx) and spreadsheets (xlsx) related to data management and research.
- Citation:** A section for adding citations, currently empty.
- Components:** A section for adding project components, currently empty.
- Tags:** A section for adding tags, with existing tags: "academic training", "data management", "open data", and "open science".
- Recent Activity:** A list of recent actions, such as linking Mendeley folders and updating wiki pages.
- Mendeley:** A section for managing citations, showing a citation for "Marx, V. (2013). Biology: The big challenges of big data. Nature, 498(7453), 255-..."

Name	Modified
OSF Storage (Germany - Frankfurt)	
00_AtelierFilRouge.pptx	2020-01-22 04:49 PM
01-En route vers l'open science.pptx	2020-01-22 02:08 PM
02- les données de la recherche et l'open data.pptx	2020-01-22 02:06 PM
03- La Vie Des Donnees.pptx	2020-01-22 03:30 PM
04-PlanGestionDonnees.pptx	2020-01-22 04:49 PM
05- gestion des données pendant le projet(2).pptx	2020-01-23 01:58 PM
07_Nommage_format.pptx	2020-01-23 01:59 PM
08_Metadata.pptx	2020-01-23 01:59 PM
09- Diffusez et partagez les données de recherche.pptx	2020-01-23 01:59 PM
10- Droit des données - Cas pratique.pptx	2020-01-24 09:13 AM
4.1_dmpLifeCycleMatrix.xlsx	2020-01-22 04:33 PM
entrepot_doc_parametres_synop_168.pdf	2020-01-23 11:29 AM

Vie pratique

Documents > 6Bioinfo > 2CC2 > 1Admin > 1Approuv

Moinfo13_CreaCompteUtilisCC2_V2.pdf 2.0
Modifié par Patricia Turquay le 15/11/2017 14:29:09

Télécharger

Filex : Echar

Cette application i applications.
Pour en savoir plu

Activités de

- 16:00, 4 juin, 2016:00 Catheri
- 15:00
- 15:41 Catheri
- 15:41 Catheri
- 15:33 Catheri
- 15:32 Catheri
- 15:20 Patricia
- 11:00
- 11:52 Catheri

- 17:00, 3 juin, 2017:00 Najate

- 16:00
- 16:12 Catheri
- 16:10 Catheri
- 16:08 Catheri
- 16:07 Catheri
- 16:07 Catheri

- 15:00
- 15:58 Catheri
- 15:57 Catheri
- 15:29 Dominic

- 16:00, 30 mai, 2016:32 Najate
- 16:13 Najate

- 14:00, 29 mai, 2014:10 Catheri
- 14:04 Catheri
- 14:04 Catheri
- 14:04 Catheri
- 14:03 Catheri

- 13:00
- 13:57 Catheri
- 13:57 Catheri
- 13:55 Catheri
- 13:25 Stépha

- 11:00
- 11:42 Françoi

Mode opératoire Equipe ID	MO-SCI-Info-013 20/09/2017 Page 1 sur 9 Moinfo13_CreaCompteUtilisCC2_V2-2.docx
Gestion compt Util CC2	

Gestion (Création/ Suppression) d'un compte ou d'un groupe utilisateur sur le cluster de Calcul CC2

Historique des versions

Version	Date de version	Gestionnaire : Responsable qualité
Version 00	30/03/2015	Historique des modifications
Version 01	09/11/2015	Création
Version 02	20/09/2017	Révision
		Modification CC2

Confidentialité : Administrateur Cluster

Sommaire

- Objectif 1
- Contacts 1
- Mots de passe 1
- Création compte utilisateur 2
- Gestion des groupes pour un compte utilisateur 4
- Suppression compte utilisateur 5
- Installation ldapbrowser sur le poste administrateur 6
- Documents associés et bibliographie 8

Fiche de visas Erreur ! Le signet n'est pas défini.

1. Objectif

L'objectif du mode opératoire est de décrire la gestion (création et suppression) des comptes et groupes utilisateurs sur le cluster de calcul.

2. Contacts

DSI :
Philippe GRANIER, philippe.granier@cirad.fr, 44 81

Secrétariat DSI :
Sylvie SATGER, sylvie.satger@cirad.fr, 56 75
Agnès CARBON, agnes.carbon@cirad.fr, 58 18

3. Mots de passe

Se référer à la procédure **PR-SCI-Info-003** «Gestion des Mots de Passe».

CONSULTATION de la dernière version du document sur Alfresco Agap-QSE

Mode opératoire Equipe ID	MO-SCI-Info-013 20/09/2017 Page 2 sur 9 Moinfo13_CreaCompteUtilisCC2_V2-2.docx
Gestion compt Util CC2	

4. Création compte utilisateur

- Si nécessaire, installer et configurer ldapbrowser sur son poste (cf § 7).
- Vérifier que l'utilisateur dispose d'un compte ldap dans l'annuaire CIRAD avec ldapbrowser.

Si tel n'est pas le cas, demander au 8855 la création d'un compte messagerie CIRAD pour l'utilisateur.

Utiliser ce formulaire : <http://8855.cirad.fr/forms/getform.asp?dform=LDAPAdd>

Attention : bien renseigner tous les champs pour éviter un retour avec demande d'explication ou demande de validation par diragap.
- Le compte utilisateur sur le cluster sera identique au champ uid de l'utilisateur dans l'annuaire ldapref.
- Sous SAGA (<http://adminannuaire.cirad.fr/saga/droitsApplication.php>), affecter le privilège « users » pour l'application « cluster ».

Droits par applications

Espace documentaire : Alfresco Share

Application cluster

Dept. : Recherche | Département de

Modifier

utilisateursgroupes

id	début	fin
1		
2		

Ajouter un nouveau droit | Supprimer un droit | Modifier les périodes de validité | Annuler

5. Se connecter root en ssh sur le nœud d'administration (cc2-login).

CONSULTATION de la dernière version du document sur Alfresco Agap-QSE

Actions sur le document

- Télécharger
- Voir dans le navigateur
- Editer les propriétés
- Importer une nouvelle version
- Editer hors-ligne
- Copier vers...
- Déplacer vers...
- Supprimer le document
- Démarrer un workflow
- Gérer les droits d'accès
- Gérer les aspects
- Changer le type
- Publier

Tags

(Aucun)

Partager

Copier ce lien pour partager la page actuelle

<https://collaboratif.cirad.fr/share/page/site/AGAPQSE/document-details?nodeRef=workspace://SpacesStore/>

Copier ce lien pour télécharger ce document

<https://collaboratif.cirad.fr/alfresco/s/d/workspace/SpacesStore/9549bd38-751a-4101-9e22-ac217e0cc742/M>

Copier ce lien pour publier ce document sur EZPublish

<https://collaboratif.cirad.fr/alfresco/s/cirad/publish-node/workspace/SpacesStore/9549bd38-751a-4101-9e22->

Propriétés

Nom: Moinfo13_CreaCompteUtilisCC2_V2.pdf
 Titre: Microsoft Word - Moinfo13_CreaCompteUtilisCC2_V2-2.docx
 Description: (Aucun)
 Type MIME: Adobe PDF Document
 Auteur: (Aucun)
 Taille: 1 Mo
 Créateur: turquay
 Date de création: 07/10/2015 16:25:36
 Modificateur: turquay
 Date de Modification: 15/11/2017 14:29:09

Droits d'accès

Coordinateurs: Gestionnaire
 Collaborateurs: Collaborateur
 Contributeurs: Contributeur
 Lecteurs: Lecteur
 Tous les autres utilisateurs: Aucun privilège

Workflow

Ce document ne fait partie d'aucun workflow.

Historique des versions

Dernière version

2.0 Moinfo13_CreaCompteUtilisCC2_V2.pdf

Patricia Turquay il y a un an environ
(Pas de commentaire)

Versions antérieures

Commentaires

Pas de commentaire

2) Les sources bibliographiques

<u>Bebop</u>	ALaRI Institute	2007-11-08	2009-11-10	1.1	Free	Yes	BSD	web-based BibTeX front-end (Apache, PHP, MySQL)
<u>BibDesk</u>	BibDesk developers	2002-04	2019-06-03	1.7.1	Free	Yes	BSD	BibTeX front-end + repository; Cocoa -based; integration with Spotlight
<u>BibSonomy</u>	University of Kassel	2006-01	2018-07-30	3.8.13	Free	Yes	AGPL , GPL , LGPL ^[2]	centrally hosted website
<u>Bibus</u>	Bibus developers	2004-06-03	2013-05-23	1.5.1	Free	Yes	GNU GPL	discontinued?
<u>JabRef</u>	JabRef developers	2003-11-29	2018-06-05	4.3.1	Free	Yes	MIT license	Java BibTeX and BibLaTeX manager
<u>KBibTeX</u>	KBibTeX developers	2005-08	2018-06-21	0.8.1	Free	Yes	GNU GPL	BibTeX front-end, using the KDE Software Compilation
<u>Pybliographer</u>	pybliographer developers	1998-10-30 (0.2)	2018-04-03	1.4.0	Free	Yes	GNU GPL	Python / GTK2
<u>refbase</u>	refbase developers	2003-06-03	2014-02-28	0.9.6	Free	Yes	GNU GPL	web-based for institutional repositories/self-archiving ^[9]
<u>RefDB</u>	refdb developers	2001-04-25	2007-11-05	0.9.9	Free	Yes	GNU GPL	network-transparent; XML / SGML bibliographies
<u>Referencer</u>	Referencer developers	2008-03-15	2014-02-27	1.2.2	Free	Yes	GNU GPL	BibTeX front-end
<u>Wikindx</u>	Mark Grimshaw	2004-02	2019-08-20	5.8.2	Free	Yes	CC-BY-NC-SA	web-based
<u>Zotero</u>	Roy Rosenzweig Center for History and New Media at GMU	2006	2019-06-14	5.0.67 ^[10]	Free / Online storage free up to 300 MB / Additional space available	Yes	AGPL	Multi-platform version with connectors for Firefox , Chrome and Safari . Web-based access to reference library also available.

Les différentes facettes de la recherche

- ❖ Analogique
- ❖ Numérique
- ❖ Algorithmique

Analogique

- ❖ échantillons
- ❖ cahiers de labo



Analogique

- ❖ Règles de nommage (partagées !)
- ❖ Plan de classement (physique)
- ❖ Règles (et contrôle) de tenue du cahier de labo
- ❖ Index des cahiers (qui, quand, quel projet ...)

- ❖ *Lien vers l'alter ego numérique*

3) CLÉ

Les cahiers de laboratoire

Version papier



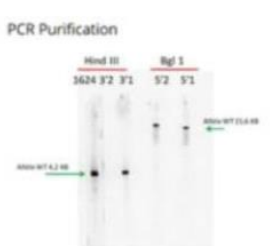
Version numérique (ELN)

Taq DNA Polymerase

The Polymerase Chain Reaction (PCR) is a powerful and sensitive technique for DNA amplification. Taq DNA Polymerase is an enzyme widely used in PCR. The following guidelines are provided to ensure successful PCR using NEB's Taq DNA Polymerase. These guidelines cover routine PCR. Amplification of templates with high GC content, high secondary structure, low template concentrations, or amplicons greater than 5 kb may require further optimization.

Step	Initial Denaturation at 95°C	95°C	95°C	95°C
1	Initial Denaturation at 95°C	95°C	95°C	95°C
2	95°C	95°C	95°C	95°C
3	95°C	95°C	95°C	95°C
4	95°C	95°C	95°C	95°C

PCR Purification



1624 32 31 32 31

1624 32 31 32 31

Journal de bord:

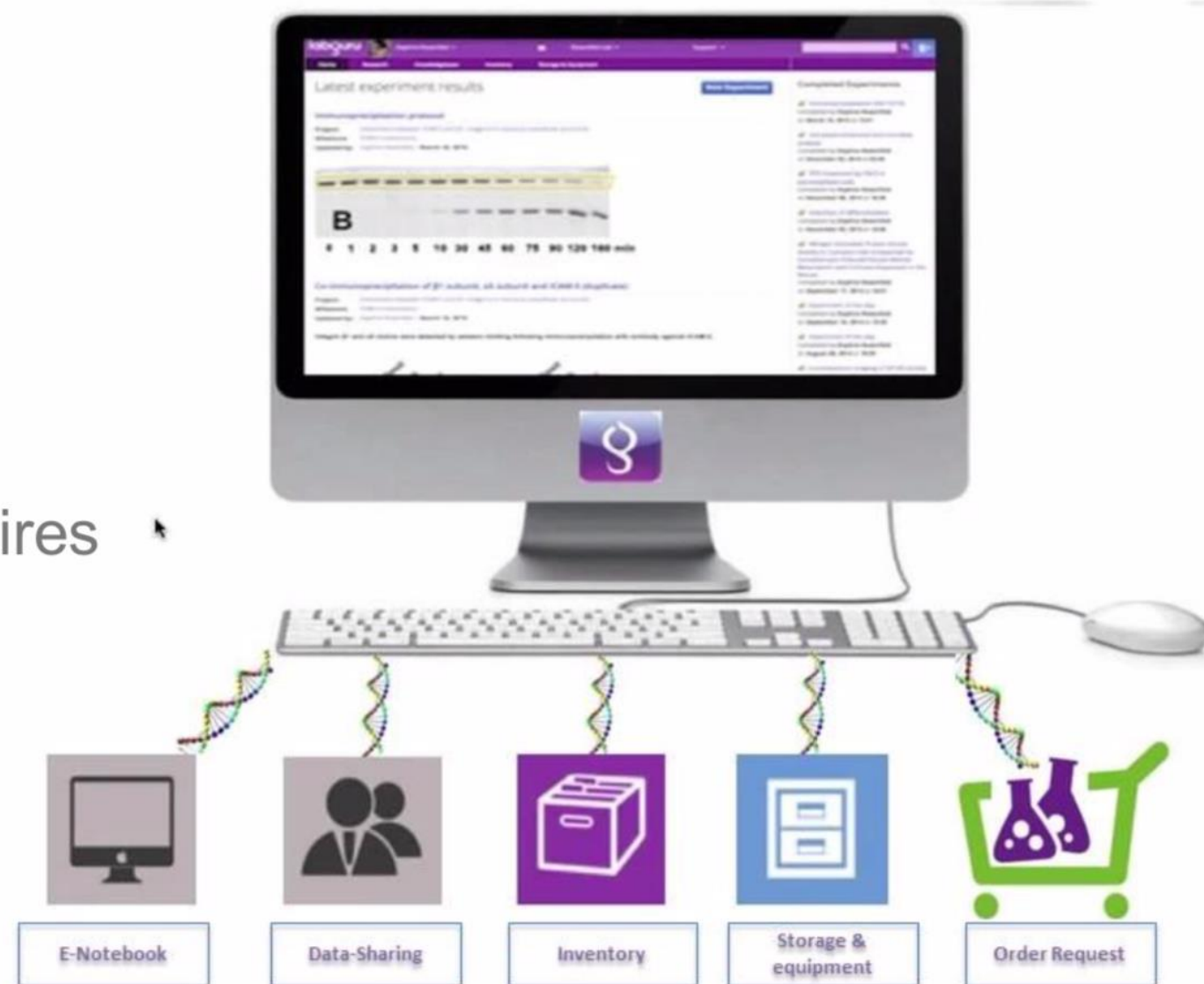
- Preuve juridique
- Traçabilité, qualité
- Gestion de la connaissance

- Horodatage, versioning, signature électronique
- Gestion des inventaires et instruments
- Indexation, moteur de recherche
- Intégration des données numériques (données de recherche)

- + *Travail collaboratif et accès distant*
- + *Management des équipes et des projets*

Présentation de CLÉ, le cahier de Laboratoire Electronique de l'Inserm

- Collaboratif
- Accessible à distance
- Dédié à la biologie
- Proposant une gestion des inventaires

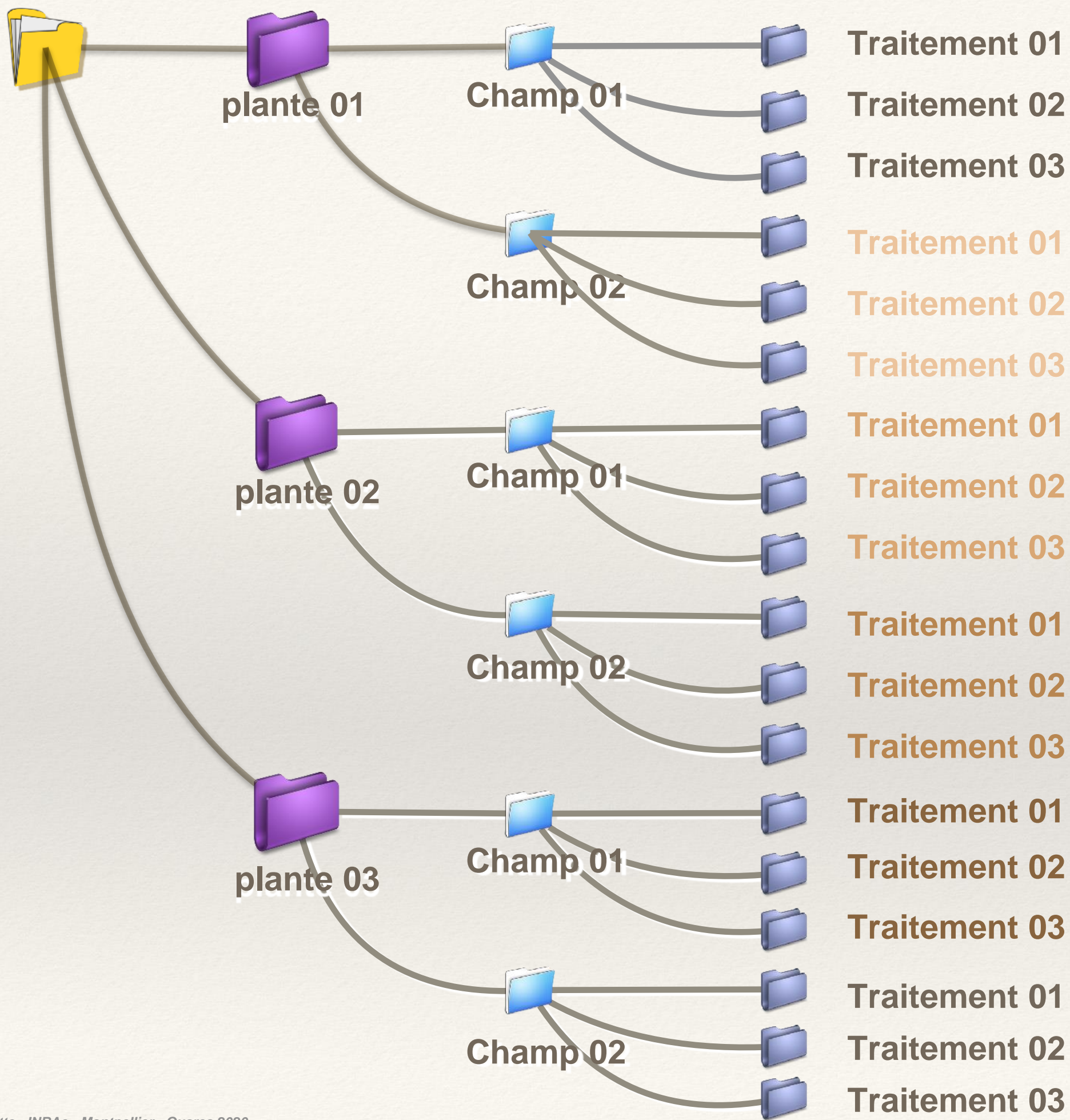


Numérique

- ❖ Bonnes pratiques pour :
 - Retrouver aisément les données
 - Rationaliser le contenu des serveurs et éviter leur saturation !
 - Pérenniser les informations et faciliter leur transmission

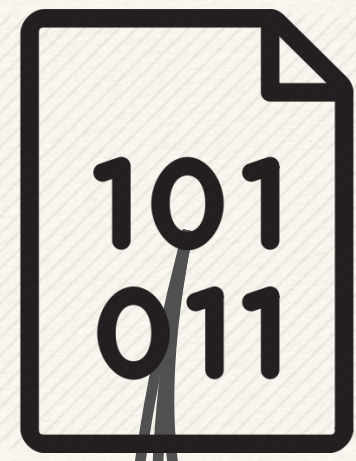
Points de vigilance

- ❖ Nommage des fichiers
 - ❖ Unique, non ambigu, caractères « légaux »
- ❖ Formats de fichiers
 - ❖ facile.cines.fr
 - ❖ Que faire des formats propriétaires ?
- ❖ Plan de classement ?

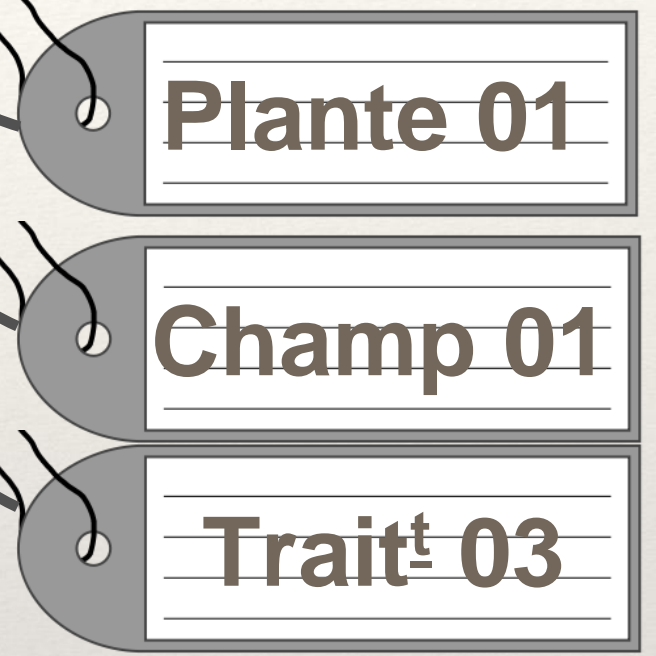


4) Les métadonnées





toto.txt



Tags
MetaData

Vocabulaire contrôlé
Défini par la communauté
Evolutif

Un vocabulaire contrôlé



Les standards de métadonnées

ABCD - Access to Biological Collection Data

The [Access to Biological Collections Data \(ABCD\) Schema](#) is an evolving comprehensive standard for the access to and exchange of data about specimens and observations (a.k.a. primary biodiversity data). The ABCD Schema attempts to be comprehensive and highly structured, supporting data from a wide variety of databases. It is compatible with several existing data standards. Parallel structures exist so that either (or both) atomised data and free-text can be accommodated.

Sponsored by Biodiversity Information Standards TDWG - the Taxonomic Databases Working Group, the current specification was last modified in 2007.

AgMES - Agricultural Metadata Element Set

A semantic standard developed by the Food and Agriculture Organization (FAO) of the United Nations, AgMES enables description, resource discovery, interoperability and data exchange of different types of information resources in all areas relevant to food production, nutrition and rural development.

Sponsored by the UN AIMS - Agricultural Information Management Standards, the current standard was issued in November 2010.

AVM - Astronomy Visualization Metadata

The [AVM](#) scheme supports the cross-searching of collections of print-ready and screen-ready astronomical imagery rendered from telescopic observations (also known as 'pretty pictures'). The scheme is compatible with the [Adobe XMP](#) specification, so the metadata can be embedded within common image formats such as JPEG, TIFF and PNG.

Such images can combine data acquired at different wavebands and from different observatories. While the primary intent is to cover data-derived astronomical images, there are broader uses as well. Specifically, the most general subset of this schema is also appropriate for describing artwork and illustrations of astronomical subject matter.

AVM is a proposed recommendation of the International Virtual Observatory Alliance and was last updated in 2011.

CERIF - Common European Research Information Format

CERIF (Common European Research Information Format) is the standard that the EU recommends to its member states for recording information about research activity. Since version 1.6 it has included specific support for recording metadata for datasets.

It is developed and maintained by EuroCRIS. The current version is 1.5, but the version 1.6 model has been available for testing and review since 24 July 2013.

CF (Climate and Forecast) Metadata Conventions

The CF standard was originally framed as a standard for data written in netCDF format, with model-generated climate forecast data particularly in mind. However, it is equally applicable to observational datasets, and can be used to describe other formats. It is a standard for "use metadata" that aims both to distinguish quantities (such as physical description, units, and prior processing) and to locate the data in space-time.

Contact us

Search

nity

Tailored support

Standards

Contribute by adding a standard

Any problems? Please tell us!

The standards in FAIRsharing are manually curated from a variety of sources, including [BioPortal](#), [MIBBI](#) and the [Equator Network](#).



Search Standards

Search

Search

Reset

Advanced

Did we miss something? Our COVID-19 Collection is growing, but not yet complete. Tell us what's missing. Email contact@fairsharing.org or add missing resources to our spreadsheet

Showing records 1 - 50 of 1428.

« 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 »

Registry	Name	Abbreviation	Type	Subject	Domain	Taxonomy	Related Database	Related Standard	Related Policy	In Collection/Recommendation	Status
	ABA Adult Mouse Brain	ABA	Standard	Neuroscience	Brain, Gene Expression, Brain Imaging	Mus musculus	NeuroMorpho.Org	None	None	None	
	Access to Biological Collection Data	ABCD	Standard	Biodiversity, Biology, Life Science	None	All	GBIF ALA IPT - GBIF Australia Repository, GBIF Spain IPT - GBIF Spain Repository, Canadensys IPT - GBIF Canadensys Repository, SIB Colombia IPT - GBIF Colombia Repository, Plus 1 more...	GGBN ABCDDNA ABCDEFG	None	TDWG Biodiversity Information Standards	
	Access to Biological Collection Databases Extended for Geosciences	ABCDEFG	Standard	Earth Science, Geology, Paleontology, Soil Science	None	All	GeoCAsE Data Portal	XML ABCD	None	None	
	Access to Biological Collection Data DNA extension	ABCDDNA	Standard	Biodiversity, Biology, Life Science	DNA Sequence Data, Experiment Metadata, Sequence, Deoxyribonucleic Acid, Polymerase Chain Reaction, Plus 1 more...	All	GenBank	GGBN MOD-CO ABCD	None	TDWG Biodiversity Information Standards	

View as Table

View as Grid

Sort by

Name

Recommended Records

Recommended

Associated Publication?

No Publication

Has Publication

Claimed?

No Maintainer

Has Maintainer

Record Status

Uncertain

Deprecated

In development

Ready

Standard Type

Terminology Artifact

789

Model/Format

423

Reporting Guideline

170

Metric

30

Identifier Schema

16

Genre	Espece	Sous espece	Groupe	Nom
Oryza	Sativa		japonica	PENTHE BLANC
Oryza	Sativa		japonica	PENTHE NOIR
Oryza	Sativa		indica	ZOGO
Oryza	Glaberrima			GBAI-GBAI
Sorghum	bicolor	bicolor	Dura	IS19453
Sorghum	bicolor	bicolor	Dura	IS19453
Sorghum	bicolor	bicolor	Dura	IS19453
Sorghum	bicolor	bicolor	Dura	IS19453
Sorghum	bicolor	bicolor	Dura	IS19453
Musa	acuminata	banksii	wild	Banksii H09

Germplasma	Origine	Collection
AG0003	Guinea	prospection 1979
AG0004	Guinea	prospection 1979

Type de sequençage	Taille insert	Longueur de read	type de machine	Lieu du sequençage
illumina		1*150	HiSeq3000	Genotoul
illumina		1*150	HiSeq3000	Genotoul

Attention

- ❖ Les métadonnées sont LA solution
- ❖ Cette solution est difficile à mettre en oeuvre

la discussion reste ouverte

On les met où toutes ces données ?

❖ Dès le début !!!

Support de stockage	Sécurité	Accès	Coût	Remarque d'utilisation
 <p>Ordinateur professionnel</p>	<p>Elle perd sa thèse dans un train Bordeaux-Nîmes, son père lance un appel</p> <p><small>A LA UNE / LOT-ET-GARONNE / MARMANDE / Publié le 03/04/2019 à 11h56. Mis à jour le 03/06/2019 par SudOuest.fr.</small></p> <p>S'ABONNER À PARTIR DE 1€    COMMENTAIRES SUSPENDUS</p>  <p><small>▲ La jeune femme a perdu son ordinateur, dans un train Intercités ©BATS PASCAL © / BATS PASCAL</small></p>			<p>Package temporaire ne crypter les données es et sensibles</p>
 <p>Support externe</p>				<p>Package temporaire ne crypter ou de sécuriser nt les données es et sensibles</p>
 <p>Serveur institutionnel</p>				<p>Package plus pérenne ur le stockage de sibles et des versions e vos données stitutions ne proposent ce</p>
 <p>Serveur Cloud</p>	<p>données, n'ce qu'elles deviennent</p>	<p>les personnes ayant été autorisées au partage</p>	<p>stockage</p>	<p>Partage avec des ternes à l'institution ette de données confidentielles</p> <p>- Pas de contrôle sur la procédure de sauvegarde des données</p>

5) Stockage le long du projet

- ❖ Sécurité
- ❖ Performance (calcul HPC; accès simultanés)
- ❖ Prix
- ❖ Accès partagés (*multi tutelles ...*)
- ❖ Comment gérer la fin du projet

Initiatives de mutualisation en cours

genossh.genouest.org (ylebras)

Terminal Sessions View X server Tools Games Settings Macros Help

Session Servers Tools Games Sessions View Split MultiExec Tunneling Packages Settings Help

Quick connect...

3. genossh.genouest.org (ylebras)

```
#####
#
# Welcome to GenQuest
#
```

/home/symbiose/ylebras/

https://data-access.cesgo.org/index.php/apps/files/?dir=/&fileid=6696

- Tous les fichiers
- Favoris
- Partagés avec vous
- Partagés avec d'autres
- Partagés par lien
- Étiquettes
- Stockage externe
- Fichiers supprimés
- Paramètres

Nom	Type	Partagé	Taille	Date
DB_VRE	Folder		26 KB	il y a 6 mois
PNDP_questionnaireSI	File	Partagé	104 KB	il y a 6 mois
_citations_2019.docx	File		18 KB	il y a 7 mois
PNDP_ateliers_CES_data	Folder	Partagé	1.5 GB	il y a 7 mois
PNDP_ecoinfoFAIR	Folder	Partagé	60.9 MB	il y a 8 mois
Fichiers SIG	Folder	Grégoire LOÏS (glois)	1.4 GB	il y a 8 mois
Projets_PNDP	Folder		70 KB	il y a 9 mois
Tableau général-liste compétences UMR.xls	File	HUGO STRUNA (hstruna)	61 KB	il y a 10 mois
PNDP_biblio	Folder	Partagé	25 KB	il y a 10 mois
PNDP_candidatures2019	Folder	Partagé	6.8 MB	il y a un an
Enseignement	Folder	Emmanuelle Porcher (eporcher)	1.4 GB	il y a un an
PNDP_Objectif_Plancton	Folder	Partagé	107 KB	il y a un an
Plages vivantes	Folder	Partagé	4 MB	il y a un an
GESTION CESCO	Folder	Christine CALVAS (ccalvas)	29.5 MB	il y a un an
PotentialRecommenders.xlsx	File	Karine PRINCE (kprince)	12 KB	il y a un an
2018_ECOSCOPE_PNDP_Rapport_Activites_IR_Fev-2018_Novembre-2018.odt	File		49 KB	il y a 2 ans
Documents officiels	Folder	Grégoire LOÏS (glois)	0 KB	il y a 2 ans
Genouest group cesco	Folder		706.7 MB	il y a 2 ans

core.galaxy ui.yaml 8

Follow terminal folder

```
default-row_fr.js
default-view_fr.js
Dockerfile
PNDB
R
options-menu_fr.js
outil_bloc_galaxy.PNG
```

16:16 01/07/2023

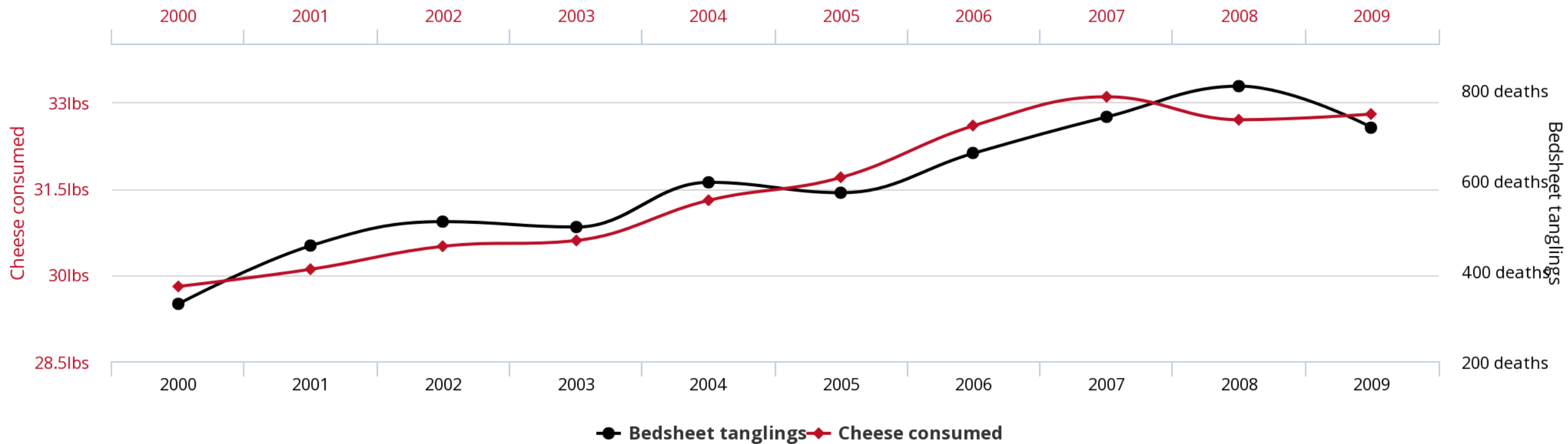
6) Qualification des données

De « ouverte » à « reproductible » : attention à la marche !

Per capita cheese consumption

correlates with

Number of people who died by becoming tangled in their bedsheets



tylervigen.com

Contrôle qualité

- ❖ Valeurs manquantes
- ❖ Valeurs manquantes remplacées par zéro
- ❖ Complétudes des séries
- ❖ Lignes ou valeurs dupliquées
- ❖ Incohérence orthographique
- ❖ Incohérence des formats de dates (1900)
- ❖ 65 536 lignes (ou 255 colonnes)



choice	count
Andy Anderson	79
Andy R. Anderson	9
Anderson, Andy	57
Beatrice Beaufort	28
Beatrice Mansfield	67
Beaufort, Beatrice	19
...	...

Résoudre certains problèmes

- ❖ Encodage
- ❖ CR / LF
- ❖ Données « PDFifiées »
- ❖ Données scannées



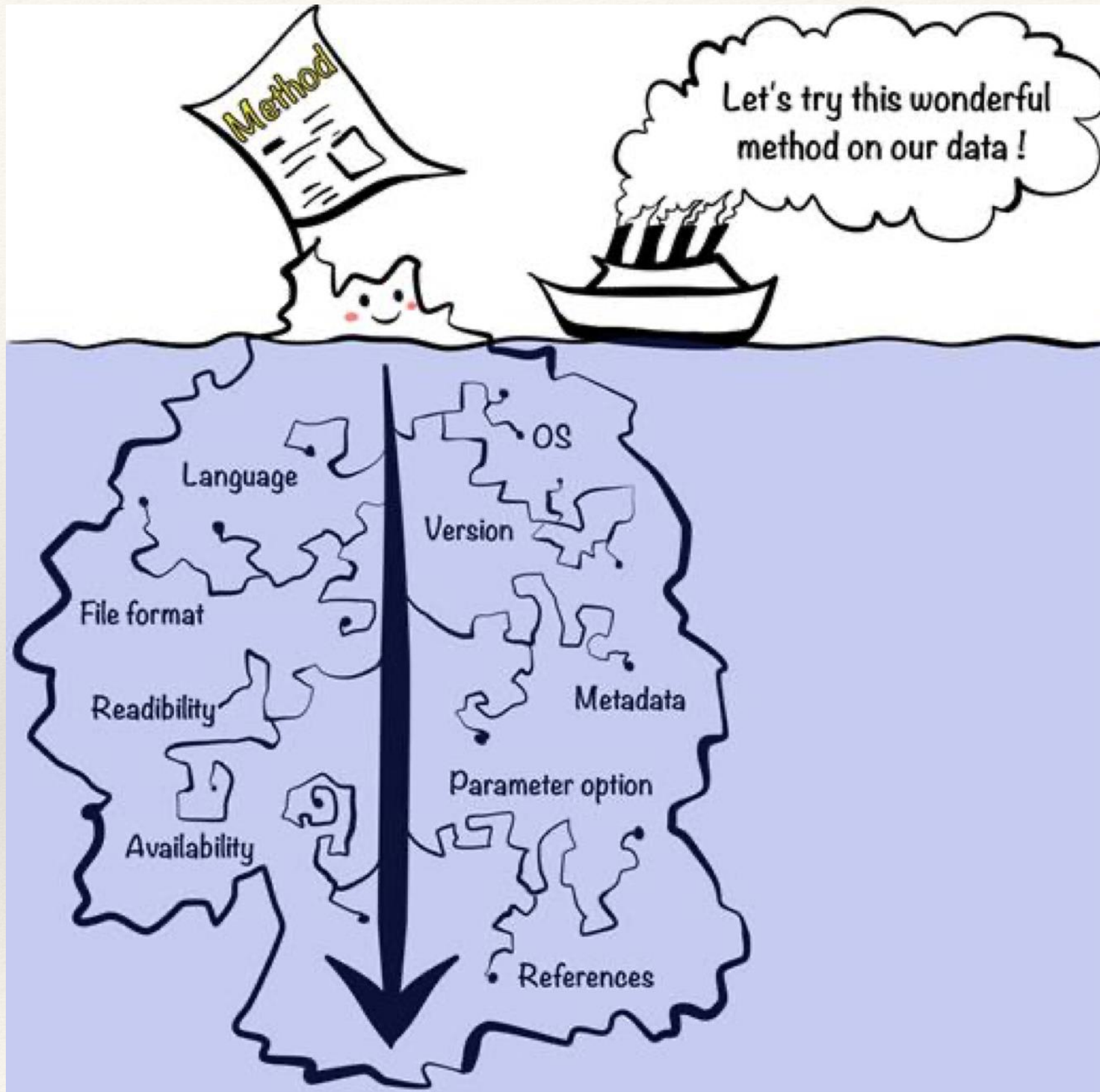
Tabula

Confier le problème à d'autres

- ❖ Source non fiable
- ❖ Opacité du processus de collecte
- ❖ Précision irréaliste des données
- ❖ Valeurs aberrantes inexplicables
- ❖ Echec de la Loi de Benford

Algorithmique

❖ Super facile !!!!! 😊



Qu'est-ce qu'un calcul ?

Input

```
100111100001001100110101101100
001010011101010111110001001101
010111101100011110111011110001
001100001110111000100100000111
110101100111001110100000100110
11011110011100001111101101111
111001001011110001100110000101
011100001000010001011110000010
110101110011101111001010100111
111000101110011001101101001001
011001010100101011000001001100
110100111001011111100001011101
011110111110001111011110101101
000001110110011001010101011100
100010110001100000111001100010
000000111011100100100101010111
000010000001100001000010110110
101111101111000111100101110101
100101010100001001110100010001
011110011010100101111011110101
100011000110110001011101100110
110100000100000011011000001101
100000011100100111101101011011
010110010001000101110111001010
```



Output

```
000000111011001010000100111011
110000110111011101111110101010
000110010101111100101110110100
001110110011010110000101011010
111101111100000100010111010111
111010001000010010111100111001
111001100101000111101000011100
10111110000011011011011110001
100100110111101111000101010100
111110011010111011010011011100
111011100011110101011111000100
010111011010100100011110100011
001111000001111110001011100111
101101100000100011100111110011
001101000010011000110011000011
101011110111101010000011010001
010111100101010010011100011011
001010101100101000001010000110
100000101001110011010000011100
001110011000111111111000001100
100100010100000110001011010000
010110010111101001000010100010
101011110001001001010010111000
011000100000010000000011100111
```

Computer by Creative Stall from the Noun Project

Input

Output

100111100001001100110101101100
001010011101010111110001001101
010111101100011110111011110001
001100001110110100100000111
11010110011100000100110
11011110011100001111101101111
111001001011110001100110000101
011100001000010001011110000010

110101110011101111001010100111
111000101110011001101101001001
011001010100101011000001001100
11010011100101111100001011101
011110111110001111011110101101
000001110110011001010101011100
100010110001100000111001100010
000000111011001010101010111
00001000001100010010110110
101111101111000111100101110101
100101010100001001110100010001
011110011010100101111011110101
100011000110110001011101100110
110100000100000011011000001101
100000011100100111101101011011
010110010001000101110111001010

Data

Code



000000111011001010000100111011
1100001101110111101111110101010
000110010101111100101110110100
001110110011010110000101011010
111101111100000100010111010111
111010001000010010111100111001
111001100101000111101000011100
101111110000011011011011110001
100100110111101111000101010100
111110011010111011010011011100
111011100011110101011111000100
0101110110110111100011110100011
001111000001111110001011100111
101101100000100011100111110011
001101000010011000110011000011
101011110111101010000011010001
010111100101010010011100011011
001010101100101000001010000110
100000101001110011010000011100
001110011000111111111000001100
100100010100000110001011010000
010110010111101001000010100010
101011110001001001010010111000
011000100000010000000011100111

Data

Input

Output

100111100001001100110101101100
001010011101010111110001001101
010111101100011110111011110001
0011000011101100100100000111
11010110011100000100110
11011110011100001111101101111
111001001011110001100110000101
011100001000010001011110000010

110101110011101111001010100111
111000101110011001101101001001
011001010100101011000001001100
110100111000101011011101
01111011111000111011110101101
000001110110011001010101011100
100010110001100000111001100010
000000111011100100100101010111

000010000001100001000010110110
101111101111000111100101110101
100101010100001001110100010001
011110011010100101111011110101
10011000101000010110010110110
110100000100000011011000001101
100000011100100111101101011011
010110010001000101110111001010

Data

Program

Environment

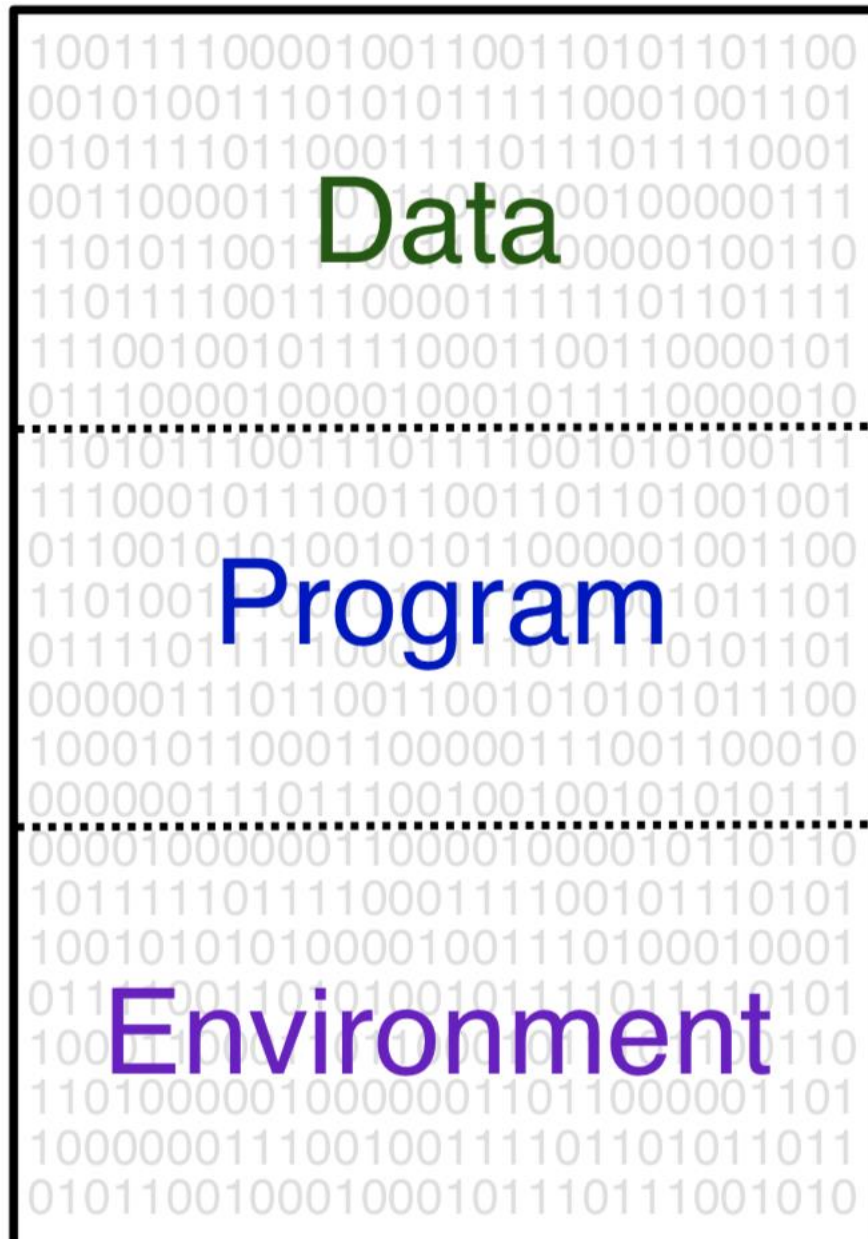


000000111011001010000100111011
110000110111011101111110101010
000110010101111100101110110100
001110110011010110000101011010
111101111100000100010111010111
111010001000010010111100111001
111001100101000111101000011100
101111110000011011011011110001
100100110111101111000101010100
1111100110110111011010011011100
111011100011110101011111000100
01011101101101111110100011
001111000001111110001011100111
101101100000100011100111110011
001101000010011000110011000011
101011110111101010000011010001
010111100101010010011100011011
001010101100101000001010000110
100000101001110011010000011100
001110011000111111111000001100
100100010100000110001011010000
010110010111101001000010100010
101011110001001001010010111000
011000100000010000000011100111

Data

à éviter ...

Input



my research

my colleagues' code

stuff I don't care about

Tools

search tools

- Get Data
- Send Data
- BASIC TOOLS
 - Text Manipulation
 - Convert Formats
- SEQUENCE ANALYSIS
 - Gene/Protein prediction
 - Fetch Sequences
 - Fetch Alignments
 - Extract Features
 - Genomics
- NGS ANALYSIS
 - NGS: Miscellaneous
 - NGS: Quality Control
 - NGS : Mapping
 - NGS: RNASeq
 - VCFtools
 - NGS: Small RNAs
- Picard Tools
- EVOLUTION/PHYLOGENY
 - Comparative Genomics
 - Genfam
 - Protein analyses
- SNP ANALYSIS
 - Tassel GBS (Version 4.0)
 - Rice Variant Analysis (Rice 3k, IRIGIN, High Density Rice Array (HDRA, 700k SNPs))
 - Varscan
 - Population structure
 - GWAS
- METAGENOMICS
 - FROGS
- STATISTICS/GRAPHICS
 - Filter and Sort
 - Join, Subtract and Group
- GENOME HARVEST
 - Phylogenetic Analysis
 - TransPo-RG Transfer of Position to Resequenced Genome
 - Cis/Trans Effect
 - Visualization
 - TraceAncestor



Welcome to Galaxy

Our pre-configured and validated workflows

1 / 6

Chromosome reconstruction

Scaffrehunter tools assemble scaffolds into pseudomolecules using markers genotyped in a population (Martin et al, 2016)

Input: Fastq + FASTA
[Access workflow](#)



These workflows as part of the services provided by South Green

Web Thread CPU



Register

Create an account to get 5Go of available space to test our Galaxy platform.



Contact

South Green members or partner benefits from 25Go (please contact admin-bioinfo@cirad.fr).



Big Data

To improve your available space please contact admin-bioinfo@cirad.fr

How to load big datasets ?



History

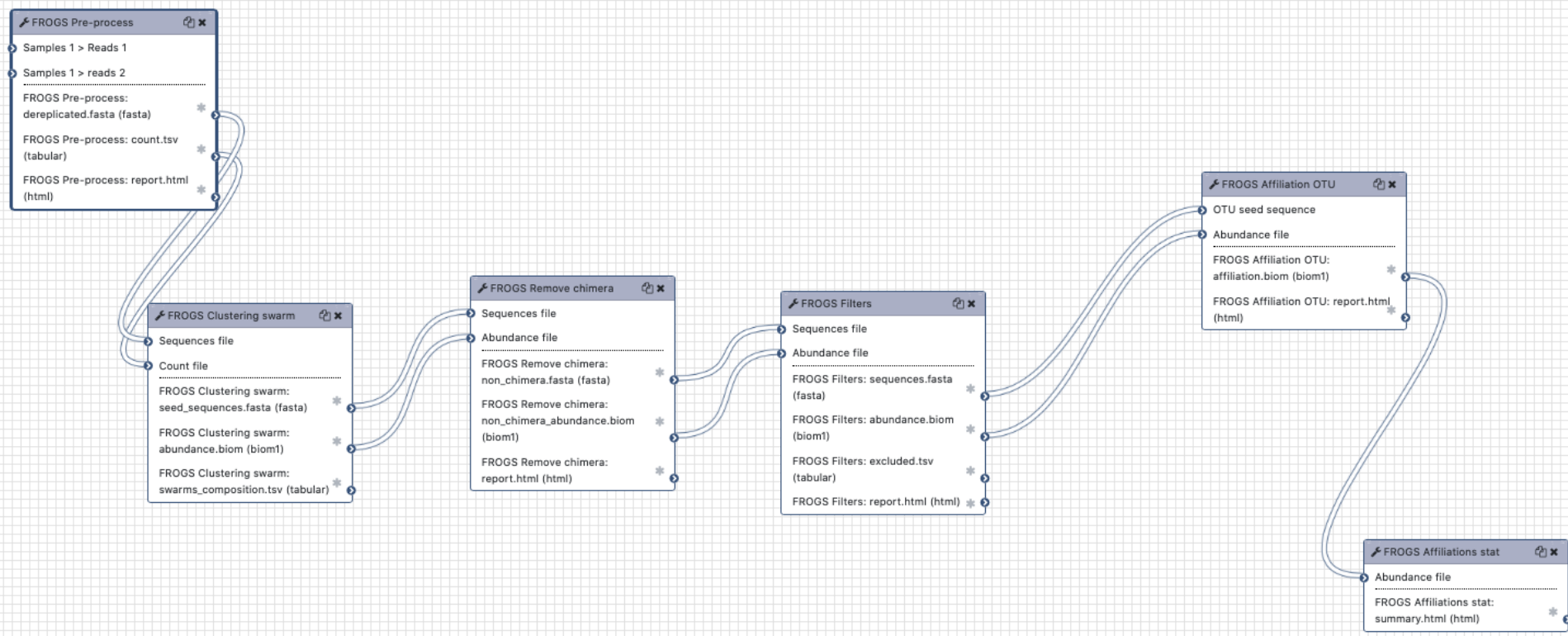
search datasets

Unnamed history

- 5 shown
- 10.12 KB
- 5: [dottup on data 1 and data 2](#)
- 4: [dottup on data 3 and data 2](#)
- 3: [Splign_waxy.fas](#)
- 2: [Ref_Waxy.fas](#)
- 1: [Aiem_Waxy.fas](#)

- Tools
- search tools
- Inputs
- Get Data
- Send Data
- BASIC TOOLS
- Text Manipulation
- Convert Formats
- SEQUENCE ANALYSIS
- Gene/Protein prediction
- Fetch Sequences
- Fetch Alignments
- Extract Features
- Genomics
- NGS ANALYSIS
- NGS: Miscellaneous
- NGS: Quality Control
- NGS : Mapping
- NGS: RNASeq
- VCFtools
- NGS: Small RNAs
- Picard Tools
- EVOLUTION/PHYLOGENY
- Comparative Genomics
- Genfam
- Protein analyses
- SNP ANALYSIS
- Tassel GBS (Version 4.0)
- Rice Variant Analysis (Rice 3k, IRIGIN, High Density Rice Array (HDRA, 700k SNPs))
- Varscan
- Population structure
- GWAS
- METAGENOMICS
- FROGS
- STATISTICS/GRAPHICS
- Filter and Sort
- Join, Subtract and Group
- GENOME HARVEST
- Phylogenetic Analysis
- TransPo-RG Transfer of Position to Resequenced Genome
- Cis/Trans Effect
- Visualization
- TraceAncestor
- TraceHaplotype
- vcfHunter
- KDE_classifier
- SOUTHGREEN PROJECTS
- SNiPlay3
- GNPAnnot Tools
- GNPAnnot Converters
- Scaffremodler
- Scaffhunter
- ESTtik
- Liftover
- Expression data

imported: FROGS metagenomics



Details

FROGS Pre-process merging, denoising and dereplication. (Galaxy Version 3.1)

Label

Add a step label.

Annotation

Add an annotation or notes to this step. Annotations are available when a workflow is viewed.

Sequencer

Illumina

Select the sequencing technology used to produce the sequences.

Input type

Files by samples

Samples files can be provided in single archive or with two files (R1 and R2) by sample.

Reads already contiged ?

No

The inputs contain 1 file by sample : R1 and R2 are already merged by pair.

Samples

1: Samples

Name

Reads 1

Data input 'R1_file' (fastq)

R1 FASTQ file of paired-end reads.

reads 2

Data input 'R2_file' (fastq)

R2 FASTQ file of paired-end reads.

Insert Samples

Reads 1 size

Reads 2 size

mismatch rate.

0.1

The maximum rate of mismatches in the overlap region

Merge software

Vsearch

Select the software to merge paired-end reads.

Would you like to keep unmerged reads?

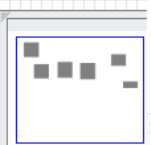
Yes No

No : Unmerged reads will be excluded; Yes : unmerged reads will be artificially combined with 100 N. (default No)

Minimum amplicon size

Maximum amplicon size

Sequencing protocol



Dans un monde idéal



<https://nbis-reproducible-research.readthedocs.io/en/latest/>

8) Le partage des données

- ❖ Les publications ne contiennent qu'une petite partie des données
- ❖ Partageons toutes nos données !

Quel entrepôt pour quelle donnée

- ❖ Institutionnel ?
- ❖ Thématique ?
- ❖ Lié à un éditeur ?



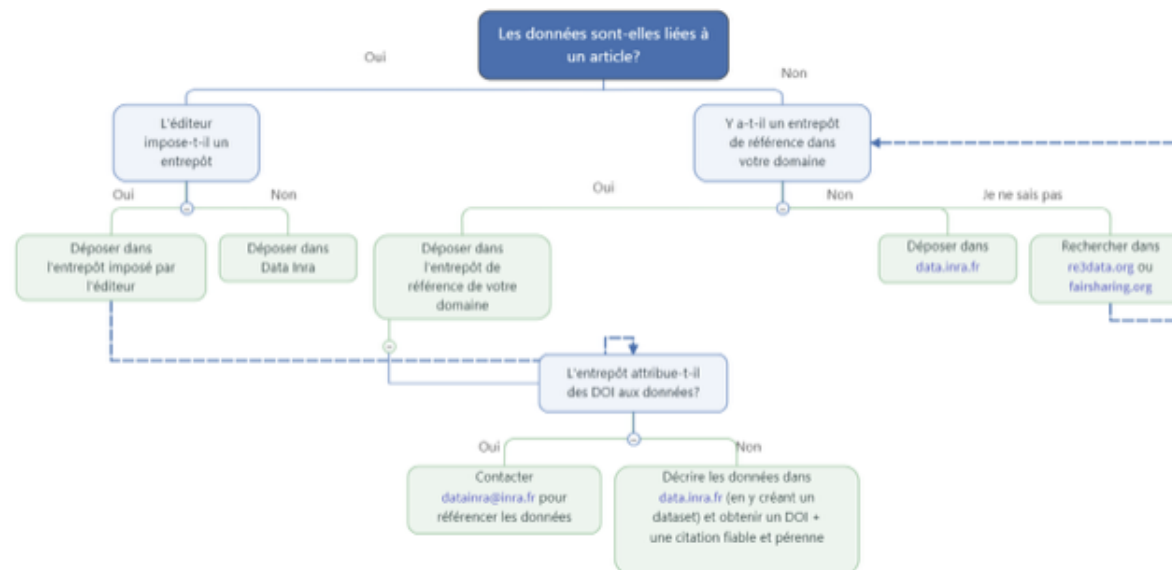
Accueil • Partager / Publier • Choisir un entrepôt

Choisir un entrepôt

Déposer vos données dans un entrepôt de données (data repository ou digital repository) contribue à les valoriser en les rendant plus visibles, accessibles, moissonnables et citables.

Selon votre situation et vos objectifs, vous avez différentes possibilités.

Si vous souhaitez partager publiquement des données, *en lien ou non* avec des articles scientifiques :



- > Les données sont liées à un article : si l'éditeur impose un entrepôt, il faut déposer dans l'entrepôt imposé, sinon il faut déposer dans **Data INRAE**
- > Les données ne sont pas liées à un article : s'il existe un entrepôt de référence dans votre domaine, il faut déposer dans cet entrepôt, sinon il faut déposer dans **Data INRAE**
 - > Accéder à une liste d'entrepôts internes attribuant des DOI aux données
 - > Accéder à une liste d'entrepôts externes attribuant des DOI aux données

Boîte à outils

- > DMP OPIDoR
- > DOI INRAE
- > Portail Data INRAE
- > Obtenir de l'aide

On vous signale

- > Vos questions, nos réponses
- > FAQ DOI
- > FAQ Datapaper
- > Formations et tutoriels
- > S'abonner aux actualités

Filter

- Subjects ⊕
- Content Types ⊕
- Countries ⊕
- AID systems ⊕
- API ⊕
- Data access ⊕
- Data access restrictions ⊕
- Database access ⊕
- Database licenses ⊕
- Data licenses ⊕
- Data upload ⊕
- Data upload restrictions ⊕
- Enhanced publication ⊕
- Institution responsibility type ⊕
- Institution type ⊕
- Keywords ⊕
- Metadata standards ⊕
- PID systems ⊕
- Provider types ⊕
- Quality management ⊕
- Repository languages ⊕
- Software ⊕
- Syndications ⊕
- Repository types ⊕
- Versioning ⊕

wheat

Search

Toggle short help

← Previous 1 Next →

Sort by ▾

Found 8 result(s)

Triticeae Toolbox

T3

Subject(s)

Plant Genetics Microbiology, Virology and Immunology Molecular Neuroscience and Neurogenetics Plant Sciences Biology

Life Sciences Medicine Neurosciences

Content type(s)

Databases Scientific and statistical data formats Structured text other Standard office documents Raw data

Country

United States

A database for plant breeders and researchers to combine, visualize, and interrogate the wealth of phenotype and genotype data generated by the Triticeae Coordinated Agricultural Project (TCAP).

CIMMYT Research Data & Software Repository Network

International Maize and Wheat Improvement Center

Subject(s)

Agriculture, Forestry, Horticulture and Veterinary Medicine Plant Genetics Life Sciences Plant Sciences Biology

Content type(s)

Standard office documents Source code Software applications Archived data other

Country

International Mexico

The International Maize and Wheat Improvement Center (CIMMYT) provides a free, open access repository of research software, studies, and datasets produced and developed by CIMMYT scientists as well as the results of the Seeds of Discovery project, which makes available genetic profiles of wheat and maize, two of mankind's three major cereal crops.

GnpIS

Genetic and Genomic Information System

Subject(s)

Engineering Sciences Computer Science, Electrical and System Engineering Computer Science

Agriculture, Forestry, Horticulture and Veterinary Medicine Life Sciences Biology Basic Biological and Medical Research

Operating, Communication and Information Systems Agriculture, Forestry, Horticulture and Veterinary Medicine Basic Forest Research

Plant Cultivation Plant Breeding Inventory Control and Use of Forest Resources Plant Systematics and Evolution Plant Genetics

Plant Sciences Bioinformatics and Theoretical Biology

Content type(s)

other Raw data Scientific and statistical data formats Databases

Country

France

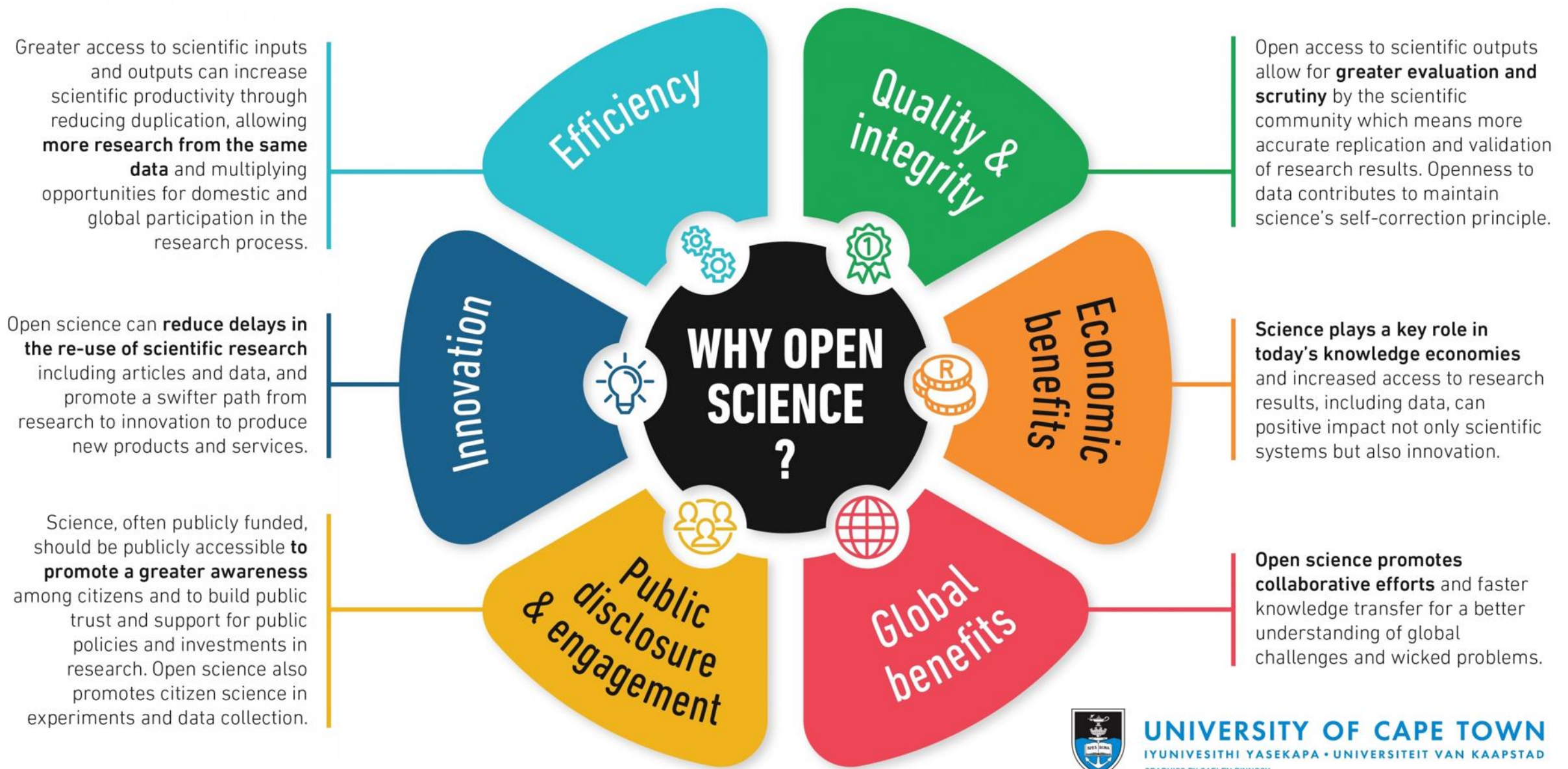
GnpIS is a multispecies integrative information system dedicated to plant and fungi pests. It bridges genetic and genomic data, allowing researchers access to both genetic information (e.g. genetic maps, quantitative trait loci, association genetics, markers, polymorphisms, germplasms, phenotypes and genotypes) and genomic data (e.g. genomic sequences, physical maps, genome annotation and expression data) for species of agronomical interest. GnpIS is used by both large international projects and plant science departments at the French National Institute for Agricultural Research. It is regularly improved and released several times per year. GnpIS is accessible through a web portal and allows to browse different types of data either independently through dedicated interfaces or simultaneously using a quick search ('google like search') or advanced search (Biomart, Galaxy, Intermine) tools.

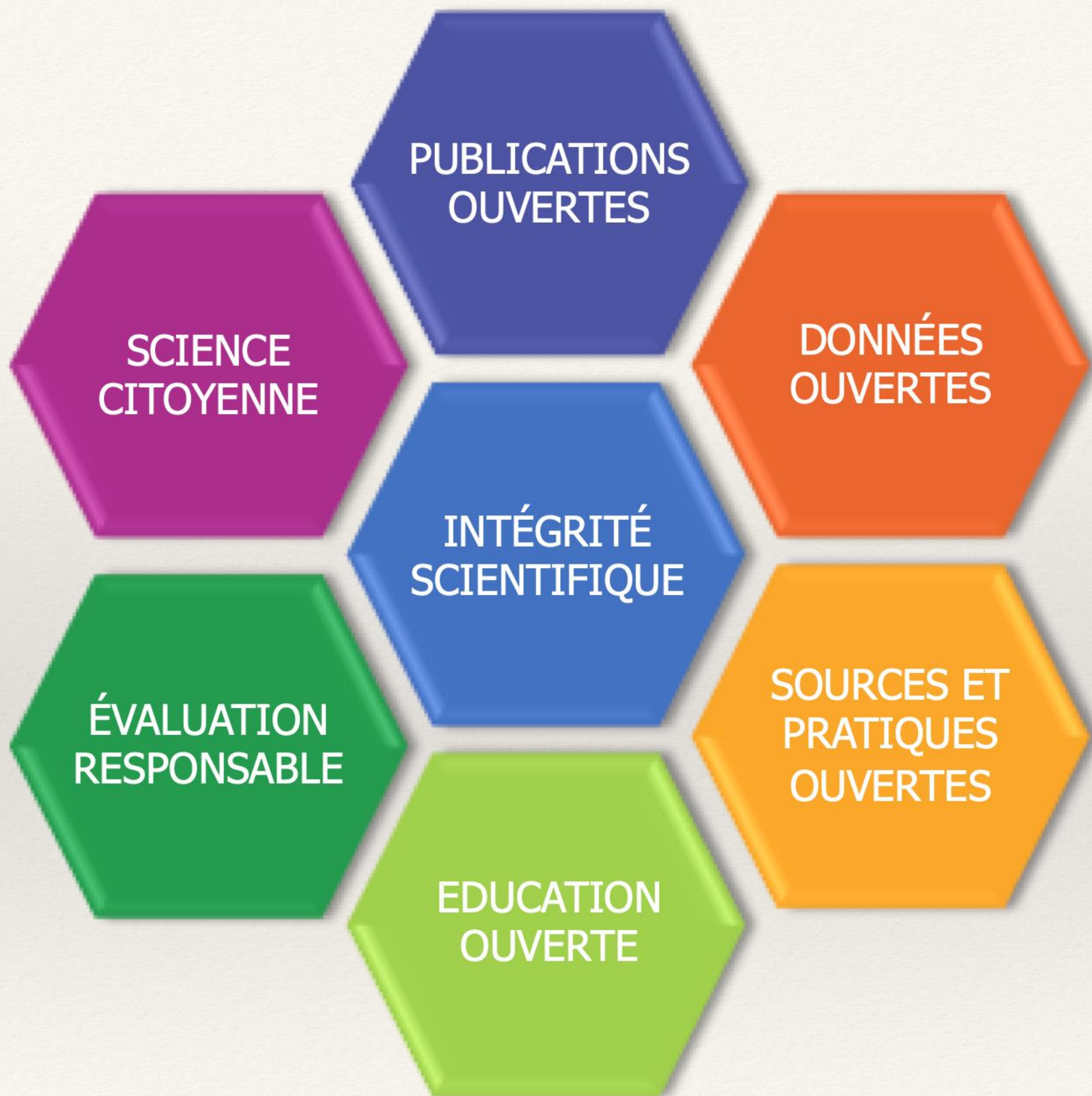
En résumé, il nous faut :

- ❖ Espace de préservation du savoir
- ❖ Lien entre l'analogique et le numérique
- ❖ Métadonnées
- ❖ Stockage le long du projet
- ❖ Qualification des données
- ❖ Gestion de code, environnement, Workflow
- ❖ Système de partage



Open Science





PUBLICATIONS
OUVERTES

SCIENCE
CITOYENNE

DONNÉES
OUVERTES

INTÉGRITÉ
SCIENTIFIQUE

SOURCES ET
PRATIQUES
OUVERTES

ÉVALUATION
RESPONSABLE

EDUCATION
OUVERTE

How OSF supports your research



Search and Discover

Find papers, data, and materials to inspire your next research project. Search public projects to build on the work of others and find new collaborators.

Design Your Study

Start a project and add collaborators, giving them access to protocols and other research materials. Built-in version control tracks the evolution of your study.

Collect and Analyze Data

Store data, code, and other materials in OSF Storage, or connect your Dropbox or other third-party account. Every file gets a unique, persistent URL for citing and sharing.

Publish Your Reports

Share papers in OSF Preprints or a community-based preprint provider, so others can find and cite your work. Track impact with metrics like downloads and view counts.

cesgo.org

Services CeSGO



Collaboration

Échanger avec votre communauté

[Plus de détails](#)



Data access

Synchroniser et partager vos données

[Plus de détails](#)



Research sharing

Stocker et publier vos données de recherche

[Plus de détails](#)



Projects

Gérer vos projets efficacement avec un kanban

[Plus de détails](#)



Instant

Discussion directe avec vos partenaires

[Plus de détails](#)

Une petite liste à emporter à la maison

DMP

<https://easydmp.eudat.eu/plan/>
<https://pgd17juin2019.sciencesconf.org/program>

Entrepôts

<https://fairsharing.org/databases/>
<https://www.nature.com/sdata/policies/repositories>
<http://about.zenodo.org>
<https://search.datacite.org/data-centers>
<https://cordis.europa.eu/project/rcn/207500/>

Gestion des données

<https://www.csc.fi/en/web/training/-/webinar-research-data-management>
<http://www.smalsresearch.be/dix-bonnes-pratiques-pour-ameliorer-et-maintenir-la-qualite-des-donnees/>
<https://mantra.edina.ac.uk>
<https://www.fairdata.fi/en/>

Choix d'une licence

<https://www6.inra.fr/datapartage/Partager-Publier/Choisir-une-licence>

Ontologie et standard de méta données

<https://bioportal.bioontology.org>
<https://isa-tools.org>
<https://datacite.org>

et aussi

<https://eudat.eu>
<http://www.dcc.ac.uk>

Qualité des données

<https://github.com/laurence001/quartz-bad-data-guide-french>
http://corist-shs.cnrs.fr/gestion_donnees_recherche_guideetformation_11

