Participer à l'organisation du management des données de la recherche, gestion de contenu et documentation des données
Session 2017

3-6 juil. 2017 Vandoeuvre-lès-Nancy (France)

www.**cnrs**.fr

Gestion des données de la recherche dans le contexte d'Open Science

Francis ANDRE– DIST/CNRS

Vandoeuvre-les-Nancy, 3 juillet 2017

Evolution des pratiques de science

Science du 21ème siècle : plus…

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

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 sociétaux

Open Science : une priorité partagée

- Conseil Européen

- G7 Science and Technology Ministers' Meeting Tsukuba, 2016 May

- Commission Européenne, DG RTD

- NIH, « Big data to knowledge » (BD2K)

- G20, sommet de Hangzu, 2016

- Déclaration de Dakar, 2016

- A surveiller...OSTP, 2013-2017

https://obamawhitehouse.archives.gov/admini
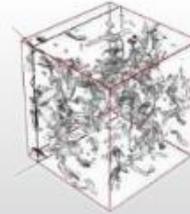stration/eop/ostp/library/archives

https://www.whitehouse.gov/ostp

La "science des données", 4ᵉ 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	**Last few hundred years**	**Last few decades**	**Today and the Future**
Description of natural phenomena	Newton's laws, Maxwell's equations…	Simulation of complex phenomena	Unify theory, experiment and simulation with large multidisciplinary Data
			Using data exploration and data mining (from instruments, sensors, humans…)
			Distributed Communities

Crédits: Dennis Gannon

Inria

D'après G. Antoniu

8

Big Data : le déluge des données

Expériences

Simulations
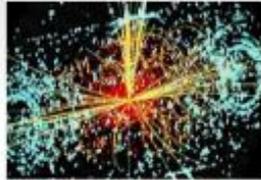
Archives

Bibliothèques
Cornell University Library
arXiv.org

Web 2.0

Défi
Accélérer les découvertes

Des péta-octets qui s'accumulent...

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

Démarches de recherche

(Roue de la science – d'ap.Wallace, cité par V.Schoeb, 2012)
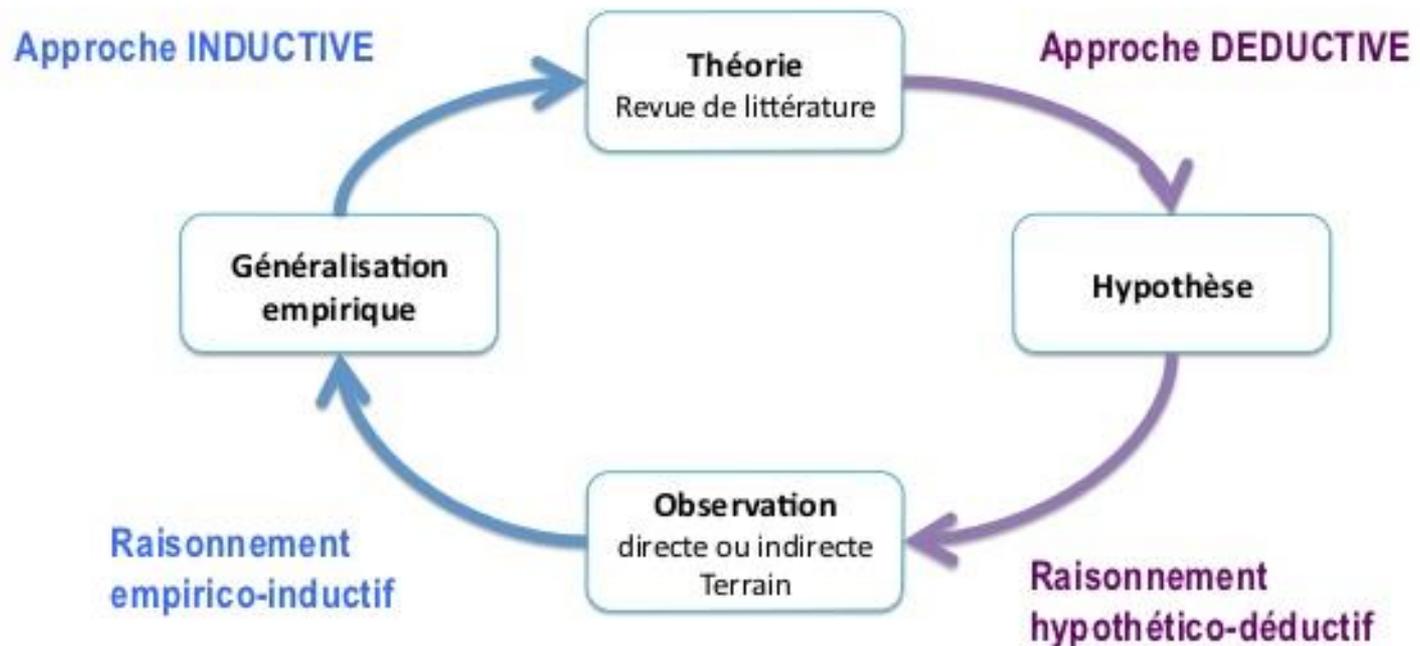
Context: a new paradigm

The scientific discovery process
= the inference cycle

rationalisme

Du général
au particulier

pragmatisme

Formulation de la
meilleure explication

HPC

Formulation of Hypothesis
as Mathematical Model

Theories

DEDUCTION
(Drawing necessary conclusions)

Execution of Model to
Generate Predictions

Generation of Explanatory
Hypothesis

HDA

ABDUCTION
(Making guesses)

Predictions

Discovery of Patterns
and Anomolies

INDUCTION
(Inferring generalizations from sampling)

empiris

HDA

Assimilation of Predictions &
Data to Evaluate Hypothesis

Du particu
au généra

Observations
From Instruments, Sensors, Records, Visualizations etc.

Science des données : composantes
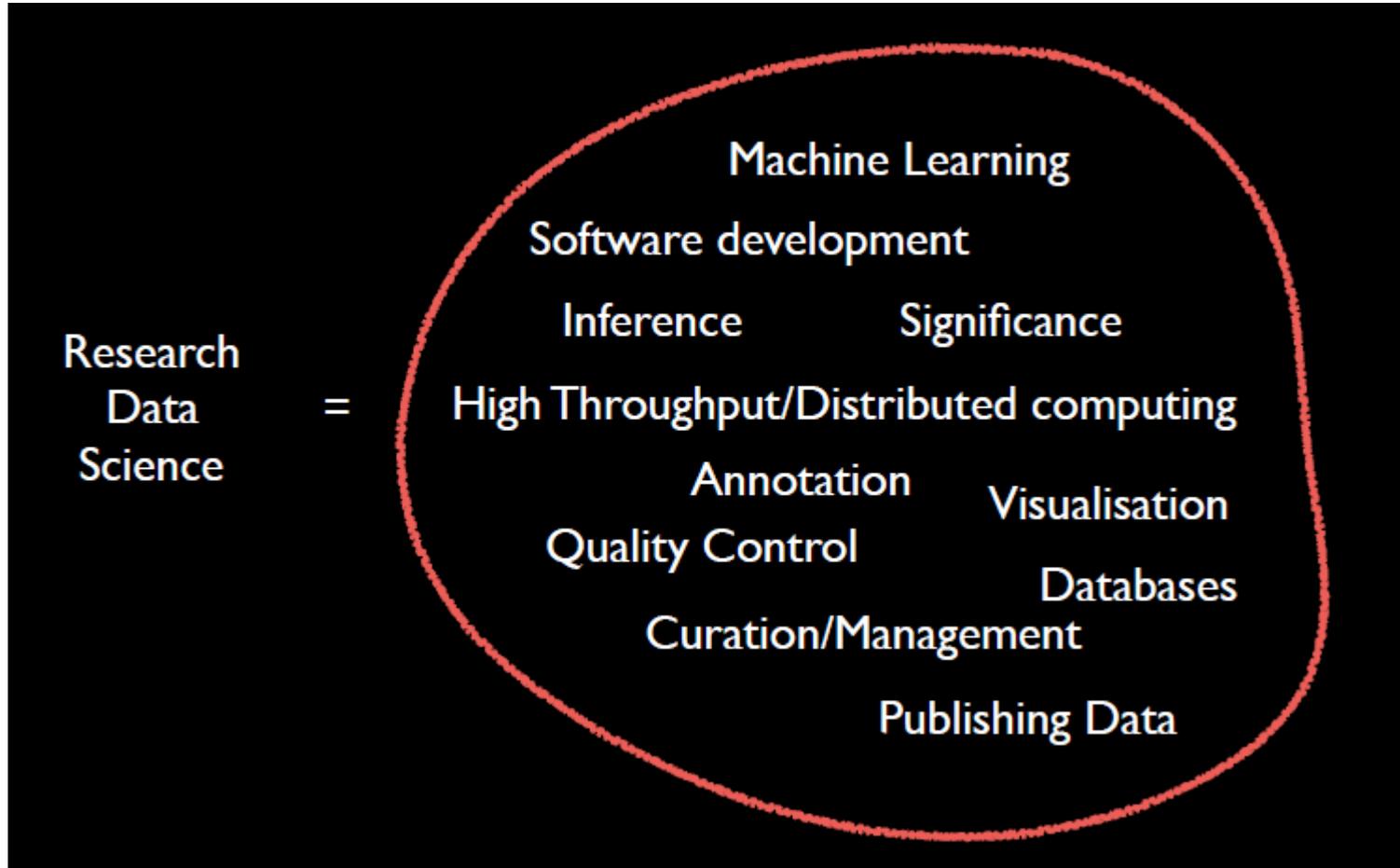
D'après Hugh Shanahan, 2017 Reflections on CODATA-RDA summer school
Royal Holloway, University of London
ANF données juillet 2017

Mais développer les connaissances à partir de l'exploration des données n'est possible que si ces données sont **partagées**

FAIR Principles

https://www.nature.com/articles/sdata201618

IL

FAUT

LE

Findable

Accessible

Interoperable

Re-usable

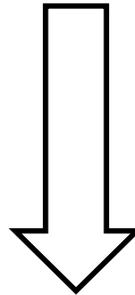# F A I R !

Soyez **FAIR** !

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

Identification
DOI : digital Object
Identifier, service
DataCite

- Unicité
- Désambiguisation
- Pérennité

PID

Metadata (intrinsic)

'provenance' (user defined)

Data (elements)
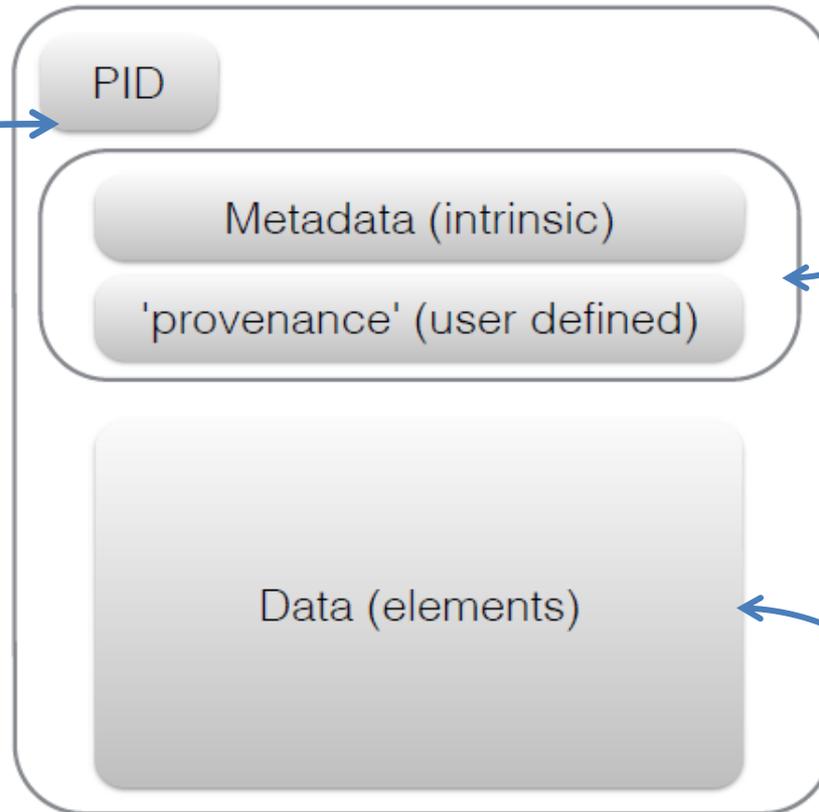
Métadonnées
descriptives/analytiques :

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

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

A simplified diagram of a Digital (data) Object irrespective of technological choices and naming

Data as increasingly FAIR Digital Objects

Totally UNFAIR

Metadata (intrinsic)
'provenance' (user defined)

Data (elements)

Findable
Usable for Humans

PID
Metadata (intrinsic)
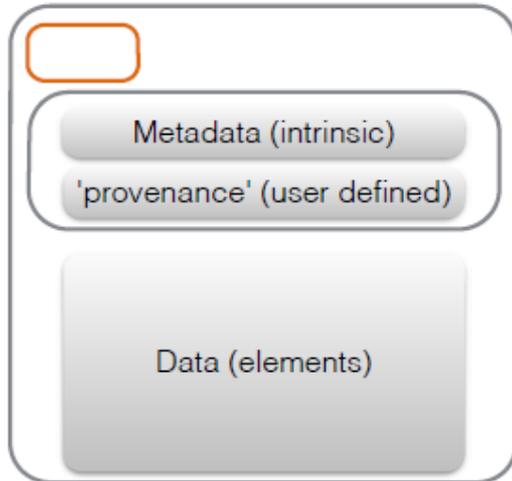'provenance' (user defined)

Data (elements)

FAIR metadata

PID
Metadata (intrinsic)
'provenance' (user defined)

Data (elements)
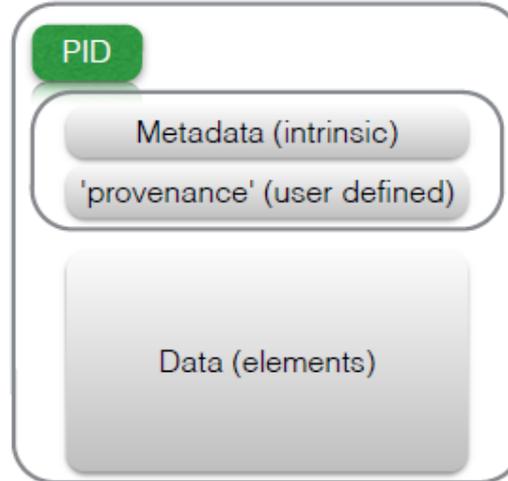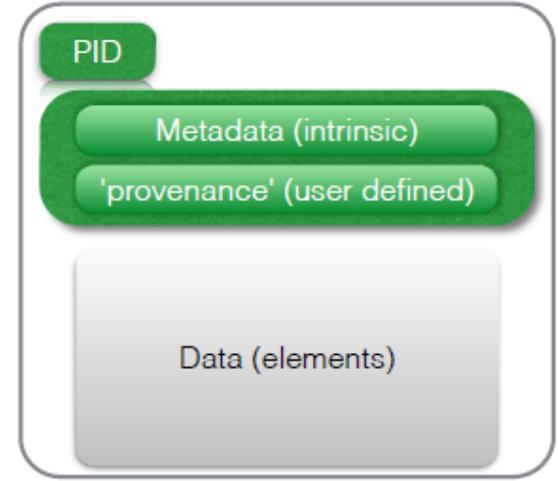
FAIR data-
restricted access

PID
Metadata (intrinsic)
'provenance' (user defined)

Data (elements)

FAIR data-
Open Access

PID
Metadata (intrinsic)
'provenance' (user defined)

Data (elements)

ANF données juillet 2017
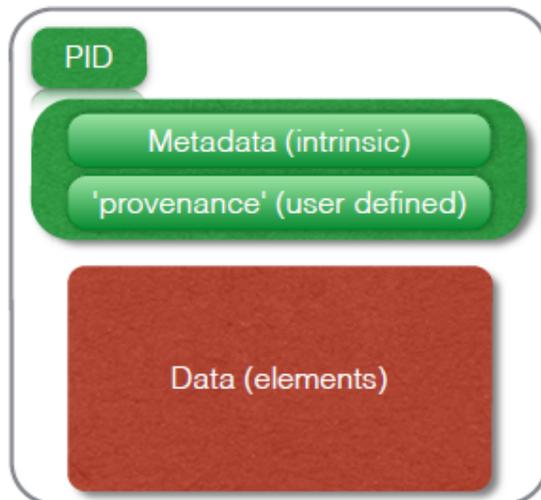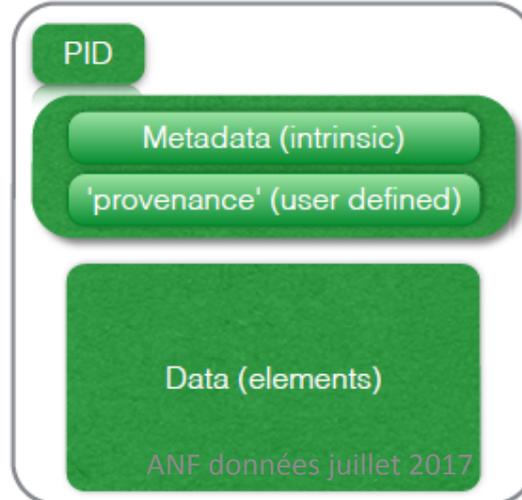
FAIR data-
Open Access/Functionally Linked

PID
Metadata (intrinsic)
'provenance' (user defined)

Data (elements)
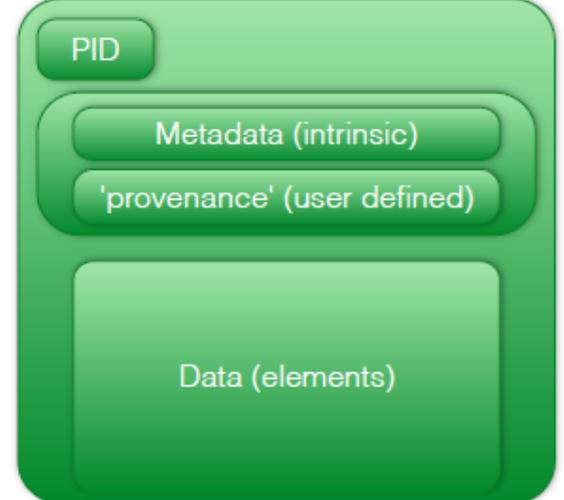
Où se niche la qualité des métadonnées ?

- 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

Which attributes do you think are most important to open data?

ANF données juillet 2017 Belmont Forum survey 2016

Conseils de lecture…

Validation of the results of the public consultation on Science 2.0: Science in Transition

Science as an open enterprise

June 2012

THE ROYAL SOCIETY

MAKING OPEN SCIENCE A REALITY

OECD
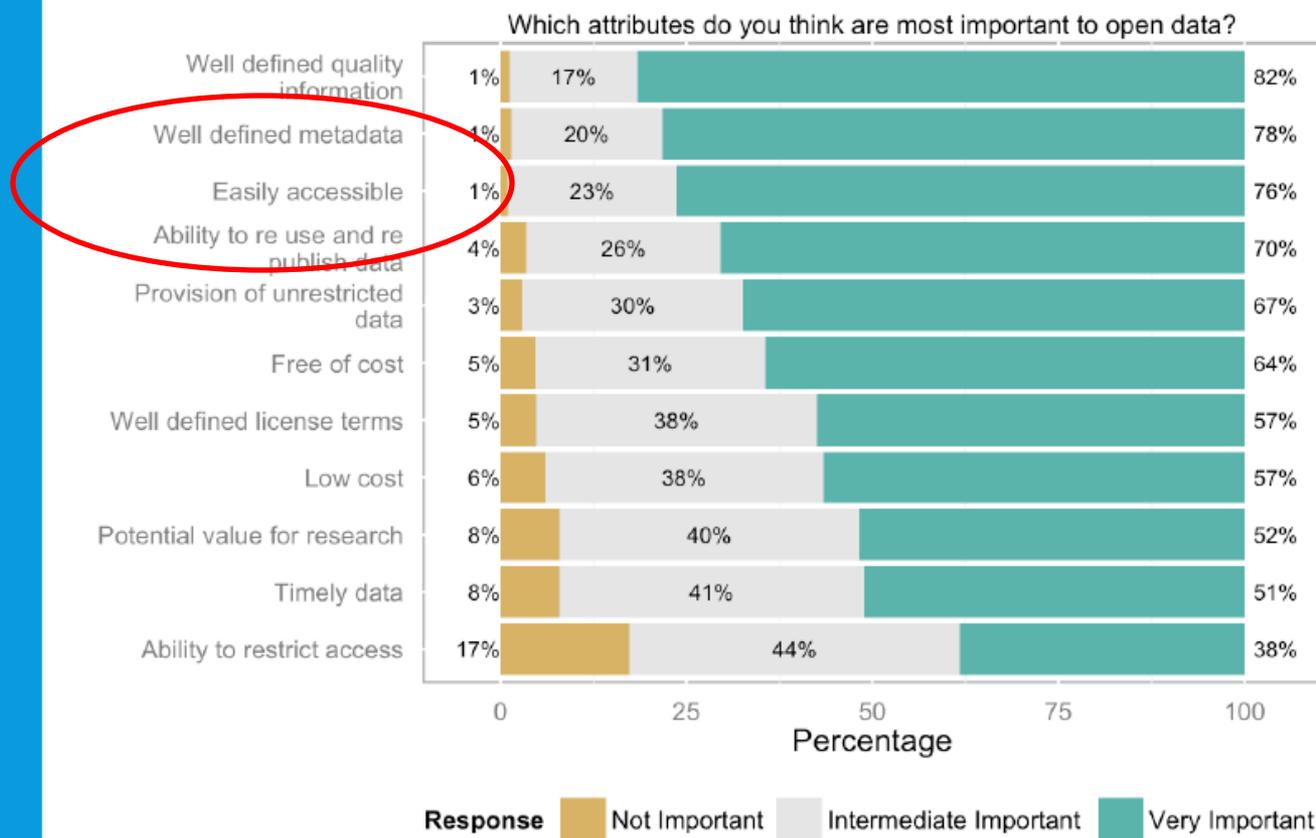
Open Science à la mode environnementale

COMMUNITY STRATEGY AND IMPLEMENTATION PLAN
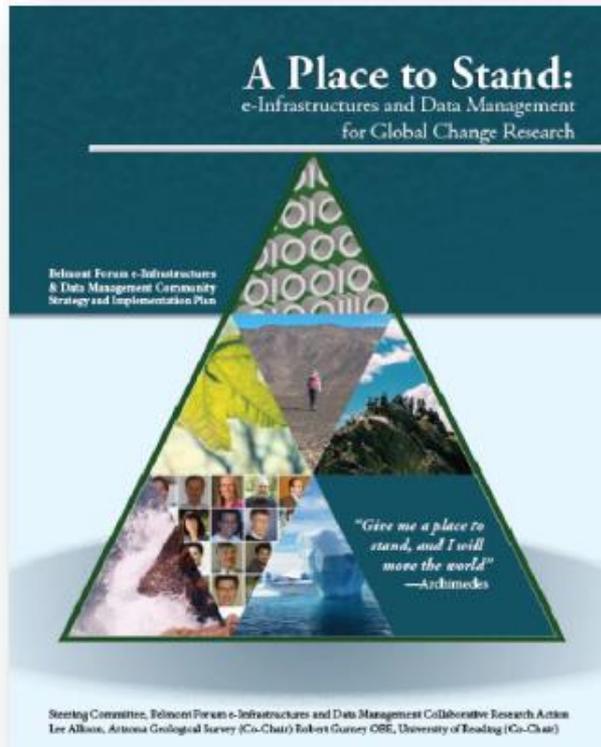
Adoption d'une politique de gestion des données en octobre 2015

réflexion internationale sur les infrastructures numériques et la gestion des données nécessaires à la recherche environnementale

ANF données juillet 2017

Arts & Humanities Research Council

SUOMEN AKATEMIA

ANR

Javna agencija za raziskovalno dejavnost Republike Slovenije

BBSRC bioscience for the future

Bundesministerium für Bildung und Forschung

bmwfw Federal Ministry of Science, Research and Economy

BOEM Bureau of Ocean Energy Management

Dipartimento Scienze del Sistema Terra e Tecnologie per l'Ambiente

cnrs

CSIRO

DFG

EPSRC Engineering and Physical Sciences Research Council

E·S·R·C ECONOMIC & SOCIAL RESEARCH COUNCIL

FAPESP FUNDAÇÃO DE AMPARO À PESQUISA DO ESTADO DE SÃO PAULO

FCT Fundação para a Ciência e a Tecnologia MINISTÉRIO DA CIÊNCIA, TECNOLOGIA E ENSINO SUPERIOR.

FFG FORSCHUNG WIRKT

Luxembourg National Research Fund

Forskningsrådet Formas

fnrs

fwo

Innovationsfonden

Innovate UK

innoviris.brussels empowering research

IRISH RESEARCH COUNCIL An Chomhairle um Thaighde in Éirinn

ISSC international social science council

JAPAN SOCIETY FOR THE PROMOTION OF SCIENCE 日本学術振興会

JST

MEXT

Ministerio de Ciencia, Tecnología e Innovación Productiva Presidencia de la Nación

Ministero dell'Istruzione dell'Università e Ricerca

Ministry of Agriculture & Rural Development

Ministry of Earth Sciences

科技部 Ministry of Science and Technology

NARODOWE CENTRUM NAUKI

NERC SCIENCE OF THE ENVIRONMENT

NordForsk

NRF National Research Foundation

NSERC CRSNG

NSF

NSFC National Natural Science Foundation of China

NWO

RANNÍS

Lietuvos mokslo taryba

The Research Council of Norway

РОССИЙСКИЙ ФОНД ФУНДАМЕНТАЛЬНЫХ ИССЛЕДОВАНИЙ РФФИ

Ίδρυμα Προώθησης Έρευνας

FNS NF SWISS NATIONAL SCIENCE FOUNDATION

Energimyndigheten

Teagasc AGRICULTURE AND FOOD DEVELOPMENT AUTHORITY

Tekes

uefiscdi

Valsts izglītības attīstības aģentūra

VINNO

Social Sciences and Humanities Research

Conseil de recherches en sciences humaines

TÜBİTAK

WAGENINGEN

BELMONT FORUM DATA POLICY AND PRINCIPLES

Data should be:

Discoverable through catalogues and search engines

Accessible as open data by default, and made available with minimum time delay

Understandable in a way that allows researchers—including those outside the discipline of origin—to use them

Manageable and protected from loss for future use in sustainable, trustworthy repositories

The Belmont Forum and its members will support and promote this data policy and principles with the intent of making these data principles enforceable over time.

DATA PLANNING

- Incorporate Data Management language into call text (October 2016)
- Establish process for review and monitoring data management plans (GEO/CODATA/RDA/WDS/Exemplars) (Spring 2017)
- Summarize data policies of all BF funders (Summer 2017)
- Establish Security Advisory Board (RDA) (Spring 2017)
- Establish Data Policy Advisory Board (CODATA/WDS) (Spring 2017)
- Establish Publishing Liaison Board (RDA/CODATA/ICSU/SCMA/COPDESS) (Summer 2017)
- Develop methods for certifying openness (CODATA/WDS) (Autumn 2017)

Data Planning

Action Theme 2 Leads: NSF (USA), JST (Japan), NERC (UK)
Staff:
- Mark Thorley (NERC) P/T
- Carrie Seltzer, Data Information Officer (NSF)

CAPACITY BUILDING

- List recommended training activities relevant to data intensive research
- Clarify intended training targets and needs: Survey followed by skills gap analysis, liaise with CODATA/WDS/GEO/RDA, others (February 2017)
- Hold curricula scoping workshop, EGU (April 2017)
- Investigate 'Belmont Certification' (Summer 2017)
- Present finalized curricula and recommendations (Winter 2017)
- Decision at 2017 Meeting of Principals on way forward

Human Dimensions

Action Theme 4 Lead: NERC UK
Staff:
- Vicky Lucas (IEA, UK) P/T
- Fiona Murphy (IEA, UK) P/T

ANF

Belmont Forum Data Management Plan Template

Draft Version 1.0

Published on *bfe-inf.org* 2017-03-03

1. What types of data, samples, physical collections, software, curriculum materials, and other materials will be collected, processed and/or generated in the course of the project?

2. Which standards will be used for data and metadata format and content? Where existing standards are absent or deemed inadequate, this should be documented along with any proposed solutions or remedies.

3. Describe plans and protocols for broad access and sharing, including provisions for appropriate protection of privacy, confidentiality, security, intellectual property, or other rights or requirements.

4. Describe plans and provisions for re-use, re-distribution, and the production of derivatives.

5. Describe plans for archiving data, samples, and other research products, and for preservation of access to them via an institutionally-supported repository.

6. What costs are required to implement the data management plan? Please provide a total along with a line item breakdown and brief justification. These costs should be included in the total budget request for your project.

Suggested limit: 1-2 pages for the full proposal. Briefly addressed in the pre-proposal.
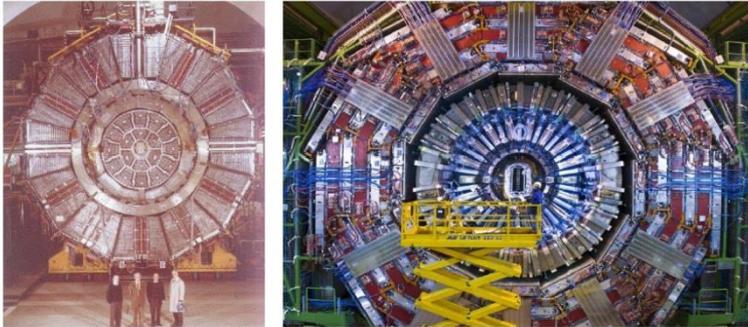
Recherche d'efficience

- Coordination des financeurs
- Simplification des circuits
- Financements coordonnés projets/infrastructures
- Des enjeux partagés ---> des approches communes

Il y a partage et partage…

Exemples disciplinaires

Données en HEP

In Big Communities
In International Labs (CERN)

Past Century collaboration
~500 Scientists

Today collaboration
~4000 Scientists

From all around the word

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)
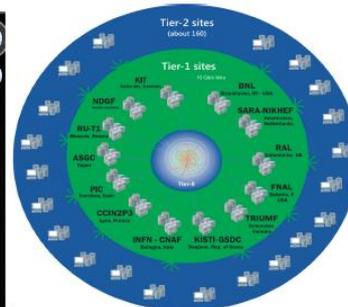Data + Replica move with a Data Management Policy
6 GB/s (600 TB/day)

LHC Disk and Tape Storage
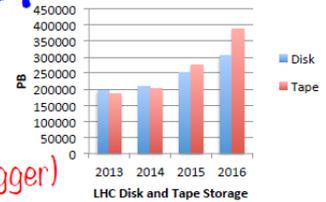
Worldwide LHC Computing Grid

Running jobs: 224651
Transfer rate: 4.11 GiB/sec

Tier-2 sites (about 160)
Tier-1 sites

Tier-0 (CERN):
•Data recording
•Initial data reconstruction
•Data distribution

Tier-1 (12 centres):
•Permanent storage
•Re-processing
•Analysis

Tier-2 (68 Federations, ~140 centres):
• Simulation
• End-user analysis

•525,000 cores
•450 PB

Data Analytics exploit data by distributed computing infrastructure of half a million cores
An average of 40M jobs/month

produces

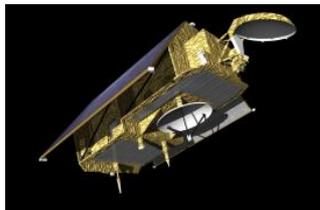"Publication Data" that are openly Shared

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

Copernicus Sentinel Data Policy — ©esa

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
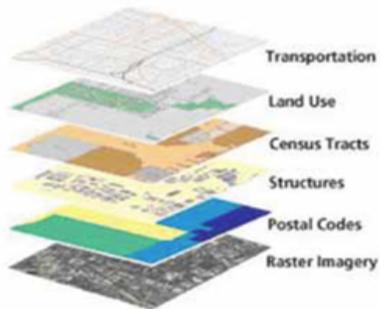
European Space Agency

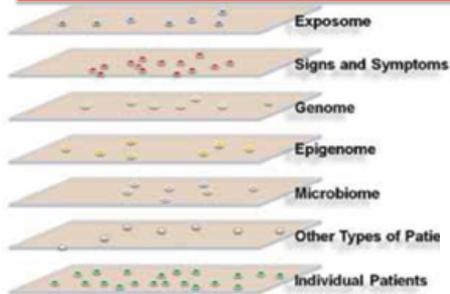- Une organisation, une série d'instruments, une politique de données

Les données de santé

Intégration des données pour une Médecine translationnelle, prédictive et personnalisée

Google Maps: GIS layers
Organized by Geographical Positioning

- Transportation
- Land Use
- Census Tracts
- Structures
- Postal Codes
- Raster Imagery

Information Commons
Organized Around Individual Patients

- Exposome
- Signs and Symptoms
- Genome
- Epigenome
- Microbiome
- Other Types of Patie
- Individual Patients

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

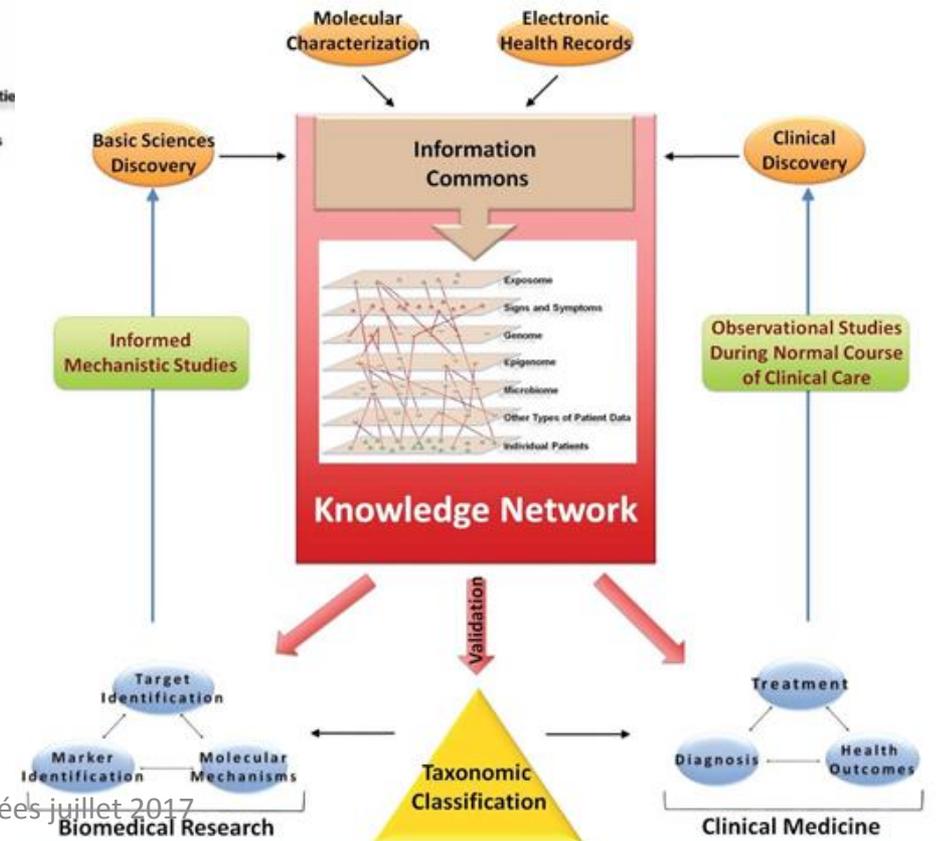- relations gène/médicament, symptômes/maladies, risques environnementaux/expression des gènes

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
- Respect des pratiques des communautés
- Lien instruments/données
- Barrières légales très diverses
- Processus de valorisation variés

Culture du partage de la donnée

- Mes données sont à moi et … tes données sont à moi également !
- Des initiatives d'acteurs variés: 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
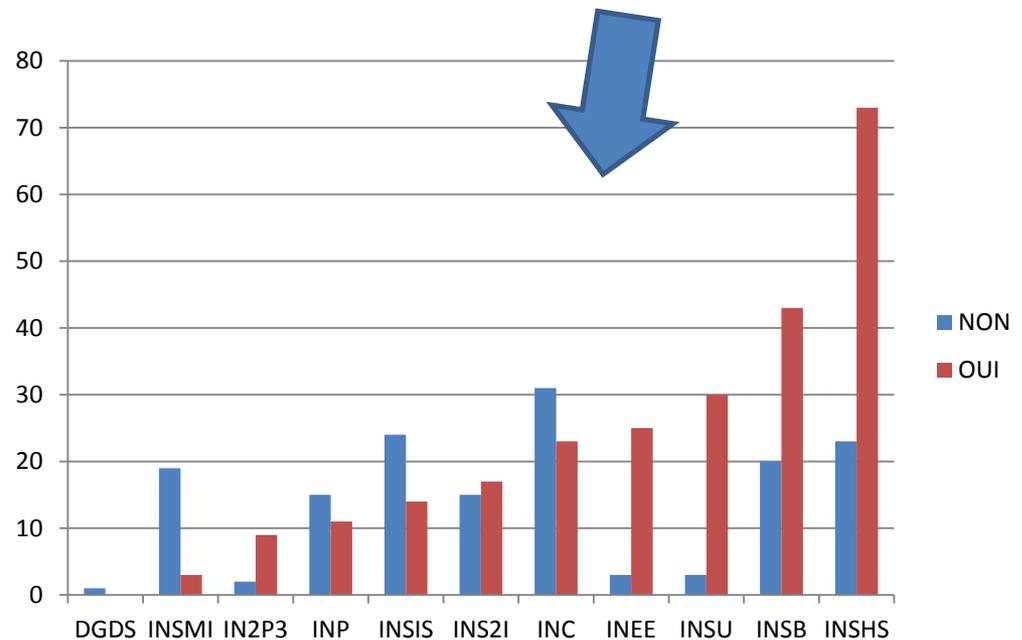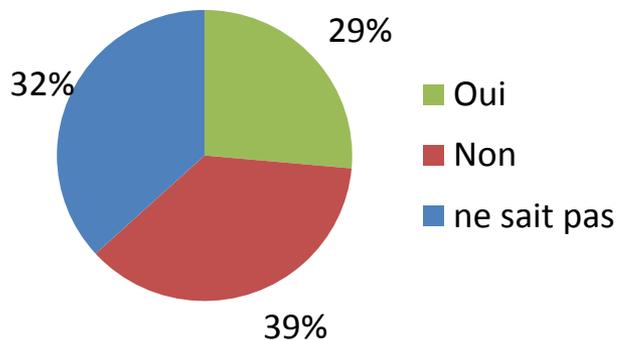
- Qu'en pensent les chercheurs ?

Enquête IST auprès des directeurs d'unité CNRS
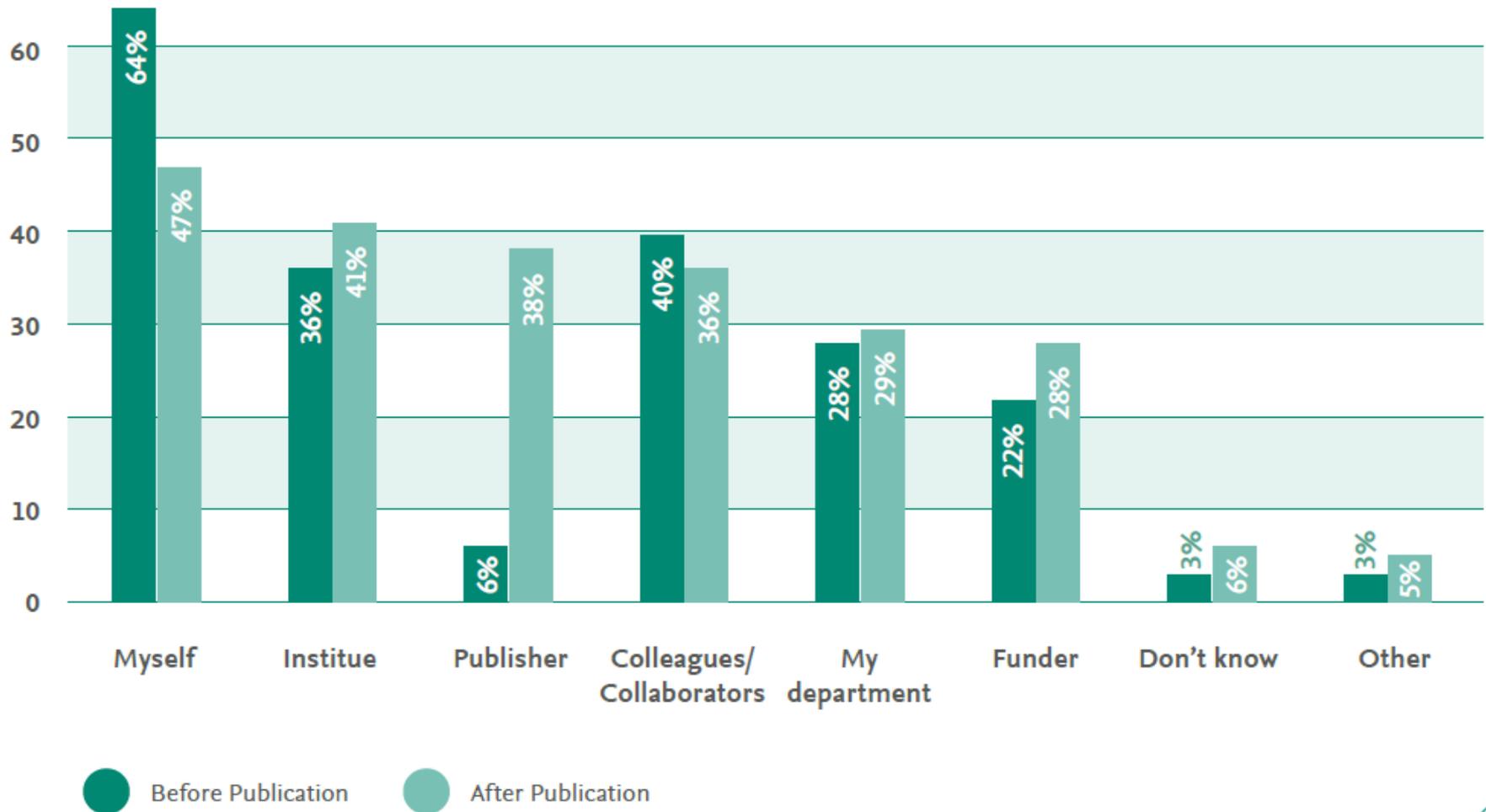
Données de la recherche

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

39%
61%
Oui
Non
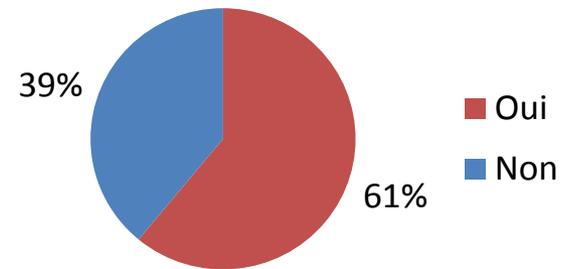
Pensez-vous que vos données de recherche soient libres de droits ?

29%
32%
39%

Oui
Non
ne sait pas

NON
OUI

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

Bar chart showing data ownership before and after publication:

Category	Before Publication	After Publication
Myself	64%	47%
Institue	36%	41%
Publisher	6%	38%
Colleagues/Collaborators	40%	36%
My department	28%	29%
Funder	22%	28%
Don't know	3%	6%
Other	3%	5%

Before Publication   After Publication

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

Diffusion des données n=1162

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

Encore de la lecture….

OPEN DATA
THE RESEARCHER PERSPECTIVE

learn

LEARN Toolkit of Best Practice
for Research Data Management

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

http://learn-rdm.eu/wp-content/uploads/RDMToolkit.pdf?pdf=RDMToolkit
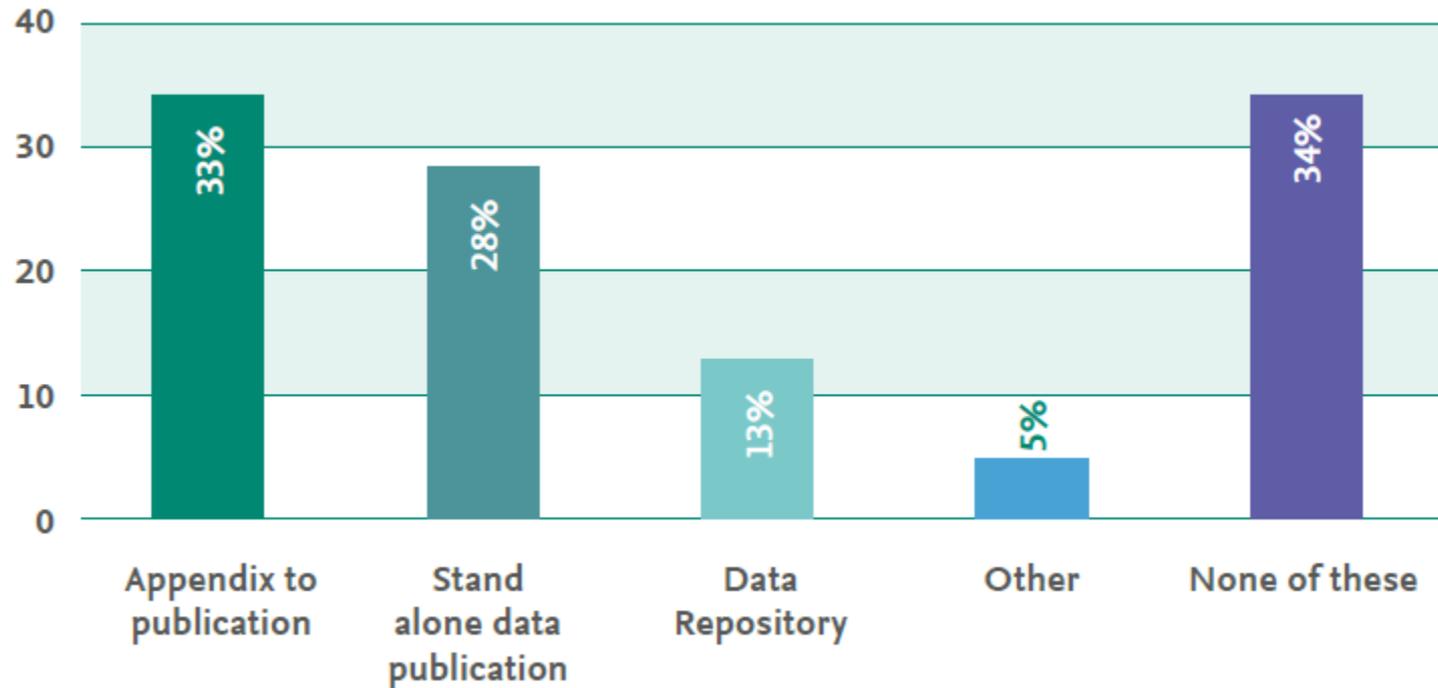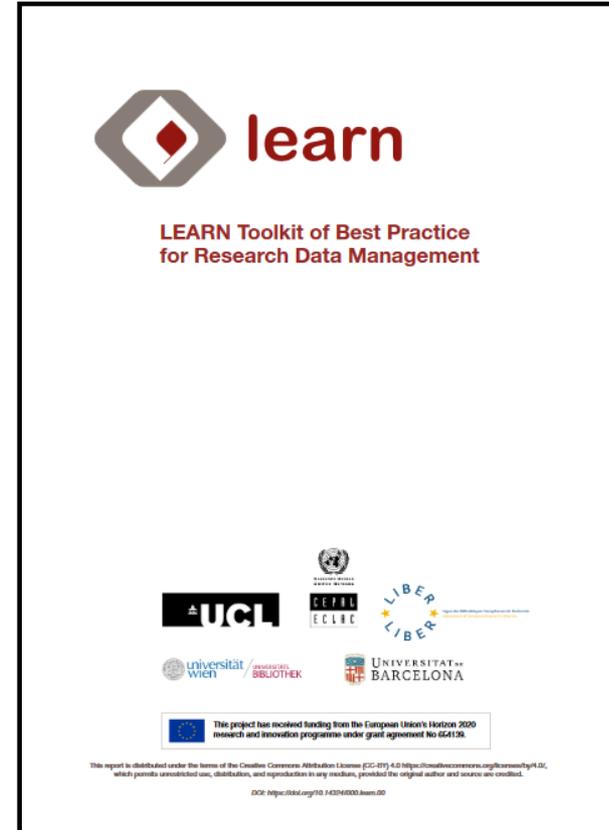
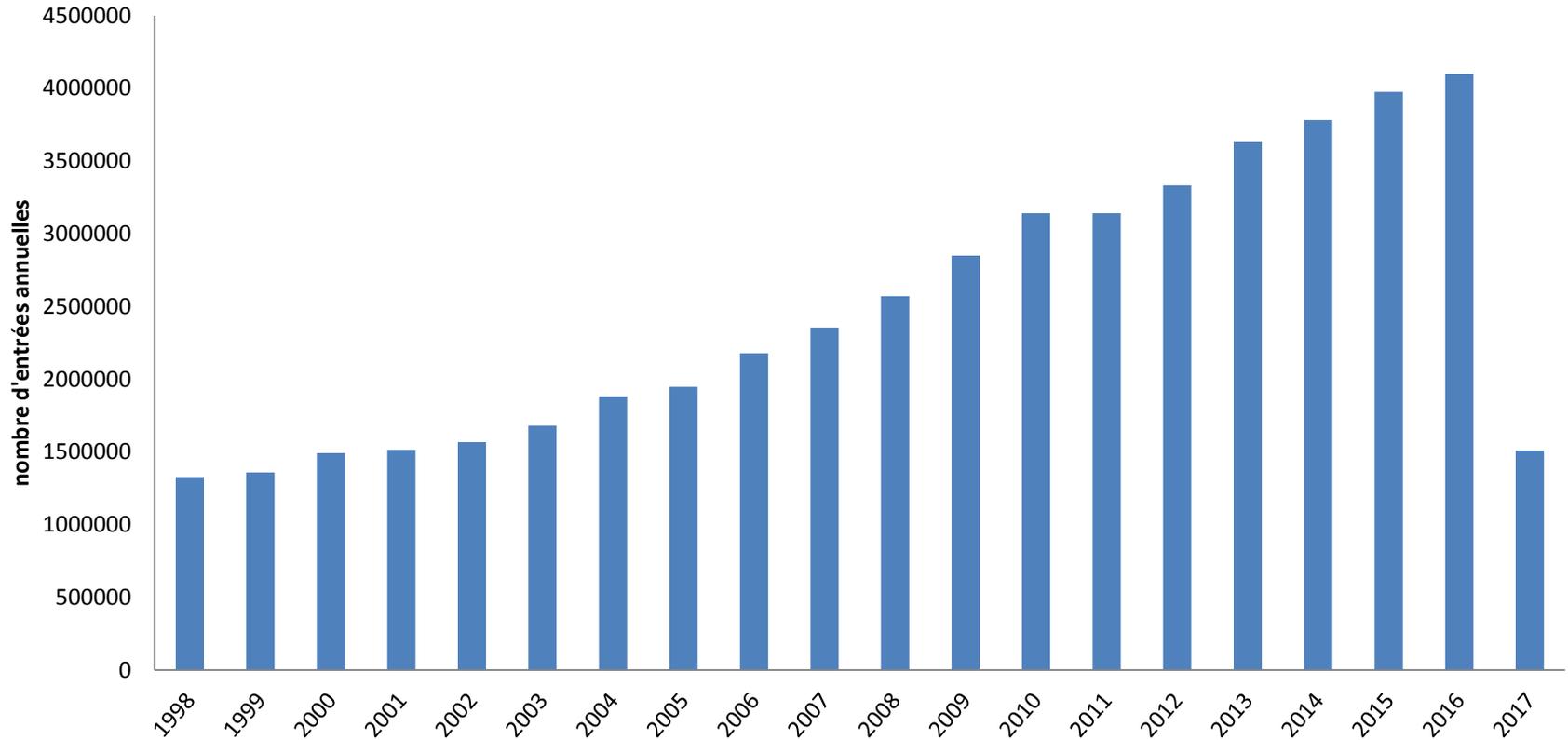# Et les publications ?
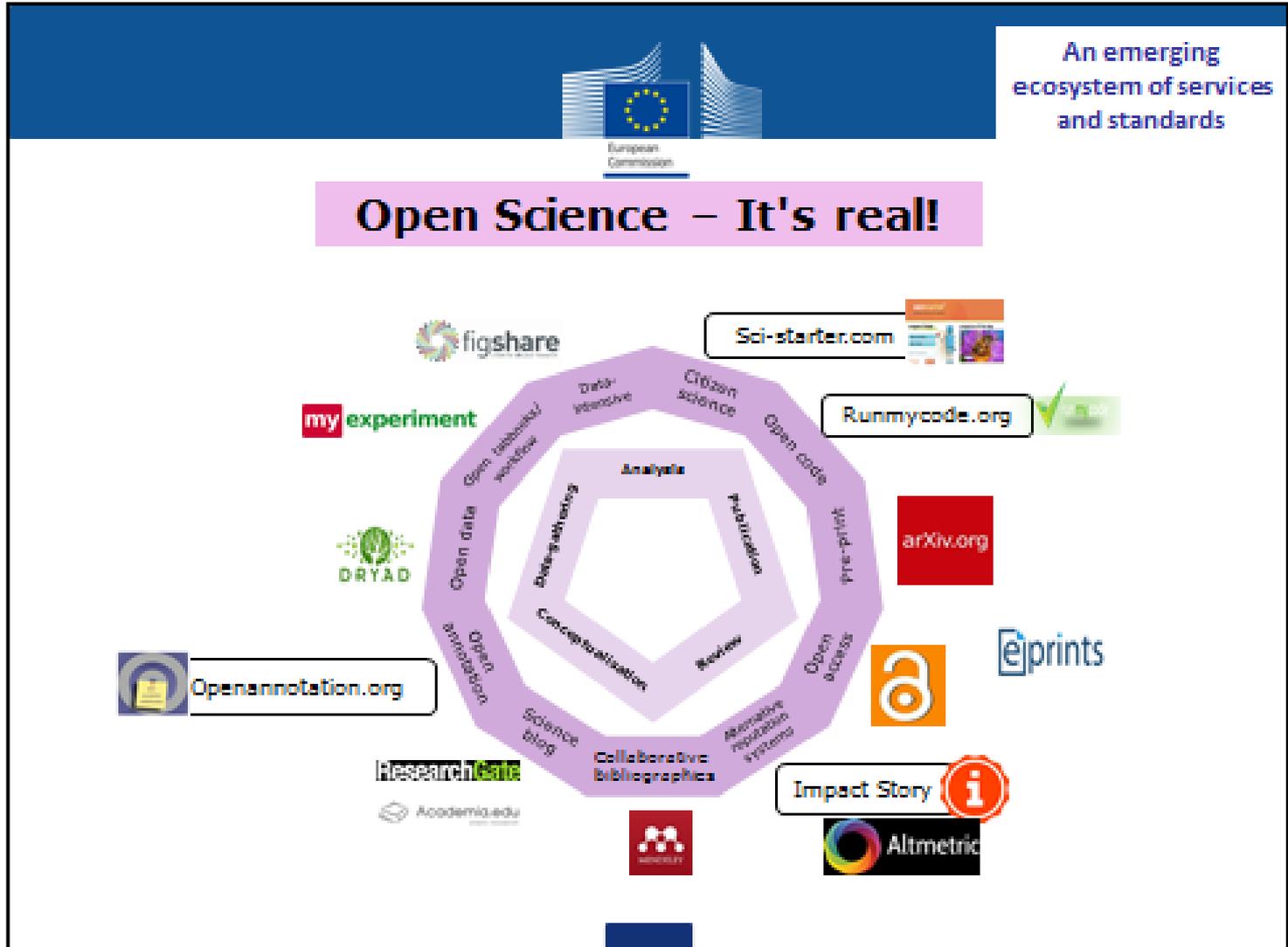
les publications dans Crossref

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

open ? Des progrès à faire....
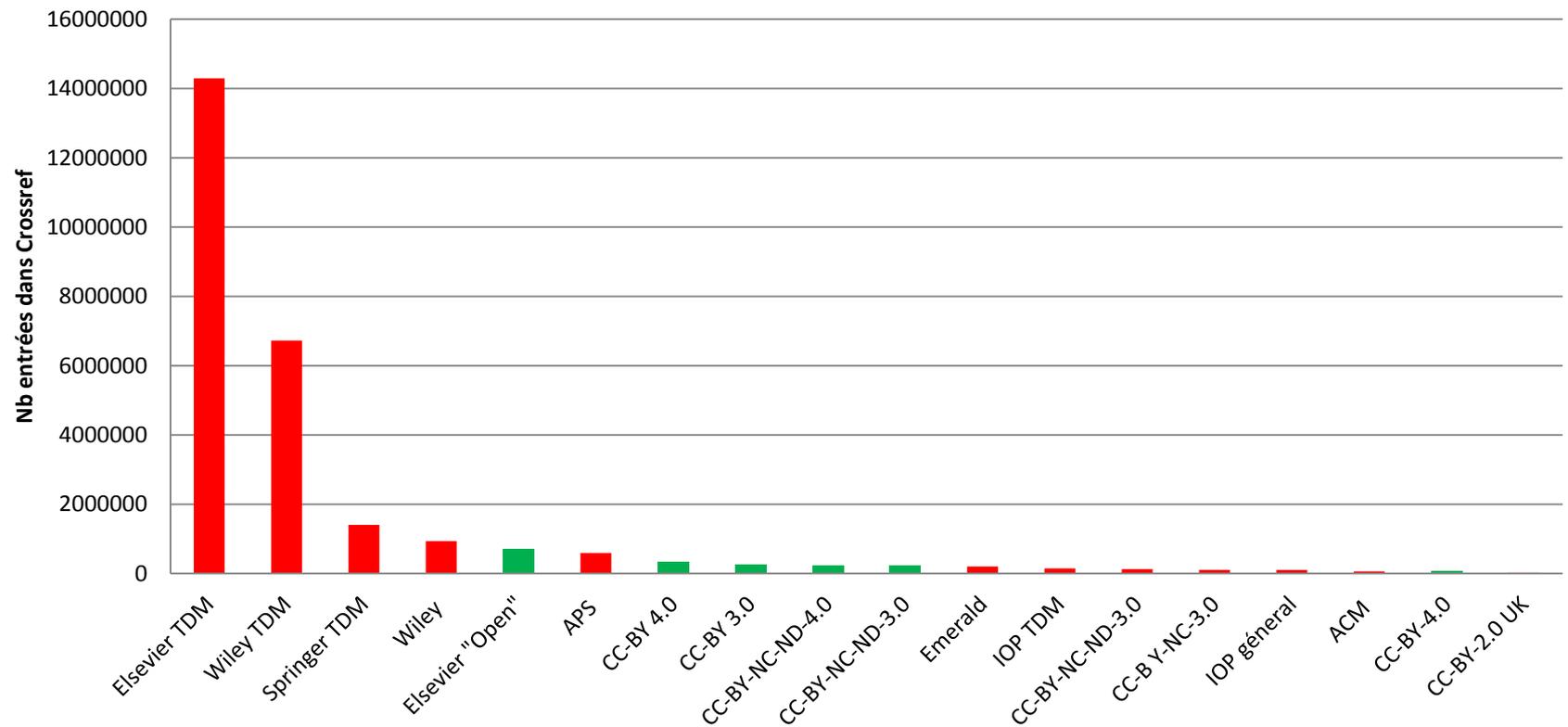
Open Science à la mode européenne

Un affichage optimiste ?

Open Science à l'échelle européenne

Amsterdam Call for Action on Open Science

nl EU 2016

- ❑ 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&section=monitor
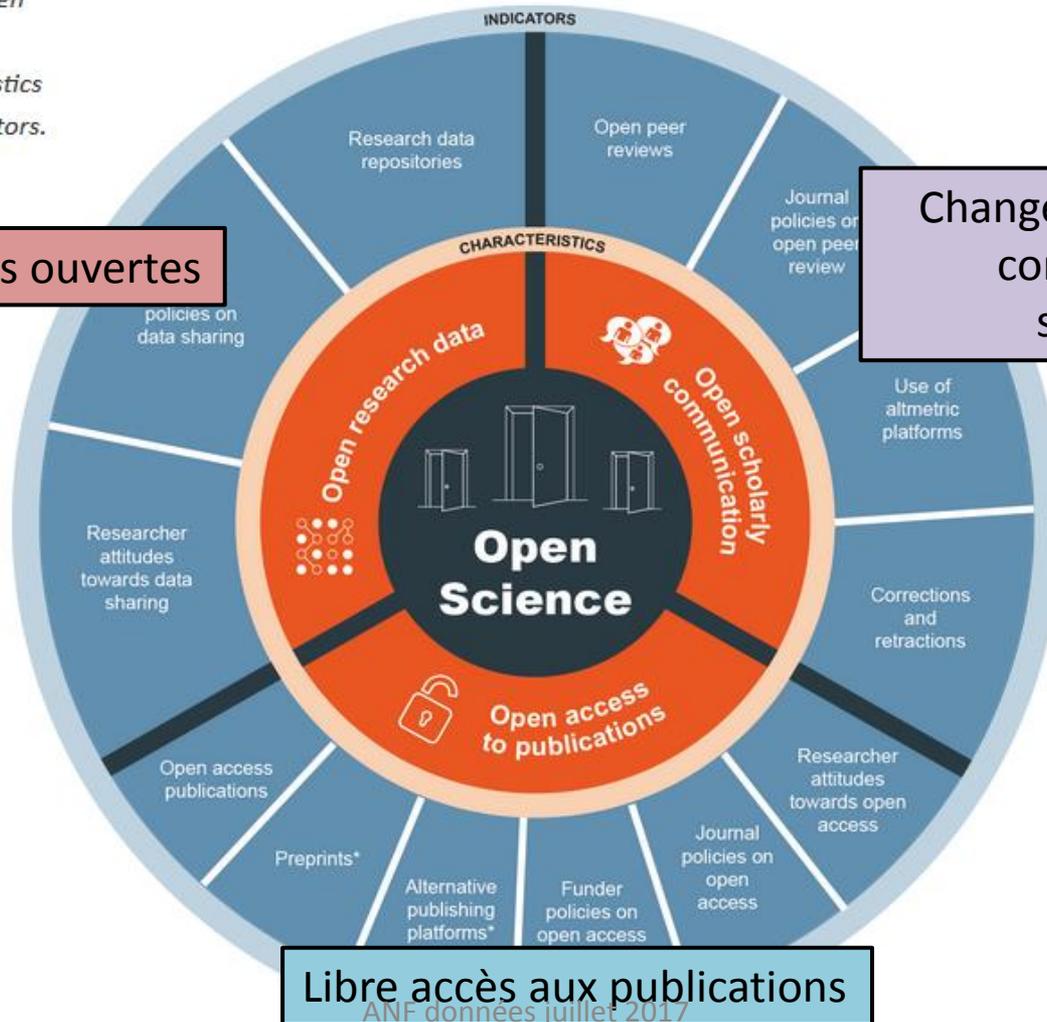

Use the wheel to explore open science characteristics and indicators.
▶ ▶

Données ouvertes

Changer les règles de la communication scientifique

Libre accès aux publications

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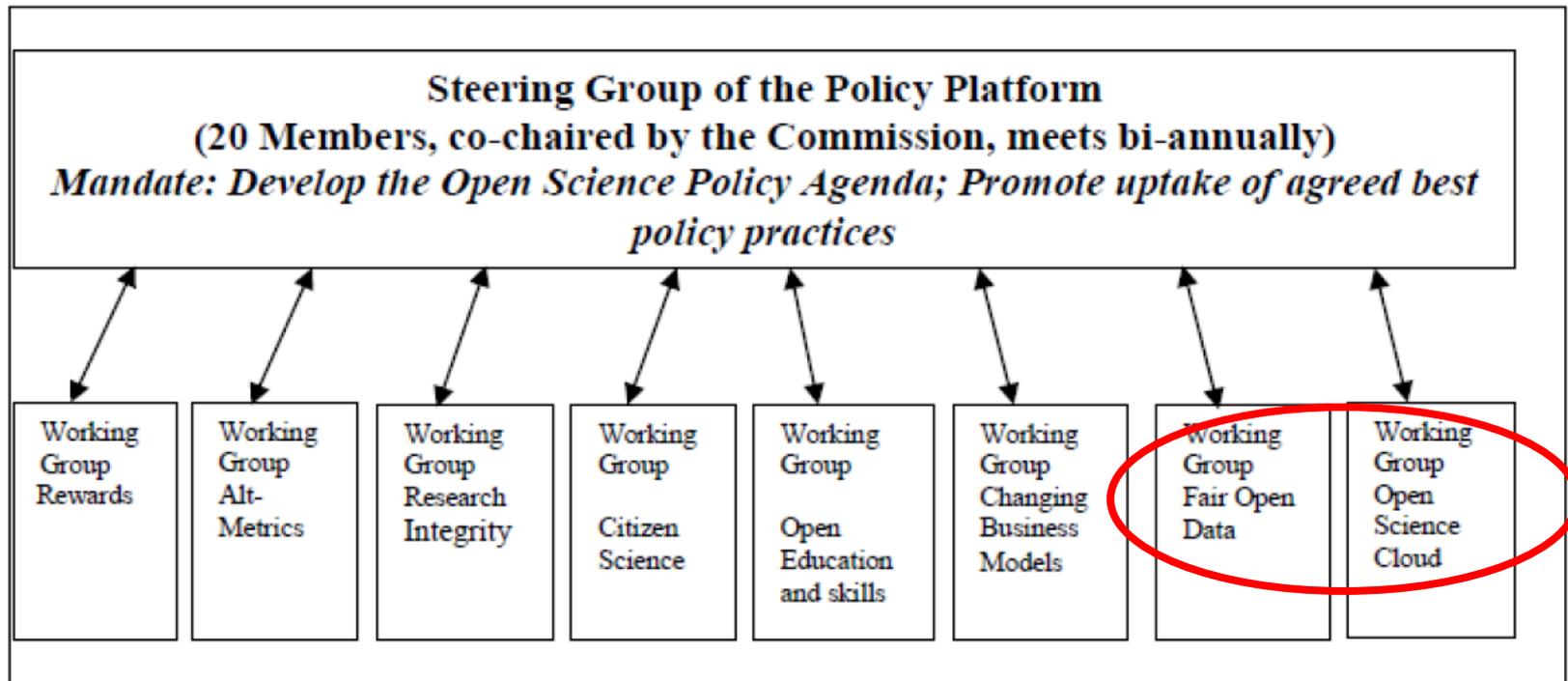
Science Europe: is an association of 47 European Research Funding Organisations and Research Performing Organisations from 27 countries, founded in October 2011. It included EUROHORCs, the European Heads of Research Councils association established in 1992 to represent the national research founding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

http://www.europarl.europa.eu/RegData/etudes/IDAN/2016/579097/EPRS_IDA%282016%29579097_EN.pdf

ANF données juillet 2017

2016

European Commission

DIRECTORATE-GENERAL FOR RESEARCH AND INNOVATION (RTD)

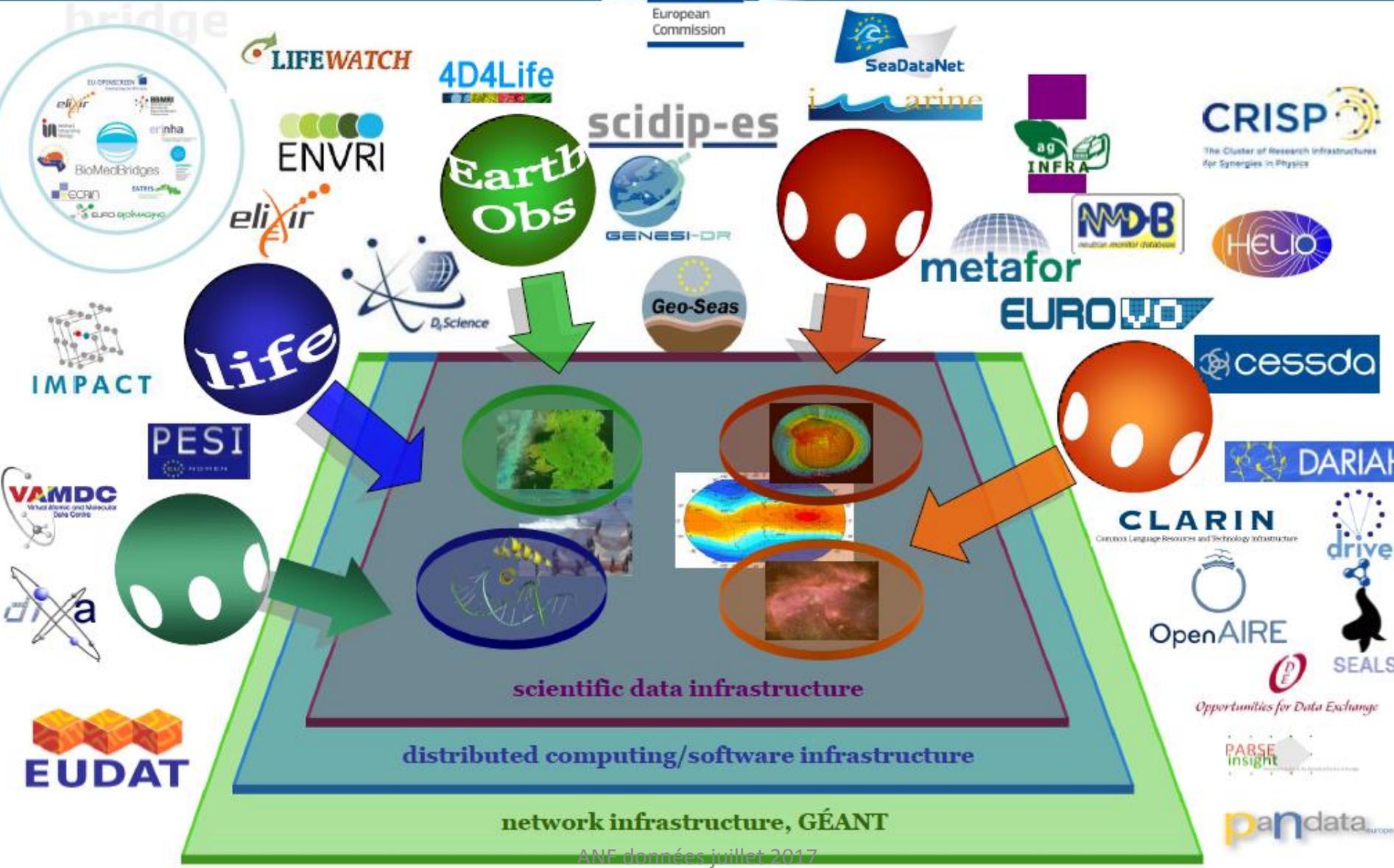**New policy initiative: The establishment of an Open Science Policy Platform**

Steering Group of the Policy Platform
(20 Members, co-chaired by the Commission, meets bi-annually)
Mandate: Develop the Open Science Policy Agenda; Promote uptake of agreed best policy practices

Working Group Rewards	Working Group Alt-Metrics	Working Group Research Integrity	Working Group Citizen Science	Working Group Open Education and skills	Working Group Changing Business Models	Working Group Fair Open Data	Working Group Open Science Cloud

FAIR : Findable, Accessible, Interoperable, Reusable

Groupes d'experts européens ad hoc

- EOSC
  - https://ec.europa.eu/research/openscience/index.cfm?pg=open-science-cloud

- Turning FAIR Data into reality
  - https://ec.europa.eu/research/openscience/pdf/fair_eg.pdf

data infrastructure: bridging islands

scientific data infrastructure

distributed computing/software infrastructure

network infrastructure, GÉANT

ANF données juillet 2017

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

ANF données juillet 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

ANF données juillet 2017
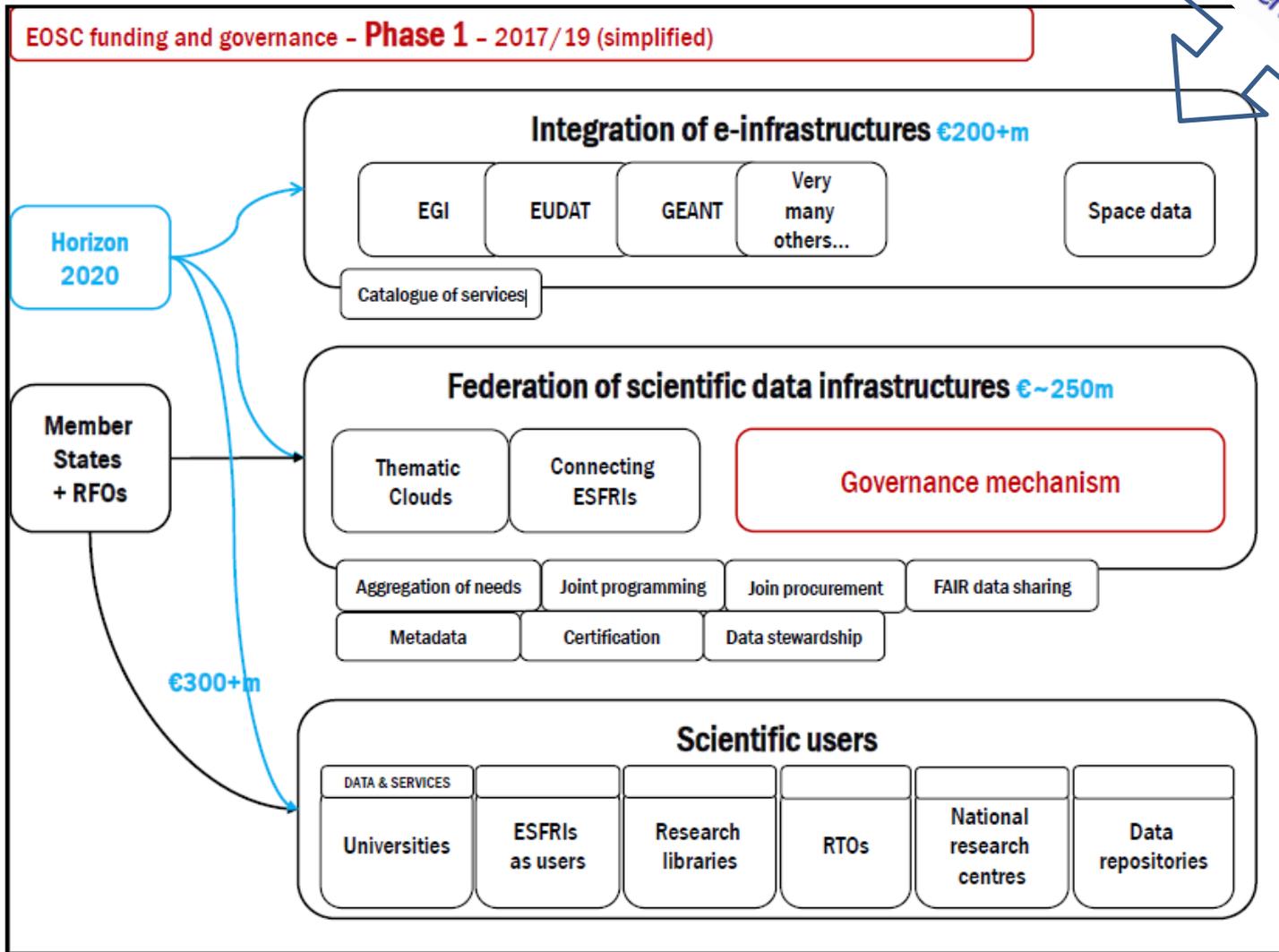
...dans un cadre de gouvernance novateur
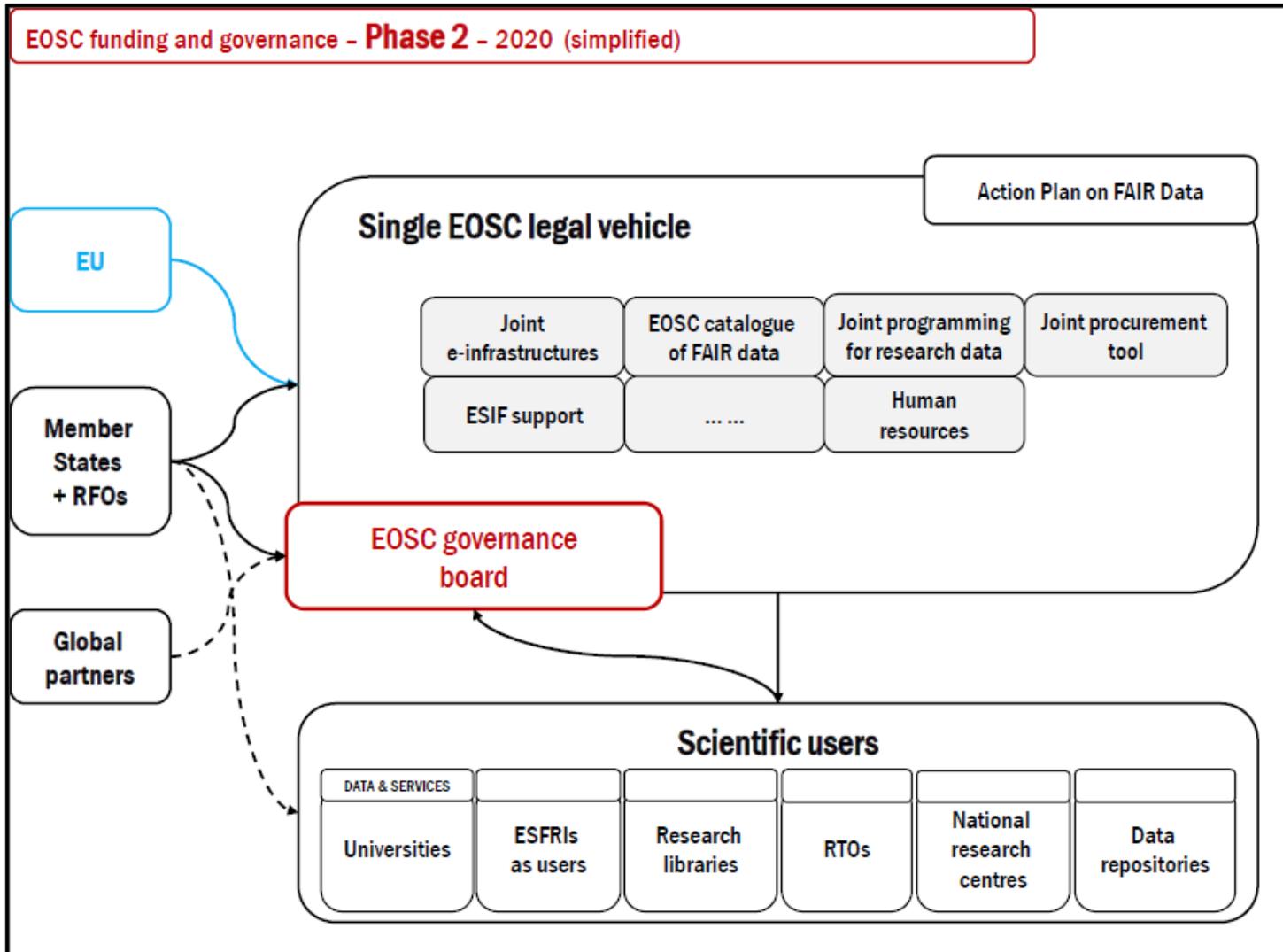
Open Data à la mode bottom-up

La Research Data Alliance

- Fondée en mars 2013 par le gouvernement australien, la CE et la NSF et NIST (USA)
- 5700 membres de ~120 pays
- Travail « bottom-up » sur tous les aspects du partage des données scientifiques, technologiques et sociologiques

http://rd-alliance.org

THE RESEARCH DATA ALLIANCE

www.rd-alliance.org

building the social and technical bridges that enable open sharing of data

18 FLAGSHIP OUTPUTS	75 ADOPTION CASES
of which 4 ICT Technical Specifications	across multiple disciplines, organisations & countries

82 GROUPS WORKING ON GLOBAL DATA INTEROPERABILITY CHALLENGES

of which 29 WORKING GROUPS & 53 INTEREST GROUPS

5,629 INDIVIDUAL MEMBERS FROM 126 COUNTRIES

66% Academia & Research
15% Public Administration
11% Enterprise & Industry

43 ORGANISATIONAL MEMBERS & 8 AFFILIATE MEMBERS
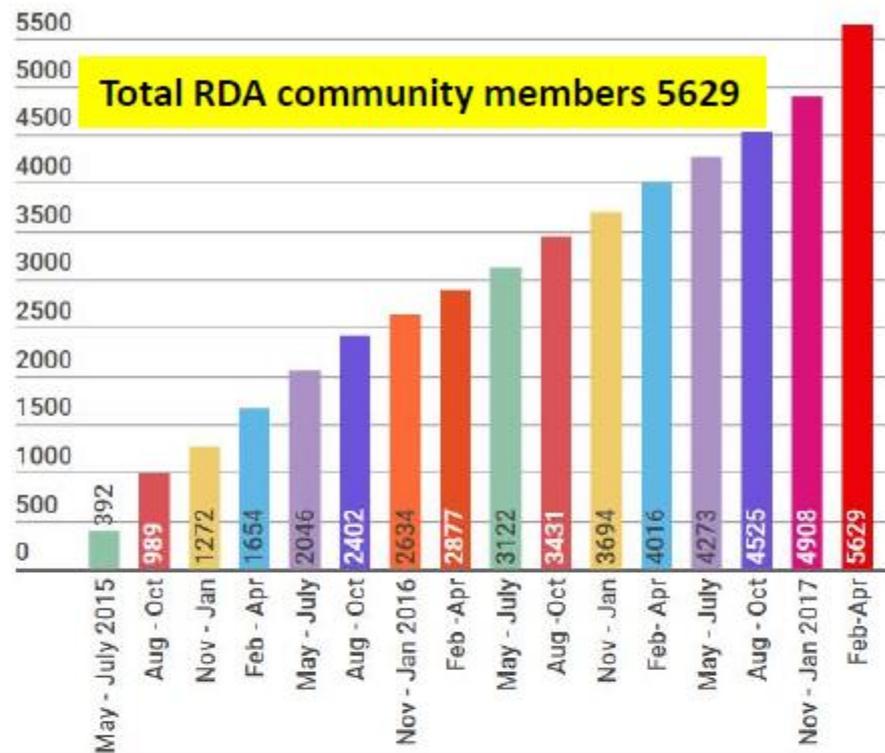
19 June 2017

Vision

Researchers and <u>innovators</u> openly share data across technologies, disciplines, and countries to address the grand challenges of society.
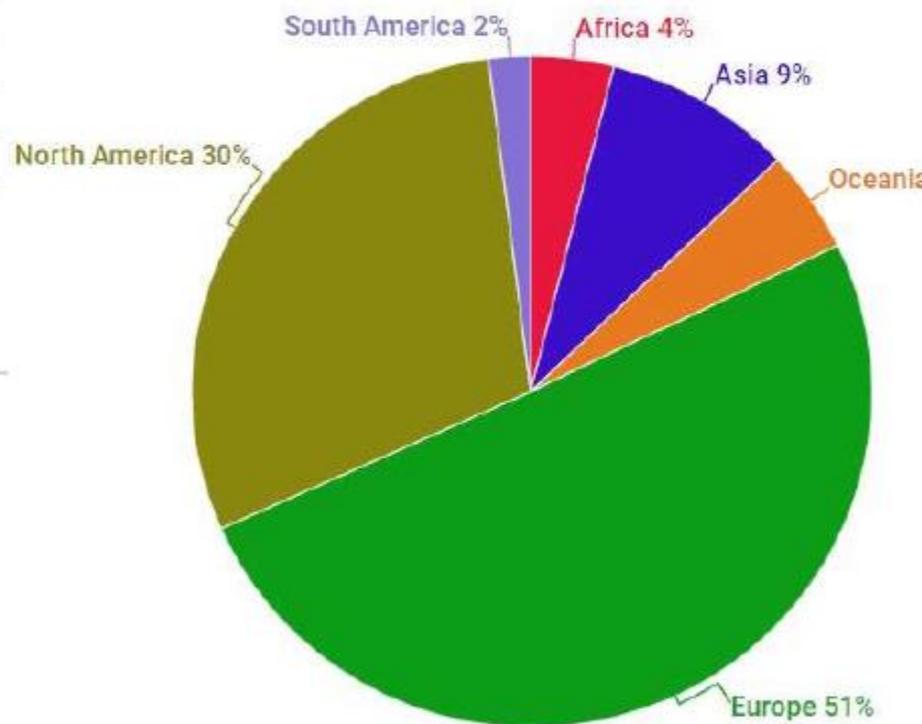
Mission

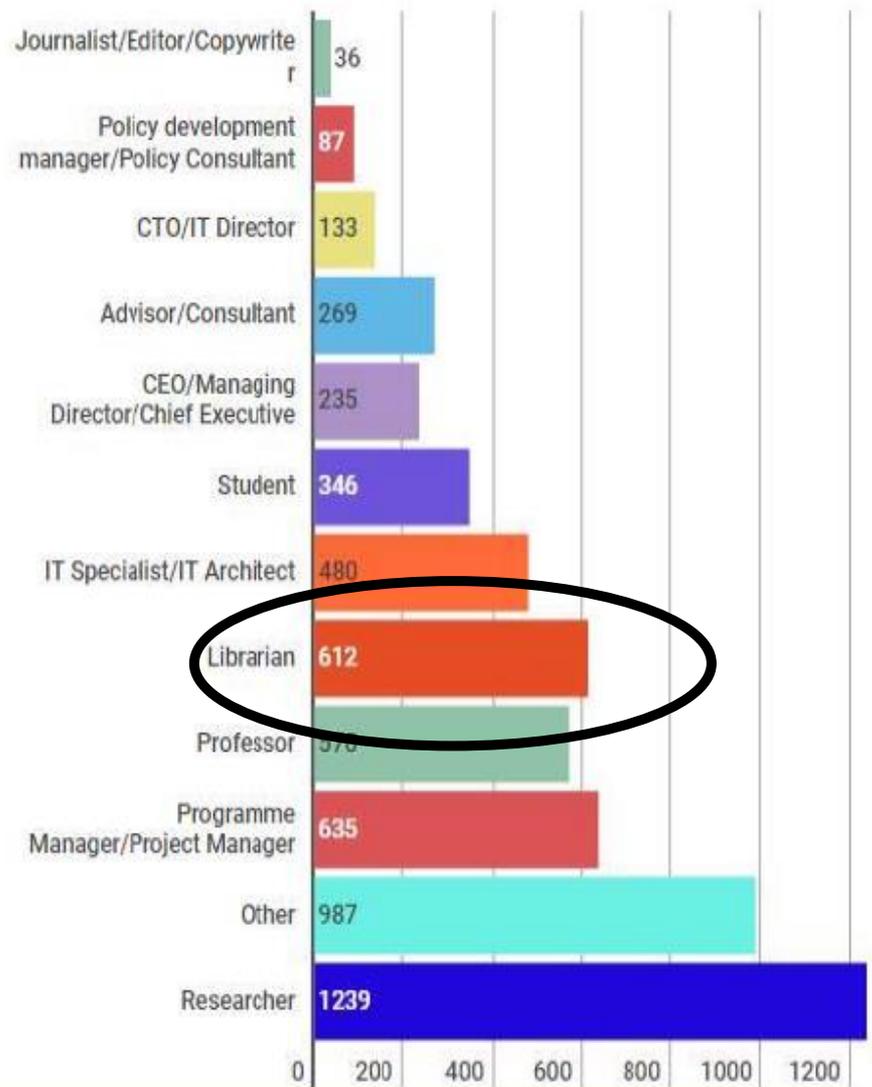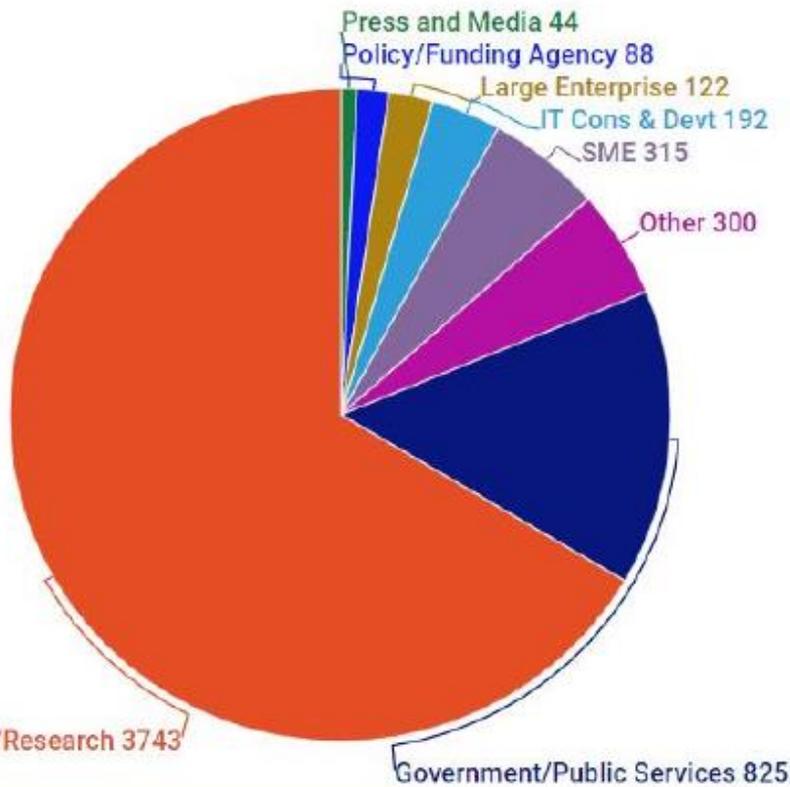RDA builds the social and technical bridges that **enable open sharing** of data.

Merci Françoise !

RDA worldwide growth

Total RDA community members 5629

392
989
1272
1654
2046
2402
2634
2877
3122
3431
3694
4016
4273
4525
4908
5629

May - July 2015, Aug - Oct, Nov - Jan, Feb - Apr, May - July, Aug - Oct, Nov - Jan 2016, Feb - Apr, May - July, Aug - Oct, Nov - Jan, Feb - Apr, May - July, Aug - Oct, Nov - Jan 2017, Feb-Apr

126 Countries

South America 2%   Africa 4%   Asia 9%
North America 30%   Oceania
Europe 51%

RDA
RESEARCH DATA ALLIANCE

Who is RDA?

Press and Media 44
Policy/Funding Agency 88
Large Enterprise 122
IT Cons & Devt 192
SME 315
Other 300
Academia/Research 3743
Government/Public Services 825

Role	Count
Journalist/Editor/Copywriter	36
Policy development manager/Policy Consultant	87
CTO/IT Director	133
Advisor/Consultant	269
CEO/Managing Director/Chief Executive	235
Student	346
IT Specialist/IT Architect	480
Librarian	612
Professor	570
Programme Manager/Project Manager	635
Other	987
Researcher	1239

Diversité!

- 82 Groupes de Travail et « Groupes d'Intérêt » sur les briques de base du partage des données
- Thèmes très divers
  - Domaines scientifiques
  - « Besoins communautaires »
  - Référencement et partage des données
  - Gestion des données et services de données
  - Technologies de base
- Producteurs de données, bibliothécaires, chercheurs, ingénieurs, « publishers », etc
- Forum de discussion neutre et international
- Recommandations et « produits »

Total 82 groups:
29 Working Groups & 53 Interest Groups

Domain Science - focused

- Agrisemantics WG
- BioSharing Registry WG
- Fisheries Data Interoperability WG
- On-Farm Data Sharing (OFDS) WG
- Rice Data Interoperability WG
- Wheat Data Interoperability WG
- Agricultural Data IG (IGAD)
- Biodiversity Data Integration IG
- Chemistry Research Data IG
- Digital Practices in History and Ethnography IG

- Geospatial IG
- Global Water Information IG
- Health Data IG
- Linguistics Data Interest Group
- Mapping the Landscape IG
- Marine Data Harmonization IG
- Quality of Urban Life IG
- RDA/CODATA Materials Data, Infrastructure & Interoperability IG
- Research data needs of the Photon and Neutron Science community IG
- Small Unmanned Aircraft Systems' Data IG
- Structural Biology IG
- Weather, Climate and air quality IG

Community Needs - focused

- Certification and Accreditation for Data Science Training and Education WG
- RDA/CODATA Summer Schools in Data Science and Cloud Computing in the Developing World WG
- Teaching TDM on Education and Skill Development WG
- Archives & Records Professionals for Research Data IG

- Data for Development IG
- Development of Cloud Computing Capacity and Education in Developing World Research IG
- Early Career and Engagement IG
- Education and Training on handling of research data IG
- Ethics and Social Aspects of Data IG
- International Indigenous Data Sovereignty IG

Total 82 groups:
29 Working Groups & 53 Interest Groups

Reference and Sharing - focused

- Data Citation WG
- Data Description Registry Interoperability WG
- Data Security and Trust WG
- Empirical Humanities Metadata WG
- International Materials Resource Registries WG
- Provenance Patterns WG
- QoS-DataLC Definitions WG

- RDA / WDS Publishing Data Bibliometrics WG
- Repository Core Description WG
- Research Data Collections WG
- Research Data Repository Interoperability WG
- Data Discovery Paradigms IG
- National Data Services IG
- RDA/CODATA Legal Interoperability IG
- Reproducibility IG

Partnership Groups

- RDA / TDWG Metadata Standards for attribution of physical and digital collections stewardship WG
- RDA/WDS Scholarly Link Exchange Working Group
- ELIXIR Bridging Force IG
- RDA/NISO Privacy Implications of Research Data Sets IG
- RDA/WDS Publishing Data IG

RDA Interest (IG) & Working Groups (WG) by Focus (3)

Total 82 groups:
29 Working Groups & 53 Interest Groups

Data Stewardship and Services – focused

- Brokering Framework WG
- WDS/RDA Assessment of Data Fitness for Use WG
- ~~RDA / WDS Publishing Data Workflows WG~~
- Active Data Management Plans IG
- Data in Context IG
- Data Rescue IG
- Data Versioning IG
- Domain Repositories IG
- Libraries for Research Data IG

- Long tail of research data IG
- Preservation e-Infrastructure IG
- Preservation Tools, Techniques, and Policies IG
- RDA/WDS Certification of Digital Repositories IG
- RDA/WDS Publishing Data Cost Recovery for Data Centres IG
- Repository Platforms for Research Data IG
- Research Data Provenance IG
- Virtual Research Environments IG

Base Infrastructure – focused

- Array Database Assessment WG
- Data Type Registries WG
- Metadata Standards Catalog WG
- PID Kernel Information WG
- Data Fabric IG
- Data Foundations and Terminology IG
- Big Data IG
- Brokering IG

- Federated Identity Management IG
- Metadata IG
- PID IG
- Vocabulary Services IG

Le DMP demain….
RDA Active DMP Interest Group

- D'abord un lieu d'échanges de bonnes pratiques puis :

- Création de WG dédiés :
  - Defining common standards for DMPs : machine actionable !
  - Publishing / depositing DMPs : réutilisation
  - Domain/infrastructure specialisation : simplification pour les chercheurs
  - Funder liaison : utilisation du DMP dans les revues
  - Software management plans : élargissement du champ du DMP

    https://www.rd-alliance.org/group/active-data-management-plans-ig/wiki/plenary-9-session

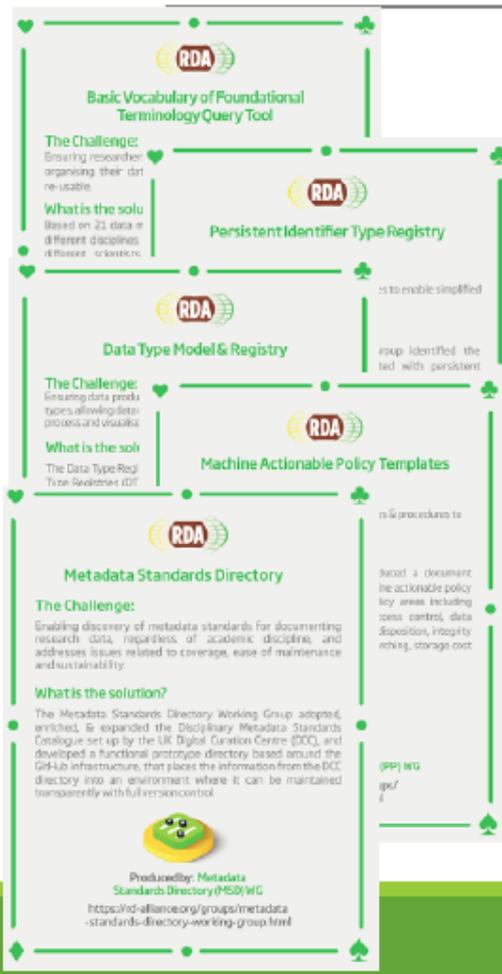- Convergence avec réflexions DANS/ScienceEurope
  - Domain Data Protocols

RDA Recommendations & Outputs

Data Foundation & Terminology: a model for data in the registered domain.

PID Information Types: a common protocol for providers and users of persistent ID services worldwide.

Data Type Registries: allowing humans and machines to act on unknown, but registered, data types.

Practical Policy: defining best practices of how to deal with data automatically and in a documented way with computer actionable policy.

Metadata standards directory: Community curated standards catalogue for metadata interoperability
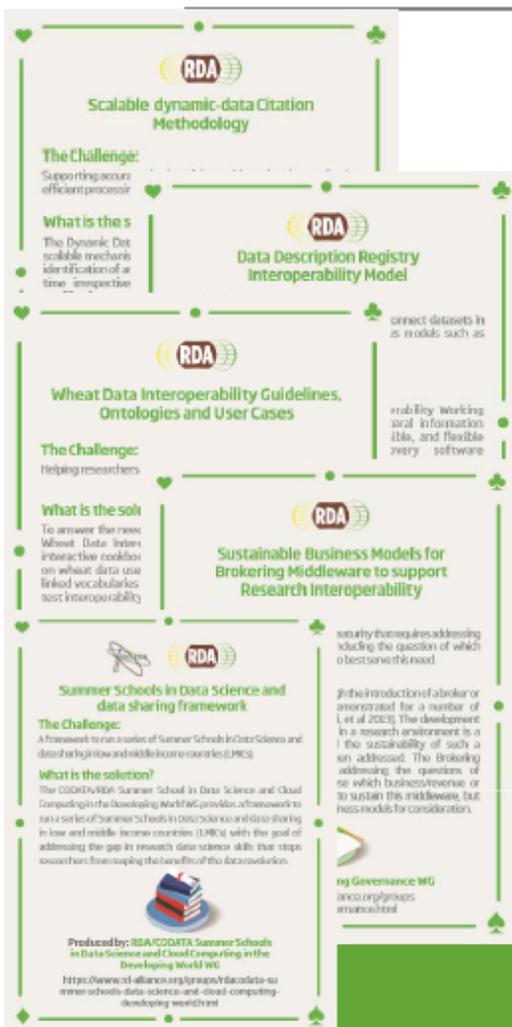
rd-alliance.org/recommendations-and-outputs/all-recommendations-and-outputs

WWW.RD-ALLIANCE.ORG
@RESDATALL

CC BY-SA 4.0

RDA Recommendations & Outputs

Data Citation: defining mechanisms to reliably cite dynamic data

Data Description Registry Interoperability solutions enabling cross platform discovery based on existing open protocols and standards

Wheat Data Interoperability impacting the discoverability, reusability and interoperability of wheat data by building a common framework for describing, representing linking and publishing wheat data

Brokering Governance WG: Sustainable Business Models for Brokering Middleware to support Research Interoperability

RDA/CODATA Summer Schools in Data Science and Cloud Computing in the Developing World WG: A framework to run a series of Summer Schools in Data Science and data sharing in low and middle income countries (LMICs)

rd-alliance.org/recommendations-and-outputs/all-recommendations-and-outputs

ANF données juillet 2017

RDA Recommendations & Outputs

Repository Audit and Certification DSA–WDS: A convergent DSA-WDS certification standard to help eliminate duplication of effort, increase certification procedure coherence and compatibility thus benefitting researchers, data managers, librarians and scientific communities.

RDA/WDS Publishing Data Bibliometrics: improved research data metrics and corresponding services, with the final goal of increasing the overall availability and quality of citations and research data itself.

RDA/WDS Publishing Data Workflows: enhance the possibilities for greater discoverability and a more efficient and reliable reuse of research data benefitting other stakeholders like publishers, libraries and data centres.

RDA/WDS Publishing Data Services: A universal interlinking service between data and the scientific literature. **The Scholix initiative** a high level interoperability framework for exchanging information about the links between scholarly literature and data. It aims to build an open information ecosystem to understand systematically what data underpins literature and what literature references data.

rd-alliance.org/recommendations-and-outputs/all-recommendations-and-outputs

Repository Audit and Certification Catalogues

Bibliometric Indicators for Data Publishing

Workflows for Research Data Publishing: Models and Key Components

An open, universal literature-data cross-linking service

The Challenge:
Sharing information about the links between the literature and research data.

What is the solution?
Building on pre-existing components and international initiatives, the RDA/WDS Publishing Data Services Working Group is one of the drivers behind the "Data Literature Interlinking Service" (DLI), developed in a synergy with OpenAIRE & PANGAEA. DLI is aimed at improving visibility, discoverability, re-use and reproducibility by bringing 2M+ existing article/data links together, normalize them using a common schema, and exposing the full set as an open service.

Produced by: RDA/WDS Publishing Data Services WG
https://rd-alliance.org/groups/rdawds-publishing-data-services-wg.html

ANF données juillet 2017

RDA Recommendations & Outputs

23 Things: Libraries For Research Data An overview of practical, free, online resources and tools that users can immediately take advantage of to incorporate research data management into the practice of librarianship.

Legal Interoperability of Research Data Principles and Implementation Guidelines: a set of principles and practical implementation guidelines offered as high-level guidance to all members of the research community —the funders, managers of data centers, librarians, archivists, publishers, policymakers, university administrators, individual researchers, and their legal counsel.

Matrix of use cases and functional requirements for research data repository platform Based on use cases, the matrix describes forty-four functional requirements identified for research data repository platforms and provides a score identifying relative importance.

BioSharing Recommendations Data repositories, standards and policies in the life, biomedical and environmental sciences

rd-alliance.org/recommendations-and-outputs/all-recommendations-and-outputs

---

Sidebar panels

RDA — 23 Things For Research Data Management

RDA — Legal Interoperability of Research Data: Principles and Implementation Guidelines

RDA — Matrix of use cases and functional requirements for research data repository platform

The Challenge: Gather and analyze research data use cases in the context of repository platform requirements

What is the solution? Based on use cases, the matrix describes forty-four functional requirements identified for research data repository platforms and provides a score identifying relative importance. These functional requirement scores can be used to assess research data repository platforms and to prioritize functional requirements for development and adoption.

Produced by: Repository Platforms for Research Data IG

https://www.rd-alliance.org/groups/repository-platforms-research-data.html

---

ANF données juillet 2017

# Inventer les métiers autour de la donnée

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*

Relations IST/Open Science

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

Un peu de lecture…

SPEC Kit 354
Data Curation
May 2017

ASSOCIATION OF RESEARCH LIBRARIES
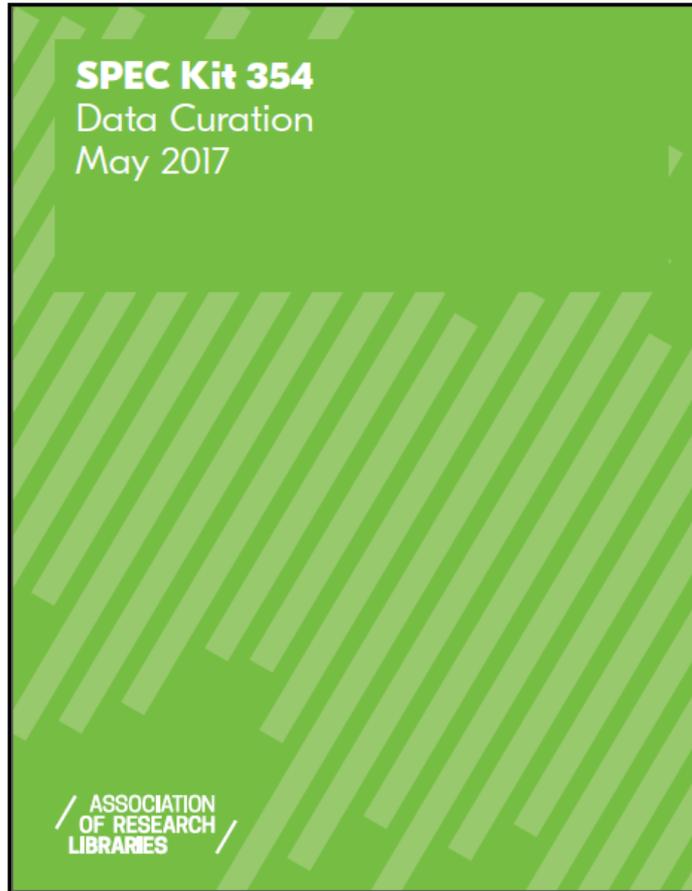
La curation de données dans les universités américaines :

- **Services**
- **Politiques**
- **Support offert par les bibliothèques dans le RDM**
- **Difficultés et aspirations…**

http://publications.arl.org/Data-Curation-SPEC-Kit-354/

Quelques réflexions pour terminer….

Les pièces du puzzle
Open Science/OpenData

- *(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 institute », « Open data Center »,…
  - Former, former, (se) 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