



www.cnrs.fr

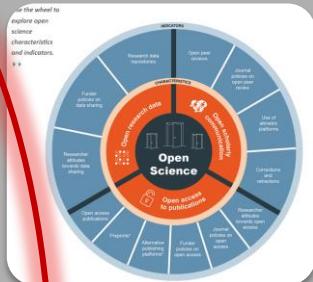


Les données de recherche ? Il y a beaucoup à dire... et à FAIR* !

Francis ANDRE – DIST/CNRS

Rennes, 18 mai 2017

* **F**indable, **A**ccessible, **I**nteroperable, **R**eusable



Ouverture ?
On saura vraiment ce que c'est quand on l'aura faite...

Démarche européenne :
Vers un partage global

Cadre éthique et juridique

Nouveaux services,
nouveaux métiers

Sujets brûlants

Evolution des pratiques de science

Science du 21^{ème} siècle : plus...

- Numérique
- Collaborative
- Interdisciplinaire
- Réactive
- Citoyenne
- Partagée

Open
Research

Science 2.0

eScience

OPEN SCIENCE

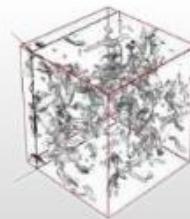
Tendances : accroissement de la production scientifique, du nombre de chercheurs, nouvelle façon de faire de la science, guidée par les données massives, importance des défis sociaux



La “science des données”, 4^e paradigme de la découverte scientifique



$$\left(\frac{\dot{a}}{a}\right)^2 = \frac{4\pi G\rho}{3} - K \frac{c^2}{a^2}$$



Experimental	Theoretical	Computational	The Fourth Paradigm
Thousand years ago <i>Description of natural phenomena</i>	Last few hundred years <i>Newton's laws, Maxwell's equations...</i>	Last few decades <i>Simulation of complex phenomena</i>	Today and the Future <i>Unify theory, experiment and simulation with large multidisciplinary Data</i> <i>Using data exploration and data mining (from instruments, sensors, humans...)</i>



Crédits: Dennis Gannon

Distributed Communities

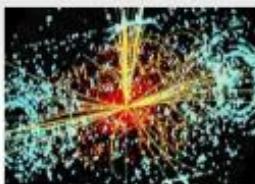
D'après G. Antoniu

Big Data : le déluge des données

Expériences



Simulations



Archives



Bibliothèques



Cornell University
Library

arXiv.org

Web 2.0



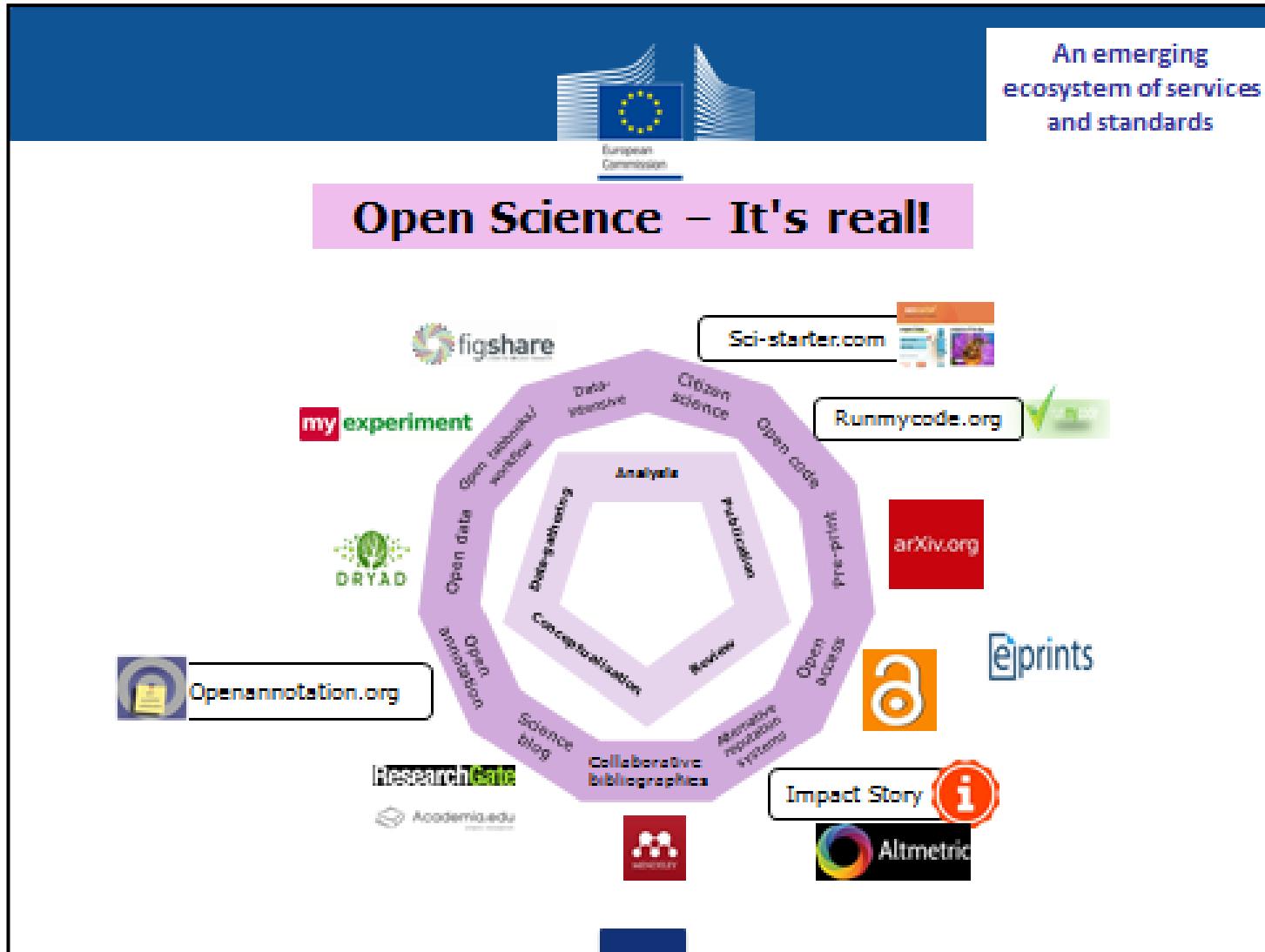
Défi

Accélérer les découvertes

Rendre possible la collecte,
la recherche et l'analyse
des données

Des péta-octets qui
s'accumulent...

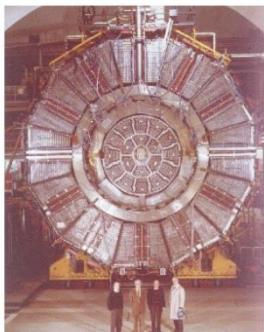
Une vision optimiste ?



Il y a partage et partage...

Exemples disciplinaires

In Big Communities In International Labs (CERN)



Past Century collaboration
~500 Scientists



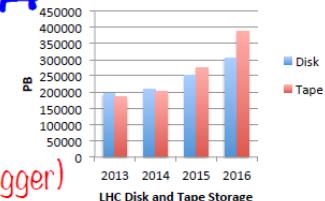
Today collaboration
~4000 Scientists

From all around the world

Données en HEP

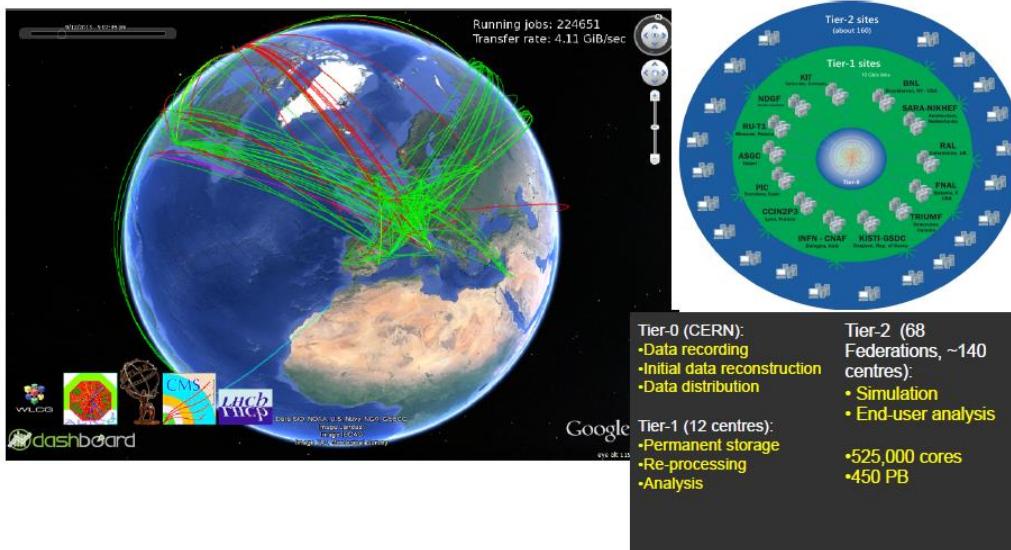
The (Big) DATA

10^7 "sensors" produce 5 PByte/sec
Complexity reduced by a Data Model



Analytics in real time filters to 0.1–1 Gbyte/sec (Trigger)
Data + Replica move with a Data Management Policy
6 GB/s (600 TB/day)

Worldwide LHC Computing Grid



Marcello Maggi
INFN Senior Researcher
Istituto Nazionale Fisica Nucleare
Bari-Italy

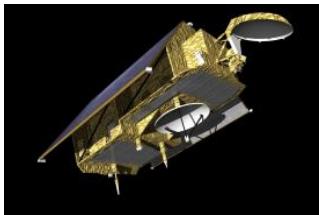


Copernicus Sentinel Data Policy



**Sentinel Data Policy =
FREE and OPEN access**

- Joint COM/ESA **Sentinel Data Policy Principles** have been prepared in 2009 - adopted by ESA MSs in Sep 2009
- **EU Delegated Act** on Copernicus Data and Information Policy has been adopted in 2013 (C(2013)4311, final)
- ESA got approval of updated **Sentinel Data Policy** from its Member States in Sep 2013. Main principles of Sentinel data policy:
 - Open access to Sentinel data by anybody and for any use
 - Free of charge data licenses
 - Restrictions possible due to technical limitations or security constraints

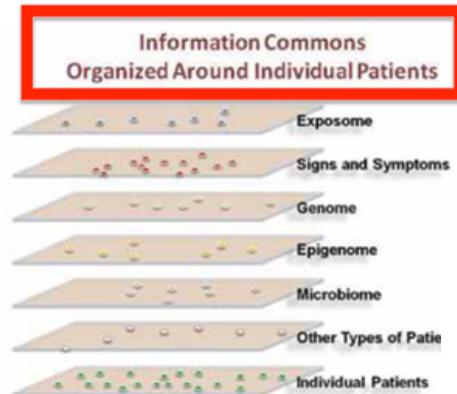
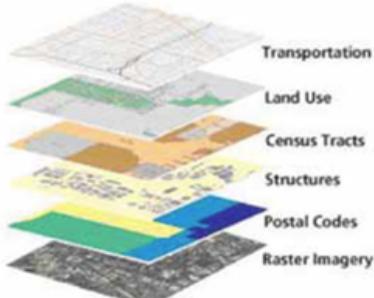


- Une organisation, une série d'instruments, une politique de données

Les données de santé

Intégration des données pour une
Médecine translationnelle, prédictive et
personnalisée

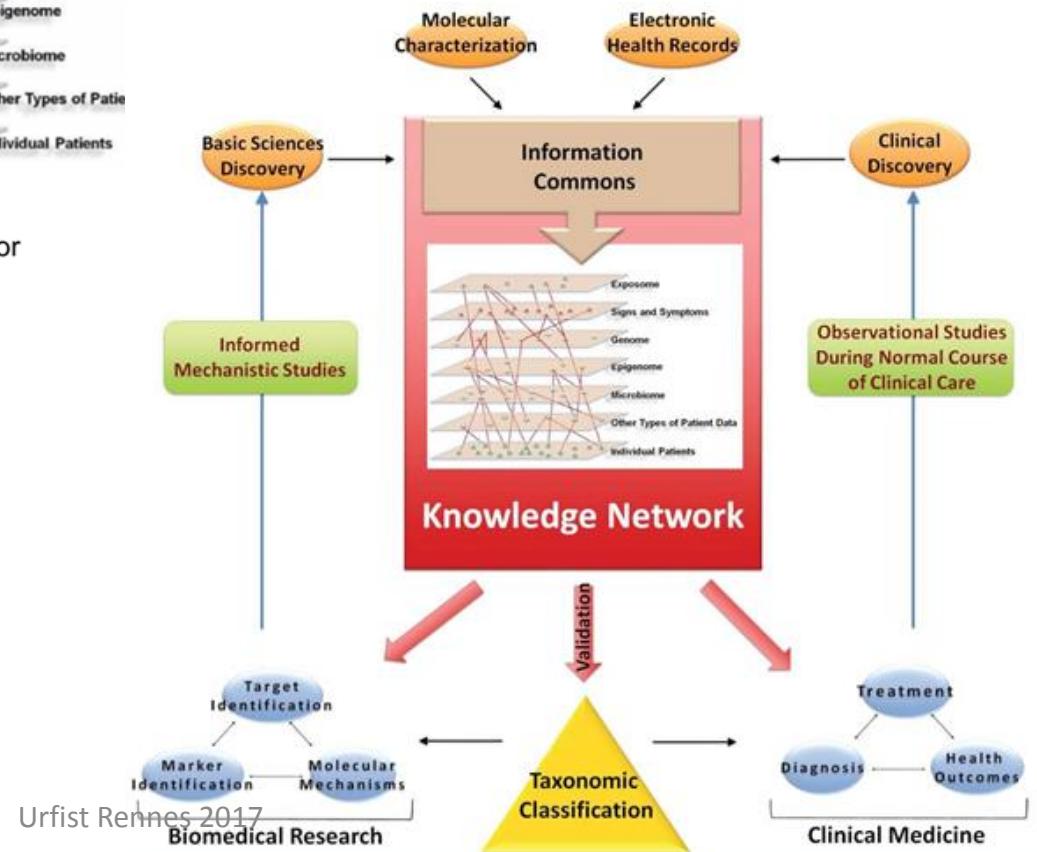
Google Maps: GIS layers
Organized by Geographical Positioning



Toward Precision Medicine: Building a Knowledge Network for
Biomedical Research and a New Taxonomy of Disease
Report from National academy of science, USA, 2011

- Utilisation des données cliniques
- Développement d'une médecine personnalisée et prédictive
- relations gène/médicament,
symptômes/maladies, risques
environnementaux/expression des gènes

Marc CUGGIA (MD,PhD)
Health Big Data team (LTSI) -
Clinical Investigation Center (CHU Rennes)
INSERM – Medical School
Université de Rennes 1 - BRITTANY



Il y a partage et partage

- Importance du facteur disciplinaire
- Pratiques de communautés
- Lien instruments/données
- Barrières légales très diverses

Culture du partage de la donnée

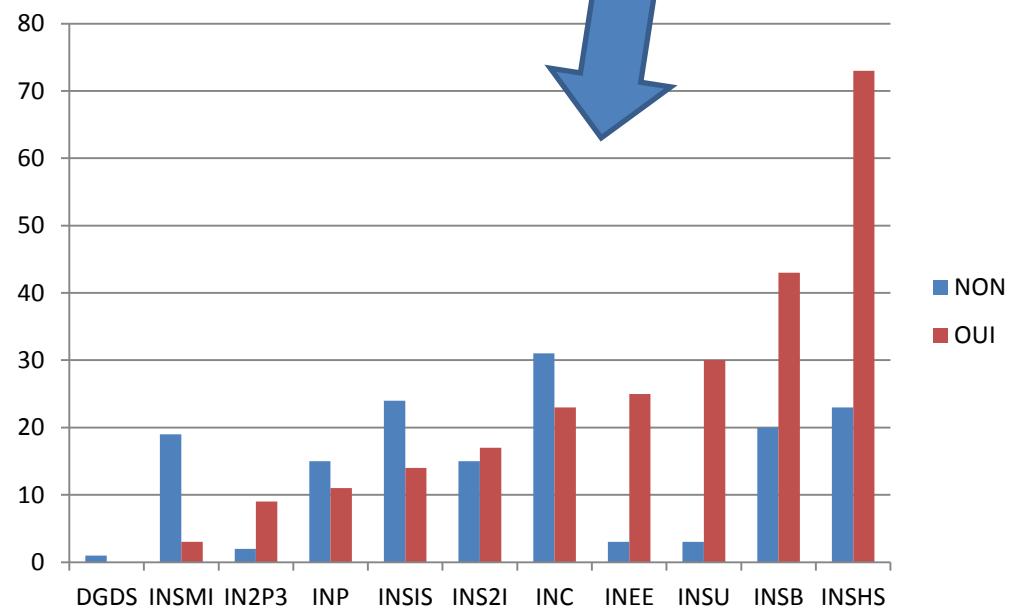
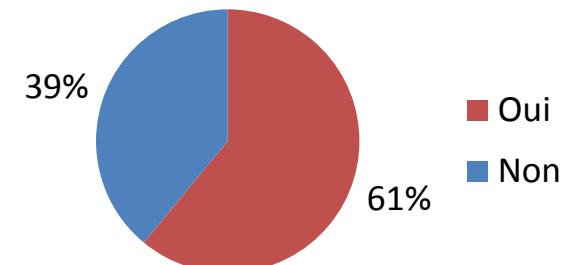
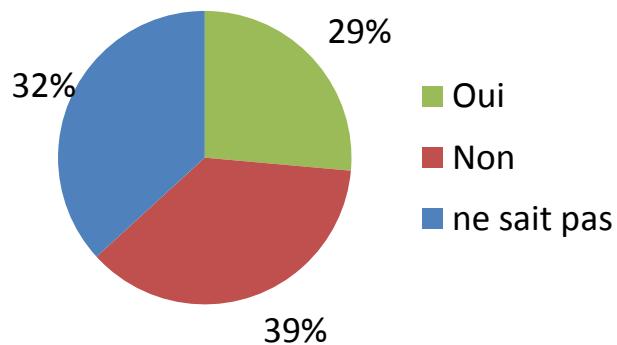
- Mes données sont à moi et ... tes données sont à moi également !
- Des initiatives : ScienceEurope, Knowledge Exchange, LERU, LIBER, DCC, Nactem, RDA,...
- Des pratiques de communautés : astrophysique, génomique,...
- Des politiques d'organismes : Inra, Irstea,...
- Des politiques d'infrastructures (nationales et/ou européennes)
- Des politiques de financeurs : ANR, CE:H2020
- Plusieurs niveaux d'intervention : besoin de cohérence
- Qu'en pensent les chercheurs ?

Enquête IST auprès des directeurs d'unité CNRS

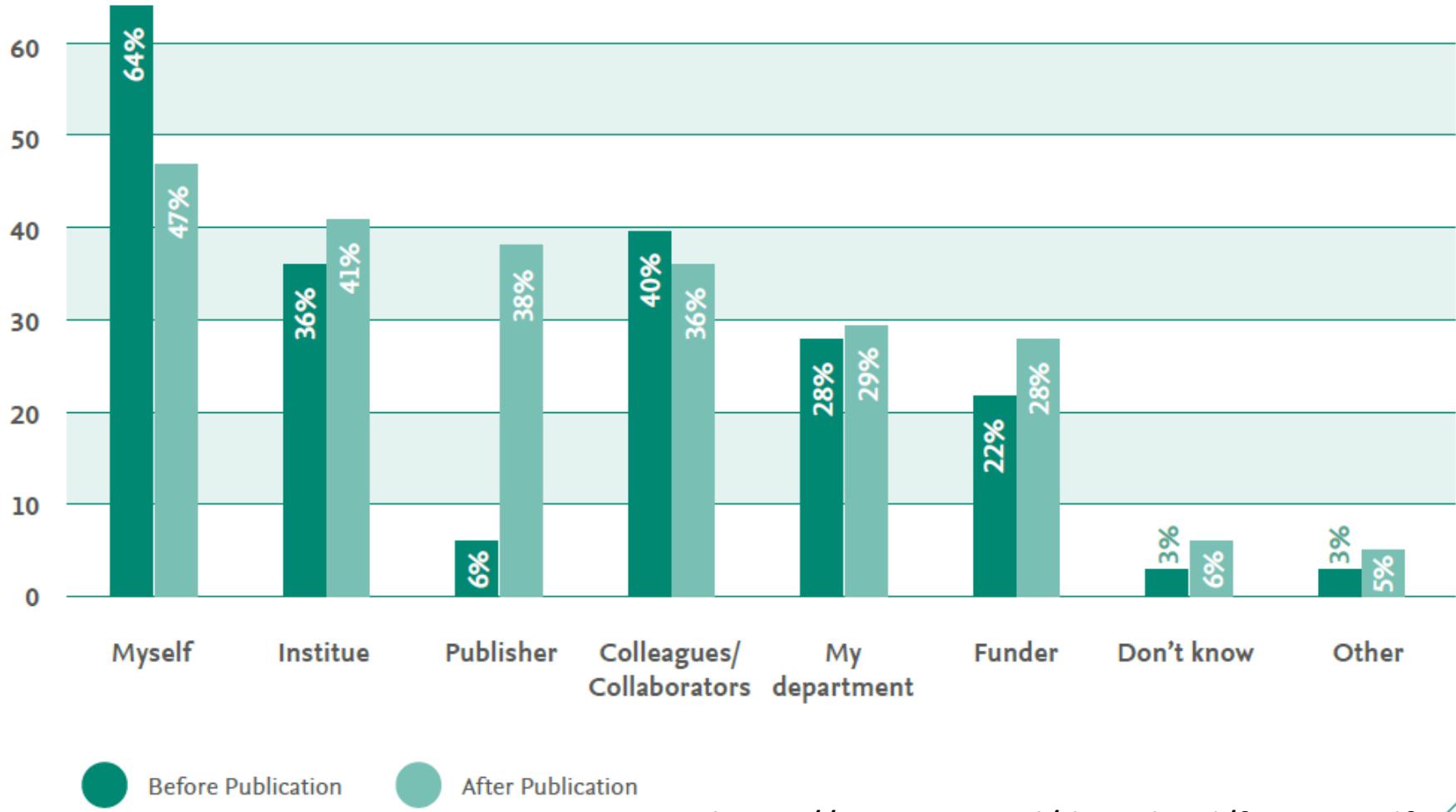
Données de la recherche

Les recherches conduites dans votre laboratoire produisent-elles des données de la recherche nécessitant des pratiques de gestion (description, archivage, diffusion...) ?

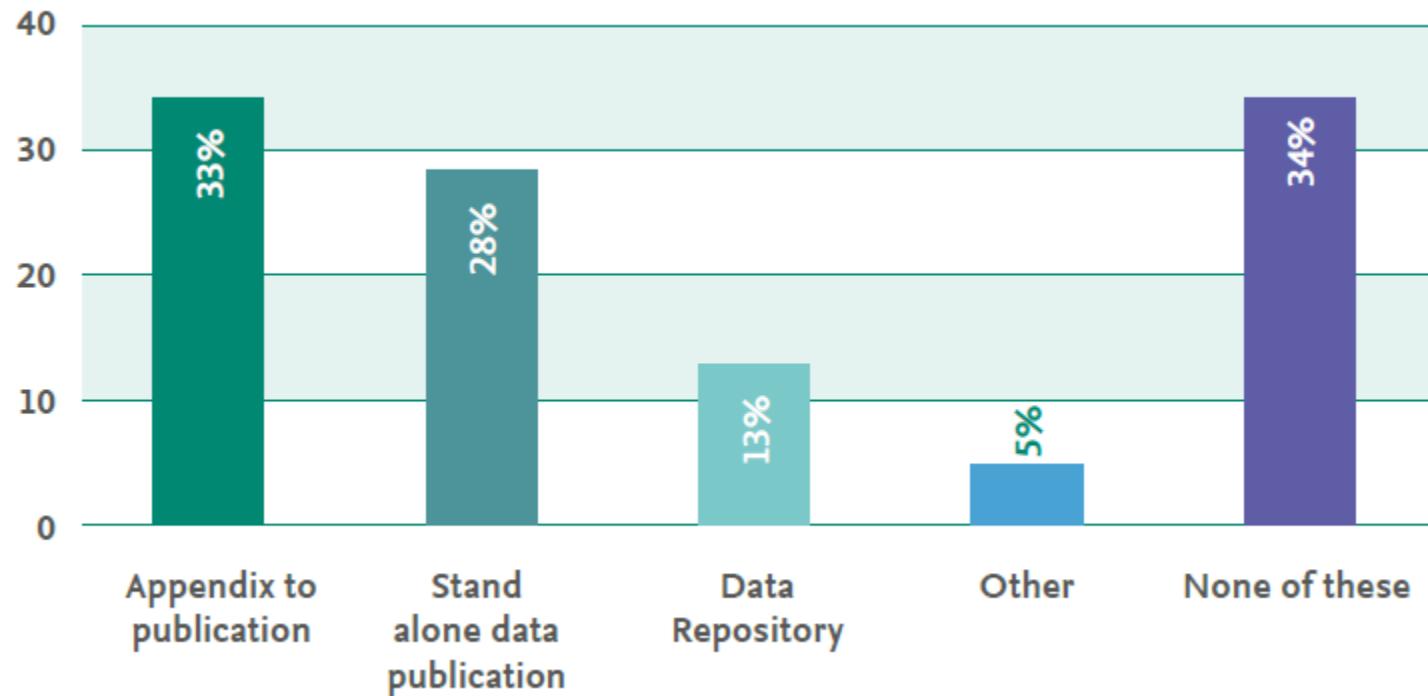
Pensez-vous que vos données de recherche soient libres de droits ?



A qui appartiennent les données, avant et après publication (n=1162)

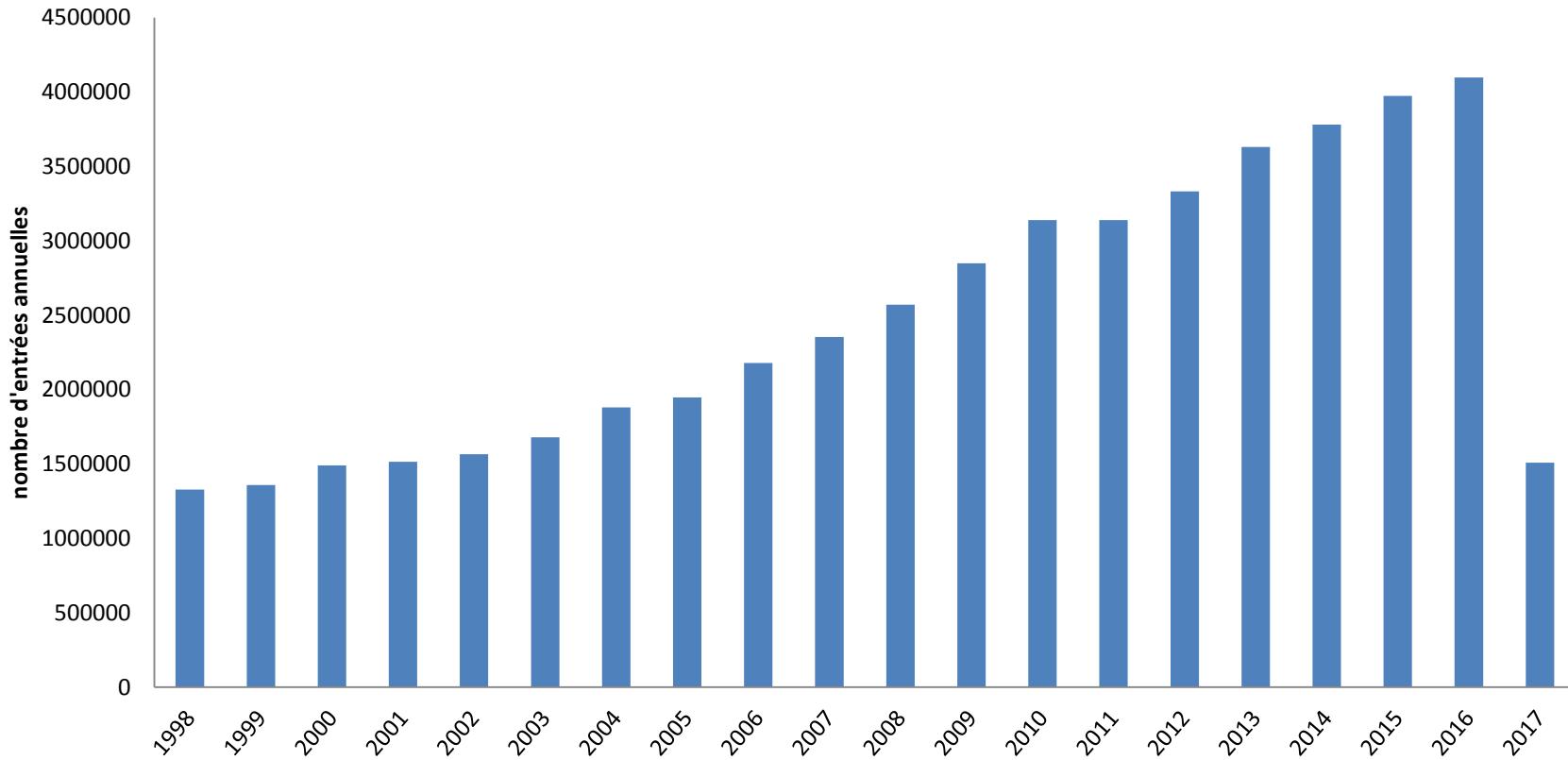


Diffusion des données n=1162



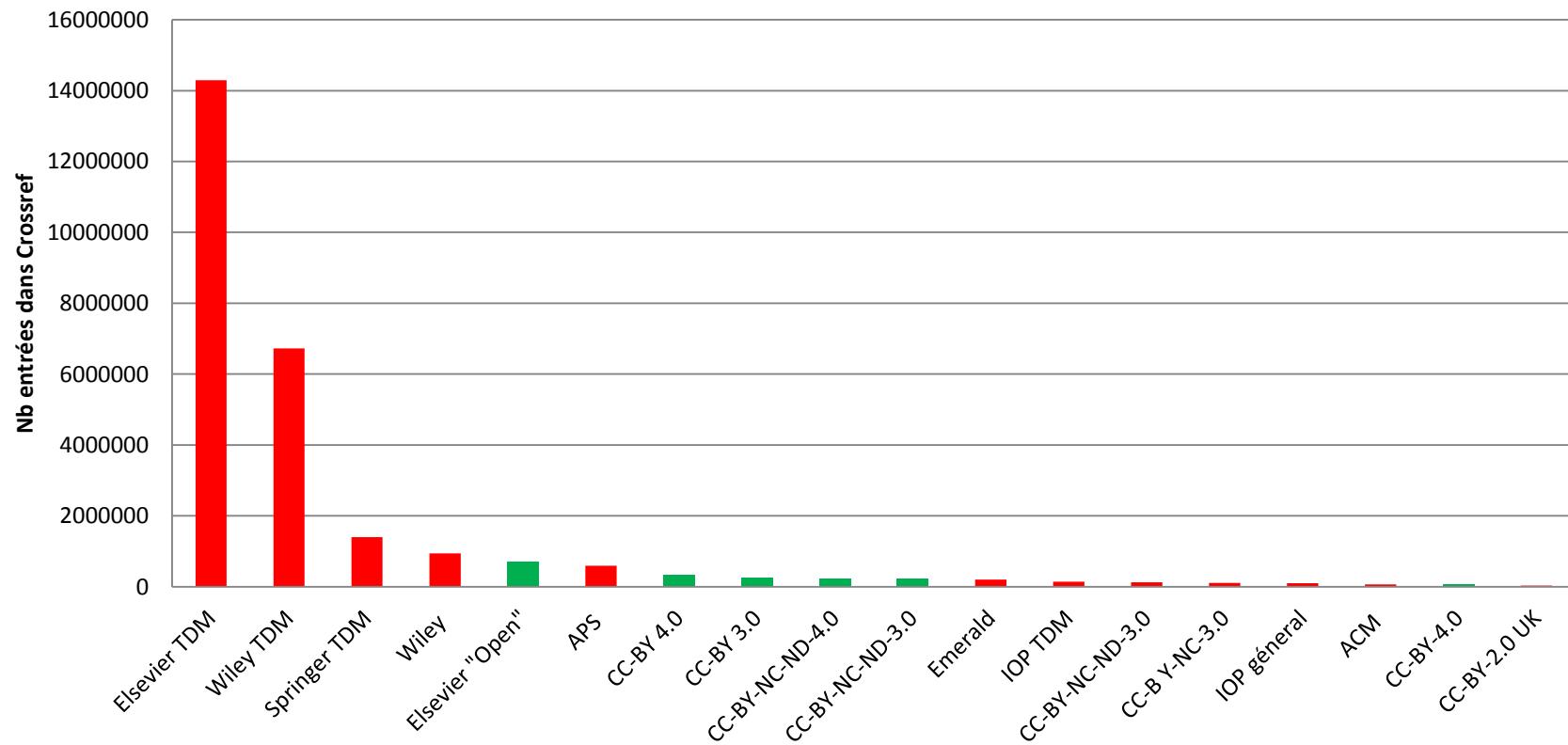
<https://www.cwts.nl/download/f-53w2.pdf>

les publications dans Crossref



Mise à jour à : <http://api.crossref.org/works?facet=t&rows=0>

open ? Des progrès à faire....



IL

FAUT

LE

Findable

FAIR !

Accessible

Interoperable

Re-usable

Soyez **FAIR** !

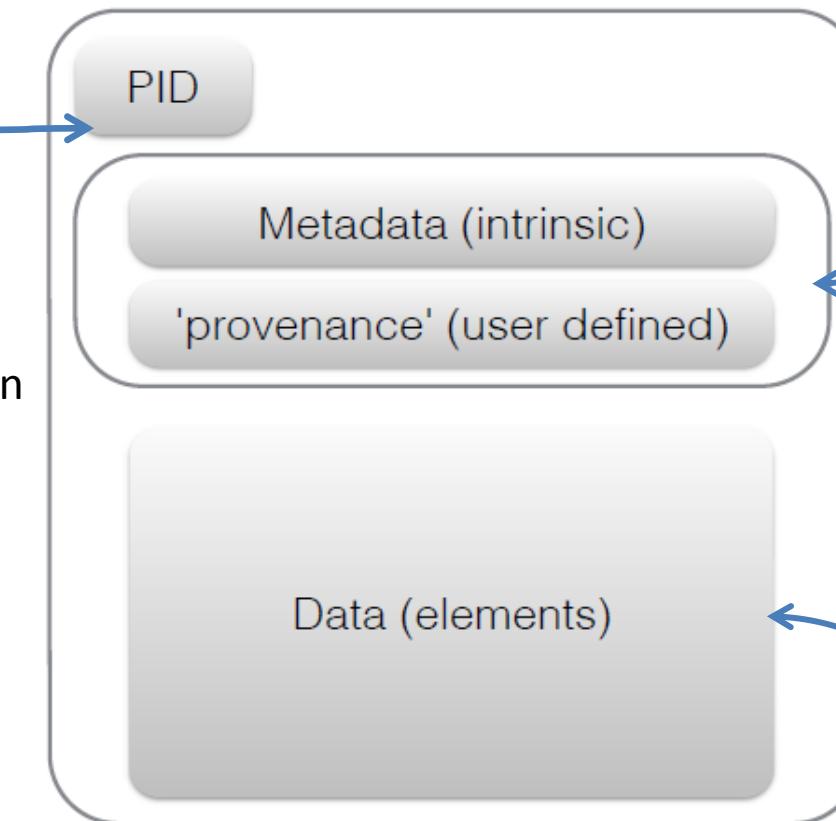
Findable, **A**ccessible, **I**nteroperable, **R**eusable

Identification

DOI : digital Object Identifier, service

DataCite

- Unicité
- Désambiguisation
- Pérennité



Métadonnées descriptives/analytiques :

- Formats standards
- Règles d'interopérabilité
- Licences d'utilisation
- Métriques de qualité
- Règles de citations

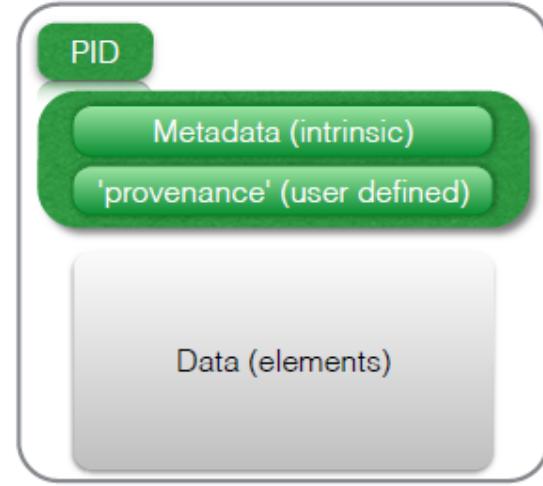
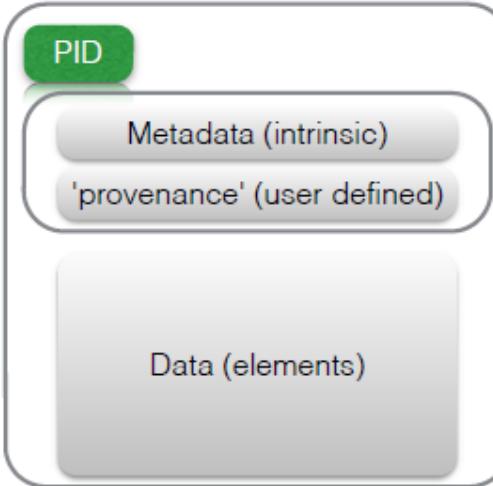
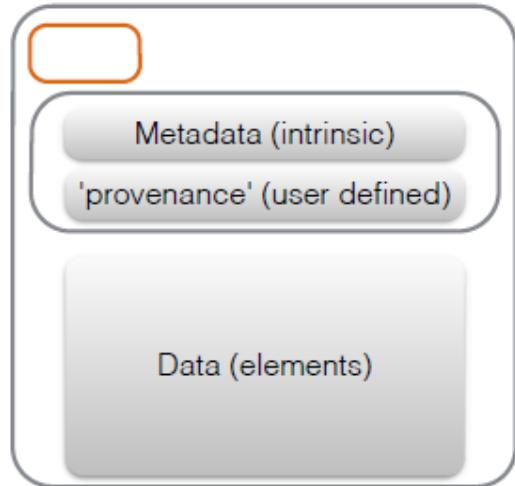
Article,
jeu de données,
vidéo, carte, logiciel...

Data as increasingly FAIR Digital Objects

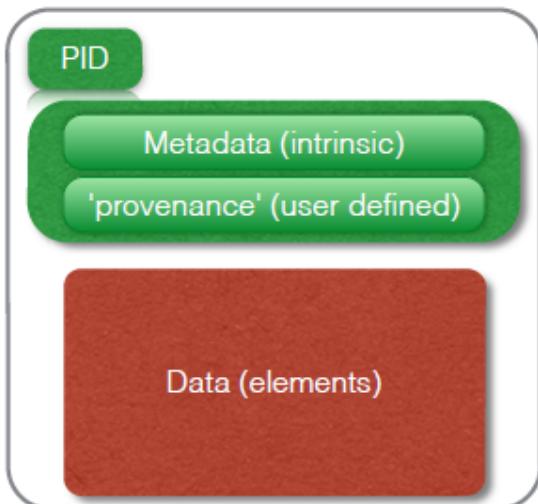
Totally UNFAIR

Findable
Usable for Humans

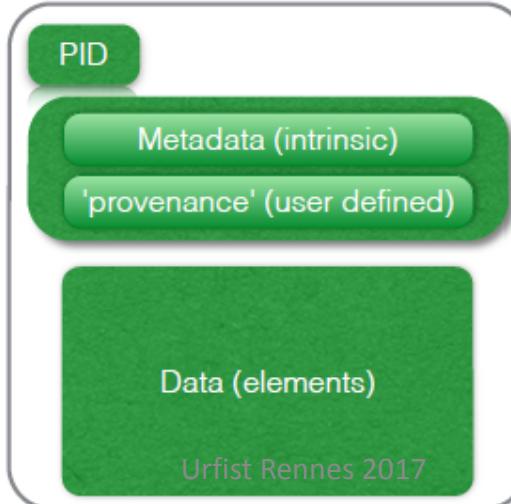
FAIR metadata



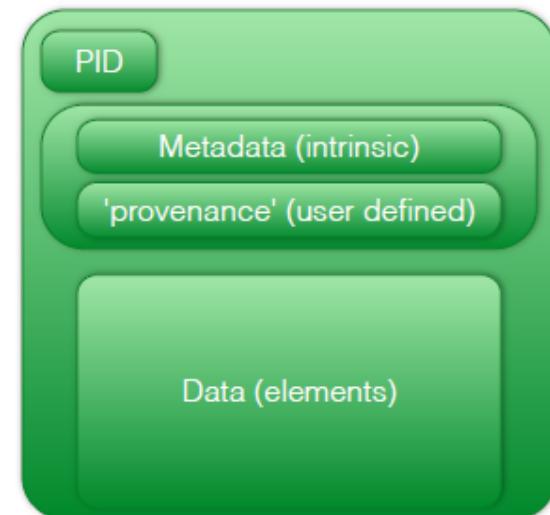
FAIR data-
restricted access



FAIR data-
Open Access

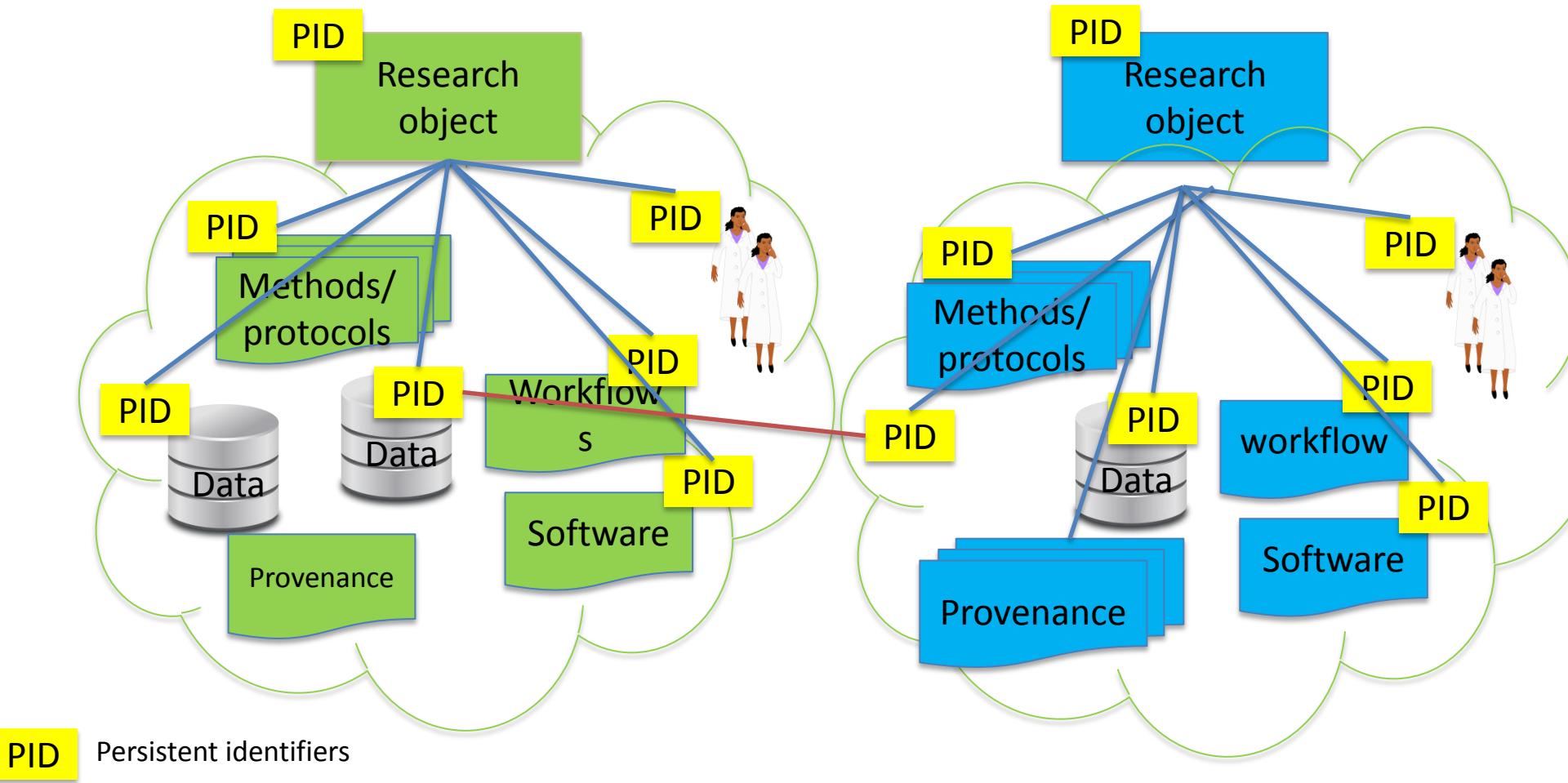


FAIR data-
Open Access/Functionally Linked



Linked research objects :

a framework to enhance discovery and reuse of data

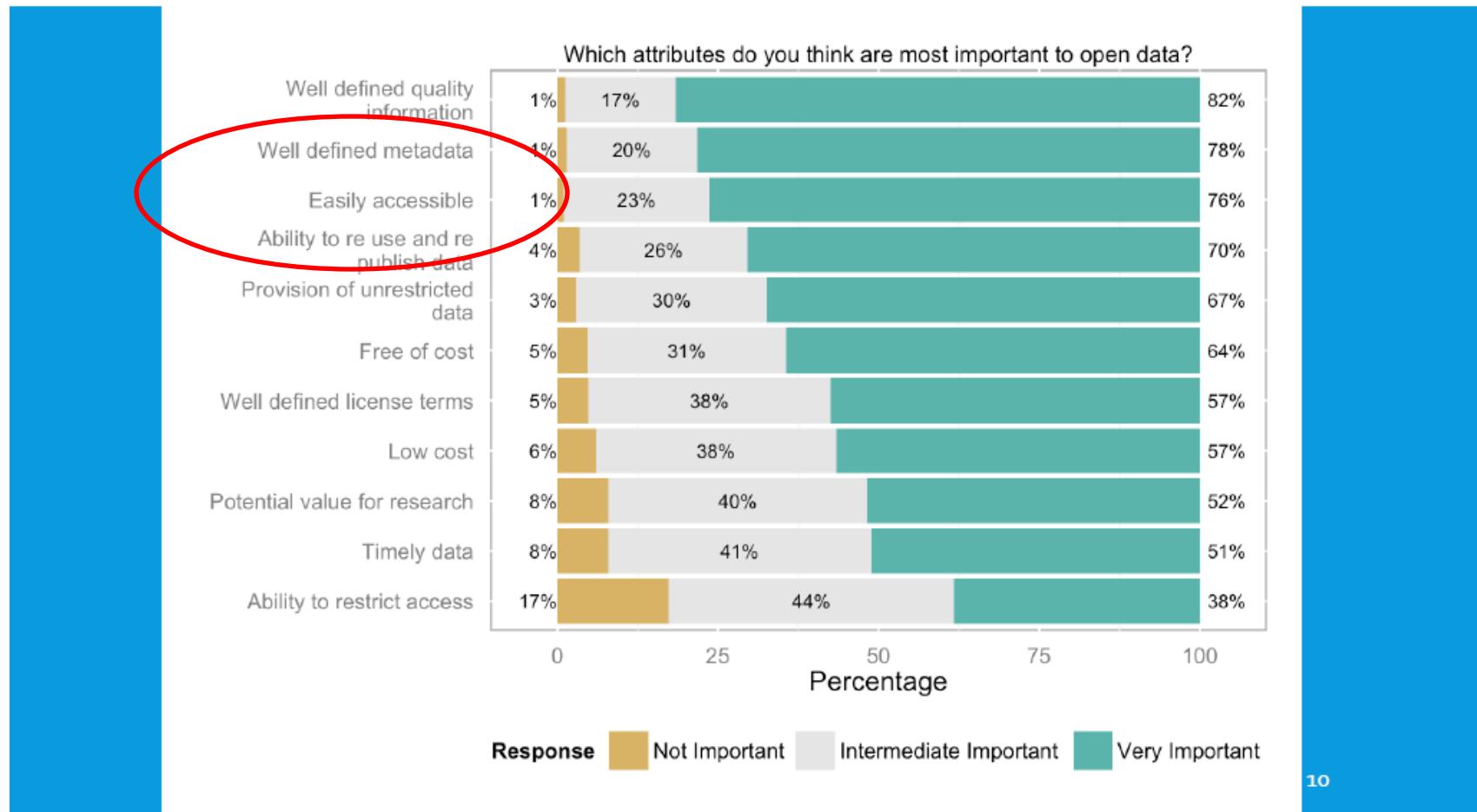


Où se niche la qualité des métadonnées ?

- Richesse des métadonnées lisibles en machine
- Identifiants
- Variété et disponibilité de formats
- Règles d'interopérabilité documentées
- Licences publiées
- Métriques de qualité affichées
- Mises à jour des métadonnées
- Règles de citations

Données ouvertes : caractéristiques attendues

Qualité des données et des métadonnées

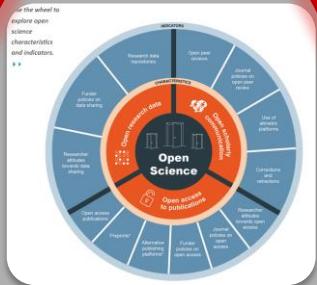


Conseils de lecture...

<https://www.innovationpolicyplatform.org/content/open-science>



<http://ec.europa.eu/digital-agenda/en/news/final-report-science-20-public-consultation>
<https://royalsociety.org/topics-policy/projects/science-publicenterprise/report/>



Ouverture ?
On saura vraiment ce que c'est quand on l'aura faite...

Démarche européenne :
Vers un partage global

Cadre éthique et juridique

Nouveaux services,
nouveaux métiers

Sujets brûlants

Open Science à l'échelle européenne



- Quitter une situation de blocage
 - changer les processus d'évaluation
 - Modifier les règles de PI
 - Faciliter le TDM
 - Changer les modèles économiques de la diffusion de la science
- Promouvoir des politiques de science ouverte
 - Adopter (et adapter !) des principes d'accès libre
 - Stimuler les pratiques de recherche et d'innovation basées sur les **données**
- Développer des infrastructures de recherche
 - Basées sur des **principes de partage**
 - Mutualisées
- Impliquer les acteurs de la recherche
 - Chercheurs, personnels de soutien, société
 - Former, former, former...

<http://www.eu2016.nl/documenten/rapporten/2016/04/04/amsterdam-call-for-action-on-open-science>

Vision CE : les bénéfices du partage des données

“Great opportunities for the **society**”

- Better value for money
 - By **strengthening the productivity of the European science** and research system through the uptake of results by businesses, in particular SMEs that may not have the resources to pay for access to research results
- More transparency, openness and collaboration
 - leading to a **higher degree of responsiveness** of the research community to societal challenges
- A sound science and society relationship
 - More openness may also lead to more **trustworthy science** from the point of view of the citizen and civil society organisations (NGOs)
- Big and open data are estimated to add 1.9% of EU-28 GDP by 2020

D'après J.F. Dechamp, CE, Directorate-General for Research & Innovation

Vision CE : les bénéfices du partage des données

“Great opportunities for **researchers**”

- Wider dissemination and sharing of the results
- Involvement in **more interdisciplinary research**
- **More visibility and credit** for those collecting and sharing underlying research data
- Involvement in **international networks** full of potential
- **New career paths** e.g. data scientists, start-ups, science diplomacy

D'après J.F. Dechamp, CE, Directorate-General for Research & Innovation

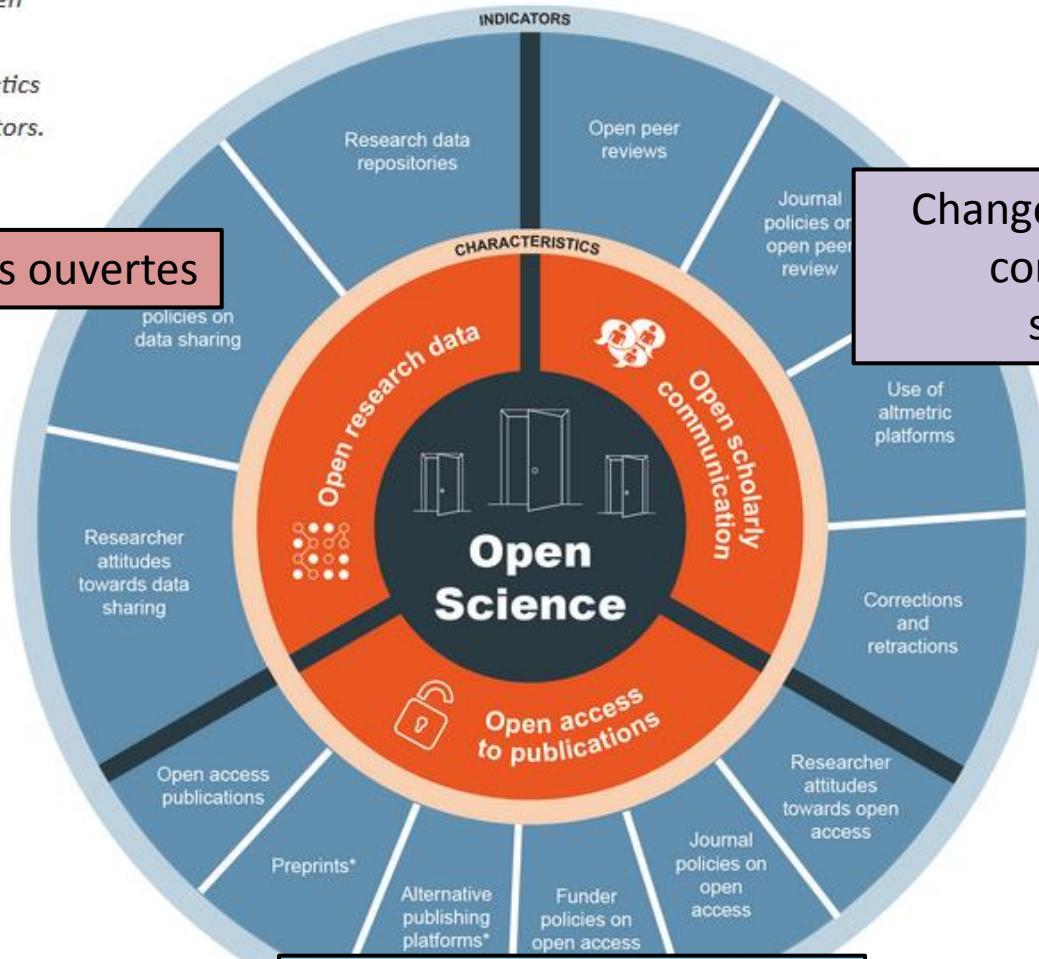
Open Science monitor

<http://ec.europa.eu/research/openscience/index.cfm?pg=home§ion=monitor>

*Use the wheel to
explore open
science
characteristics
and indicators.*



Données ouvertes



Changer les règles de la
communication
scientifique

Libre accès aux publications
Urfist Rennes 2017

Une déjà longue histoire...

ERA n'est pas tombé du ciel bruxellois...



The European Research Area

Evolving concept,
implementation challenges



IN-DEPTH ANALYSIS

EPRS | European Parliamentary Research Service
Author: Vincent Bellon
Members' Research Service
March 2016 — PE 579.097

EN

Key European stakeholders for research policies

[ALLEA](#): the federation of All European Academies was founded in 1994 and currently brings together 56 academies in more than 40 countries from the Council of Europe region.

[Business Europe](#): is an association founded in 1958 of 40 national business federations from 34 countries representing the interests of enterprises.

[CESAER](#): the Conference of European Schools for Advanced Engineering Education and Research – is a non-profit international association of 50 European universities of technology and engineering schools/faculties from 24 countries, founded in 1990.

[EARTO](#): the European Association of Research and Technology Organisations is an association of 91 members representing more than 350 research and technology organisations across the European Union and countries associated with the framework programme founded in 2000.

[EIROForum](#): this forum of eight European Intergovernmental Research Organisations was established in November 2002 by the signature of a Charter.

[ESF](#): the European Science Foundation was established in 1974 to act as a coordinating body for Europe's main research funding and research performing organisations. Its mandate was progressively modified by the creation of Science Europe and terminated in 2016.

[EUA](#): the European University Association founded in 2001 represents 850 universities from 47 countries.

[Eurodocs](#): the European Council of Doctoral Candidates and Junior Researchers is a federation of 35 national associations representing early stage researchers founded in 2002.

[Euroscience](#): is the non-profit grassroots researchers' association in Europe founded in 1997.

[LERU](#): the League of European Research Universities is an association of 21 research-intensive universities from 10 countries founded in 2002.

[Nordforsk](#): NordForsk was established in 2005 by the Nordic Council of Ministers and represents the actors of the research systems of the Nordic countries.

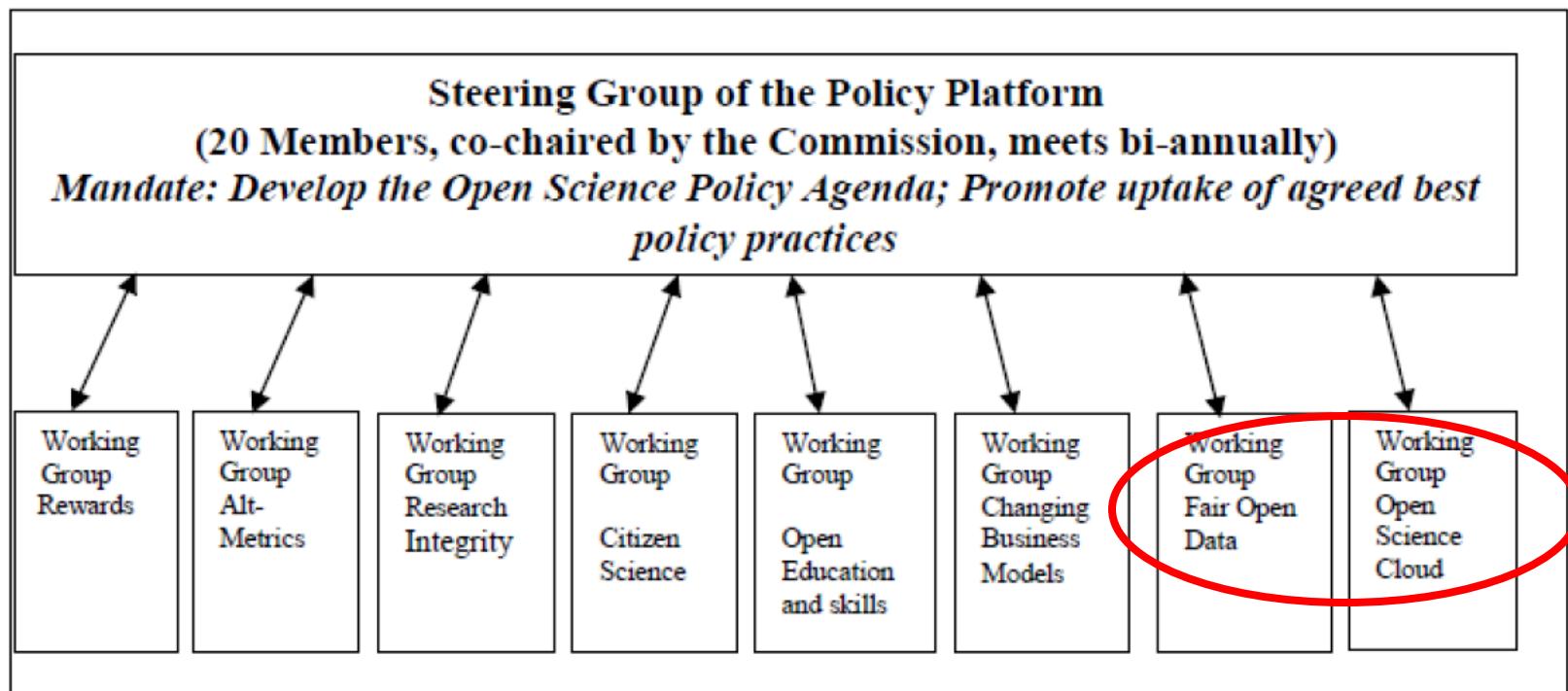
[Science Europe](#): is an association of 47 European Research Funding Organisations and Research Performing Organisations from 27 countries, founded in October 2011. It included EUROHORCs, the European Heads of Research Councils association established in 1992 to represent the national research funding organisations (national research councils).

2016



DIRECTORATE-GENERAL FOR RESEARCH AND INNOVATION (RTD)

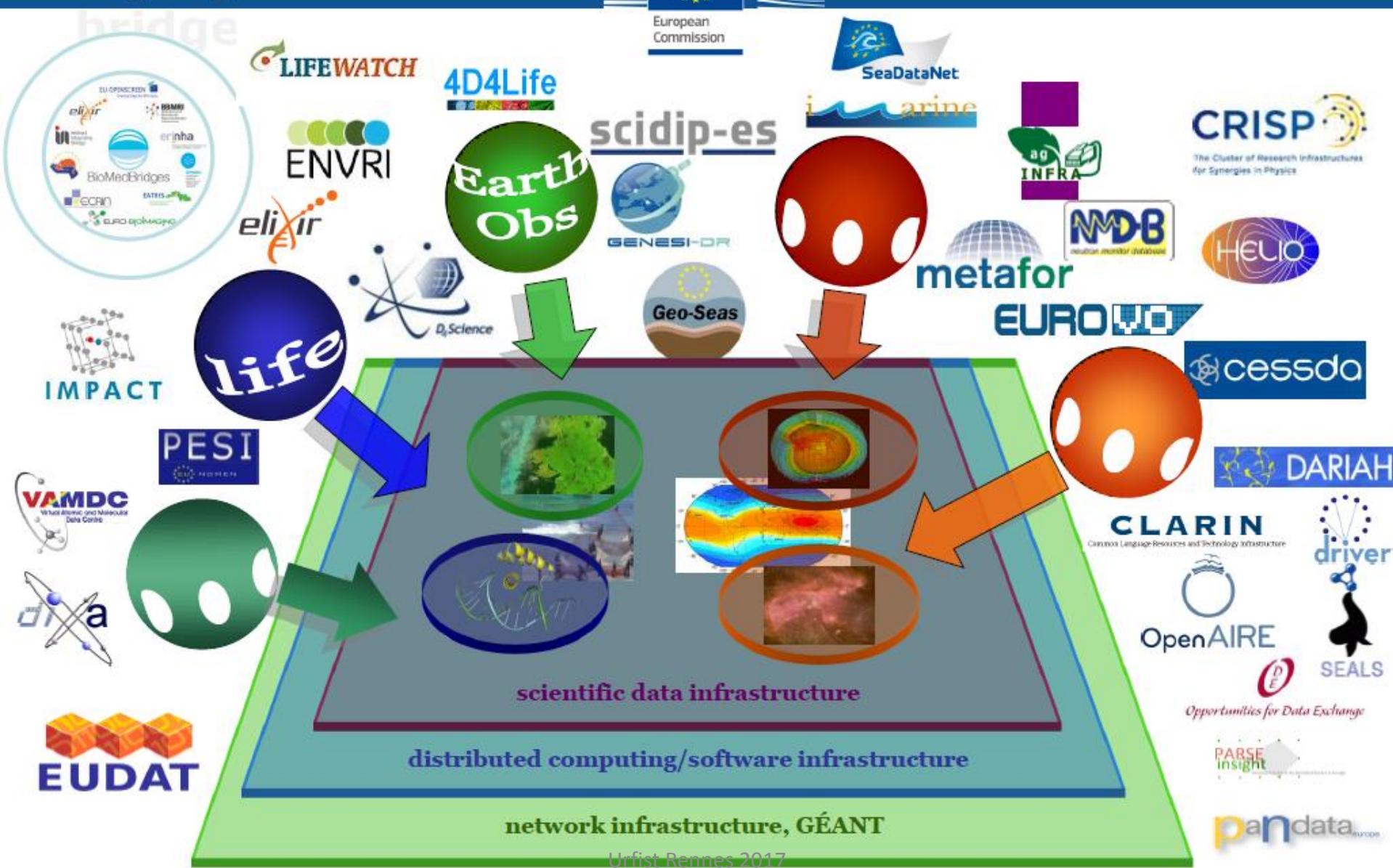
New policy initiative: The establishment of an Open Science Policy Platform



FAIR : Findable, Accessible, Interoperable, Reusable

Unisys Rennes 2017

data infrastructure: bridging islands



EUROPEAN OPEN SCIENCE CLOUD

BRINGING TOGETHER CURRENT AND FUTURE DATA INFRASTRUCTURES



A trusted, open environment
for sharing scientific data



Open and seamless
services to analyse and
reuse research data

<http://ec.europa.eu/research/openscience/index.cfm?pg=open-science-cloud>



Linking data



Connecting across borders
and scientific disciplines



Connecting scientists
globally



Improving science



Long term
and sustainable

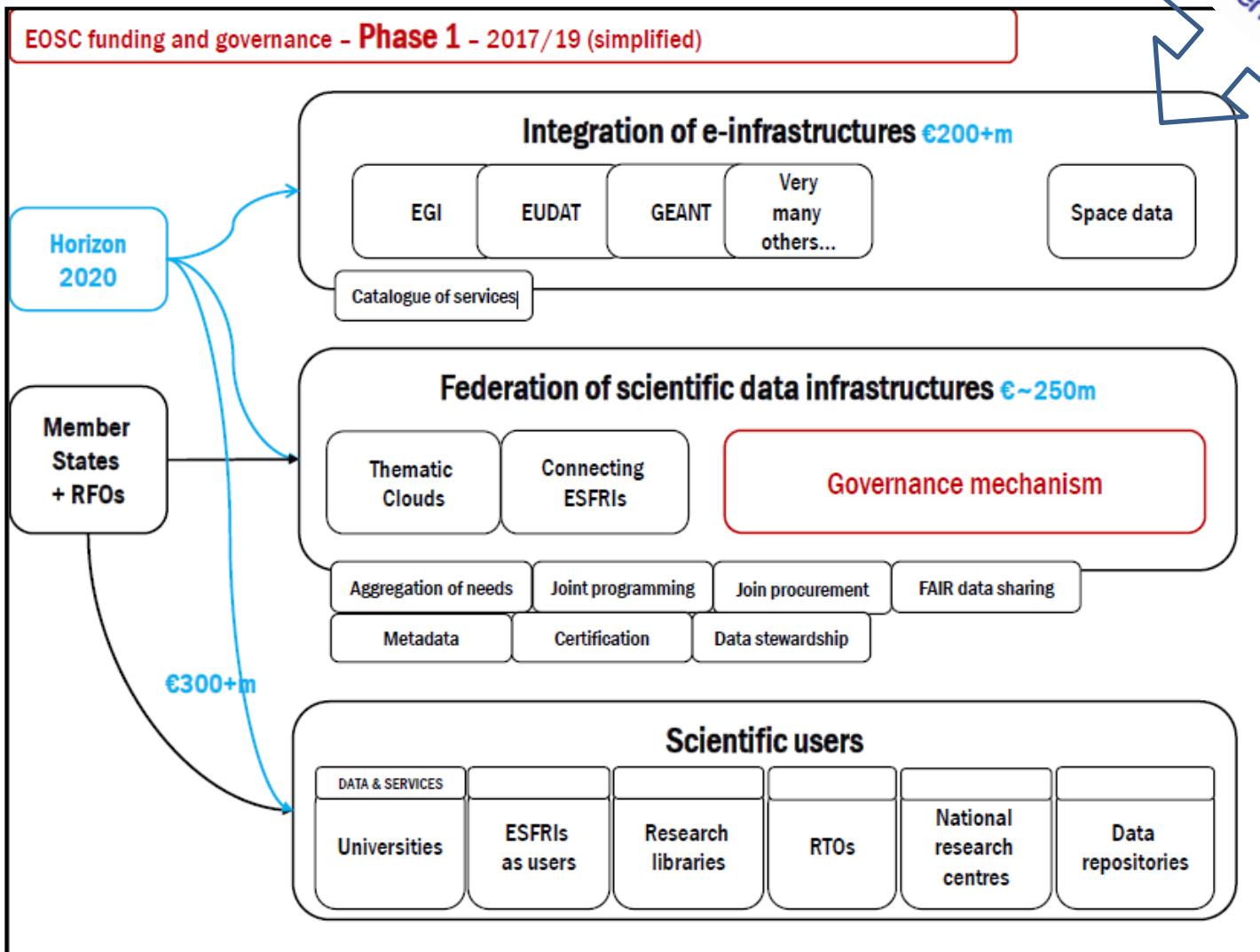
Urfist Rennes 2017



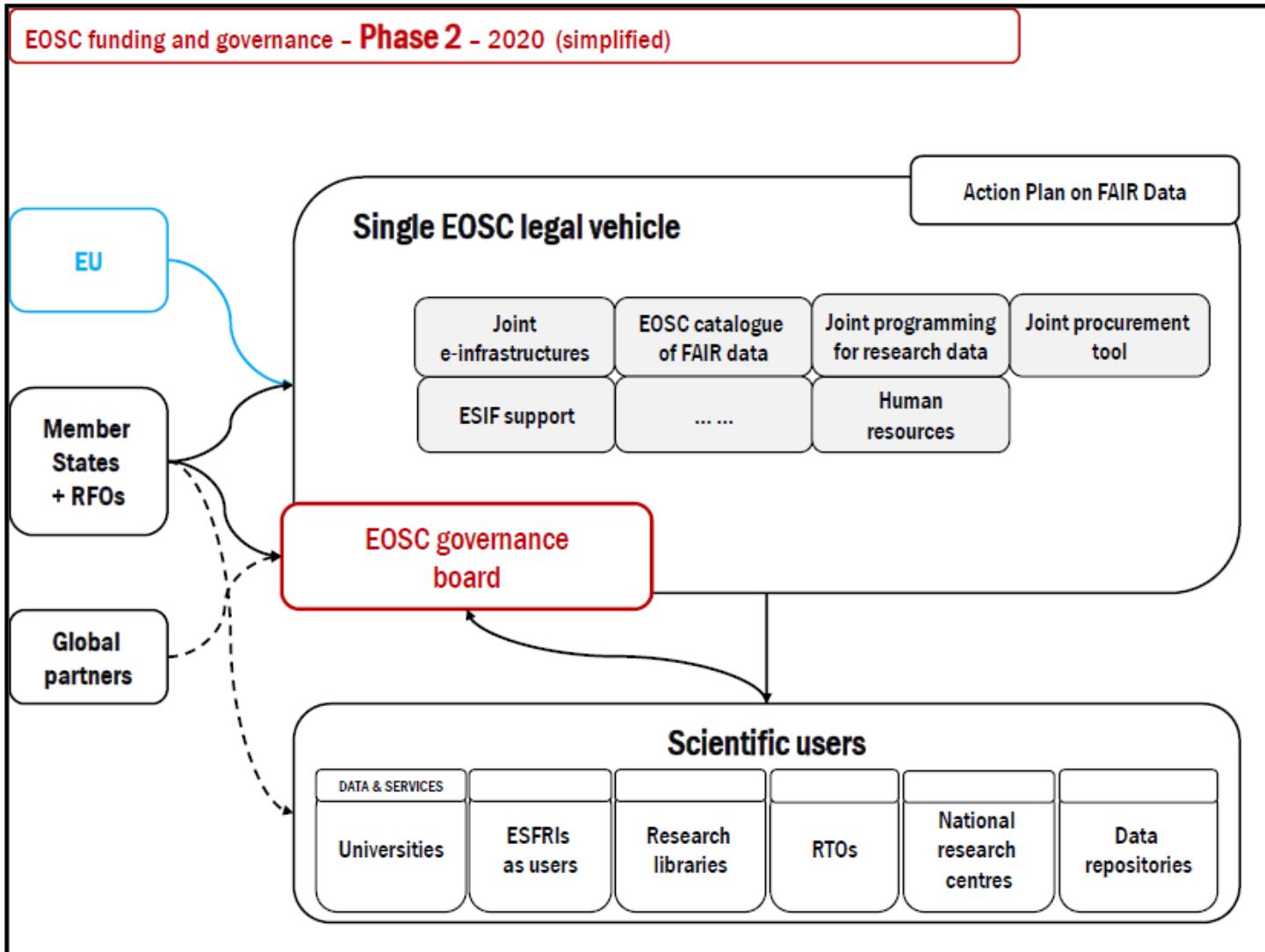
Proposed solutions: the European Cloud Initiative communication (April 2016)

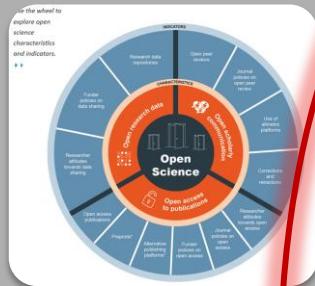
- **European Open Science Cloud (EOSC)**
 - Integration and consolidation of e-infrastructures
 - Federation of existing research infrastructures and scientific clouds
 - Development of cloud-based services for Open Science
 - Connection of ESFRIs to the EOSC
- **European Data Infrastructure (EDI)**
 - Development and deployment of large-scale European HPC, data and network infrastructure
- **Widening access and building trust**
 - SMEs, Industry at large, Government

EOSC, s'appuyer sur l'existant



...dans un cadre de gouvernance novateur





Ouverture ?
On saura vraiment ce que c'est quand on l'aura faite...

Démarche européenne :
Vers un partage global

Cadre éthique et juridique

Nouveaux services,
nouveaux métiers

Sujets brûlants



Projet de loi pour une République numérique



Gouvernement, le 26 septembre 2015

147709 8501 21411
votes contributions participants

Partager ▾

- Consultation citoyenne : 5 articles de loi rajoutés !
- Demande de la communauté scientifique de voir inscrit dans la loi le sujet des productions scientifiques dans le contexte numérique
- Des réticences fortes exprimées par les éditeurs sur les orientations d'ouverture et de partage
- Des avancées pour la recherche mais vote positif : groupe écologiste, groupe radical, républicain, démocrate et progressiste (2RDP), groupe gauche démocrate et républicaine, groupe socialiste, républicain et citoyen, union des démocrates et indépendants
- Abstention : groupe les républicains
- Mais... les communs informationnels ont disparu du projet de loi

Cadre juridique national pour l'ouverture des données

- Un contexte en évolution...
- Loi CADA, modifiée par la loi Valter
- Transposition de la directive PSI
- Loi Pour une république numérique
- A venir, modification de la directive européenne sur les droits d'auteurs et droits voisins

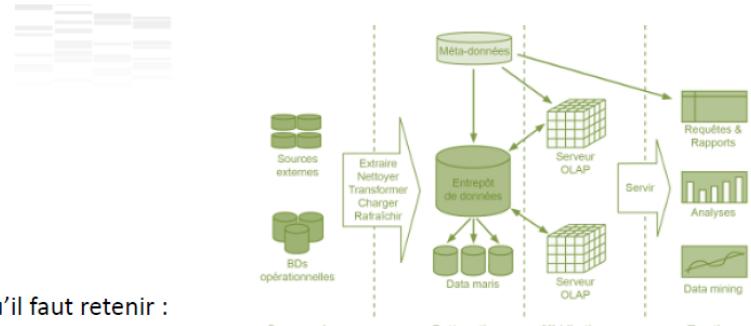
Ouverture des données de recherche

Guide d'analyse du cadre juridique en France*


Contenu sous licence Creative Commons Attribution 4.0 International (CC BY 4.0).

* Le présent guide est issu des réflexions d'un groupe de travail inter-organismes. Il ne prétend pas à l'exhaustivité et est fourni uniquement à titre d'information. Il ne saurait en tout état de cause se substituer aux politiques d'établissements, au respect des dispositions législatives ou réglementaires et au respect de la jurisprudence applicable en la matière. Ce guide peut évoluer.
Membres du groupe de travail : BECARD Nicolas (INRA), CASTETS-REINARD Céline (UT1), CHASSANG Gauthier (Inserm, Membre de la Plateforme Genotoul Société), COURTOIS Mery-Audrey (INRA), DANTANT Martin, GONDON Nathalie (co-animateur, INRA), MARTIN Caroline (Inria), MARTELLETTI Andrea (stagiaire à INRA, M2 droit et informatique), MENOCZI-CAMINADE Alexandra (UT1), MORCRETTE Nathalie (co-animateuse, INRA), NEIRAC Claire (Gres) - Guide redoré par MARTELLETTI Andrea dans le cadre de son stage de fin d'études de M2 Droit et informatique.

Données de la recherche, essai de définition



Ce qu'il faut retenir :

- ⇒ Pas de distinction entre données brutes, élaborées ou métadonnées d'un point de vue juridique.
- ⇒ **Pas de droit de propriété** dans la plupart des cas sur la donnée (données machine, etc.). Elle est considérée comme une information « de libre parcours ». A ce titre, l'établissement du producteur de la donnée peut restreindre ou non sa diffusion.
- ⇒ Mais il existe deux exceptions où une « propriété » peut s'exercer.

Des exceptions en fonction de la nature des données



Des interdictions totales de diffusion :

- Les documents réalisés en exécution d'un contrat de prestation de services exécuté pour le compte d'une ou de plusieurs personnes déterminées (non publiques)
- Les données relevant du secret défense
- Les données relatives aux secrets professionnels : secret des procédés, secret des informations économiques et financières, secret des stratégies commerciales ou industrielles
- Données portant atteinte à la sécurité du SI de l'administration (NOUVEAU).

Extraits...

Nathalie Gandon, Nathalie Morcrette,Inra

Exceptions : « données propriétaires »

- ⇒ Données sont soumises au **droit d'auteur** : textes, plans, photographies, etc. et notamment les publications scientifiques.
 - ⇒ **Condition** : **originalité** de la forme (pas de l'idée).
 - ⇒ **Conséquence** : pour utiliser ces données, **l'accord de l'auteur** est indispensable (donc attention!! au text mining) sauf exception de courte citation .
Le droit revient à l'auteur et non à l'établissement (sous réserve de conditions « d'autonomie »).
- ⇒ Données organisées en bases de données : **Droit sui generis** qui peut s'appliquer sous réserve de la preuve d'un investissement substantiel (le plus souvent financier). Le droit revient à l'investisseur (le plus souvent l'établissement).



Obligation d'ouverture...

Deux lois nouvelles :

- Loi Valter du 28 décembre 2015 relative à la gratuité et aux modalités de réutilisation des informations du secteur public (transposition de la directive PSI)
- Loi Lemaire du 7 octobre 2016 pour une République numérique (article 7, 30 et 38 notamment)
- les **établissements de l'ESR** sont soumis à l'**obligation** de rendre accessibles et librement réutilisables leurs données
 - => **les données de la recherche = *open data***

Ouverture par défaut mais...

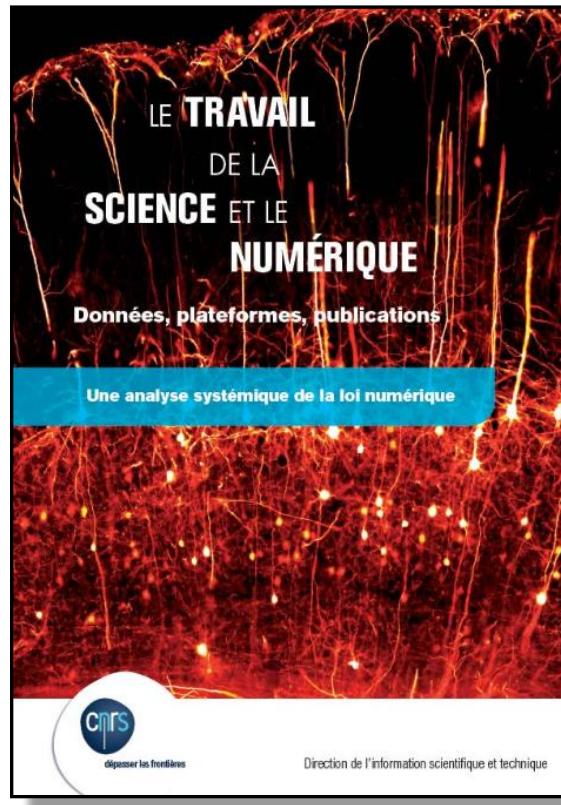
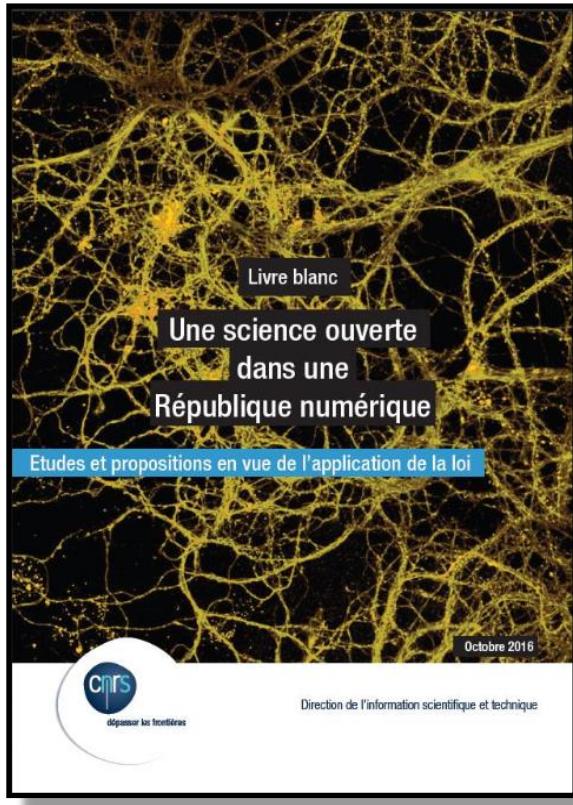
- Des restrictions, limitations et obligations
 - respecter des interdictions de diffusion
protection du potentiel scientifique et technique de la nation (PPSN), Zone à régime restrictif (ZRR)
secret défense, secret médical, secret des affaires
 - respecter les droits des tiers
respect de la vie privée, données à caractère personnel
droits de propriété intellectuelle : droits d'auteur, des bases de données
 - obligation de rendre accessibles certaines données
Données géographiques, environnementales (INSPIRE), projets H2020 (obligations contractuelles)

Rien n'est simple...

- Fouille de textes et de données
 - Article 38 : « The right to read is the right to mine »
 - rejet du décret d'application par le conseil d'état
- Licences d'utilisation
 - La liste prévue par décret ne mentionne pas les licences Creative Commons
- Révision de la « directive copyright »
 - Quid du TDM ?
 - Quid de l'exploitation commerciale ?

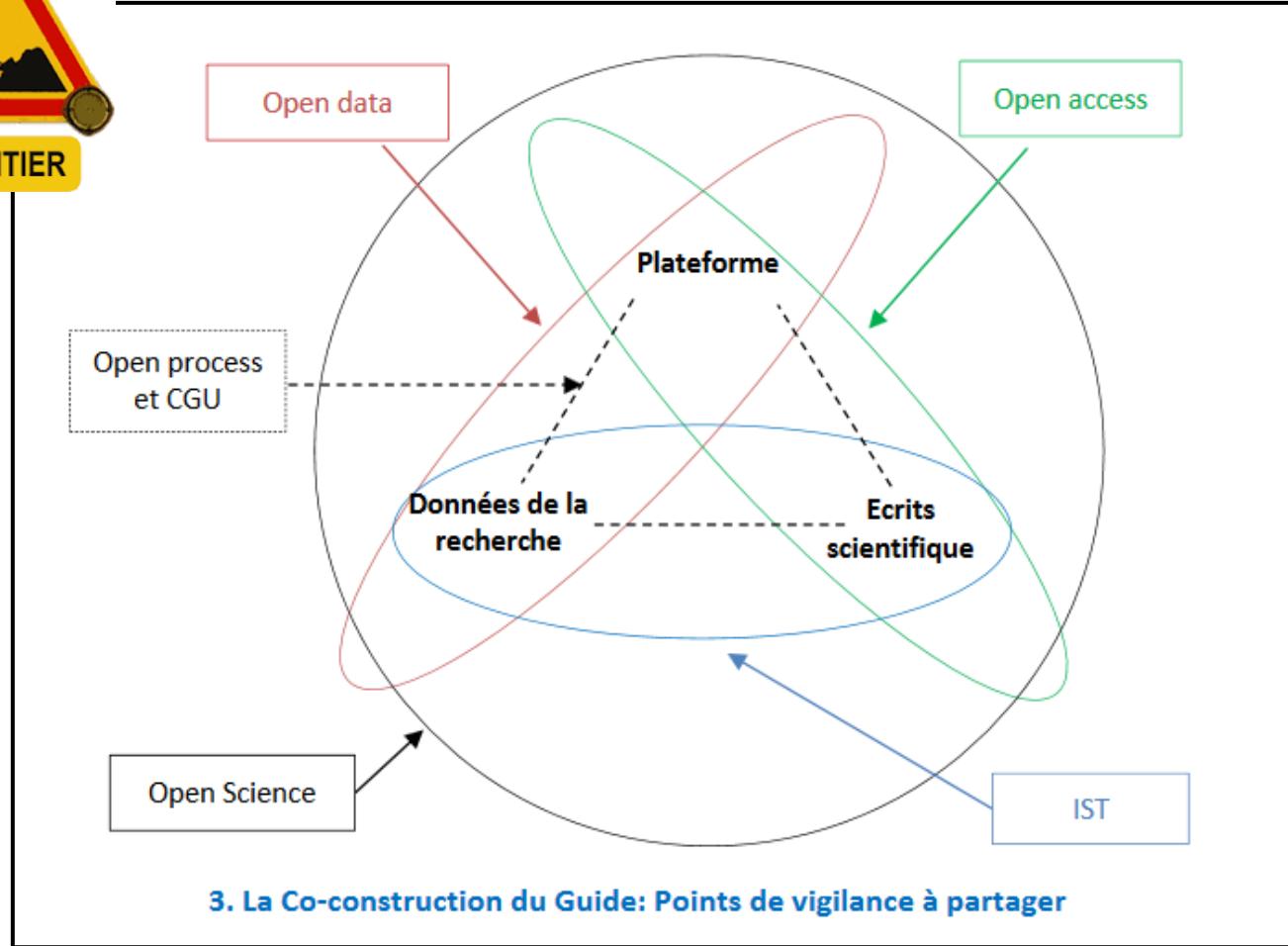
A suivre...

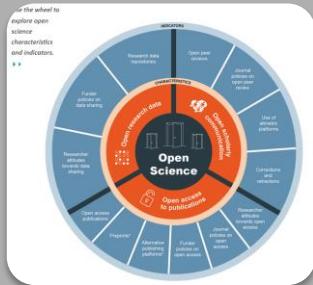
Cadre éthique et juridique de l'Open Data dans le contexte Open Science



http://www.cnrs.fr/dist/z-outils/documents/20170203_analyse_systémique_vf.pdf

L'APPLICATION DE LA LOI "POUR UNE REPUBLIQUE NUMERIQUE" UN "GUIDE PARTAGÉ POUR LE TRAVAIL DE LA SCIENCE"





Ouverture ?
On saura vraiment ce que c'est quand on l'aura faite...

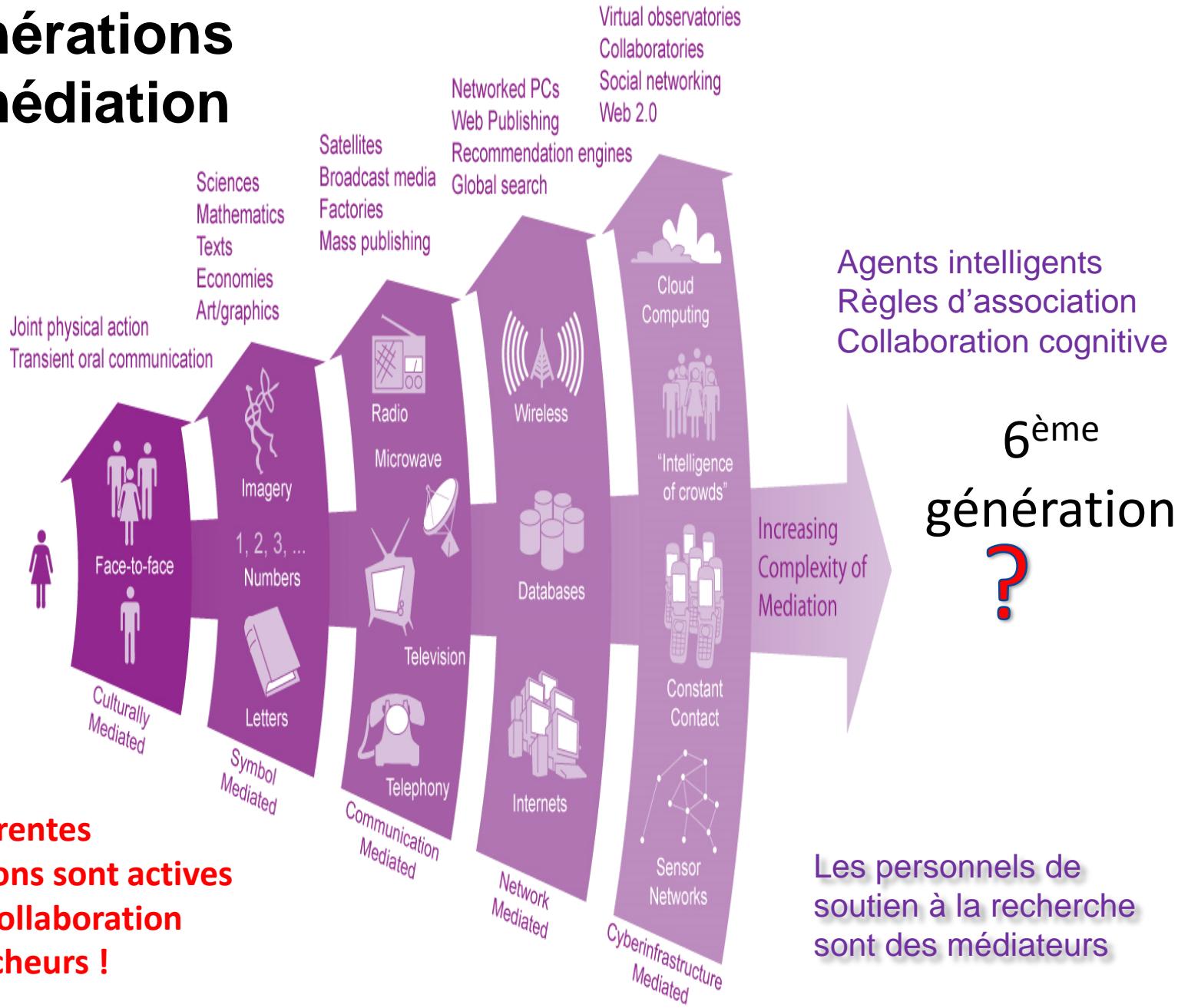
Démarche européenne :
Vers un partage global

Cadre éthique et juridique

Nouveaux services,
nouveaux métiers

Sujets brûlants

5 générations de médiation



Des services à façon

- Les cibles :
 - Recherche
 - Enseignement
 - Pilotage scientifique
 - Évaluation
 - Gestion
-
- Couplage publications données processus (dont logiciels)
- Services d'informations entrantes
- Services de valorisation de données sortantes
- Forte composante technologique :
agrégation,
harmonisation,
traitement,
représentation

1. Reward systems
2. Measuring quality and impact: altmetrics
3. Future of scholarly publishing
4. FAIR open data
5. Open Science Cloud
6. Research integrity
7. Citizen Science
8. Open education and skills

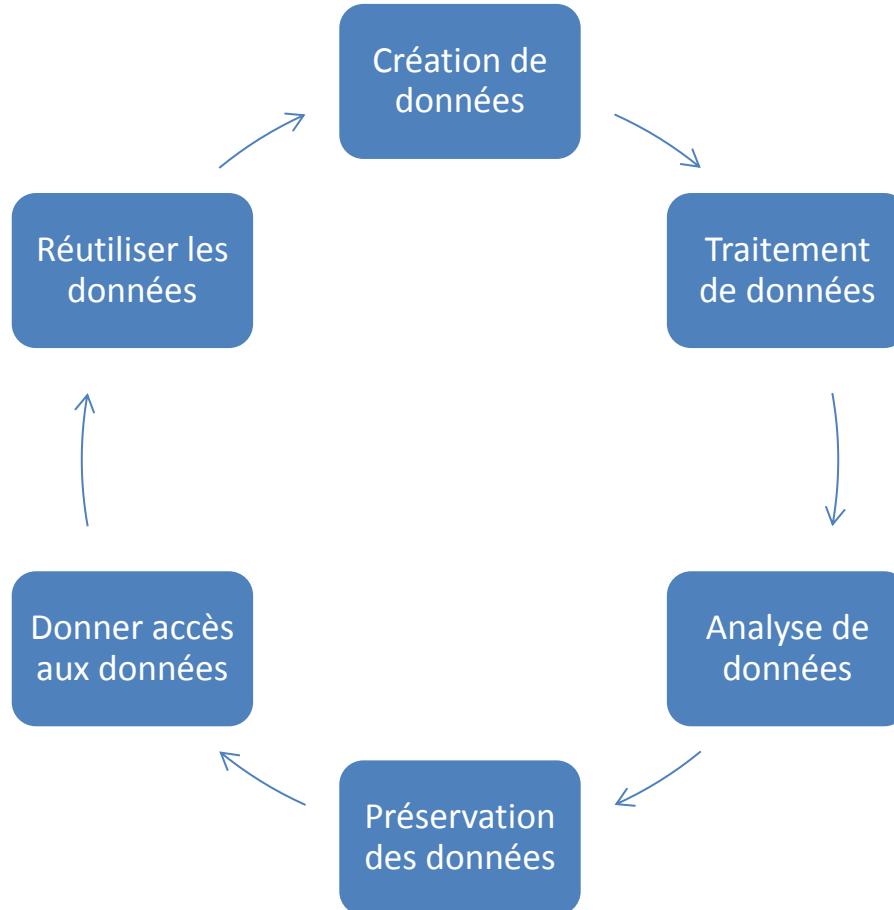


- **Bibliométrie**
- **Valorisation institutionnelle**
- **Métriques alternatives**
- **Nouvelles formes d'édition : épijournaux, data journaux**
- **Modifications des modèles économiques**
- **Curation des données de recherche**
- **DMP**
- **Accès aux publications, archives, bases de données**
- **Recherche d'informations et TDM**
- **Administration de la preuve, réplicabilité**
- **Gestion des processus**
- **Formations à la gestion/valorisation des productions scientifiques**

Les données sont vivantes

Rôle des acteurs

Cycle des données



Cycle des données

**Enseignement,
nouvelle recherche,
évaluation**

*Collecte : expériences, mesures,
observations, simulations*

*Création
de métadonnées*

*Saisir, formater, nettoyer,
organiser, vérifier, valider,
décrire, stocker*

Data journals

Donner accès
aux données

*Curation
de données*

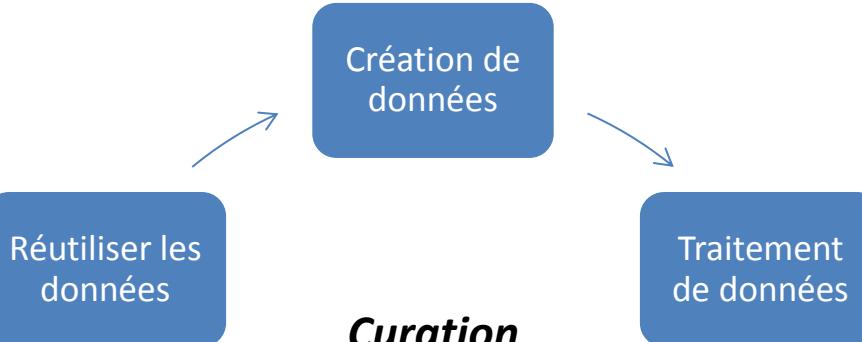
*Distribution, référencement,
Signalement, gestion des droits*

Préservation
des données

Analyse de
données

Métadonnées, documentation, certification

Urfist Rennes 2017



Complémentarité IST, chercheur, informaticien

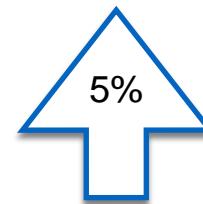
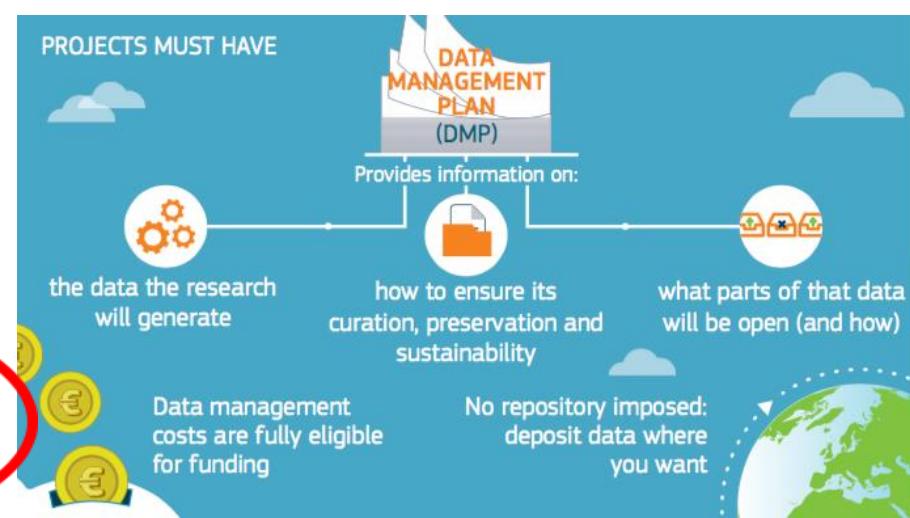
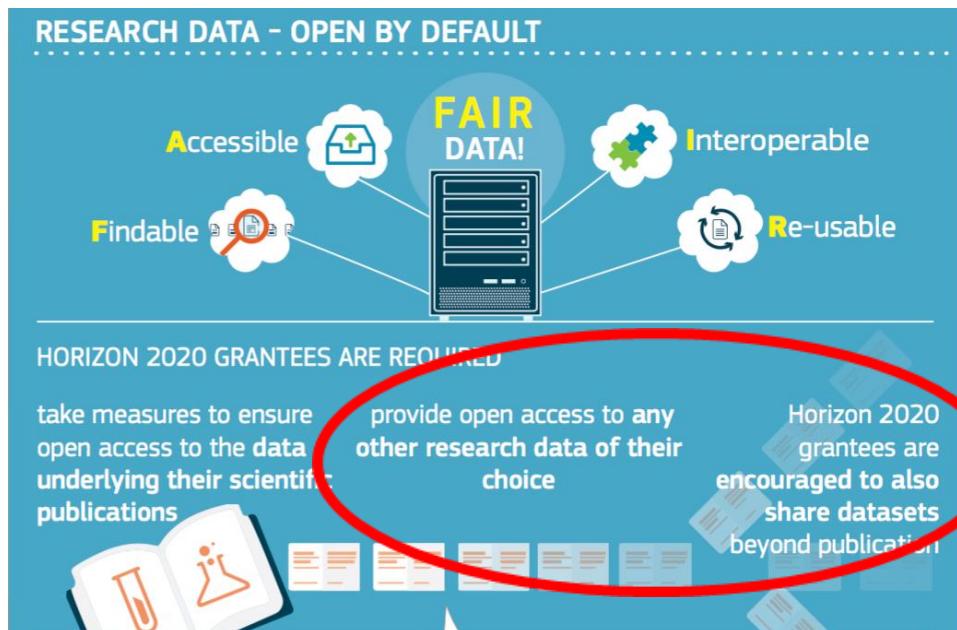
*Interprétation, visualisation,
mise en forme, publication*

*Migration, reformatage,
back-up, stockage pérenne,*

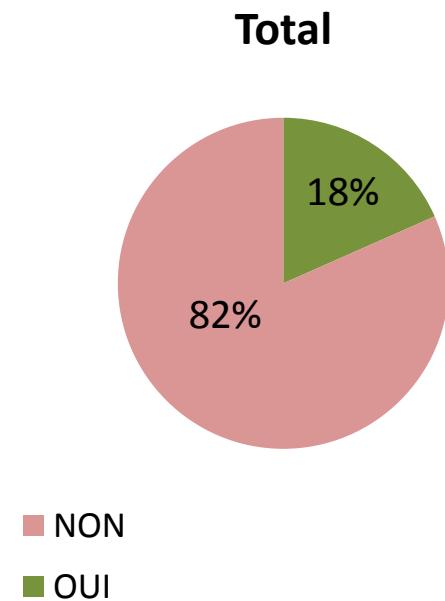
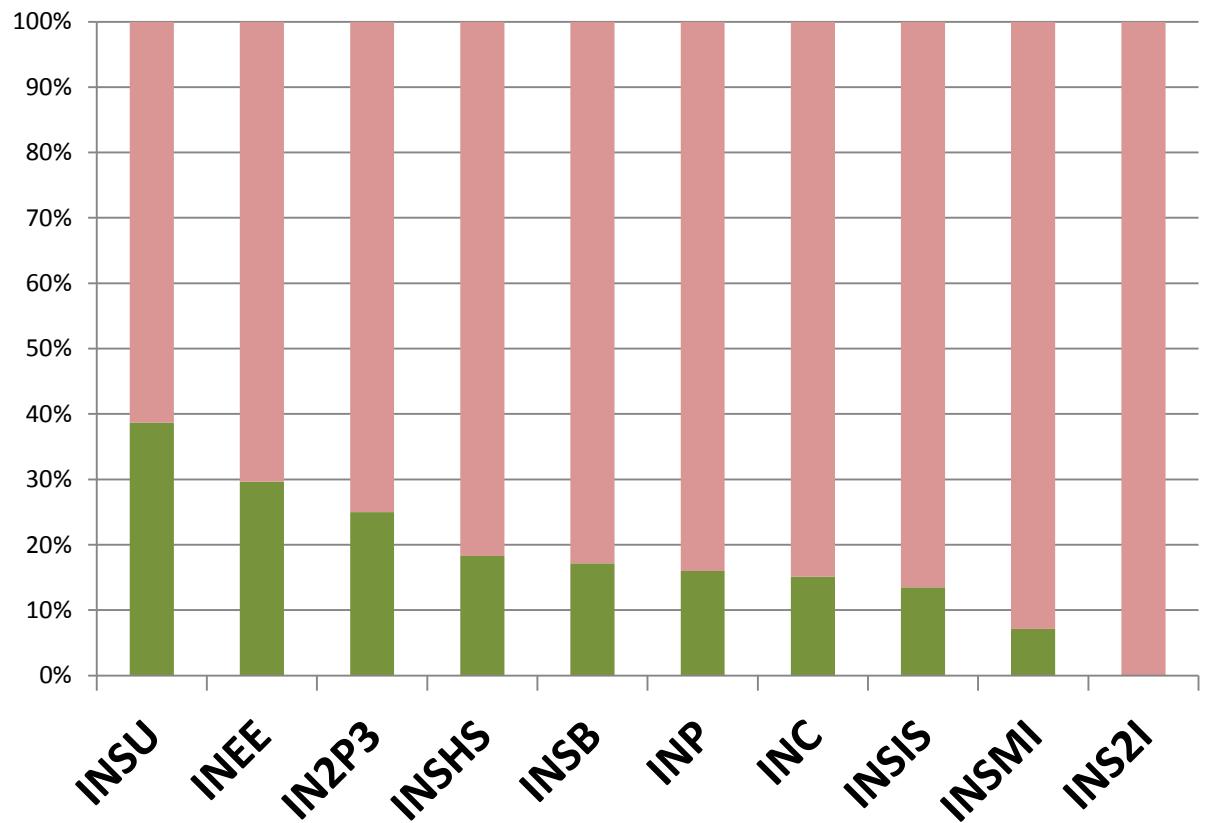


Research data : Open by default

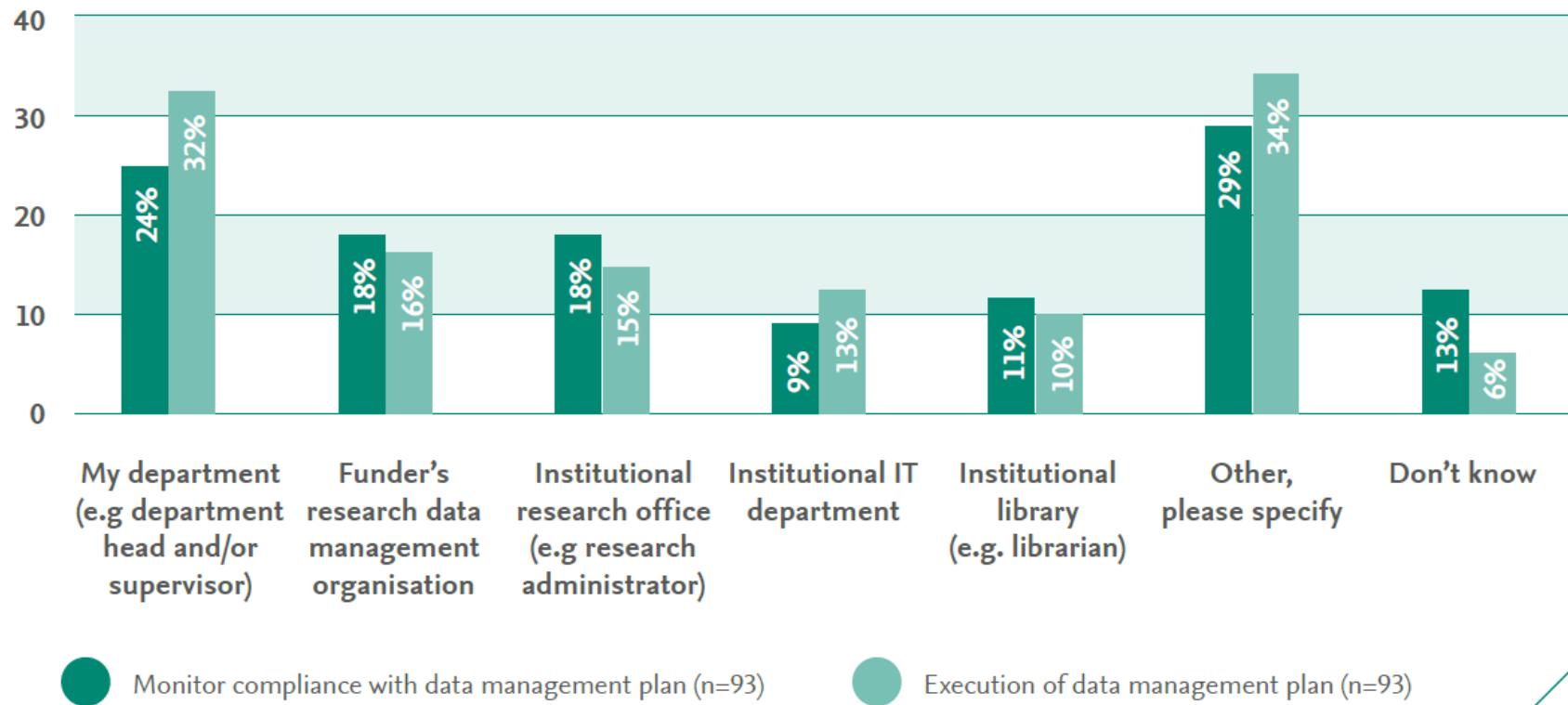
D'après Barend Mons, EOSC



"Avez- vous élaboré un plan de gestion de données ?" (enquête unités CNRS 2014)



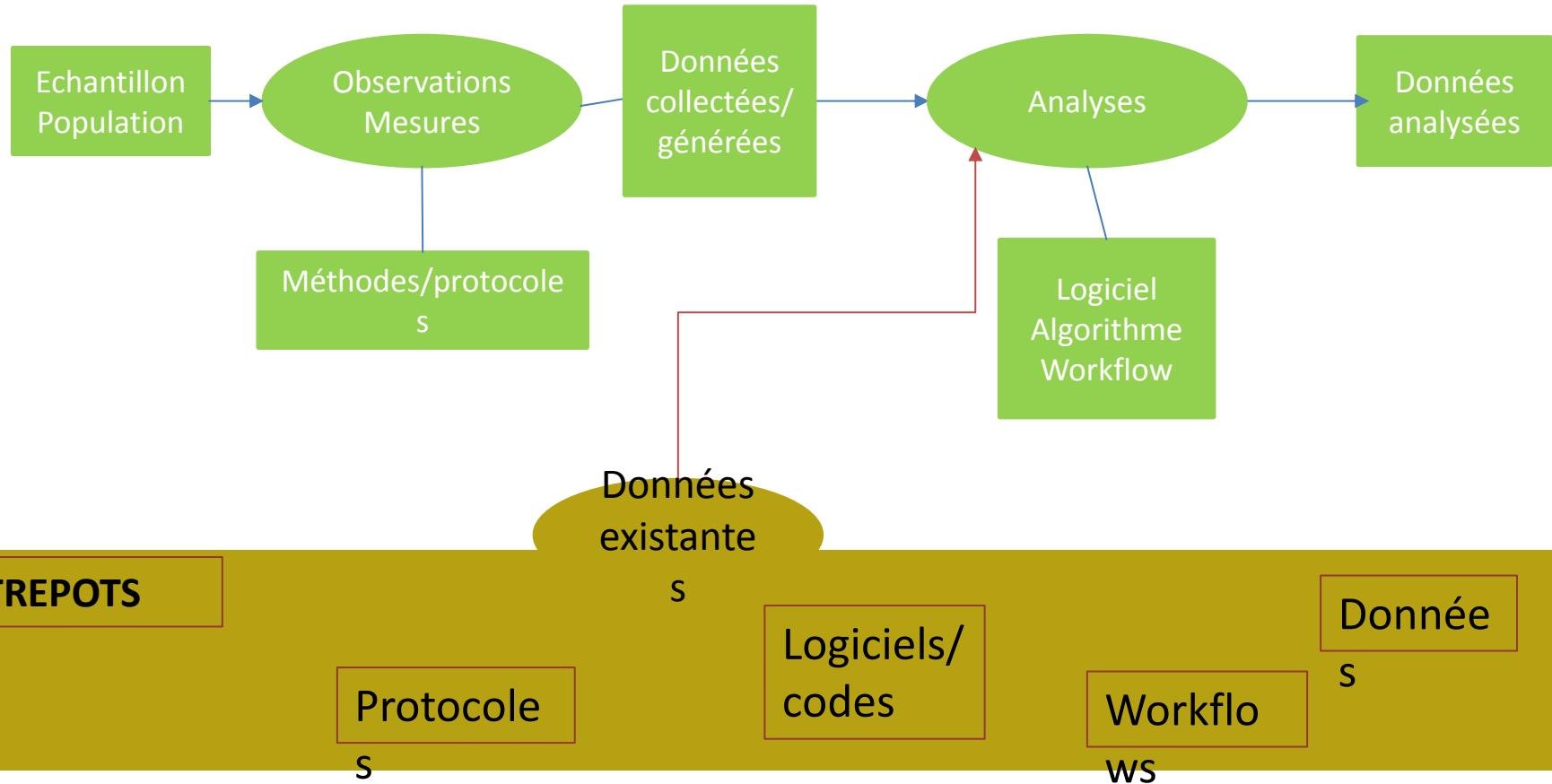
Plan de gestion de données : qui fait quoi ?



<https://www.cwts.nl/download/f-53w2.pdf>

Anticiper la publication des données

Se poser ces questions sous l'angle IST« gestion d'objets numériques »





On recherche...



Data Experts: this term refers to a distinct and largely novel class of research professional. They are not traditional core computer or data scientist, but embedded data specialists that are able to support domain specific researcher throughout the entire knowledge discovery cycle. They typically do not end up with high impact factors in traditional systems but should become indispensable core partners in any modern data driven research team with a solid perspective.

HLEG EOSC report oct. 2016

Où sont les « data librarians » et « data curators » ?!!

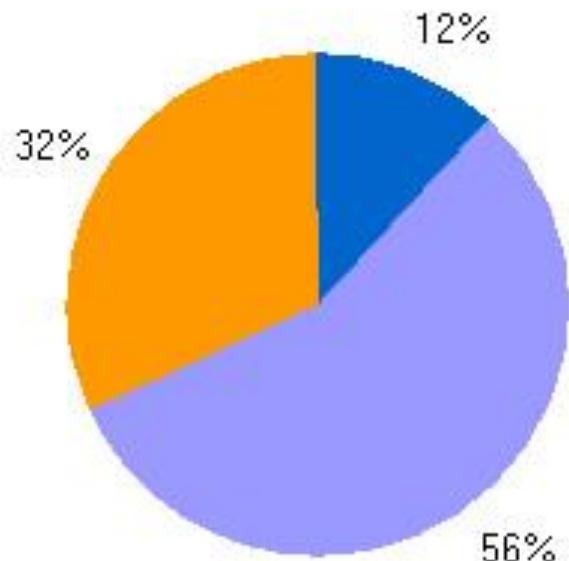
<http://ec.europa.eu/research/openscience/index.cfm?pg=open-science-cloud>

Libraries' Opportunities

Data Issue:	Libraries and data centres opportunities (Chapter 4):
Availability	<ul style="list-style-type: none">✓ Lower barriers to researchers to make their data available.✓ Integrate data sets into retrieval services.
Findability	<ul style="list-style-type: none">✓ Support of persistent identifiers.✓ Engage in developing common metadescription schemas and common citation practices.✓ Promote use of common standards and tools among researchers
Interpretability	<ul style="list-style-type: none">✓ Support crosslinks between publications and datasets.✓ Provide and help researchers understand metadescriptions of datasets.✓ Establish and maintain knowledge base about data and their context.
Re-usability	<ul style="list-style-type: none">✓ Curate and preserve datasets.✓ Archive software needed for re-analysis of data.✓ Be transparent about conditions under which data sets can be re-used (expert knowledge needed, software needed).
Citability	<ul style="list-style-type: none">✓ Engage in establishing uniform data citation standards.✓ Support and promote persistent identifiers.
Curation/Preservation	<ul style="list-style-type: none">✓ Transparency about curation of submitted data.✓ Promote good data management practice.✓ Collaborate with data creators✓ Instruct researchers on discipline specific best practices in data creation (preservation formats, documentation of experiment,...)

Developing our own workforce!

Do you feel that your library has the right skills to be prepared for such activities?



■ Yes

- Not yet, but the library is investing in developing data managers, data curators, or similar skills
- No, and my library hasn't begun to develop such skills

Activities ?

availability, findability, interpretability, re-usability,
catability, curation/preservation





www.cnrs.fr

DMP OPIDoR : un outil pour faciliter la rédaction d'un plan de gestion

Une équipe : Benjamin Faure, Marie-Christine Jacquemot-Perbal, Anne Busin, Françoise Cosserat, Laurent Rassinoux

Inist-CNRS, info-opidor@inist.fr



Un outil à disposition de l'ESR

- ➊ Facilite la rédaction
 - Outil collaboratif : option de partage
 - Modèles de plan de gestion de données + conseils + exemples + réponses suggérées
 - Mises à jour

- ➋ Facilite la mise en œuvre
 - Guide/liens sur les outils à disposition pour la réalisation du workflow de recherche - adaptation à l'environnement
 - De la collecte/acquisition à la publication : application des principes FAIR
 - Structuration du DMP et de la documentation (métadonnées) : simplifie les procédures de dépôt dans un entrepôt de données, élaboration d'un « data paper »

DMP OPIDoR



Logged in as DMP Administrator ▾

[View plans](#) [Create plan](#) [About](#) [Help](#)

Create a new plan

Please select from the following drop-downs so we can determine what questions and guidance should be displayed in your plan.

If you aren't responding to specific requirements from a funder or an institution, [select here to write a generic DMP](#) based on the Horizon 2020 FAIR DMP

If applying for funding, select your research funder.

Otherwise leave blank.

To see institutional questions and/or guidance, select your organisation.

You may leave blank or select a different organisation to your own.

Tick to select any other sources of guidance you wish to see.

DCC

https://dmp.opidor.fr/

À consommer sans modération !

European Commission (Horizon 2020)

Organisation

Organisation
INRA
IRSTEA
Université de Strasbourg
CNRS

[Create plan](#)

Contact us | Terms of use



<https://dmp.opidor.fr/>

Powered by: FDMP
ONLINE



My plan (Horizon 2020 FAIR DMP)

[Plan details](#)
[Initial DMP](#)
[Share](#)
[Export](#)
1. Data summary (1 question, 0 answered)
 +
2. FAIR data (4 questions, 0 answered)
 -

In general terms, your research data should be 'FAIR' that is findable, accessible, interoperable and re-usable. These principles precede implementation choices and do not necessarily suggest any specific technology, standard or implementation-solution.

2.1 Making data findable, including provisions for metadata:

- Outline the discoverability of data (metadata provision)
- Outline the identifiability of data and refer to standard identification mechanism. Do you make use of persistent and unique identifiers such as Digital Object Identifiers?
- Outline naming conventions used
- Outline the approach towards search keyword
- Outline the approach for clear versioning
- Specify standards for metadata creation (if any). If there are no standards in your discipline describe what metadata will be created and how

B **I**

[Save](#)

[Guidance](#) [Share note](#)

CE Guidance

The Research Data Alliance provides a [Metadata Standards Directory](#) that can be searched for discipline-specific standards and associated tools.

Digital Curation Centre guidance on Documentation
+
Digital Curation Centre guidance on Metadata
+

Travailler à ...

- Définir des stratégies institutionnelles
- Mobiliser toutes les compétences : communauté scientifique, professionnels IST, informaticiens, archivistes, juristes,...
- Adapter les cadres éthique, juridique, économique.
- Adapter le dispositif d'évaluation des scientifiques et le cadre RH des fonctions d'appui
- Allouer les moyens de mise en œuvre de la politique en mobilisant des ressources pour :
dresser un état des lieux, harmoniser les pratiques , clarifier les spécificités de chaque communauté, **produire un guide de bonnes pratiques**, rendre obligatoires les **plans de gestion de données (DMP)**, définir une stratégie de lobbying,...
- Développer des services de RDM pour les chercheurs

Actions clés

- Cartographier des acteurs dans les différentes composantes ESR (Cat OPIDoR)
- Mener des auditions : scientifiques pour clarifier les spécificités disciplinaires, groupe juridique Inra,....
- Participation au GT données du pilier INFRANUM



Cat OPIDoR

Catalogue pour une Optimisation
du Partage et de l'Interopérabilité
des Données de Recherche

Anne CIOLEK-FIGIEL

Membre de l'équipe Valorisation des Données de Recherche (Inist-CNRS)

Violaine REBOUILLAT

Chargée de mission BSN10



BIBLIOTHÈQUE SCIENTIFIQUE
NUMÉRIQUE
DIGITAL SCIENTIFIC
LIBRARY



Contact : infocatopidor@inist.fr

UNIVERSITÉ DE STRASBOURG





Objectifs

Quoi : un catalogue pointant vers les services dédiés à la gestion des données de recherche en France

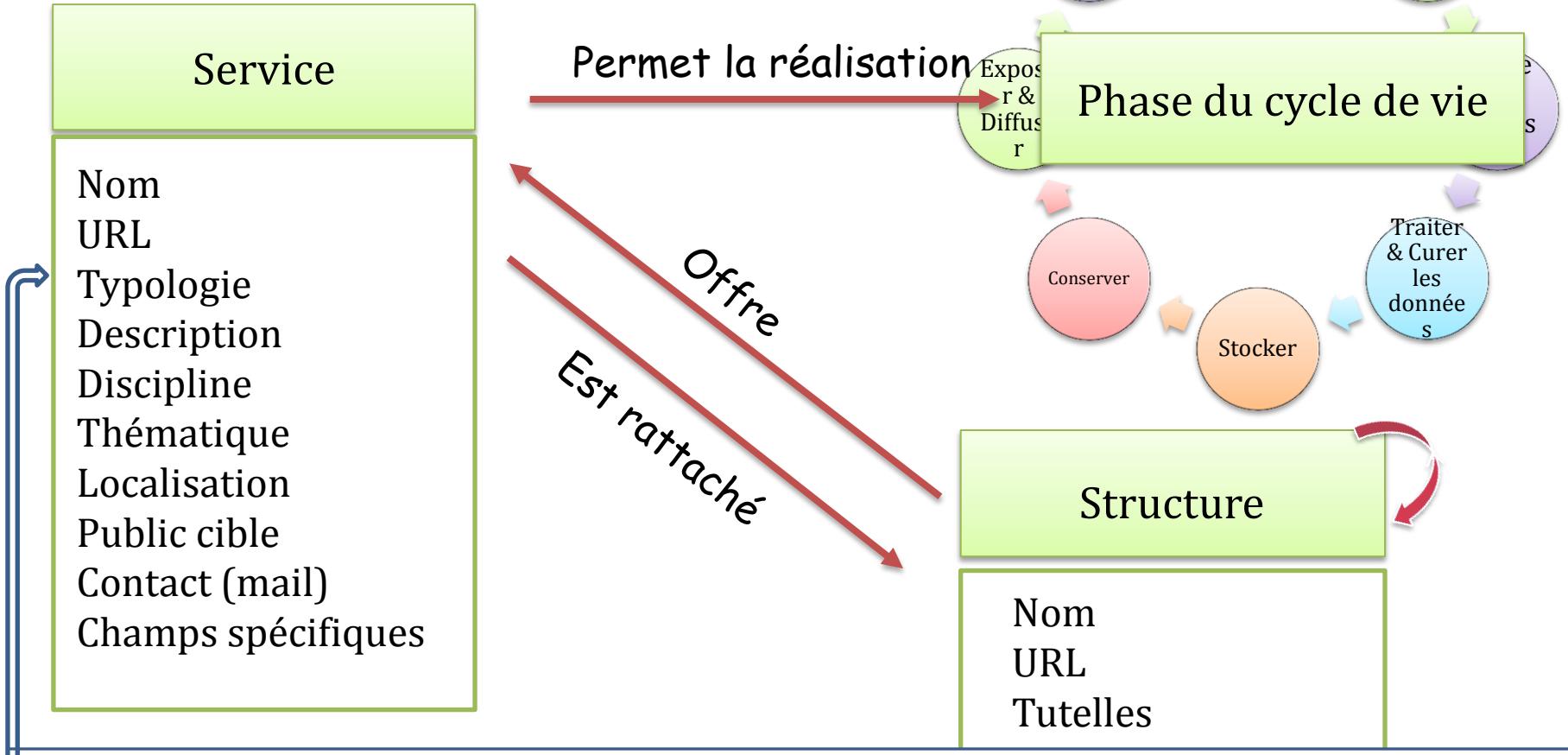
Pour qui ?

- Producteurs et utilisateurs de données
- Services d'appui à la recherche (informaticiens, professionnels de l'IST, juristes ou archivistes)
- Décideurs et financeurs de la recherche

Comment ?

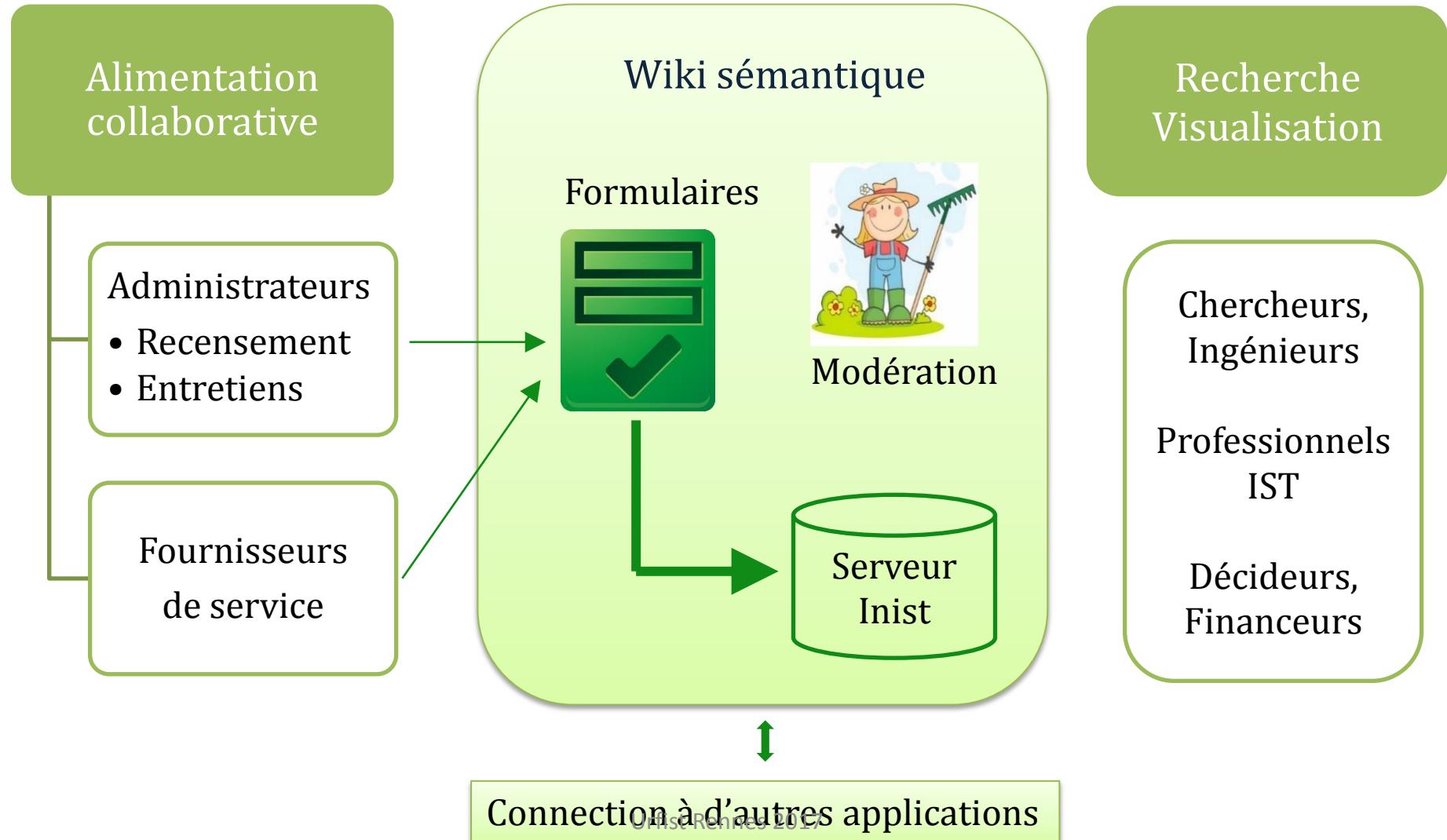
- Un wiki sémantique en libre accès
- Hébergé et modéré par l'Inist-CNRS
- Compatibilité avec les répertoires de rang international ([Re3data](#), [MERIL](#),...)

Modèle de données



Information, Formation, Accompagnement, Outils de gestion des données, Plateforme d'acquisition, Plateforme de calcul, Entrepôt de données, Annuaire de données, Plateforme d'archivage

Représentation synthétique de Cat OPIDoR



9

Partenariats

Apports



à intégrer



Modalités Pédagogiques

Développement
Ressources Pédagogiques

Infrastructure informatique

Ingénierie de développement
de modules de formation

Co-construction



Urfist

Publications de données

Entrepôt de données

Identification
pérenne, DOI

Archivage des données

Plan de gestion de
donnéesAccès, consultation et visualisation des
données

Aide au dépôt

Aspects juridiques et
éthiquesEnjeux et bénéfices de la gestion et
du partage

Construction des thématiques

Le dépôt des données

Catalogue thématique

En bref



La fiche synthétique

Fiche 2 min PDF



La minute entrepôt

Vidéo 1 min PDF

Le dépôt des données de recherche répond à une logique de partage et nécessite une gestion rigoureuse des données pour que celles-ci soient accessibles et réutilisables dans le temps. Vous voulez faire le point sur vos connaissances dans ce domaine ? Cliquez sur le bouton !

TESTEZ VOS CONNAISSANCES !

L'essentiel



Le dépôt des données en 5 questions

Infographie 10 min PDF



Du choix de l'entre�ot au dépôt des données

Infographie 10 min PDF



Les critères de choix d'entre�ot

Fiche 1 min PDF



Vérifier ses données

Checklist 10 min PDF

Pour aller plus loin

Cours

Suivez ce cours regroupant toutes les ressources sur la thématique du dépôt des données de recherche.

ACCÉDER AU COURS

Classe virtuelle

Inscrivez-vous à nos prochaines classes virtuelles, animées par nos partenaires formateurs.

ACCÉDER À L'AGENDA

Autres ressources



Plan de Gestion de Données : DMP

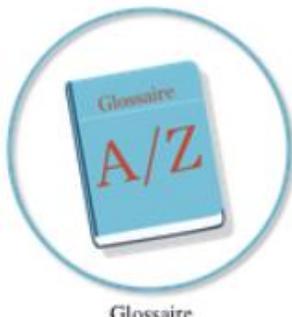
En bref



L'essentiel



Autres Ressources



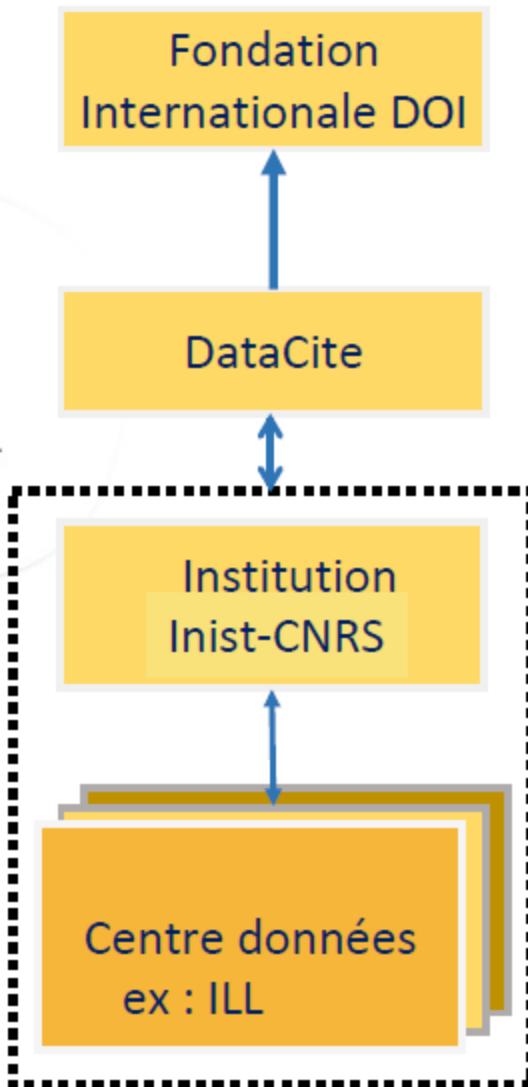
10.xxxxx/abcd_x-x_efg

Préfixe

Suffixe

Inist-CNRS

Agence française d'attribution de DOI



Creator (PublicationYear): Title. Publisher. Identifier

Kalaydjian Regis, Girard Sophie (2014). Civil marine research effort in main French research organisations: data and estimates. SEANOE
<http://doi.org/10.17882/44428>

- Lien fort avec le consortium (participation aux groupes de travail, assemblée générale, stratégie...)
- Sensibilisation aux services DataCite
- Fourniture des préfixes de DOI
- Assistance à la création et à la conversion de fichiers de métadonnées
- Aide au paramétrage de l'API

Des services d'accès, de publications et d'analyses des productions scientifiques

http://www.cnrs.fr/dist/z-outils/documents/DIST_annuaire-services-IST-CNRS.pdf

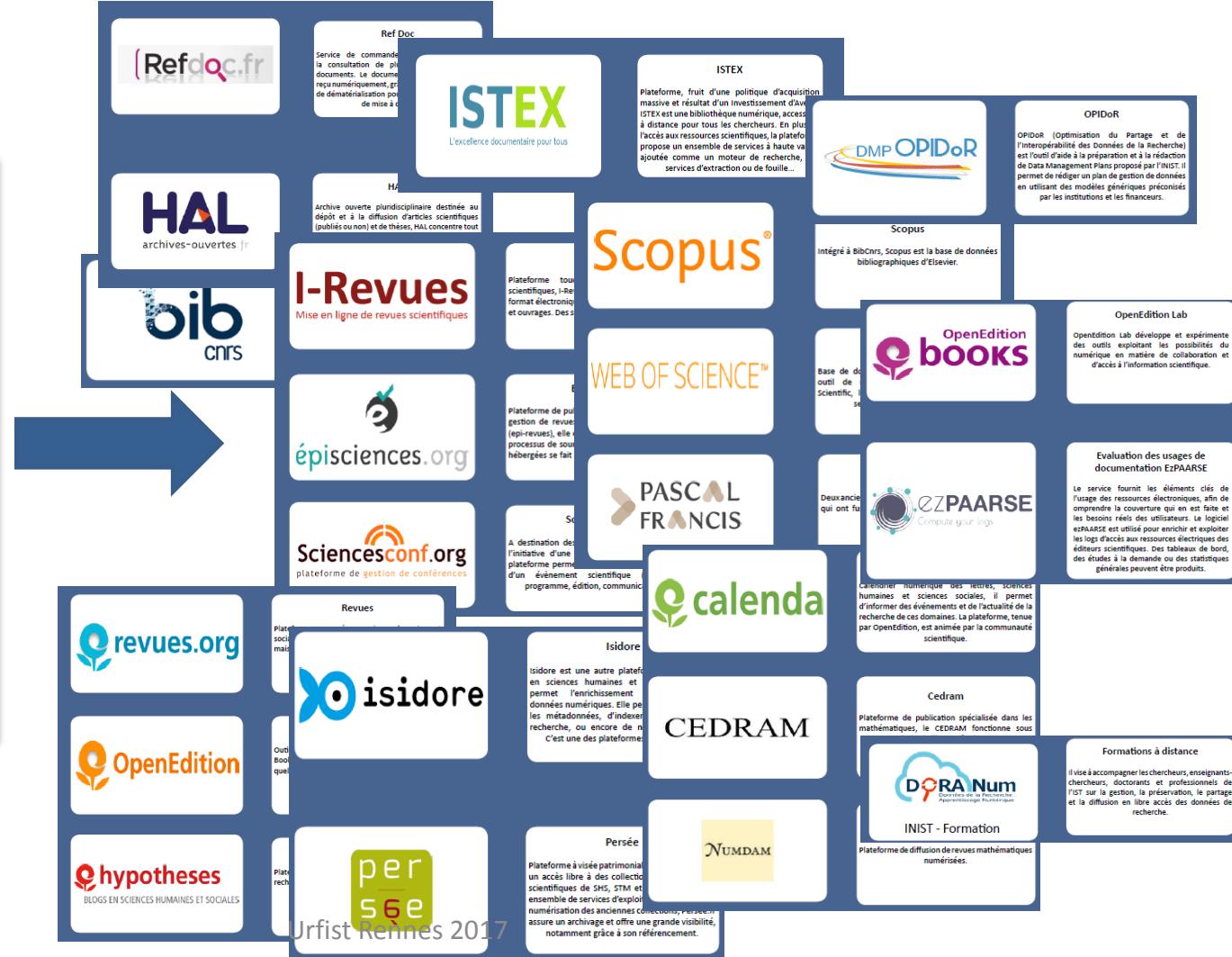
Offre de services aux chercheurs

Pourquoi ce document ?
 Les services et outils d'information Scientifique et Technique se développent à l'heure où l'offre nationale évolue et se partage avec le Catalogue d'Offre Partagé (COPEST), l'offre de service du CNRS est présentée ci-après. Cet espace permettra aux chercheurs d'avoir une vue globale des outils numériques proposés par la Direction de l'Information Scientifique et Technique (DIST).

Vos commentaires et propositions sont les bienvenus sur les espaces dédiés. Cette offre est évolutive : elle met en action la stratégie IST du CNRS «Mieux partager les connaissances», en s'adaptant à vos besoins.

Mars 2017

Retrouvez tous ces services en accès direct sur le site de la Direction de l'Information Scientifique et Technique (DIST) du CNRS.

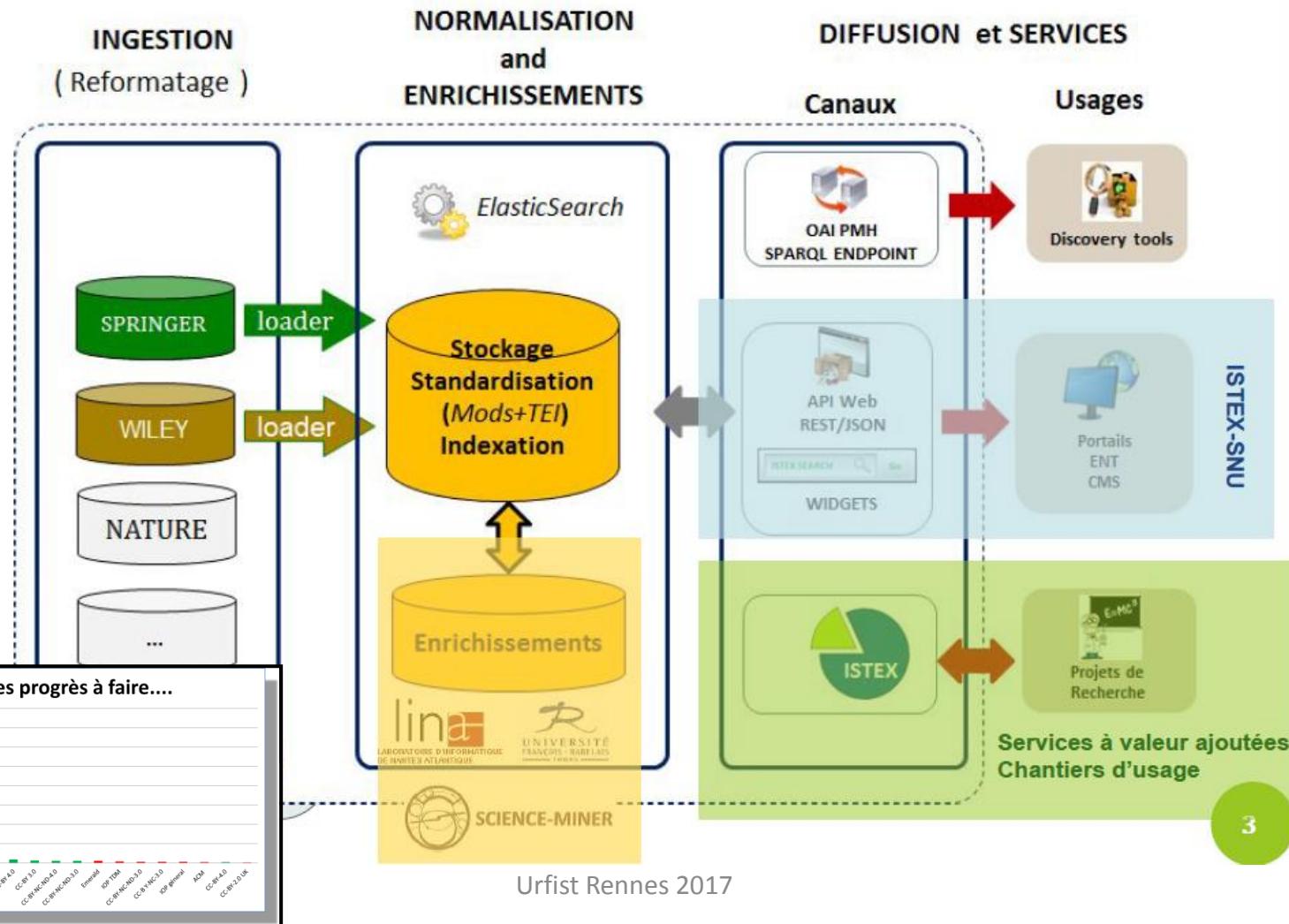


Infrastructures d'accès aux données ISTEX



L'ISTEX ET VOUS

ISTEX
L'excellence documentaire pour tous



Explorer des contenus informationnels

TDM

- Sources hétérogènes, structurées ou non :
 - Formatage
 - Agrégation de données hétérogènes
 - « nettoyage », enrichissement sémantique
 - Repérage d'entités nommées
 - vocabulaires controlés, ontologies
- Fouille de textes
 - Constitution d'un corpus validé à partir de sources hétérogènes
 - TDM sur texte intégral : Xmlisation du full text
 - Choix d'outils d'exploration, des méthodes de traitement
 - Graphes, représentations, navigation sémantique, cartographie



Quelques projets en cours ou à venir

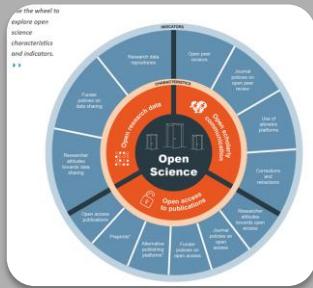
TDM

- Les services à valeur ajoutée ISTEK
- Les chantiers d'usage ISTEK
- Projet EasISTEK : Inscription de Gargantext et des logiciels ISC dans l'API ISTEK, pour un usage facilité de fouilles des métadonnées et du TDM
- VISA TM : déploiement national Inra/CNRS de la plateforme de TDM issue du projet européen Openminted (<http://openminted.eu/>)



Lever les barrières au TDM

- Cadre juridique national/international se modifie
 - Lois Valter, Lemaire, directive européenne
- Développement nécessaire des métiers dataX (**X= scientist, analyst, librarian, curator, ...**)
- Manque d'infrastructures et d'outils d'analyses TDM
- Le ticket d'entrée au TDM est élevé (**formation, infrastructures, outils logiciels, ...**)
- Faible prise de conscience par les chercheurs de l'apport de la fouille de textes et de données



Ouverture ?
On saura vraiment ce que c'est quand on l'aura faite...

Démarche européenne :
Vers un partage global

Cadre éthique et juridique

Nouveaux services,
nouveaux métiers

Sujets brûlants



Les pièces du puzzle...

- (*La réflexion sur*) l'organisation se met en place à différentes échelles :
 - internationale, européenne, nationale, institutionnelle,...
- Dans ses différentes composantes :
 - technique, organisation des infrastructures, politique, économique, humaine,...
- Pas d'angélisme : il manque encore beaucoup de pièces au puzzle

Organiser...

- S'approprier la gestion des données
 - « maison de la donnée », « data insitute », « Open data Center »,...
 - Former, former, former !
- Respecter les cultures et approches disciplinaires (dont certaines ont fait leurs preuves depuis plus de 40 ans !) –Astro
- Traiter de façon concertée les développements de la « science des données » et de la « gestion des objets numériques »
- Stimuler le dialogue chercheurs/IST/informaticiens





Merci de votre attention

francis.andre@cnrs-dir.fr