

*Collecter, préserver, annoter,
organiser, réutiliser nos données*

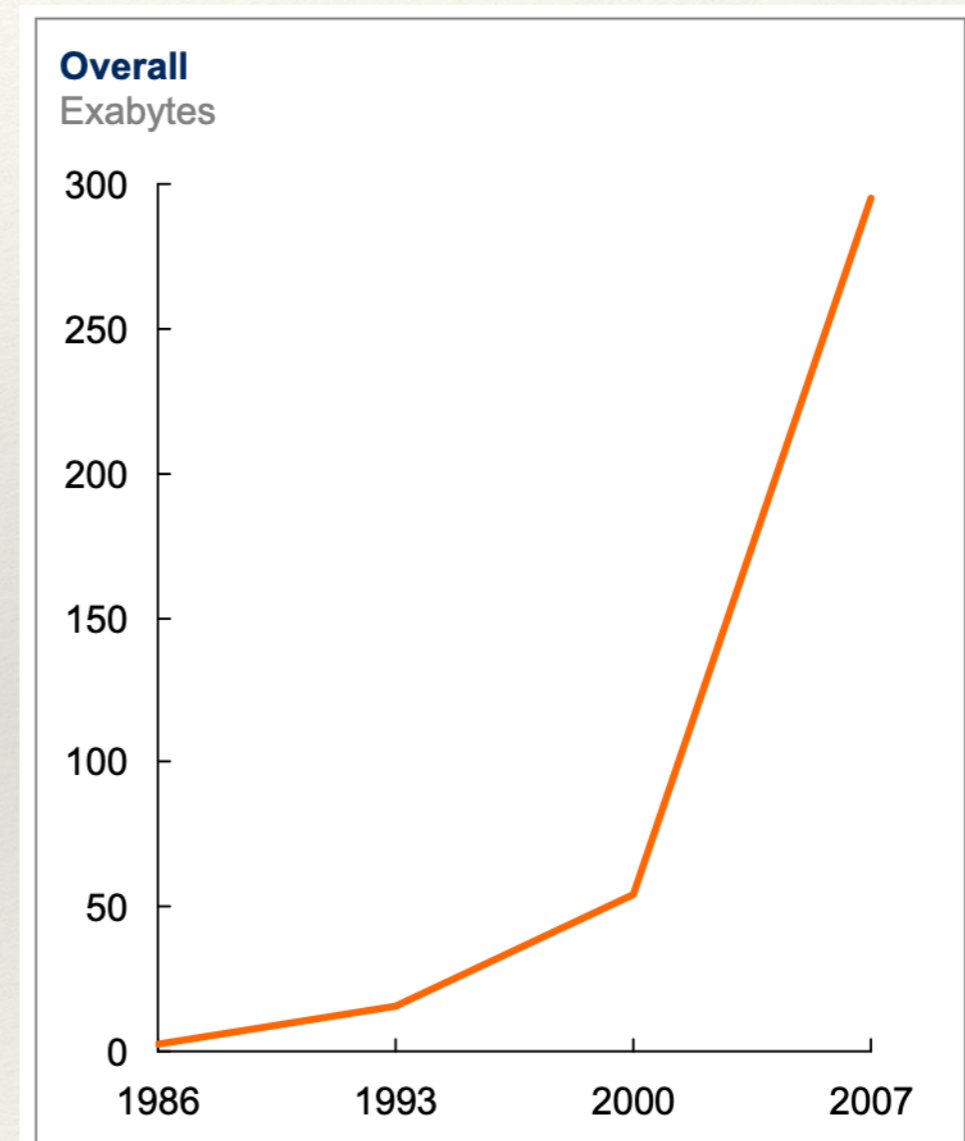
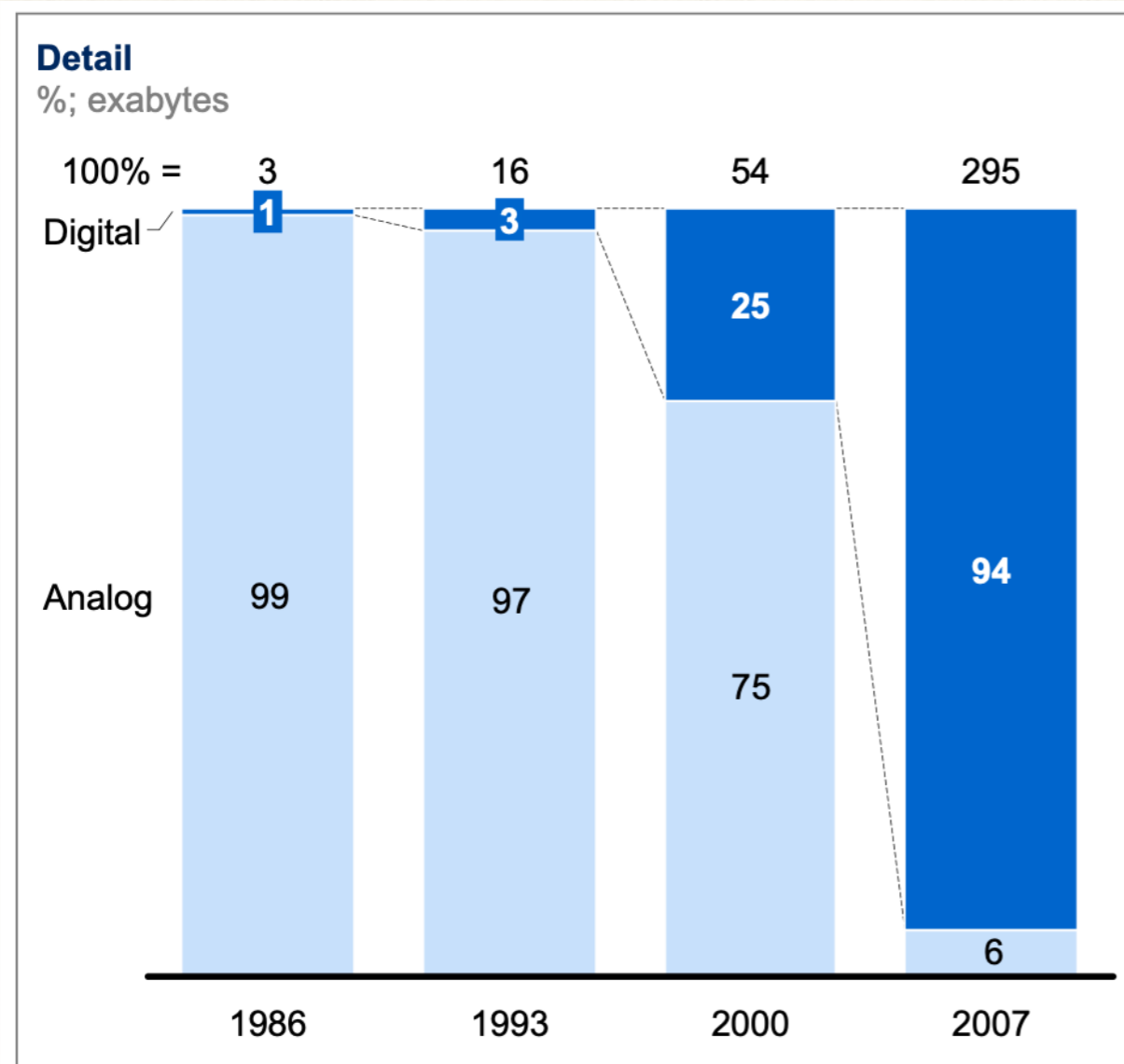
Essayons !



C'est complexe !

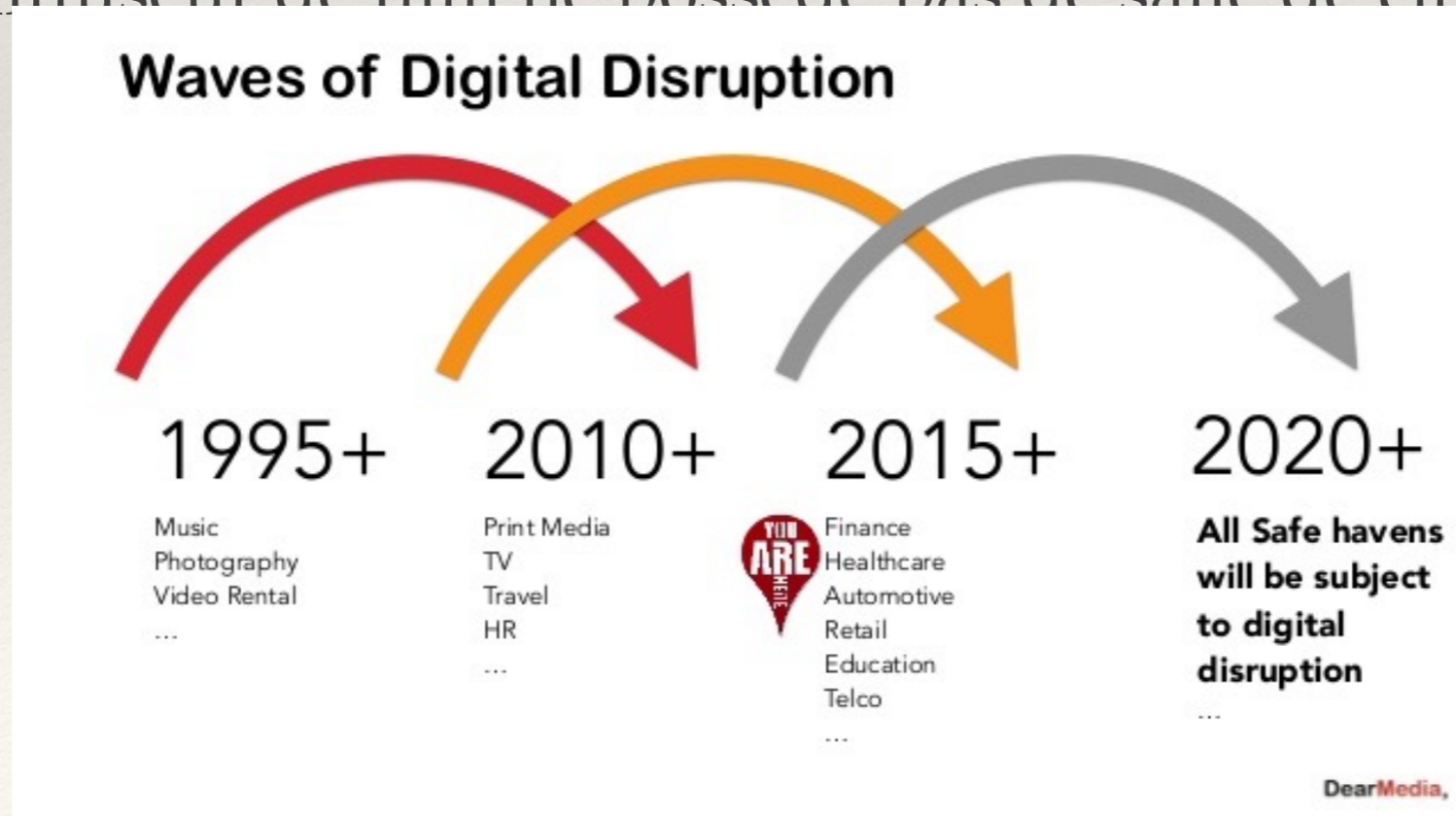
- ❖ Contexte
- ❖ éléments de réflexion
- ❖ quelques pistes

La disruption numérique : une bascule brutale



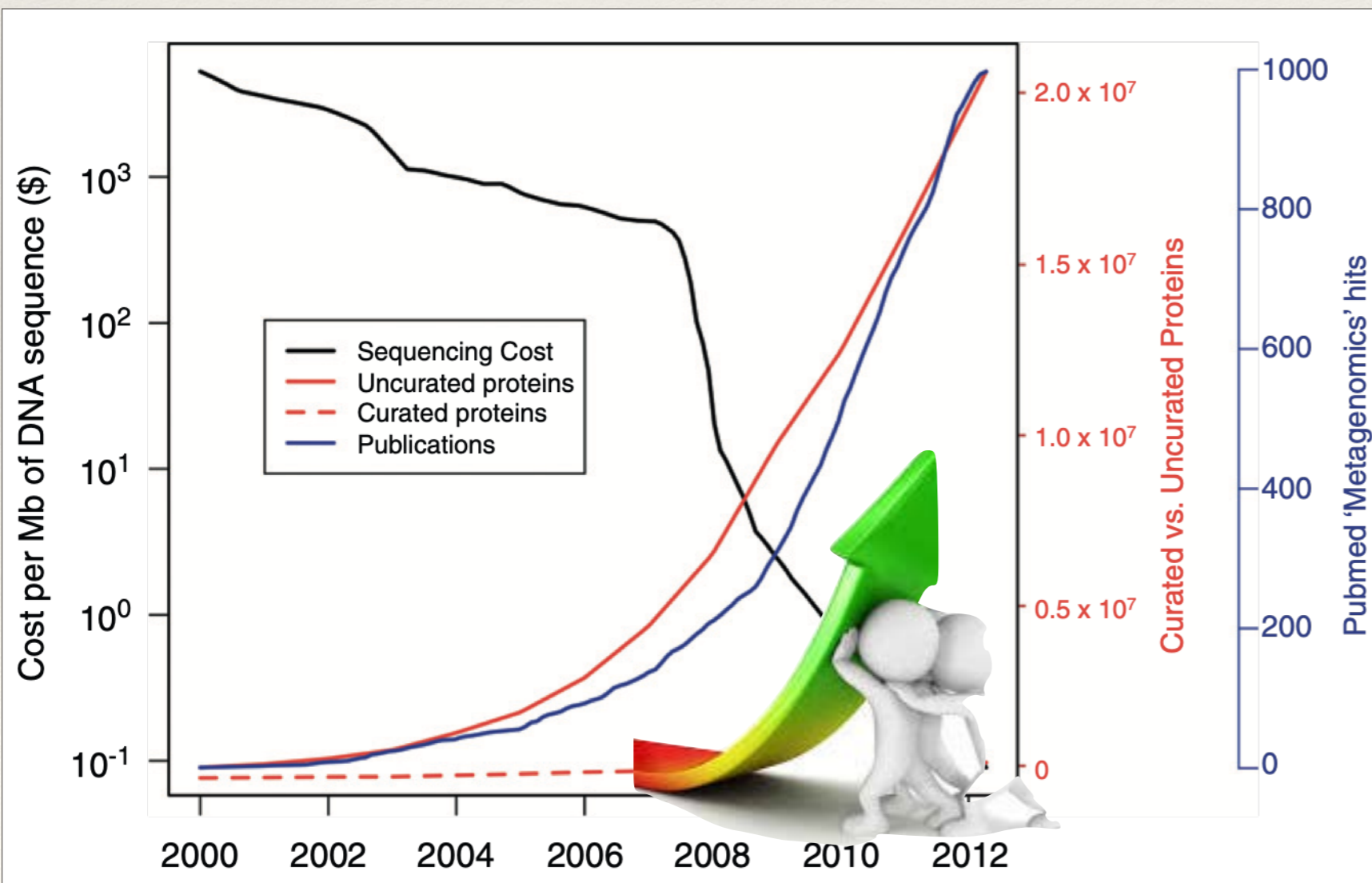
Qui change notre monde

- ❖ 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 diffuseur de film ne possède pas de salle de cinéma (Netflix)



Et en Science ?

- Les techniques à haut débit, une révolution qui provoque un déluge 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

Répondre aux questions de Science

Avant

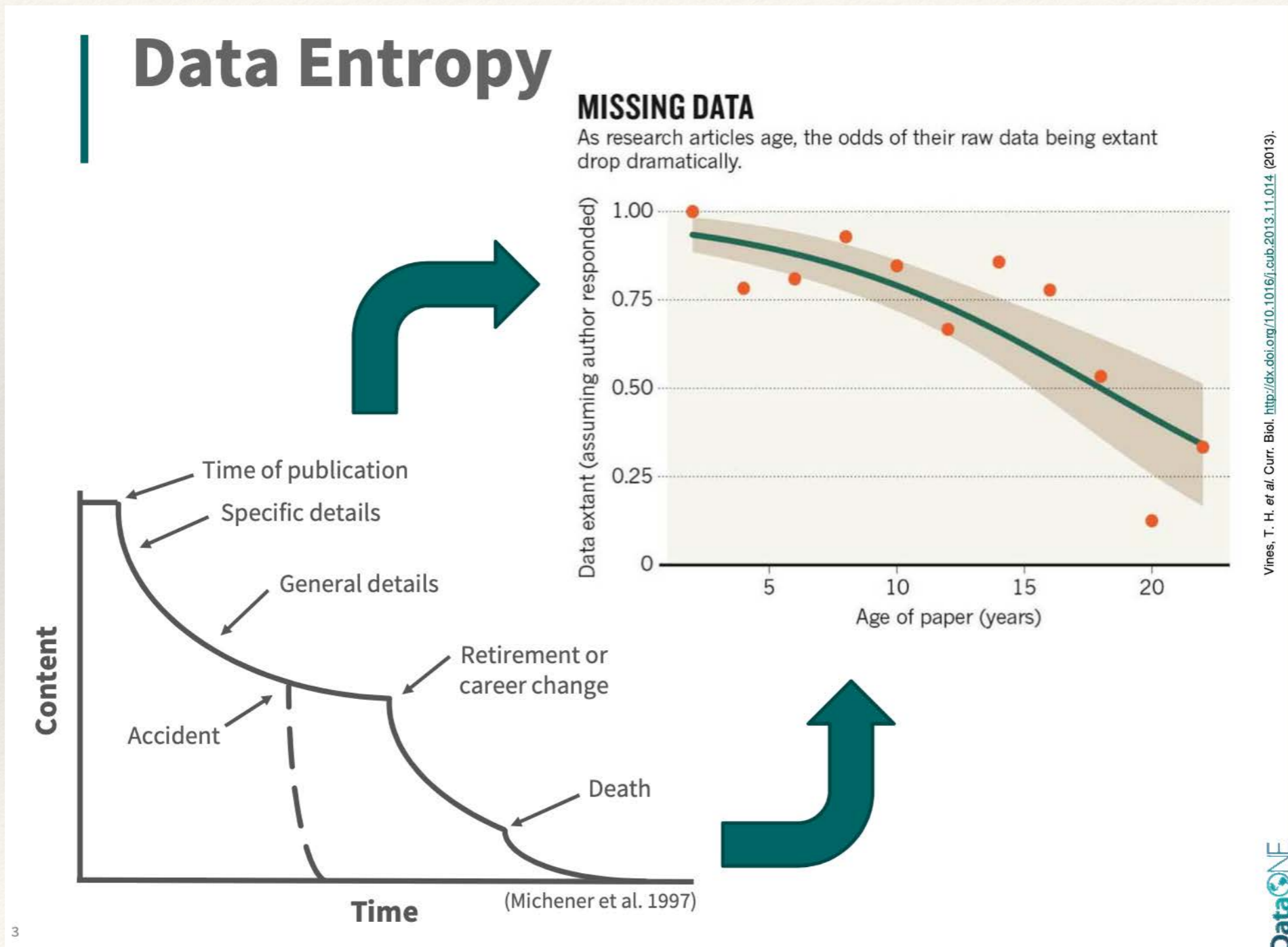
- 1 Concevoir l'expérimentation
- 2 Collecter des résultats
- 3 Analyser des résultats

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



Les défis de la reproductibilité

RESEARCH ARTICLE

Estimating the reproducibility of psychological science

Open Science Collaboration^{*,†}

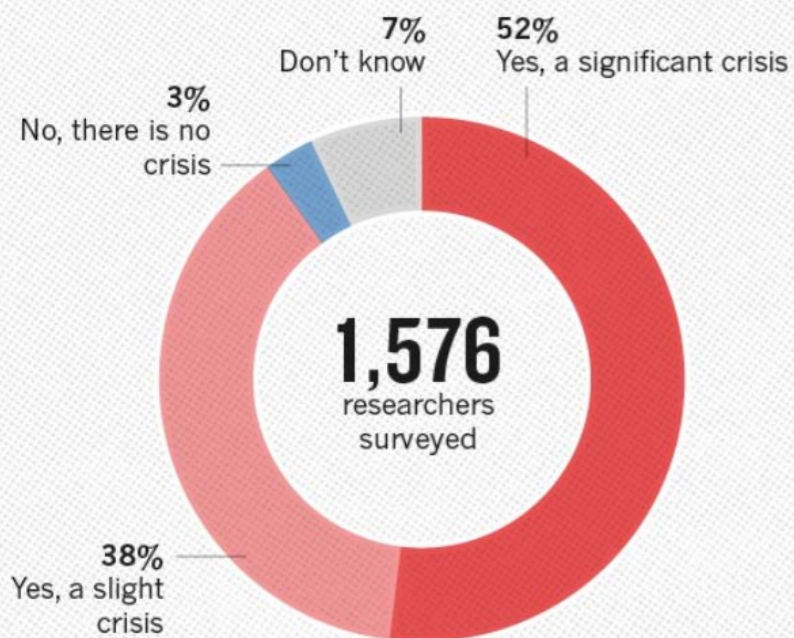
[†] See all authors and affiliations

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

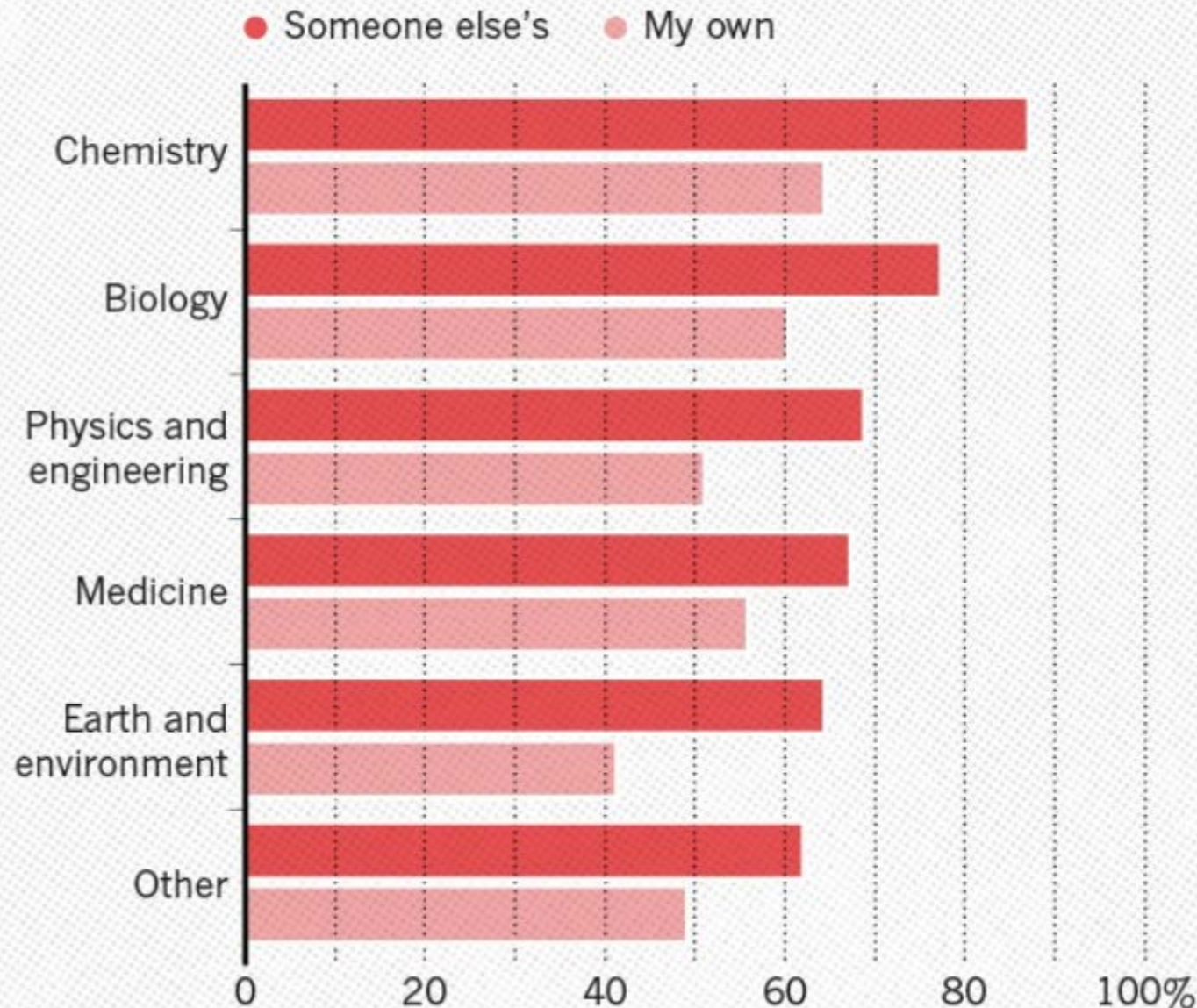
IS THERE A REPRODUCIBILITY CRISIS?



©nature

HAVE YOU FAILED TO REPRODUCE AN EXPERIMENT?

Most scientists have experienced failure to reproduce results.

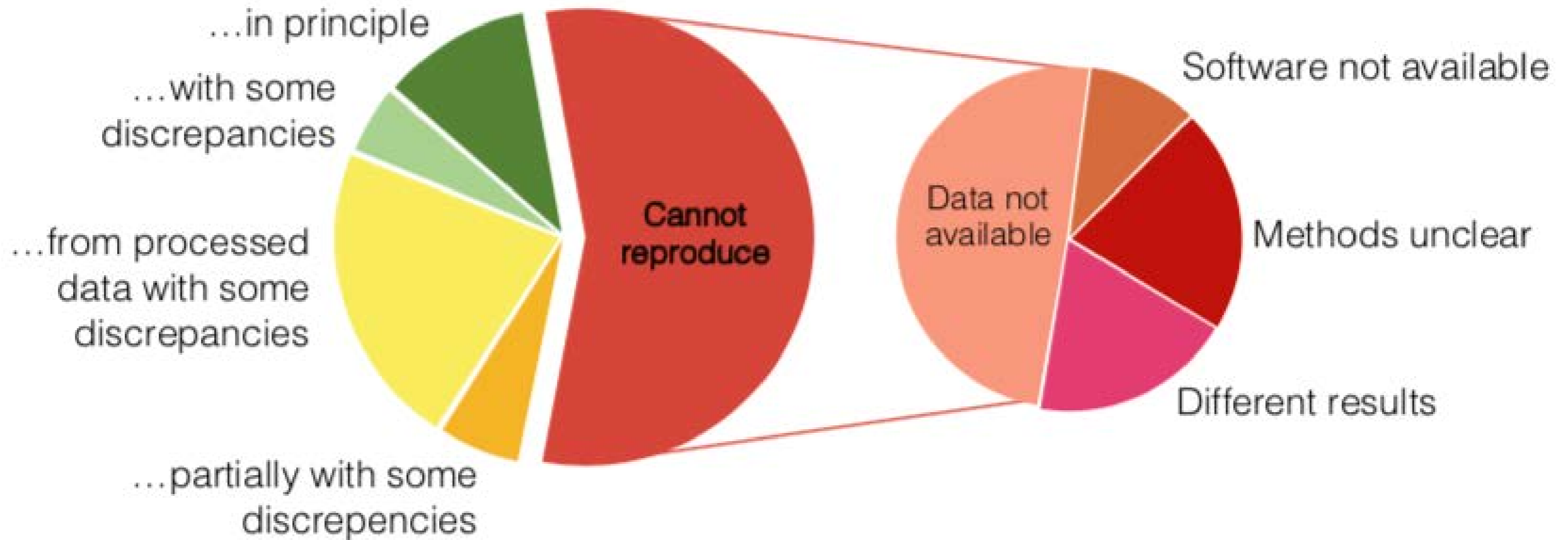


1,500 scientists lift the lid on reproducibility". Nature. 533: 452–454 - 2016

F. de Lamotte - INRA - Montpellier

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: implications for epidemiology and public health
Nature Genetics **41** (2009) doi:10.1038/ng.295

Open Science

Open Data

Open Source

Open Methodology

Open Peer Review

Open Access

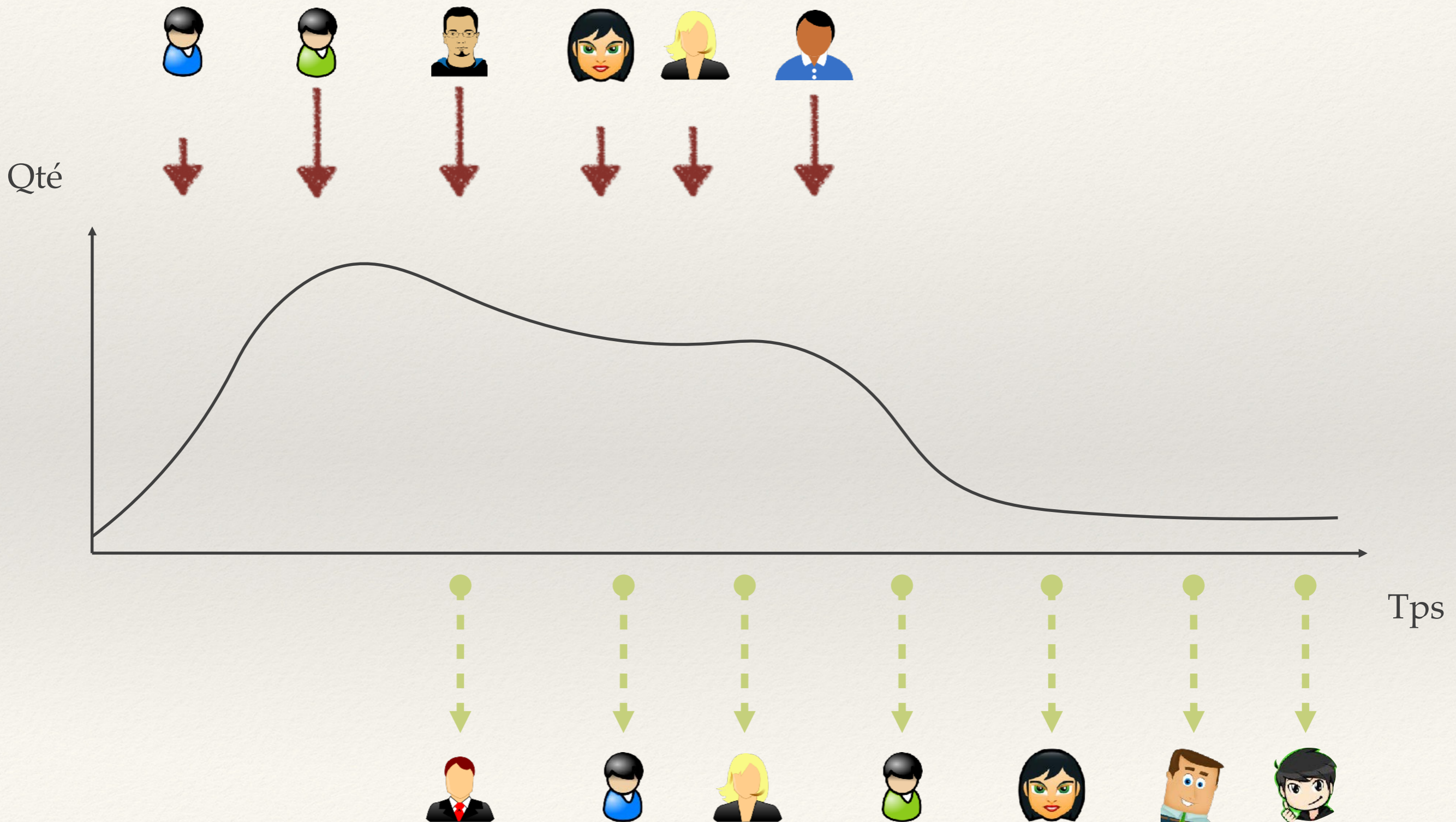
Open Educational Resources

Management éthique et efficace de la recherche

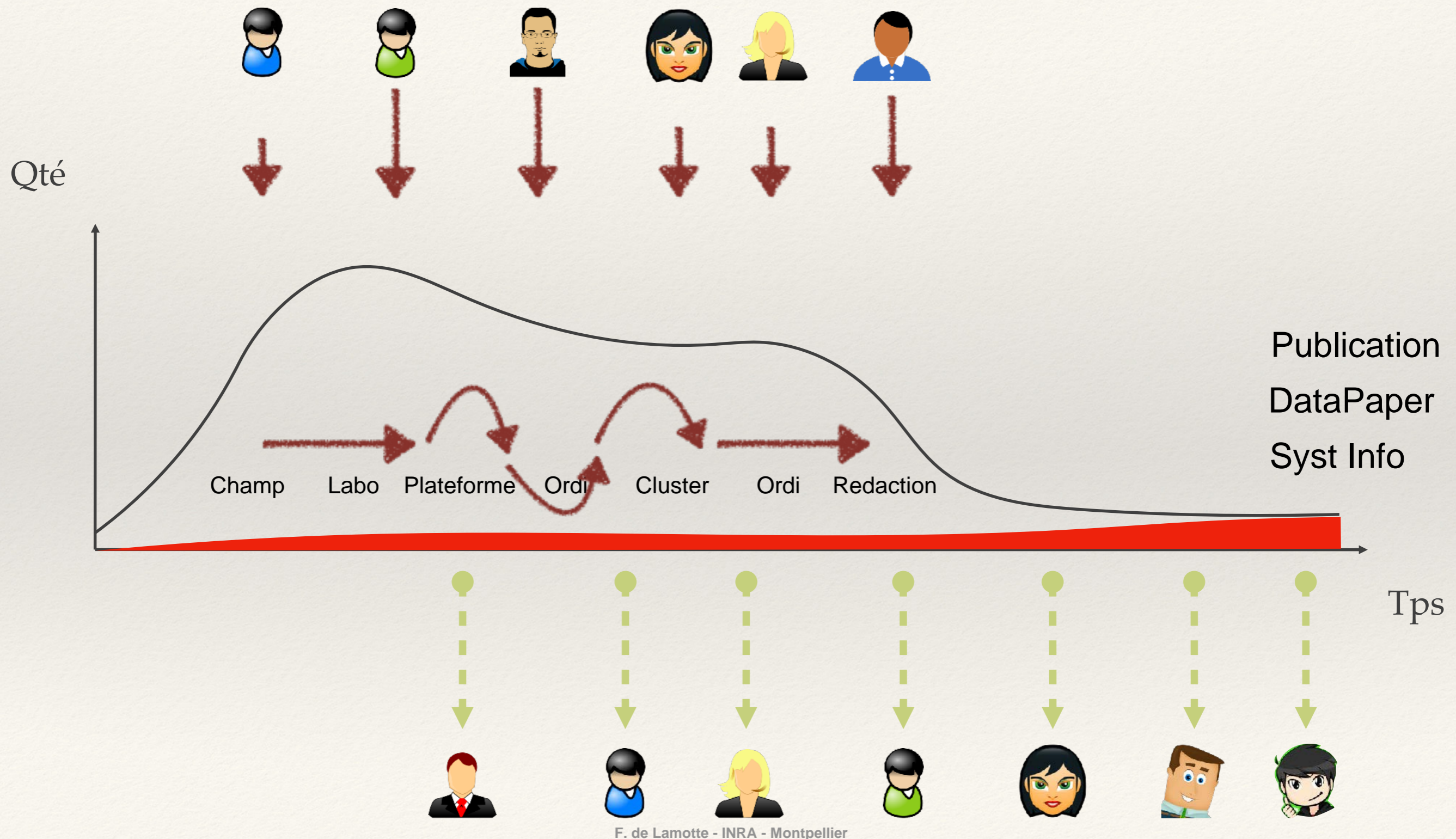
Nommez les exigences pour une recherche intègre et de qualité

- 1/ Traçabilité
- 2/ Honnêteté, Sincérité, Transparence, Impartialité
- 3 et 4 ex aequo
Méthodologie, Protocole validé
Objectivité, Pertinence, Traitement des résultats

Un projet sur la durée



Un enjeux de traçabilité



Traçabilité et qualité pour

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

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

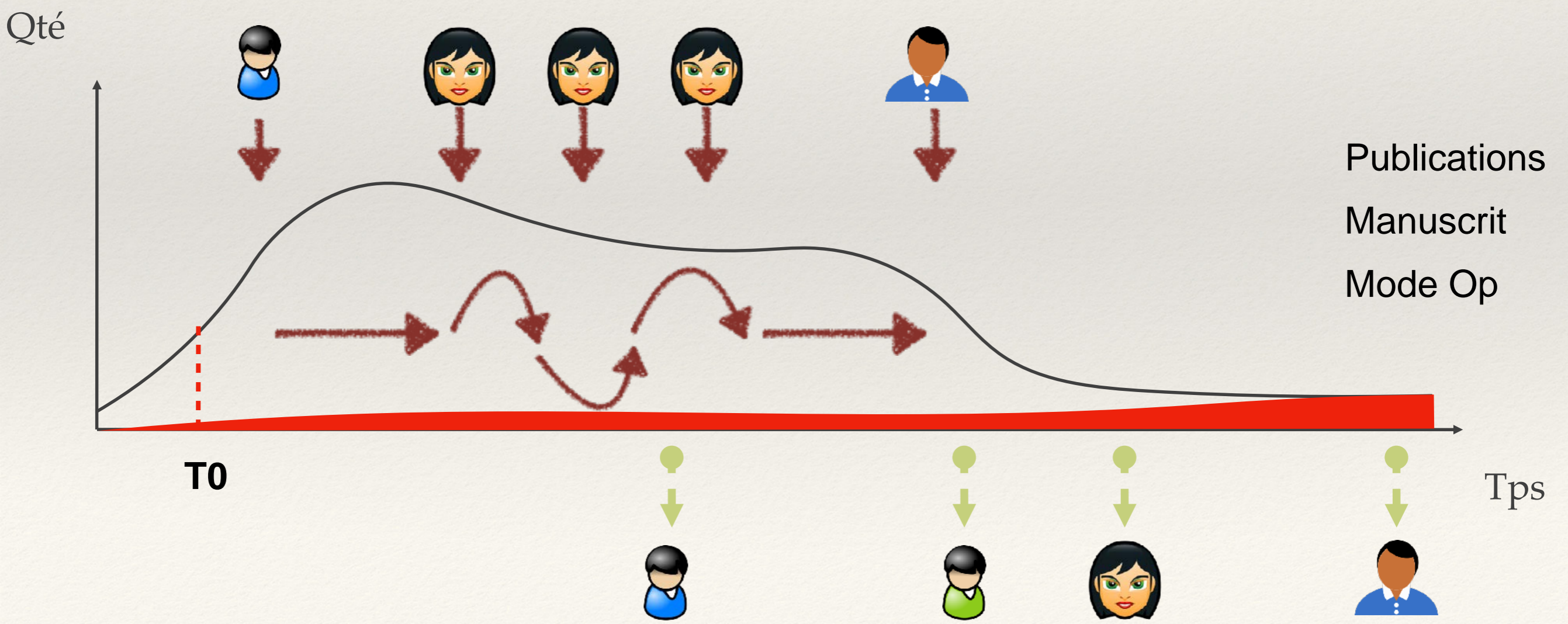


La traçabilité -

- ❖ Personnel
- ❖ Collectif
- ❖ Confiance
- ❖ Efficience
- ❖ Ré-exploitation



Tracer quoi ? Comment ?



Tracer 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, com^o ...)

- ❖ **Le futur**

- ❖ Le manuscrit
- ❖ Les publications (leur contenu)

Le passé

- ❖ Comment faites (faisiez) vous ?

Accès rapide

- Tous les documents
- Modifiés par moi
- Modifiés par d'autres
- Récemment modifiés
- Récemment ajoutés
- Mes favoris

Bibliothèque

- Documents
 - _INTERNET AGAPQSE
 - _INTRANET AGAPQSE
 - _INTRANET PUBLIC AGAPQS
 - 1ManagementQualite
 - 4RH
 - 5Genot
 - 6Bioinfo
 - 7CommunLabo
 - 8Equipmt
 - 9Achat
 - modification documentaire
 - PARAPHEUR
 - xAccueil2014
 - XDocDivers
 - XHorsPerimetre
 - zAjeter












Catégories

- Catégorie racine

Tags

Sélectionner Créer... Ajouter Eléments sélectionnés...

Documents

- 
zAjeter
 Modifié il y a un an environ par [Frédéric De Lamotte](#)
 En attente de poubellisation
 Aucun tag
 ★ Favori | 👍 Aimer 0 | 🗨 Commentaire
- 
XHorsPerimetre
 Modifié il y a 9 mois par [Patricia Turquay](#)
 Autres labo (consigne atelier), Robot, Serre, Ditam, GdT2009
 Etiquette Porte, Responsable atelier
 PR Publi
 Aucun tag
 ★ Favori | 👍 Aimer 0 | 🗨 Commentaire
- 
XDocDivers
 Modifié il y a 5 mois par [Philippe Carbonneill](#)
 Docs CRB et BGPI, Doc qual Extern, Prez, Documents anciens, veille biblio...
 Aucun tag
 ★ Favori | 👍 Aimer 0 | 🗨 Commentaire
- 
xAccueil2014
 Modifié il y a plus de 5 ans par [Frédéric De Lamotte](#)
 Espace du chantier Accueil Agap 2014
 Aucun tag
 ★ Favori | 👍 Aimer 0 | 🗨 Commentaire
- 
PARAPHEUR
 Modifié il y a plus de 4 ans par [Thierry Le Signor](#)
 répertoire nécessaire pour faire fonctionner le workflow qualité
 Aucun tag
 ★ Favori | 👍 Aimer 0 | 🗨 Commentaire
- 
modification documentaire
 Modifié il y a 27 jours par [Catherine Carasco-Lacombe](#)
 Aucune description
 Aucun tag
 ★ Favori | 👍 Aimer 0 | 🗨 Commentaire
- 
9Achat
 Modifié il y a 4 mois par [Catherine Carasco-Lacombe](#)
 Suivi Fournisseurs, MO Boutiq, PR Achat,
 Aucun tag
 ★ Favori | 👍 Aimer 0 | 🗨 Commentaire
- 
8Equipmt
 Modifié il y a plus de 2 ans par [Patricia Turquay](#)
 Balances,Pipettes, Réfrigérants, Thermocycleurs, Tout Matériel, Etalon, Formation, GenralMétro,
 Aucun tag
 ★ Favori | 👍 Aimer 0 | 🗨 Commentaire
- 
7CommunLabo
 Modifié il y a 4 mois par [Catherine Carasco-Lacombe](#)
 Instruction utilisation Pipettes, Balances
 Aucun tag
 ★ Favori | 👍 Aimer 0 | 🗨 Commentaire
- 
6Bioinfo
 Modifié il y a un an environ par [Patricia Turquay](#)
 Mo et In Bioinformatiq (privé et public)
 Aucun tag
 ★ Favori | 👍 Aimer 0 | 🗨 Commentaire
- 
5Genot
 Modifié il y a un an environ par [Catherine Carasco-Lacombe](#)
 Consigne Atelier,
 Form : habilitat instrument, list Agent habilité,
 Mo, In, génotypage
 Aucun tag
 ★ Favori | 👍 Aimer 0 | 🗨 Commentaire

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

MolInfo13_CreaCompteUtilisCC2_V2.pdf 2.0

Modifié par Patricia Turquay le 15/11/2017 14:29:09

Télécharger

Précédent Suivant 1 / 9 108% Télécharger

Actions sur le document

	Mode opératoire	MO 20/
	Equipe ID	
Gestion compt Util CC2		

Gestion (Création/ Suppression) d'un utilisateur sur le cluster de

Historique des versions

Version	Date de version	Gestionnaire
Version 00	30/03/2015	Création
Version 01	09/11/2015	Révision
Version 02	20/09/2017	Modification

Confidentialité : Administrateur Cluster

Sommaire

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

1. Objectif

L'objectif du mode opératoire est de décrire la gestion 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

CONSULTATION de la dernière version du document:

Commentaires

Ajouter un commentaire

Pas de commentaire

Workflow

Ce document ne fait partie d'aucun workflow.

Historique des versions

Dernière version

2.0 MolInfo13_CreaCompteUtilisCC2_V2.pdf



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

Versions antérieures

1.4 MolInfo13_CreaCompteUtilisCC2_V2.pdf



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

1.3 MolInfo13_CreaCompteUtilisCC2_V2.pdf



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

1.2 MolInfo13_CreaCompteUtilisCC2_V1.pdf



Patricia Turquay il y a plus de 4 ans
(Pas de commentaire)

1.1 MolInfo13_CreaCompteUtilisCC2_V1.pdf



Patricia Turquay il y a plus de 4 ans
(Pas de commentaire)

1.0 MolInfo13_CreaCompteUtilisCC2_V0.pdf



Patricia Turquay il y a plus de 4 ans
(Pas de commentaire)

Ce document ne fait partie d'aucun workflow.

Historique des versions

Dernière version

2.0 MolInfo13_CreaCompteUtilisCC2_V2.pdf



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

Versions antérieures

Les sources bibliographiques

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

Le présent

- ❖ Analogique
- ❖ Numérique
- ❖ Algorithmique

Analogique

- ❖ échantillons
- ❖ cahiers de labo



Analogique

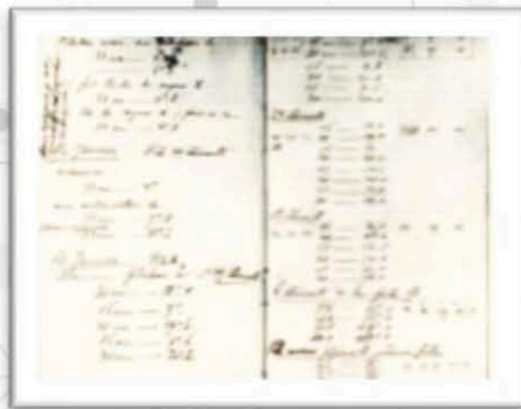
- ❖ Règles de nommage (partagées !)
- ❖ Plan de classement (physique)
- ❖ Règles (et contrôle) de tenue du cahier due labo
- ❖ Index des cahiers (qui, quand, quel projet ...)

- ❖ *Lien vers l'alter ego numérique*

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	Temperature	Time
1. Initial Denaturation at 95°C	95°C	30 sec
2. 30 Cycles of Amplification at 94°C	94°C	30 sec
3. 30 Cycles of Amplification at 55°C	55°C	30 sec
4. 30 Cycles of Amplification at 72°C	72°C	30 sec

PCR Purification



1624 32 31 32 31

1624 32 31 32 31

1624 32 31 32 31

1624 32 31 32 31

Journal de bord:

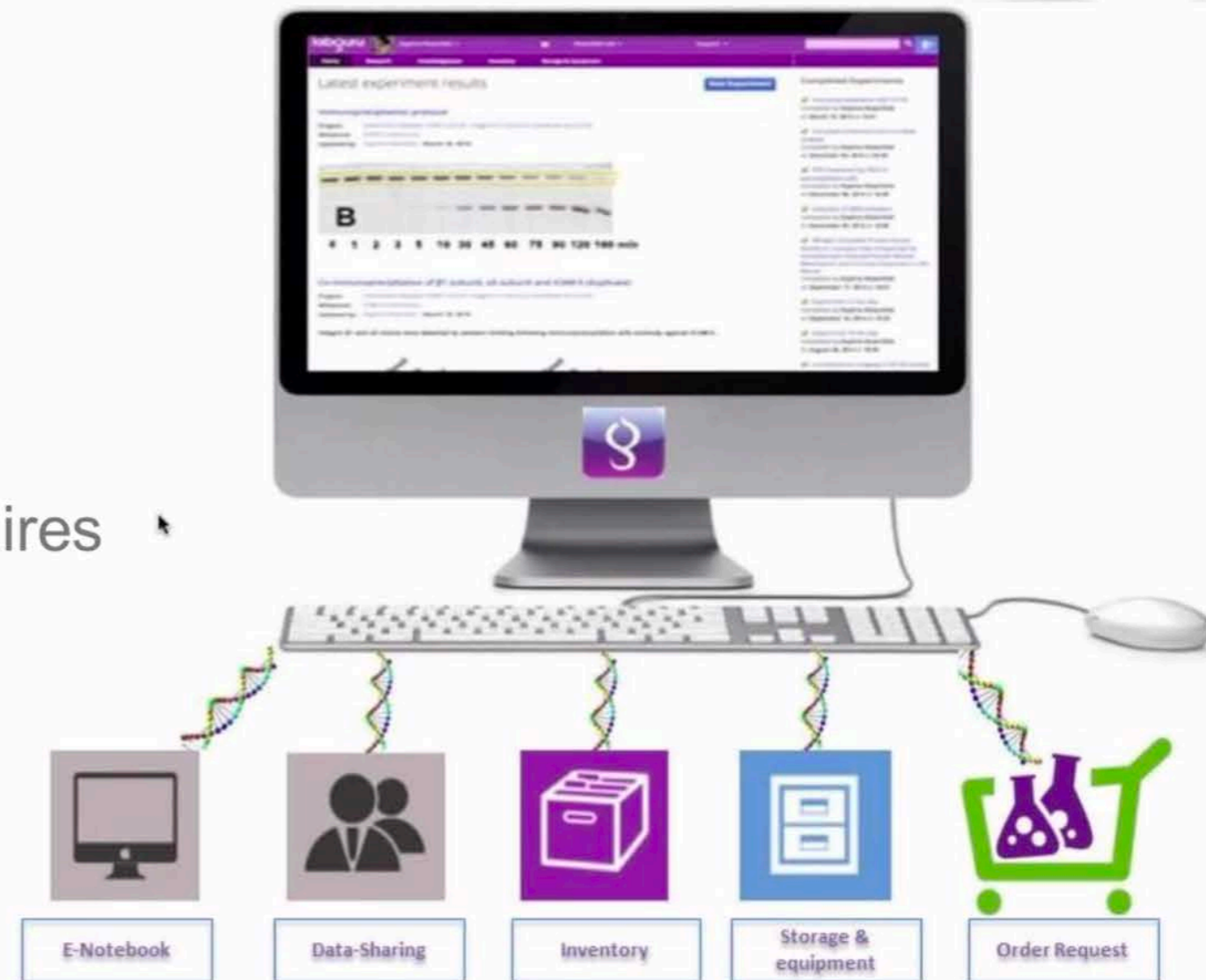
- Preuve juridique
- Traçabilité, qualité
- Gestion de la connaissance

- 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

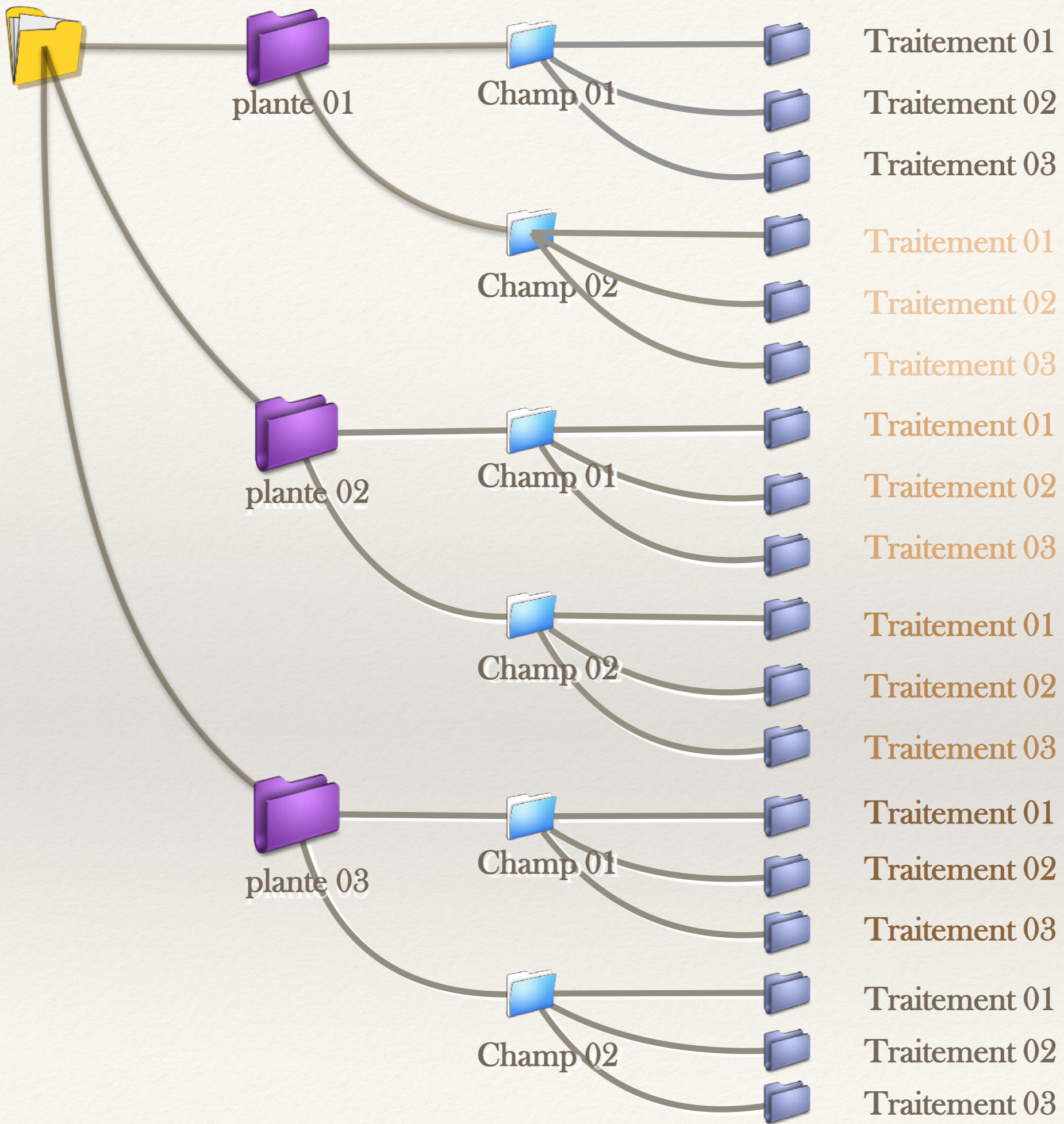


Numérique

- ❖ 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



Les métadonnées



Le moteur de recherche qui respecte votre vie privée. [En savoir plus >>](#)

[Installer Qwant](#)

Frederic



[Paramètres](#)



Que recherchez-vous ?



Qwant soutient le financement participatif d'une centrale solaire dans le département de l'Hérault

[» Investissons ensemble dans l'énergie verte !](#)

Partenariat

-
-
-
-
- Social**
-
-
-
-
-



Confidentialité

À propos

Conditions d'utilisation

Mentions légales

Contactez-nous

F.A.Q.

Qwick

Jobs

Presse

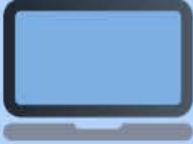

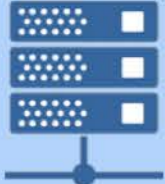

Blog

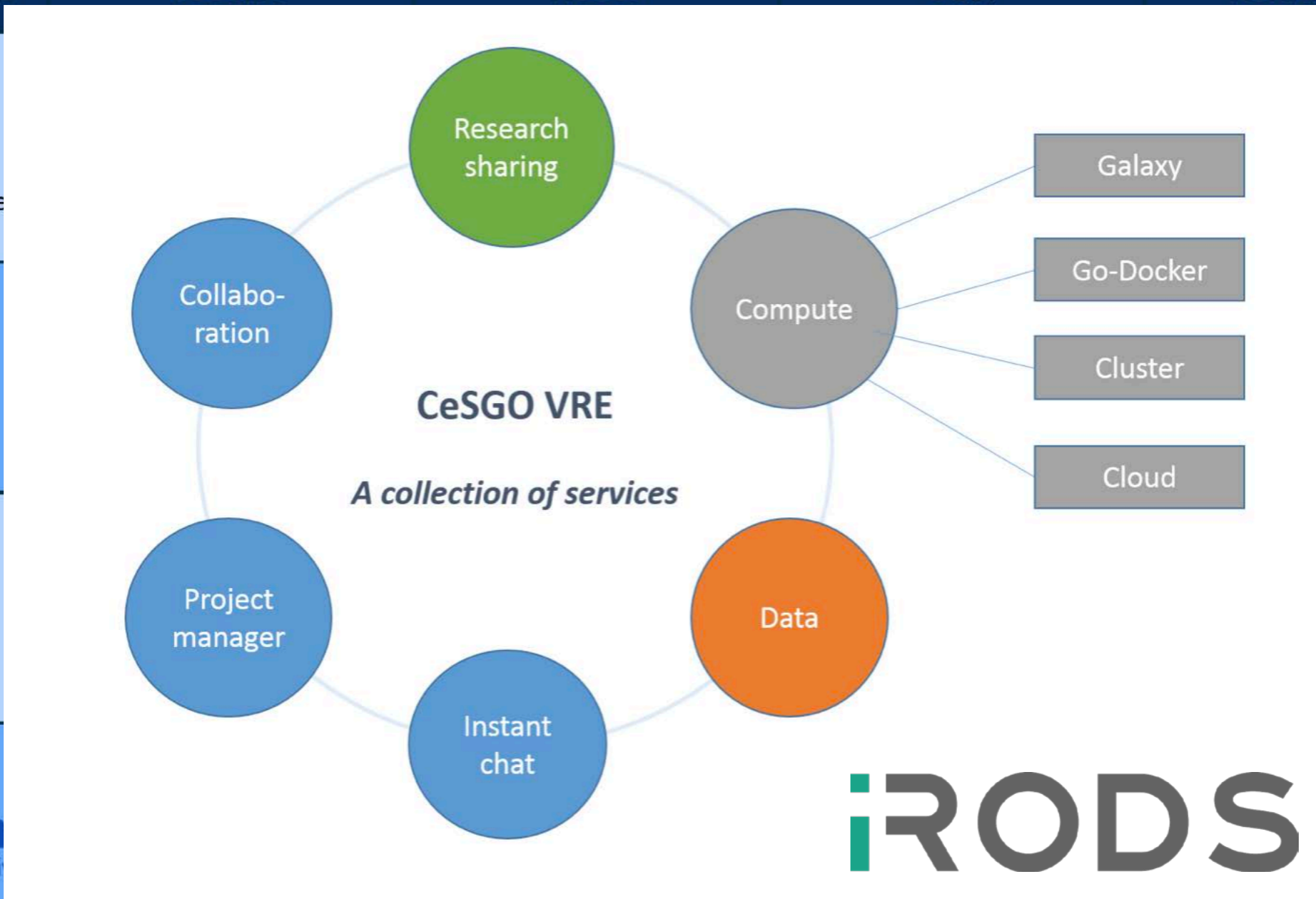
Lite

auté

On les met où ces données ?

❖ Dès le début !!!

Support de stockage	Sécurité	Accès	Coût	Remarque d'utilisation
 Ordinateur professionnel				Stockage temporaire ne crypter les données sensibles et sensibles
 Support externe				Stockage temporaire ne crypter ou de sécuriser les données sensibles et sensibles
 Serveur institutionnel				Stockage plus pérenne pour le stockage de sensibles et des versions vos données Institutions ne proposent pas
 Serveur Cloud				Stockage avec des termes à l'institution pour être de données confidentielles à définir le rôle sur la procédure de sauvegarde des données

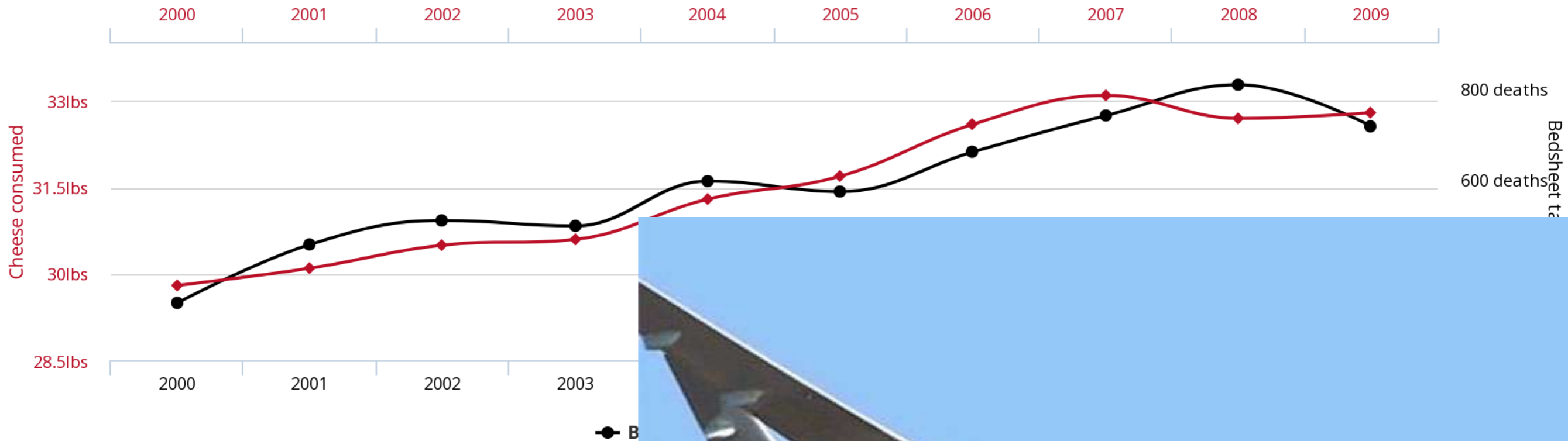


Qualification des données

Per capita cheese consumption

correlates with

Number of people who died by becoming tangled in their bedsheets



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 !!!!!
 - On va y consacrer la prochaine journée

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
101111110000011011011011110001
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
0011000011101100100000111
11010110011100000100110
11011110011100001111101101111
111001001011110001100110000101
011100001000010001011110000010

110101110011101111001010100111
111000101110011001101101001001
011001010100101011000001001100
11010011100101111100001011101
011110111110001111011110101101
000001110110011001010101011100
100010110001100000111001100010
000000111011001010101010111
0000100000110010010110110
101111101111000111100101110101
100101010100001001110100010001
011110011010100101111011110101
100011000110110001011101100110
110100000100000011011000001101
100000011100100111101101011011
010110010001000101110111001010

Data

Code



000000111011001010000100111011
1100001101110111101111110101010
000110010101111100101110110100
001110110011010110000101011010
111101111100000100010111010111
111010001000010010111100111001
111001100101000111101000011100
101111110000011011011011110001
100100110111101111000101010100
111110011010111011010011011100
111011100011110101011111000100
0101110110111110100011110100011
001111000001111110001011100111
101101100000100011100111110011
001101000010011000110011000011
101011110111101010000011010001
010111100101010010011100011011
001010101100101000001010000110
100000101001110011010000011100
001110011000111111111000001100
100100010100000110001011010000
010110010111101001000010100010
101011110001001001010010111000
011000100000010000000011100111

Data

Input

Output

100111100001001100110101101100
001010011101010111110001001101
010111101100011110111011110001
00110000111000100100000111
1101011001100000100110
11011110011100001111101101111
111001001011110001100110000101
011100001000010001011110000010

110101110011101111001010100111
111000101110011001101101001001
011001010100101011000001001100
1101001101101101101101
01111011111000111011110101101
000001110110011001010101011100
100010110001100000111001100010
000000111011100100100101010111

000010000001100001000010110110
101111101111000111100101110101
100101010100001001110100010001
011110011010010111011110101
1001100110110110110110110110
110100000100000011011000001101
100000011100100111101101011011
010110010001000101110111001010

Data

Program

Environment

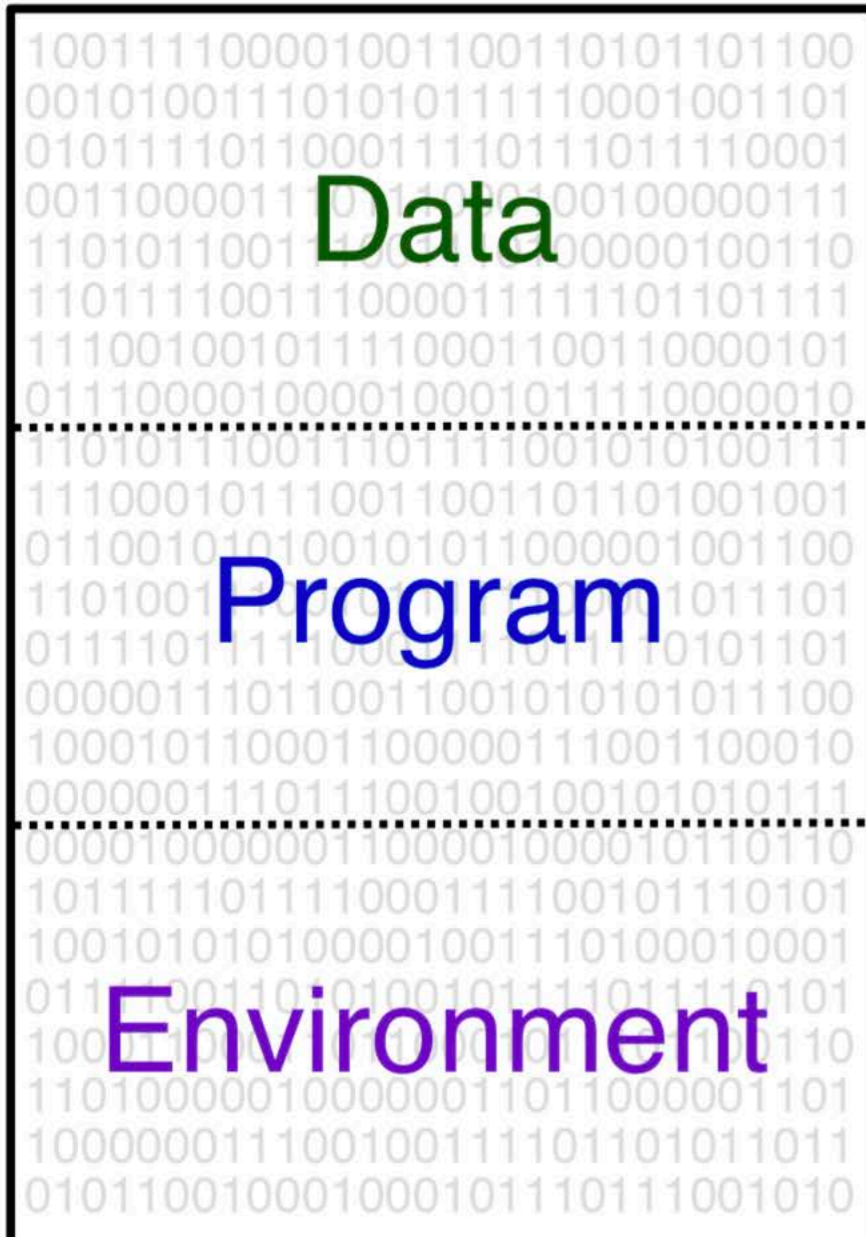


000000111011001010000100111011
110000110111011101111110101010
000110010101111100101110110100
001110110011010110000101011010
111101111100000100010111010111
111010001000010010111100111001
111001100101000111101000011100
101111110000011011011011110001
100100110111101111000101010100
111110011010111011010011011100
111011100011110101011111000100
010111011011011110110100011
001111000001111110001011100111
101101100000100011100111110011
001101000010011000110011000011
101011110111101010000011010001
010111100101010010011100011011
001010101100101000001010000110
100000101001110011010000011100
001110011000111111111000001100
100100010100000110001011010000
010110010111101001000010100010
101011110001001001010010111000
011000100000010000000011100111

Data

à éviter ...

Input



my research

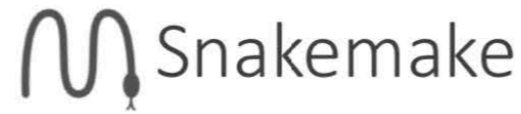
my colleagues' code

stuff I don't care about

Tools for Reproducible Research



Managing dependencies



Managing and executing analysis workflow



Versioning and collaborating on code (and some other files)



Connecting code and reporting

and...



Isolating and exporting environment

		Data	
		Same	Different
Code	Same	Reproducible	Replicable
	Different	Robust	Generalisable

"The foundations of knowledge should be constituted by experimentally produced facts, which can be made believable to a scientific community by their reproducibility."

- Robert Boyle, 1627-1691

Une petite liste à emporter à la maison

DMP

<https://easydmp.eudat.eu/plan/>

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



Traçabilité : définition

La **fiabilité des données** réside dans la possibilité de suivre les **différents états de celles-ci tout au long de leur cycle de vie**. Pour cela, la présente charte préconise l'utilisation de techniques dites «*reproductibles*». L'objectif est de tracer et de conserver les opérations effectuées sur les données lors des étapes de vérification, de correction, d'amélioration et d'enrichissement, avant leur mise à disposition aux utilisateurs.

La personne ou le collectif signataire de cette Charte s'engage à conserver les fichiers contenant les scripts et les procédures appliqués aux jeux de données et à ne pas recourir à des opérations manuelles ne laissant pas de traces (type copier/coller).