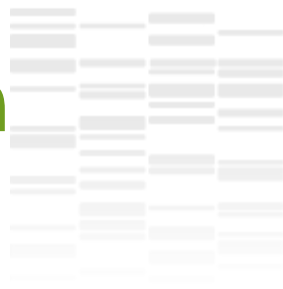


# Plan



- ❖ Contexte politique international
- ❖ Exemple anglais
- ❖ Définir les « Données de la recherche » ?
- ❖ Editeurs, données, publications, évaluation
- ❖ Apport des professionnels IST
- ❖ Démarche de l'Inra



\_01

# Un contexte politique en faveur de l'ouverture

Déclaration de Berlin, Rapport de l'OCDE, Recommandations de la CE et autres



# La déclaration de Berlin – Octobre 2003

## Goals

<http://oa.mpg.de/lang/en-uk/berlin-prozess/berliner-erklarung/>

Our mission of disseminating knowledge is only half complete if the information is not made widely and readily available to society. New possibilities of knowledge dissemination not only through the classical form but also and increasingly through the open access paradigm via the Internet have to be supported. We define open access as a comprehensive source of human knowledge and cultural heritage that has been approved by the scientific community.

In order to realize the vision of a global and accessible representation of knowledge, the future Web has to be sustainable, interactive, and transparent. Content and software tools must be openly accessible and compatible.

## Definition of an Open Access Contribution

Establishing open access as a worthwhile procedure ideally requires the active commitment of each and every individual producer of scientific knowledge and holder of cultural heritage. Open access contributions include original scientific research results, raw data and metadata, source materials, digital representations of pictorial and graphical materials and scholarly multimedia material.

# Travaux de l'OCDE – 2004-2007

Un accès efficace aux données de la recherche, utilisé de façon responsable et efficiente, est nécessaire pour tirer pleinement parti des nouvelles possibilités et retombées offertes par les TIC. L'accessibilité aux données de la recherche est devenue une importante condition pour :

- Une gestion avisée de l'investissement public dans l'information factuelle ;
- Le développement **Valeur économique, chaîne de la valeur**
- L'accroissement de la valeur procurée par la coopération internationale. Plus spécifiquement, l'amélioration de l'accès aux données et du partage de celles-ci :
  - Renforce la liberté de l'inv **Liberté d'investigation**
  - Encourage la diversité d'analyse et d'opinion ;
  - Stimule de nouvelles recherches ; **Preuve**
  - Rend possible l'expérimentation d'hypothèses et de méthodes d'analyse nouvelles ou différentes ; **Créativité**
  - Contribue aux études sur les méthodes de collecte de données et leur mesure ;
  - Facilite la formation des nouveaux chercheurs ;
  - Permet l'exploration de thèmes qui n'avaient pas été envisagés par les chercheurs d'origine ; **Innovation**
  - Conduit à la création de nouveaux ensembles de données par la combinaison de données provenant de multiples sources.

L'accès ouvert aux données de la recherche financée sur fonds publics et leur partage contribuent non seulement à maximiser l'impact des nouvelles technologies et des nouveaux réseaux numériques sur le potentiel de recherche, mais permettent aussi un retour plus important sur l'investissement public dans la recherche.

Extrait de [ Principes et lignes directrices de l'OCDE pour l'accès aux données de la recherche financée sur fonds publics] publié en 2007

# Recommandations de la CE – juillet 12

## *Financement d'infrastructures et de projets*

- Continuer à financer les projets pertinents dans le cadre d'Horizon 2020, à partir de 2014.
- Consacrer 45 millions d'EUR à des dépenses dans le domaine des infrastructures permettant le libre accès à des articles et données de recherche et dans celui de la recherche sur la conservation numérique, 2012-2013.

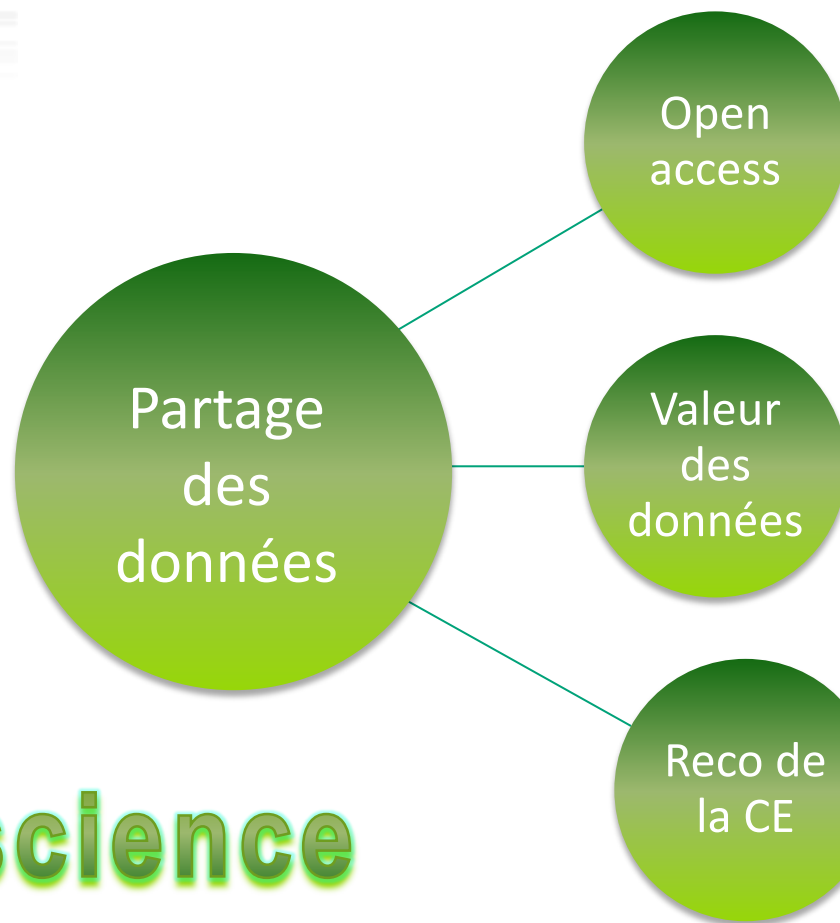
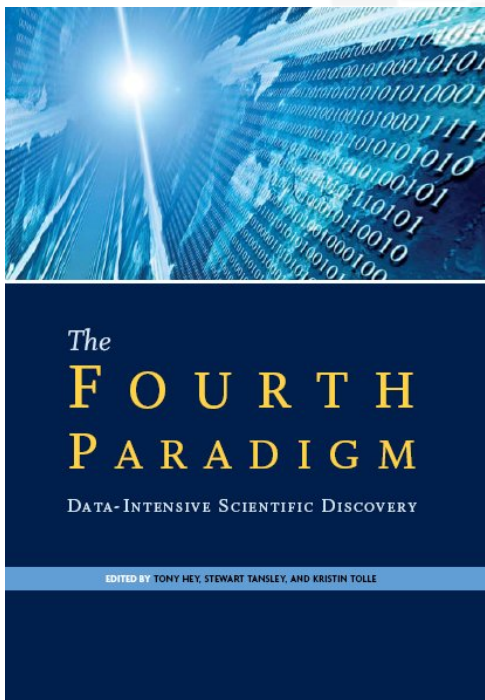
## *Coordination au-delà de l'UE*

- Promouvoir les politiques de libre accès et l'interopérabilité des infrastructures de données auprès des partenaires internationaux.

## **Objectifs :**

- D'ici à 2014, des politiques de libre accès aux articles et données scientifiques auront été établies dans tous les États membres à tous les niveaux pertinents.
- D'ici à 2016, le pourcentage d'articles scientifiques sur des travaux financés par des fonds publics disponibles en libre accès dans l'UE sera passé de 20 % à 60 %.
- 100 % des publications scientifiques issues d'Horizon 2020 seront disponibles en libre accès.

# Les faits marquants



- Déclaration de Berlin 2003 : Libre accès à l'IST publi/données

- OCDE 2007 : Recherche financée sur fonds publics

- Politique à l'horizon 2014, H2020

## Open science Data driven science

<http://research.microsoft.com/en-us/collaboration/fourthparadigm/>

# Des catalyseurs récents

## ❖ Révision du « Grant Proposal Guide » de la NSF : janvier 2013

Chapter II.C.2.f(i)(c), Biographical Sketch(es), has been revised to **rename the “Publications” section to “Products”** and amend terminology and instructions accordingly. This change makes clear that products may include, but are not limited to, publications, data sets, software, patents, and copyrights.

## ❖ Position de la Maison Blanche (février 2013)

## ❖ G8+5 Open data for Agriculture

Obtain commitment and action from nations and relevant stakeholders to **promote policies and invest in projects that open access to publicly funded global agriculturally relevant data streams**, making such data readily accessible to users in Africa and world-wide, and ultimately supporting a sustainable increase in food security in developed and developing countries



## ❖ Research data Alliance

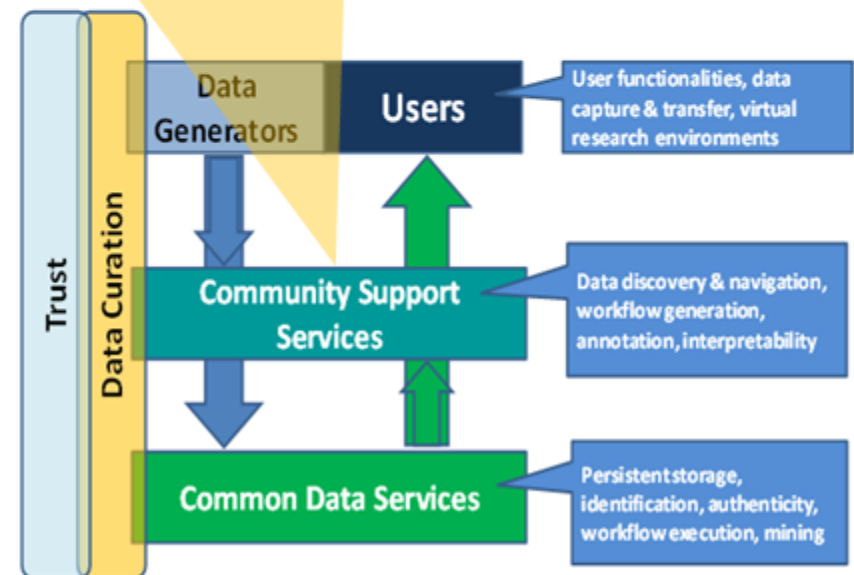
The Research Data Alliance aims to **accelerate and facilitate research data sharing and exchange**



# implementing interoperable data infrastructure

- (a) **data generators**; research projects, big research infrastructure, installations or medium size laboratories, simulation centres, surveys or individual researchers
- (b) **discipline-specific data service providers**, providing data and workflows as a service
- (c) **providers of generic common data services** (computing centres, libraries)
- (d) **researchers as users**, using the data for science and engineering

community driven data infrastructure, including ESFRI, ESFRI clusters and others





# Partage de données - Enjeux

## ❖ Enjeux patrimoniaux

- ✓ Preuve
- ✓ Mémoire

## ❖ Enjeux économiques

- ✓ Valeur économique de la donnée
- ✓ Ré-utilisation gratuite ou payante des données, exploitation des résultats de recherches antérieures
- ✓ Open data : accélérer l'innovation et le retour sur investissement dans la R&D

## ❖ Enjeux scientifiques

- ✓ Data driven Science
- ✓ Economies
- ✓ Qualité de la recherche

## ❖ Enjeux sociétaux

- ✓ Participation des citoyens et de la société civile (transparence accrue du processus scientifique, production de données)



\_03

## L'exemple anglais

RCUK, BBSRC, Université / JISC et DCC



# les principes des RCUK

<http://www.rcuk.ac.uk/research/Pages/DataPolicy.aspx>

- ❖ Faciliter l'accès aux données issues de recherches financées sur fonds publics ;
- ❖ Adopter des standards et des normes, renseigner les métadonnées pour faciliter la réutilisation ;
- ❖ Prendre en compte les aspects légaux, éthiques et commerciaux à tous les stades ;
- ❖ Avoir un embargo pour conserver la priorité d'exploitation aux producteurs des données ;
- ❖ Les données doivent être citables, les utilisateurs doivent citer les producteurs.

# Politique du BBSRC

## Data Sharing Plans

BBSRC recognises that data sharing plans will vary according to the type of data collected. Data sharing should be driven by scientific benefit and should also be cost effective. Data should be shared using established standards and existing resources where this is possible. Applicants may wish to include details of:

- *Data areas and data types* - the volume, type and content of data that will be generated e.g. experimental measurements, models, records and images;
- *Standards and metadata* - the standards and methodologies that will be adopted for data collection and management, and why these have been selected;
- *Relationship to other data available in public repositories*;
- *Secondary use* - further intended and/or foreseeable research uses for the completed dataset(s);
- *Methods for data sharing* - planned mechanisms for making these data available, e.g. through deposition in existing public databases or on request, including access mechanisms where appropriate;
- *Proprietary data* - any restrictions on data sharing due to the need to protect proprietary or patentable data;
- *Timeframes* - timescales for public release of data;
- *Format of the final dataset*.

 DCC because good research needs good data

<https://dmponline.dcc.ac.uk/>

# Politique de l'Université de Manchester 1/3

1. All data created or owned by the University and its wholly subsidiary companies are the property of the University of Manchester and are regarded as corporate assets.
2. Research data will be managed to the highest standards throughout the research data lifecycle as part of the University's commitment to research excellence.
3. Responsibility for research data management, through a sound research data management plan, during any research project or programme lies primarily with Principal Investigators (PIs).
4. All new research proposals [from date of adoption] must include research data management plans that explicitly address data capture, management, integrity, confidentiality, retention, sharing, publication and state explicitly the availability for access and re-use (including safeguards).

# Politique de l'Université de Manchester 2/3

5. The University will support researchers by providing:
  - a. Mechanisms and services for storage, backup, registration, deposit and retention of research data assets in support of current and future access, during and after completion of research projects.
  - b. Guidelines and templates for research data management and research data management plans.
  - c. Training, support and advice on Research Data Management.
  - d. Help in identifying national/international data archives for the external storage of research data.
  - e. Professional support for researchers who receive FOI/EIR requests. It is obligatory for researcher to notify the University Records Manager if such a request is received.

# Politique de l'Université de Manchester 3/3

6. The legitimate interests of the subjects of research data must be protected.
7. Research data of future historical interest, and all research data that represent records of the University, including data that substantiate research findings, will be offered and assessed for deposit and retention in an appropriate national or international data service or domain repository, or a University repository.
8. Any data which is retained elsewhere, for example in an international data service or domain repository should be registered with the University.
9. Exclusive rights to reuse or publish research data should not be handed over to commercial publishers or agents without retaining the rights to make the data openly available for reuse, unless this is a condition of funding.

# Pourquoi le chercheur doit partager ses données

<http://www.data-archive.ac.uk/media/2894/managingsharing.pdf>

- ❖ encourage la recherche scientifique et le débat
- ❖ favorise l'innovation
- ❖ conduit à de nouvelles collaborations entre les utilisateurs de données et les créateurs de données
- ❖ maximise la transparence et la responsabilité
- ❖ permet l'examen des résultats de recherche
- ❖ encourage l'amélioration et la validation des méthodes de recherche
- ❖ réduit le coût de duplication des collection de données
- ❖ augmente l'impact et la visibilité de la recherche
- ❖ favorise la recherche qui a créé les données et les résultats
- ❖ Valorise le chercheur (produit de la recherche à part entière)
- ❖ fournit des ressources importantes pour l'enseignement et la formation



# En France

- ❖ Pas de politique au niveau de l'ANR
- ❖ Perspective «BSN 10»
- ❖ Mais il existe plusieurs infrastructures « disciplinaires », ex :
  - ✓ Astronomie (CDS de Strasbourg)
  - ✓ SHS : Adonis / Darjah (digital infrastructure)
  - ✓ SDV : réseaux de plateformes séquençage ...
  - ✓ Mer : Marine Knowledge 2020
  - ✓ ....



Mission d'archive numérique des données



\_02

# Données, data ???

Tentative de définition

# Seeking win + win + win + win + win.....

Where do I safely keep my data from my fieldwork, as I travel home?

How do we ensure we have access to our research data after some of the team have left?

How can our research collaborations share data, and make them available once complete?

How can I best keep years worth of research data secure and accessible for when I and others need to re-use it?

How do we ensure compliance to funders' requirement for several years of open access to data?

PhD student

individual researcher

research team

university

supra-university

LEVEL

Simon Hodson - Research Integrity, London - Sept 2011

# Définitions

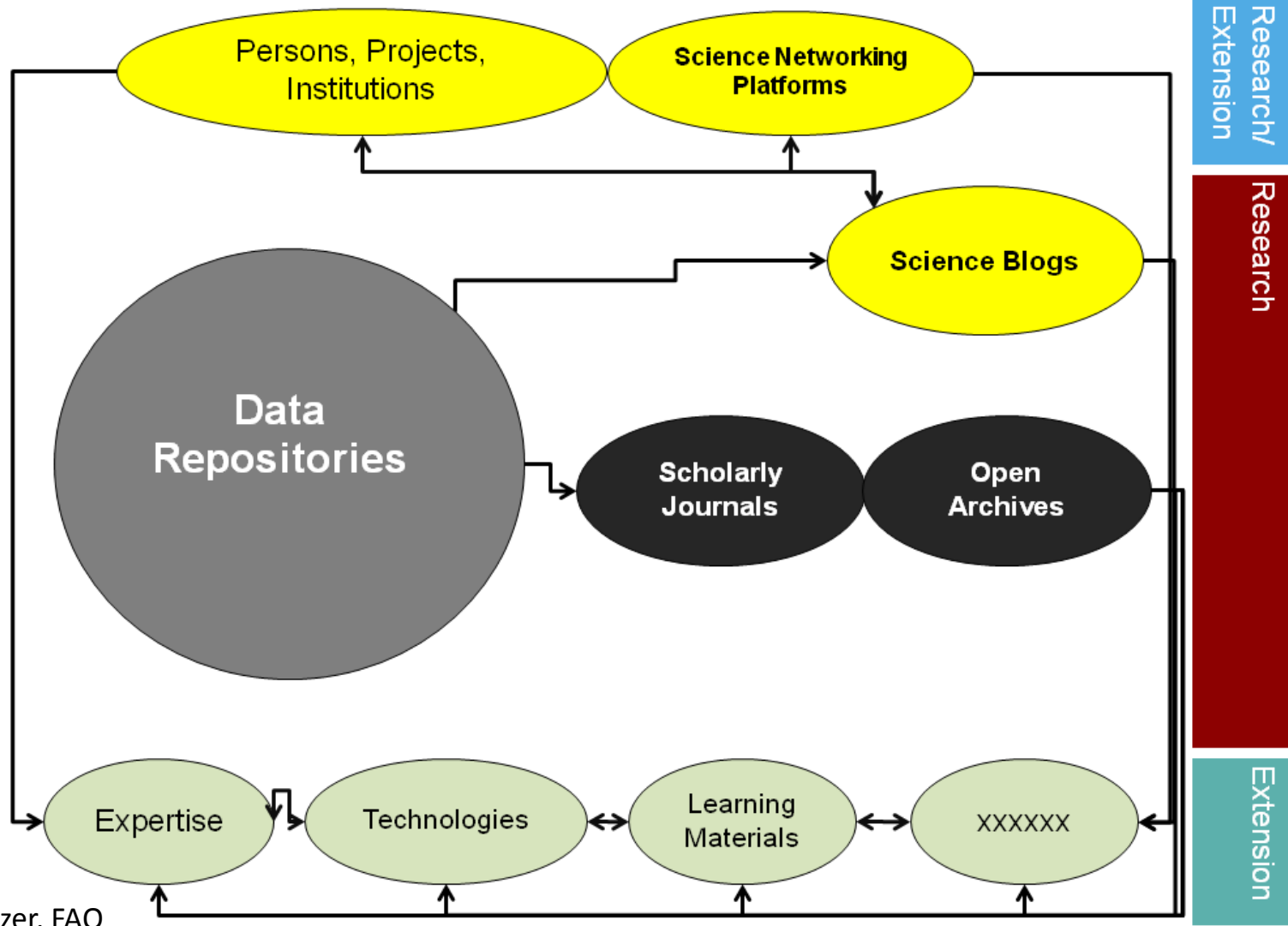
**Data:** digital recorded factual material commonly accepted in the scientific community as necessary to validate research findings

(not include lab notebooks, preliminary analysis, drafts of scientific papers, plans for future research, peer review reports, communication with peers, physical objects, lab specimens)

*[c.f. White House Memo on "Increasing Access to the Results of Federally Funded Scientific Research"]*

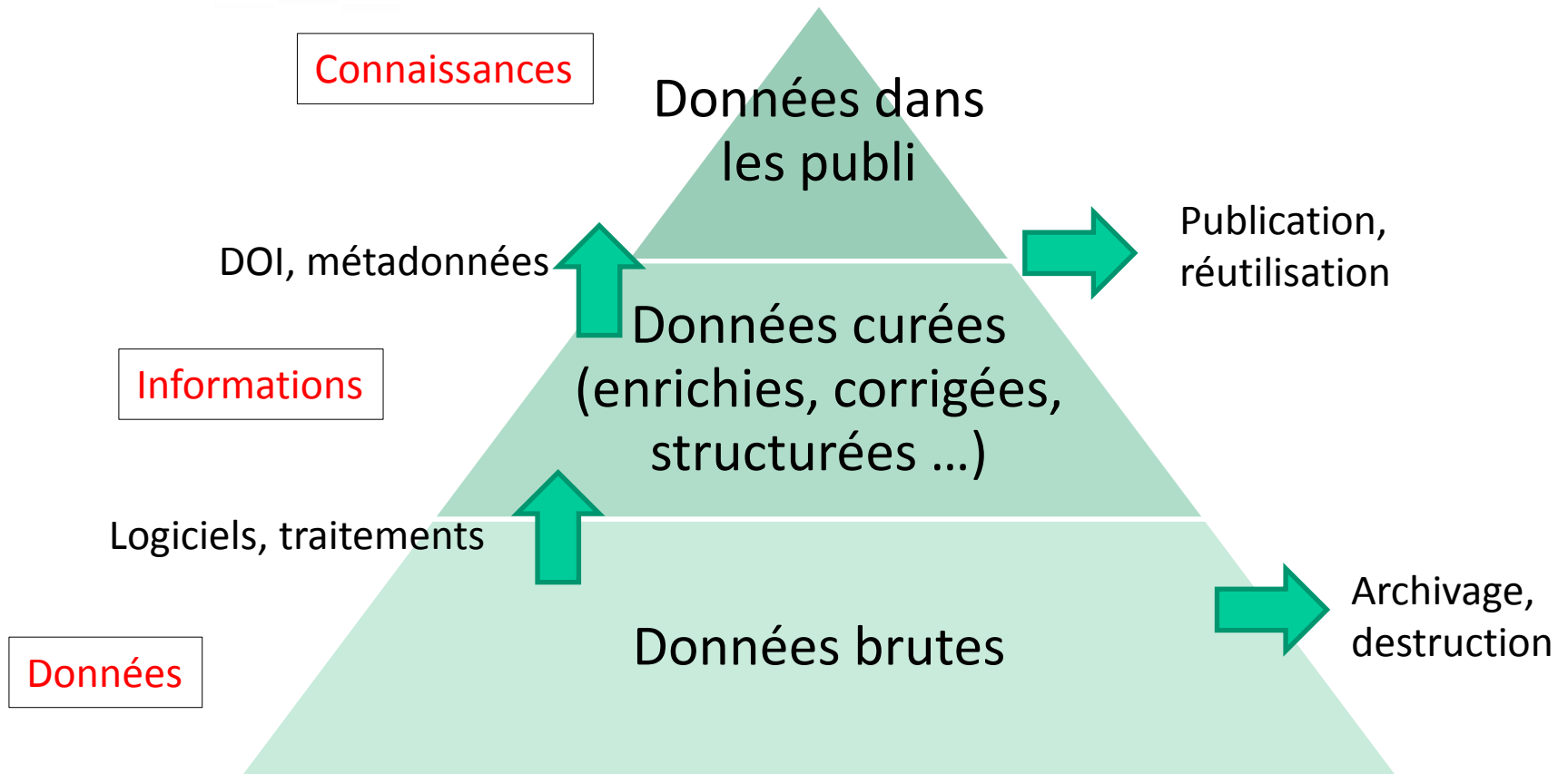
**Data infrastructures:** services, applications, tools, knowledge and policies for research data to be discoverable, understandable, accessible, preserved and curated... and available 24/7

# Dans le SI scientifique



Johannes Keizer, FAO

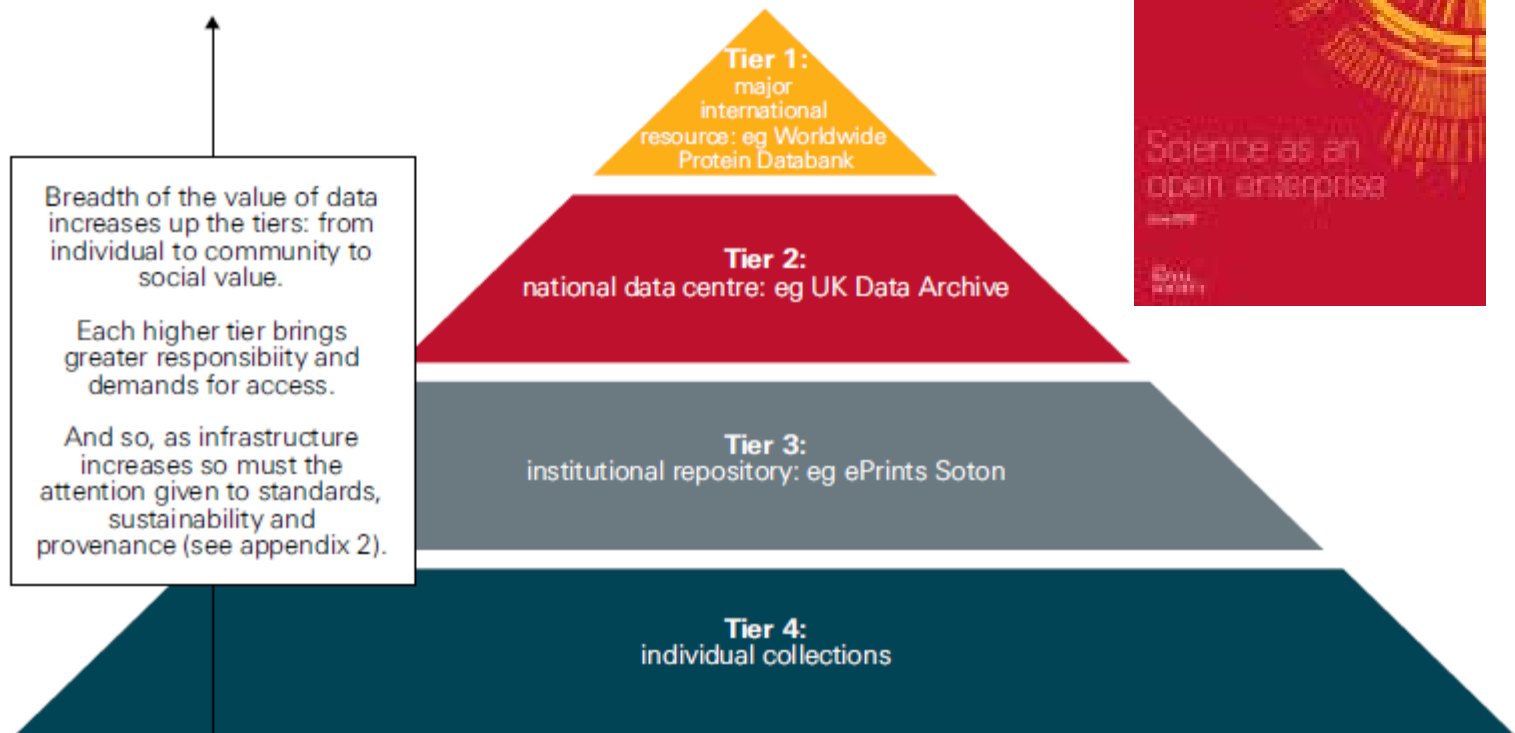
# Les états de la donnée – vision processus



# Data pyramid

## Box 4.1 The Data Pyramid – a hierarchy of rising value and permanence

Details of examples given in appendix 3.



<http://royalsociety.org/policy/projects/science-public-enterprise/report/>

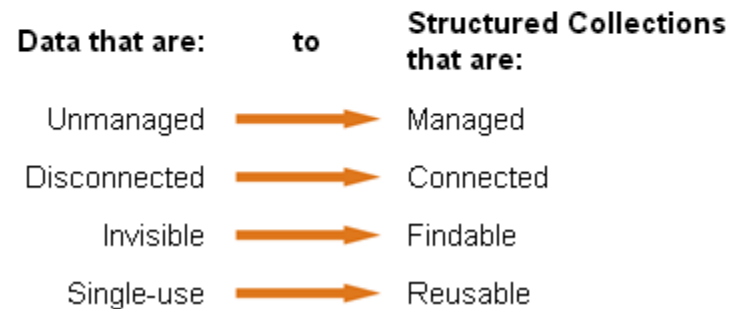
# Vision australienne ....



## Australian National Data Service

***Our Vision: More Australian researchers reusing research data more often***

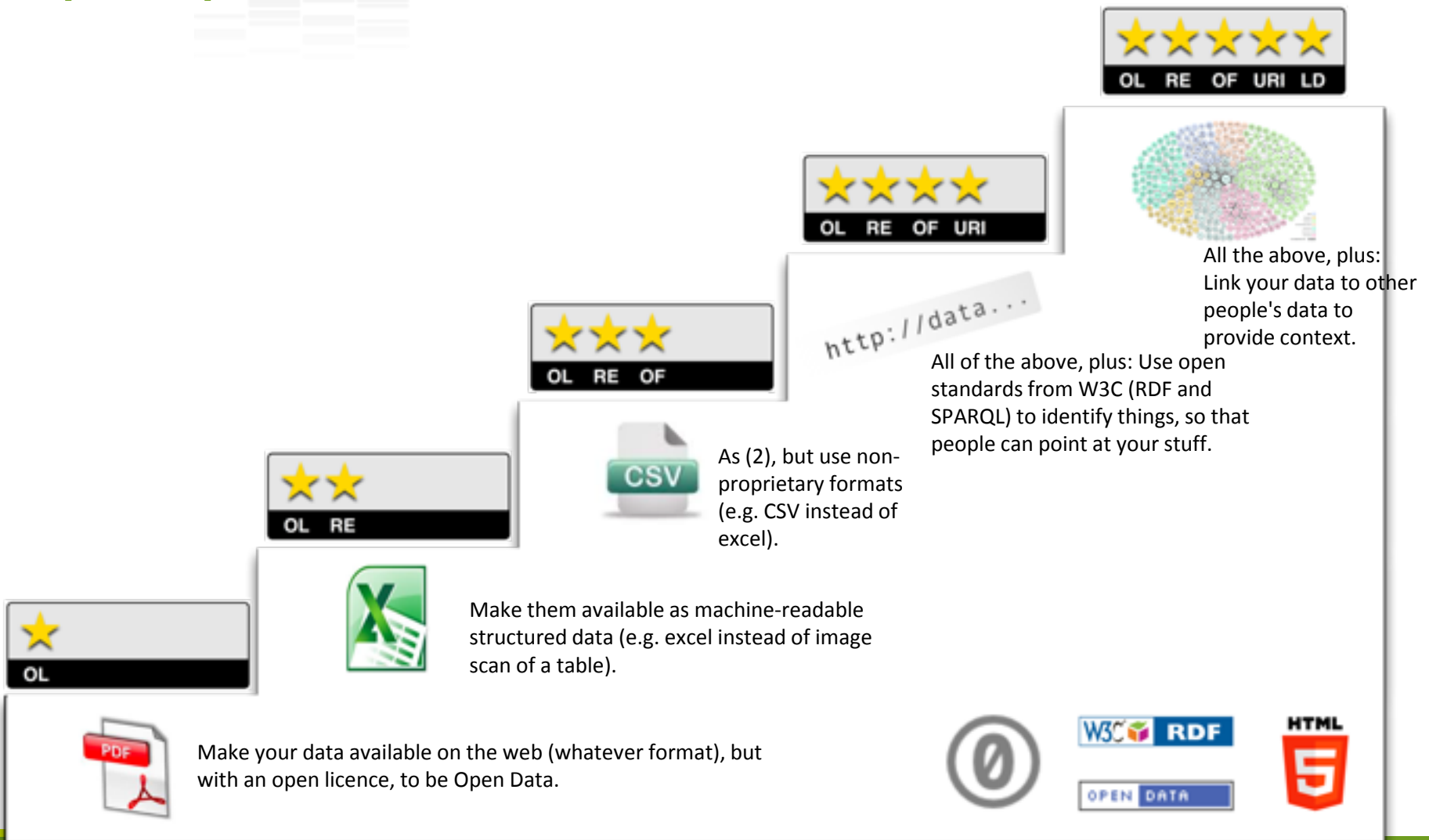
**ANDS is enabling the transformation of:**



[http://www.nhmrc.gov.au/files\\_nhmrc/publications/attachments/r39.pdf](http://www.nhmrc.gov.au/files_nhmrc/publications/attachments/r39.pdf)



# Tim Berners-Lee's Five Stars for Open Data (2009)



# Aspects juridiques des données : à préciser

- ❖ La donnée brute n'existe pas en droit : document administratif = propriété de l'institution, potentiellement diffusable
- ❖ Droit des bases de données : protection du producteur
- ❖ Droits d'auteur sur les données dans les publis
- ❖ licence sur les données (ex : CC)



LICENCE OUVERTE  
OPEN LICENCE

Vous pouvez réutiliser « l'Information » rendue disponible par le « Producteur » dans les libertés et les conditions prévues par la présente licence.

### La réutilisation de l'Information diffusée sous cette licence

Le « Producteur » garantit au « Réutilisateur » le droit personnel, non exclusif et gratuit, de réutilisation de « l'Information » soumise à la présente licence, dans le monde entier et pour une durée illimitée dans les libertés et les conditions exprimées ci-dessous.

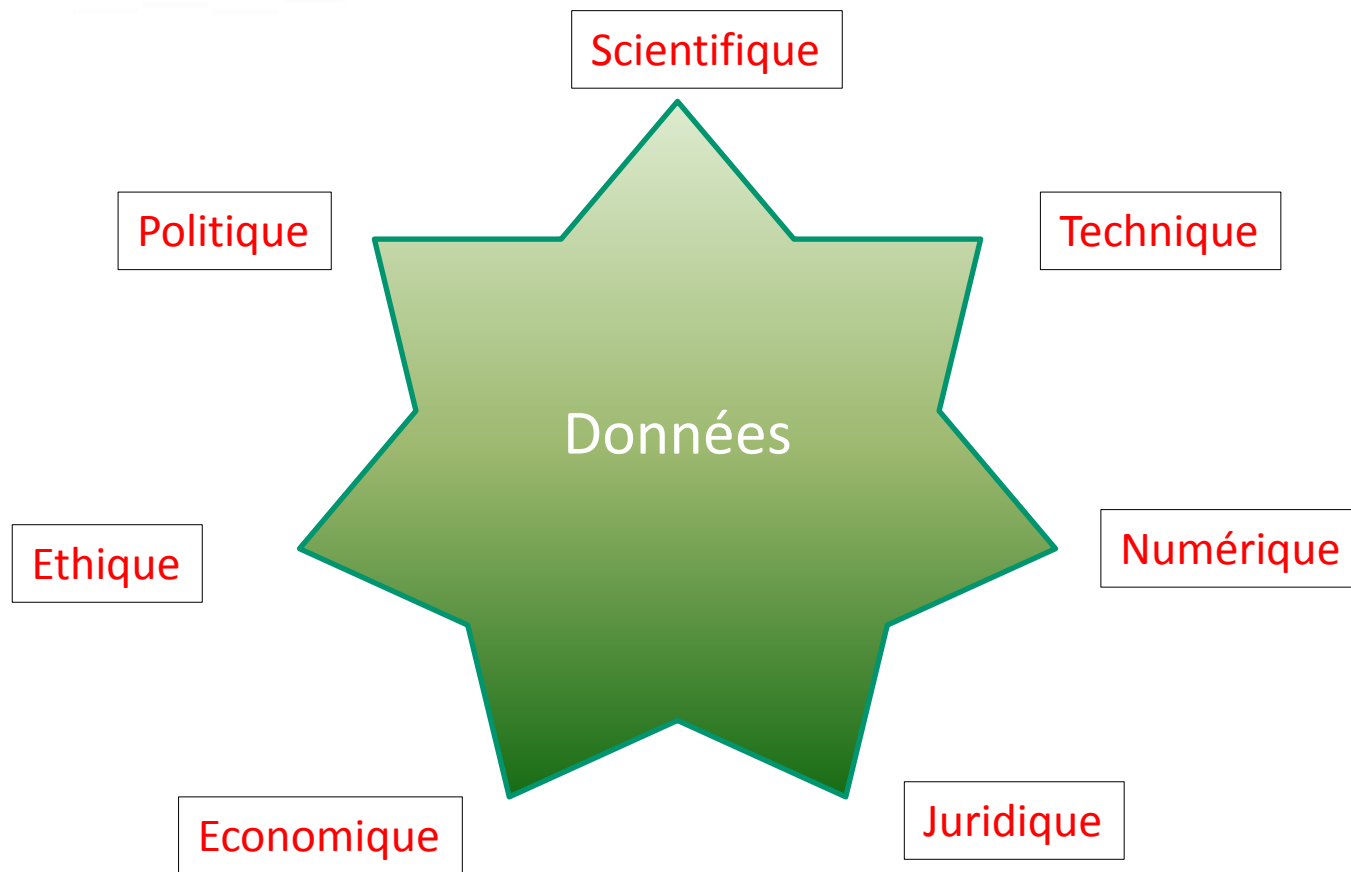
### Vous êtes libre de réutiliser « l'Information » :

- Reproduire, copier, publier et transmettre « l'Information » ;
- Diffuser et redistribuer « l'Information » ;
- Adapter, modifier, extraire et transformer à partir de « l'Information », notamment pour créer des « Informations dérivées » ;
- Exploiter « l'Information » à titre commercial, par exemple en la combinant avec d'autres « Informations », ou en l'incluant dans votre propre produit ou application.

### Sous réserve de :

- Mentionner la paternité de « l'Information » : sa source (*a minima* le nom du « Producteur ») et la date de sa dernière mise à jour.

# Conclusion : un objet complexe





# Editeurs, Données, publications et évaluation

# En lien avec les publications

## ❖ Des nouvelles revues :



BGI Shenzhen and  
BioMed Central

GigaScience aims to revolutionize data dissemination, organization, understanding, and use. An online open-access open-data journal, **we publish 'big-data' studies from the entire spectrum of life and biomedical sciences.** To achieve our goals, the journal has a novel publication format: one that **links standard manuscript publication with an extensive database that hosts all associated data and provides data analysis tools and cloud-computing resources.**

**Not just 'omic' type data** ... imaging, neuroscience, ecology, cohort data, systems biology and other new types of large-scale sharable data.

## ❖ Note aux auteurs – revues classiques

The logo for the journal Science, consisting of the word "Science" in white text on a red rectangular background.

**Data and materials availability** All data necessary to understand, assess, and extend the conclusions of the manuscript must be available to any reader of *Science*. ]...[ *Science* supports the efforts of databases that aggregate published data for the use of the scientific community. Therefore, appropriate data sets (including microarray data, protein or DNA sequences, atomic coordinates or electron microscopy maps for macromolecular structures, and climate data) must be deposited in an approved database, and an accession number or a specific access address must be included in the published paper. We encourage compliance with MIBBI guidelines (Minimum Information for Biological and Biomedical Investigations).

## ❖ De nouveaux entrepôts de données



**Dryad is an international repository of data** underlying peer-reviewed articles in the basic and applied biosciences. Dryad enables scientists to validate published findings, explore new analysis methodologies, repurpose data for research questions unanticipated by the original authors, and perform synthetic studies. Dryad is governed by a consortium of journals that collaboratively promote data archiving and ensure the sustainability of the repository.

## Respectful re-use

*Nature Genetics* **44**, 1073 (2012) | doi:10.1038/ng.2433

Published online 26 September 2012

**The impact of the papers we publish depends increasingly on the data they describe. In insisting on data access for referees and readers, we prioritize scientific integrity above all and place the interests of research participants before impact.**

The Nature research journals insist that materials, methods and data be made available and that authors detail any conditions for access where these exist. In our shared guide to authors (<http://www.nature.com/authors/policies/availability.html>), we state:

*The preferred way to share large data sets is via public repositories. Some of these repositories offer authors the option to host data associated with a manuscript confidentially, and provide anonymous*



[print](#)



[email](#)



[pdf](#)



[download citation](#)



[order reprints](#)



[rights and permissions](#)

# Lien avec les citations des articles

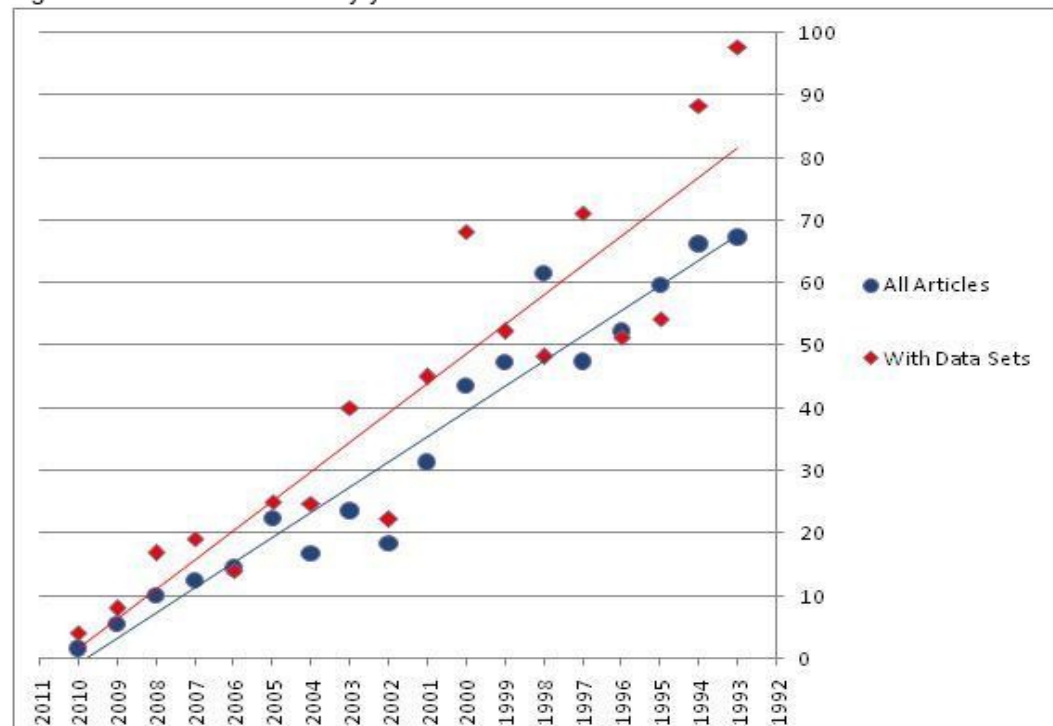
## Data Sharing Effect on Article Citation Rate in Paleoceanography

posted Nov 27, 2011 4:11 AM by Michael Diepenbroek [ updated Nov 27, 2011 11:21 PM ]

Jon Sears (AGU) - Abstract for the AGU 2011:

The validation of scientific results requires reproducible methods and data. Often, however, data sets supporti PANGAEA® data library measurably increases the citation rate of articles published between 1993 and 2010 i available supporting data sets received 19.94% (8,056) of the aggregate citations (40,409). Publicly available c sampled over the 18-year study period (1,331), and the increase is fairly consistent over time (14 of 18 years).

Figure: Plot of mean citations by year to articles with data sets vs. citations to all articles in Paleoceanography betw

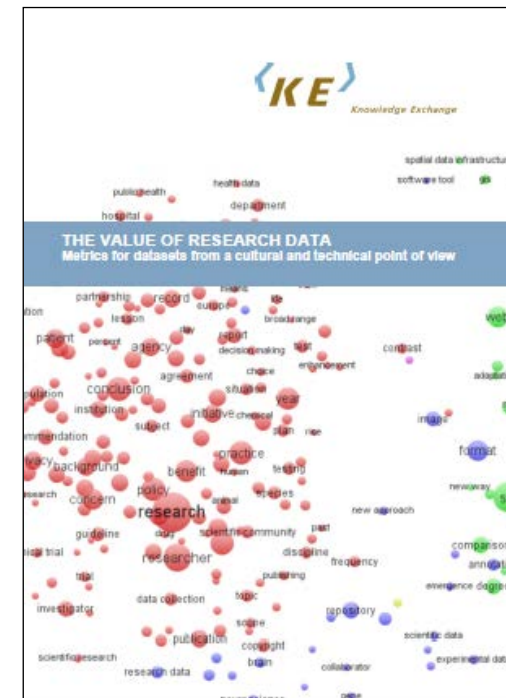




# Evaluer les données

❖ The value of research data – rapport du Knowledge Exchange, avril 2013

<http://www.knowledge-exchange.info/datametrics>



# Building a Culture of Data Citation




# THE DATA CITATION INDEX<sup>SM</sup>

## DEFINITIONS:

**Data repository:** a database or collection comprising data studies, data sets and/or microcitations which stores and provides access to the raw data. Constituent data studies, and sometimes individual data sets, are marked up with metadata providing a context for the available raw data.

**Data study:** description of studies or experiments held in repositories with the associated data which have been used in the data study. (Includes serial or longitudinal studies over time). Data studies can be a citable object in the literature and may have cited references attached in their metadata, together with information on such aspects as the principal investigators, funding information, subject terms, geographic coverage etc. The level of metadata provided varies between repositories.


**Data set:** a single or coherent set of data or a data file provided by the repository, as part of a collection, data study or experiment. Data sets may present in a number of file formats and media types: they may be number based files such as spreadsheets, images, video, audio, databases etc. Data sets can be a citable object in the literature and may have cited references attached in their metadata, but more commonly they inherit the metadata of the overall study in which they are used.



- 1. Title: **ESTerases and alpha/beta Hydrolase Enzymes and Relatives.**  
Editor(s): Hotelier, Thierry; Renault, Ludovic; Cousin, Xavier; et al.  
Source: ESTerases and alpha/beta Hydrolase Enzymes and Relatives  
Source URL: <http://bioweb.ensam.inra.fr/ESTHER/general?what=index>  
Document Type: **Repository** Times Cited: **2** (from All Databases)  
[  [View abstract](#) ]



- 1. Title: **Enzymatic Activity and Protein Interactions in Alpha/Beta Hydrolase Fold Proteins: Moonlighting Versus Promiscuity**  
Author(s): Marchot, Pascale; Chatonnet, Arnaud  
Source: PROTEIN AND PEPTIDE LETTERS Volume: **19** Issue: **2** Pages: **132-143** Published: **FEB 2012**  
Times Cited: **3** (from All Databases)  
 [  [View abstract](#) ]
- 2. Title: **ESTHER, the database of the alpha/beta-hydrolase fold superfamily of proteins**  
Author(s): Hotelier, T; Renault, L; Cousin, X; et al.  
Source: NUCLEIC ACIDS RESEARCH Volume: **32** Special Issue: **SI** Pages: **D145-D147** DOI: **10.1093/nar/gkh141** Published: **JAN 1 2004**  
Times Cited: **79** (from All Databases)  
  [  [View abstract](#) ]


# Figshare – Digital Science (MacMillan/Nature)



My data   [Browse](#) [Upload](#)  **O. Hologne** ▼

My data 0% of private storage used

[Add to Fileset](#) [Batch edit](#)

	Type <small>mouseover</small>	Date	Status	Statistics <small>public items only</small>
<input type="checkbox"/> esi18012012.xlsx	 DATASET	25.01.2012 13:31	<b>DRAFT</b>	<a href="#">Add info</a>

showing all 1 items

## Helping you publish, discover, and reuse research data



### Welcome to *Scientific Data*

*Scientific Data* is a new open-access, online-only publication for descriptions of scientifically valuable datasets. It introduces a new type of content called the Data Descriptor, which will combine traditional narrative content with curated, structured descriptions of research data, including detailed methods and technical analyses supporting data quality. *Scientific Data* will initially focus on the life, biomedical and environmental science communities, but will be open to content from a wide range of scientific disciplines. Publications will be complementary to both traditional research journals and data repositories, and will be designed to foster data sharing and reuse, and ultimately to accelerate scientific discovery.

*Scientific Data* will launch in Spring 2014 and be open for submissions in Autumn 2013. Sign up for our



\_04

# Apport des professionnels IST

# Nouveaux challenges pour l'IST

NATURE | NEWS FEATURE

## Publishing frontiers: The library reboot

As scientific publishing moves to embrace open data, libraries and researchers are trying to keep up.

Richard Monastersky

27 March 2013

### The new data wranglers

... At Johns Hopkins and many other top universities, libraries are aiming to become **more active partners in the research enterprise** — altering the way scientists conduct and publish their work. **Libraries are looking to assist with all stages of research, by offering guidance and tools for collecting, exploring, visualizing, labelling and sharing data.** “I see us moving up the food chain and being **co-contributors to the creation of new knowledge,**” says Sarah Thomas, the head of libraries

## THE DATA GAP

A survey of more than 1,300 US scientists in 2010 showed an appetite for sharing data but significant hurdles that kept many from doing it.

67%

Said lack of access to others' data is a major impediment to scientific progress.

39%

Said their organization has an established process for long-term data storage.

22%

Said their organization or project provides funds for long-term data management.

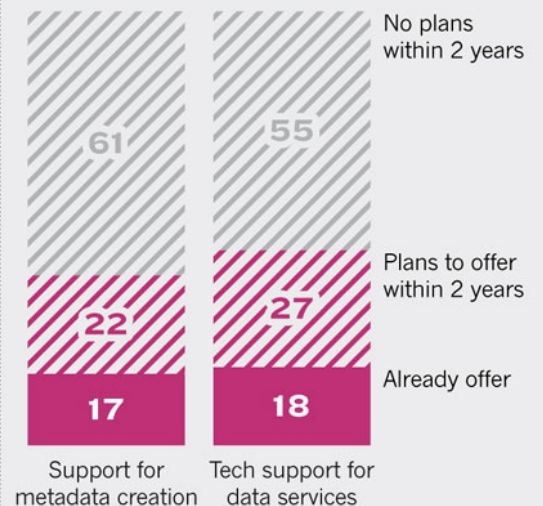


Scientists who said that they would use other researchers' data sets if they were easily available.



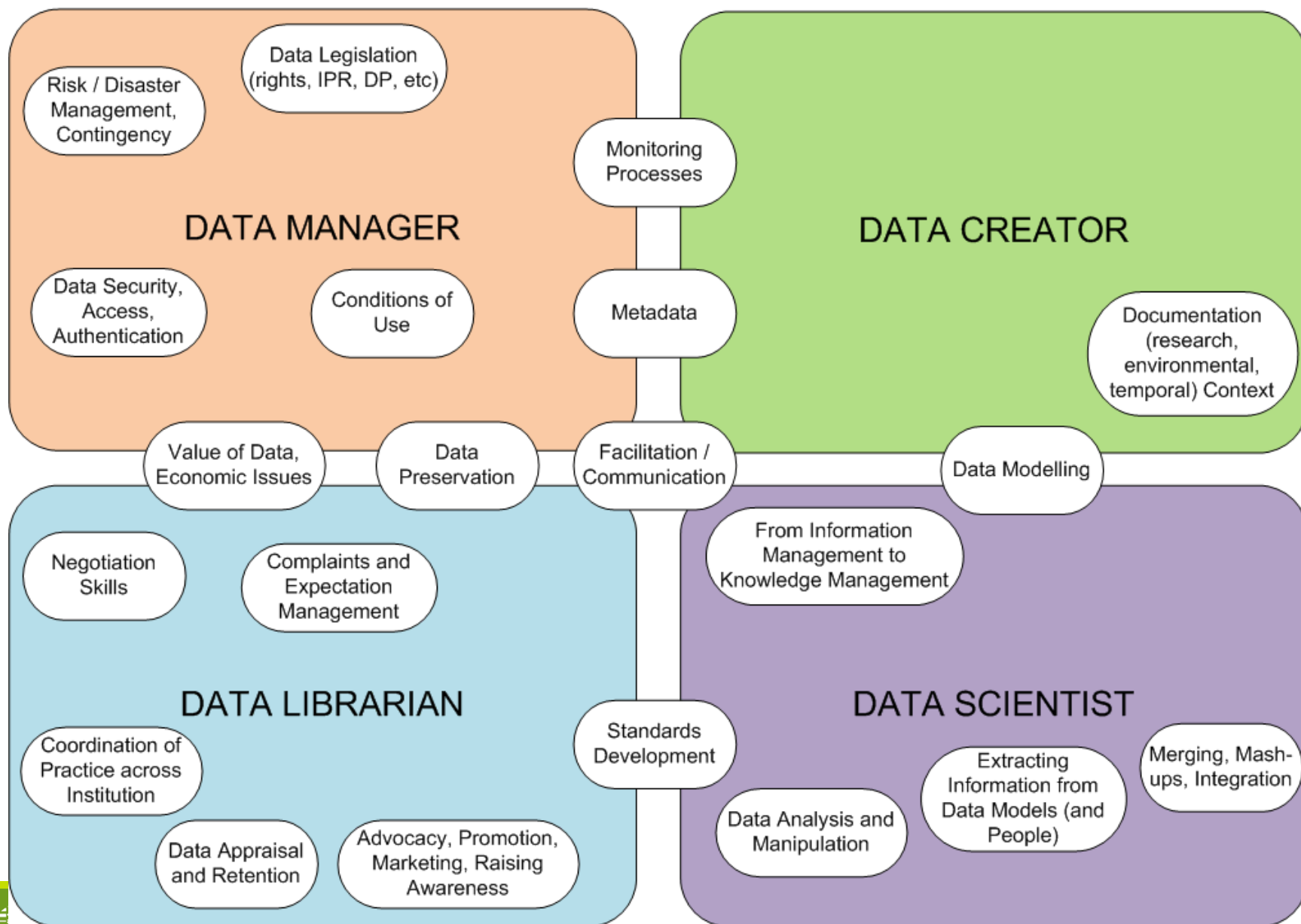
Scientists who said that others can access their data easily.

A survey of US research universities showed that many libraries are gearing up to provide data services.



# CORE SKILLS FOR DATA MANAGEMENT

A follow-up from the second DCC Research Data Management Forum (November 2008)





# Data Scientist : the sexiest job ...

<http://hbr.org/2012/10/data-scientist-the-sexiest-job-of-the-21st-century/ar/1>

## HIGH-PERFORMANCE ANALYTICS

Decisions at the speed of right.

LEARN MORE ABOUT HIGH-PERFORMANCE ANALYTICS.

**sas** | THE POWER TO KNOW.



• SUBSCRIBE US/CANADA  
• SUBSCRIBE OUTSIDE US  
• RENEW

Subscribe Now >>



SEARCH

FOLLOW HBR:   
DIGITAL & MOBILE

THE MAGAZINE

BLOGS

AUDIO & VIDEO

BOOKS

WEBINARS

COURSES

STORE

Guest | limited access

Register today and save 20%\* off your first order! [Details](#)

[Subscribe](#)

[Sign in / Register](#)

[My Account](#)



THE MAGAZINE

October 2012

[Buy Reprint >>](#)

## Data Scientist: The Sexiest Job of the 21st Century

by Thomas H. Davenport and D.J. Patil

Comments (35)



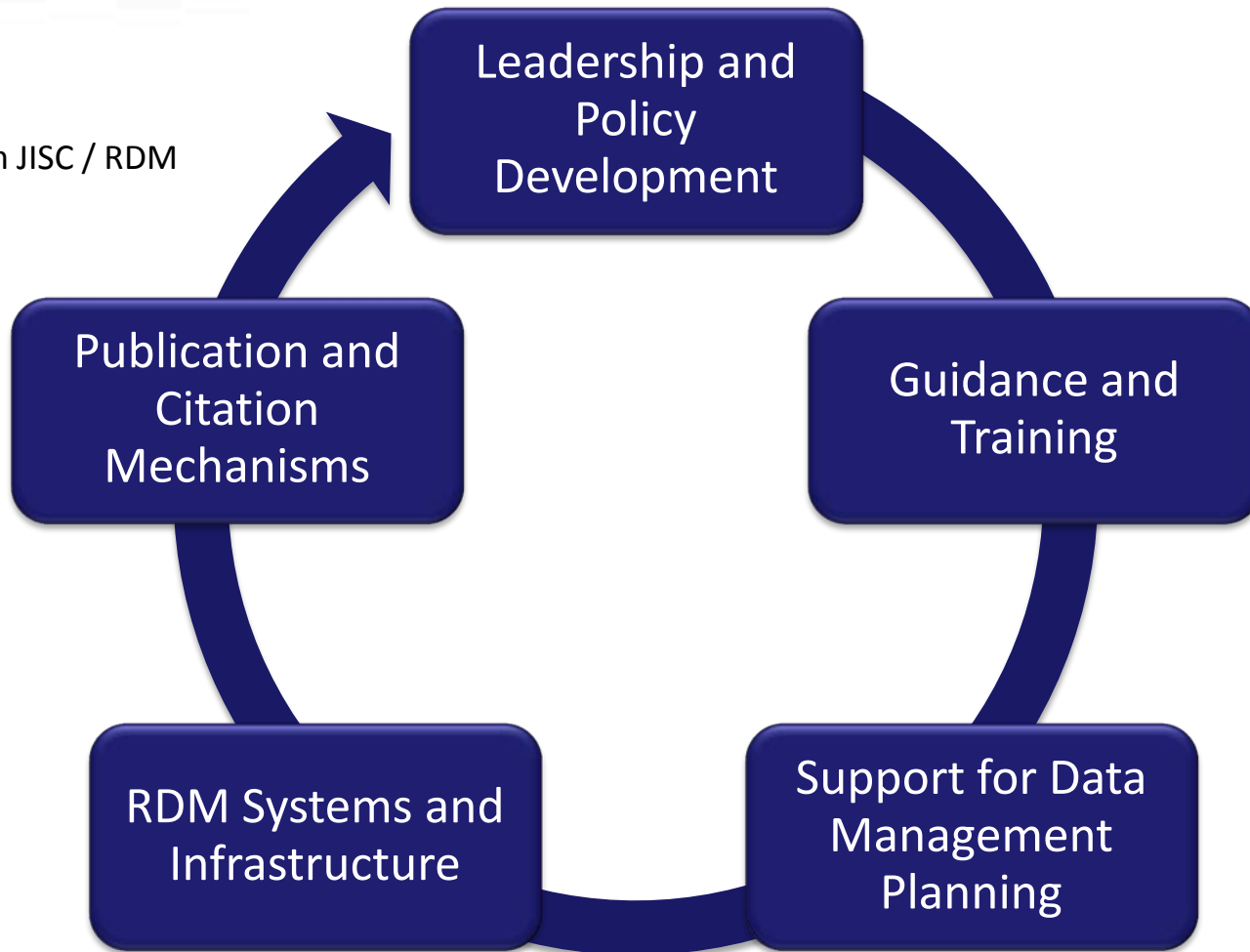
ALL NEW  
Print, Digital & iPad  
Edition

SUBSCRIBE TODAY →

## HIGH-PERFORMANCE ANALYTICS

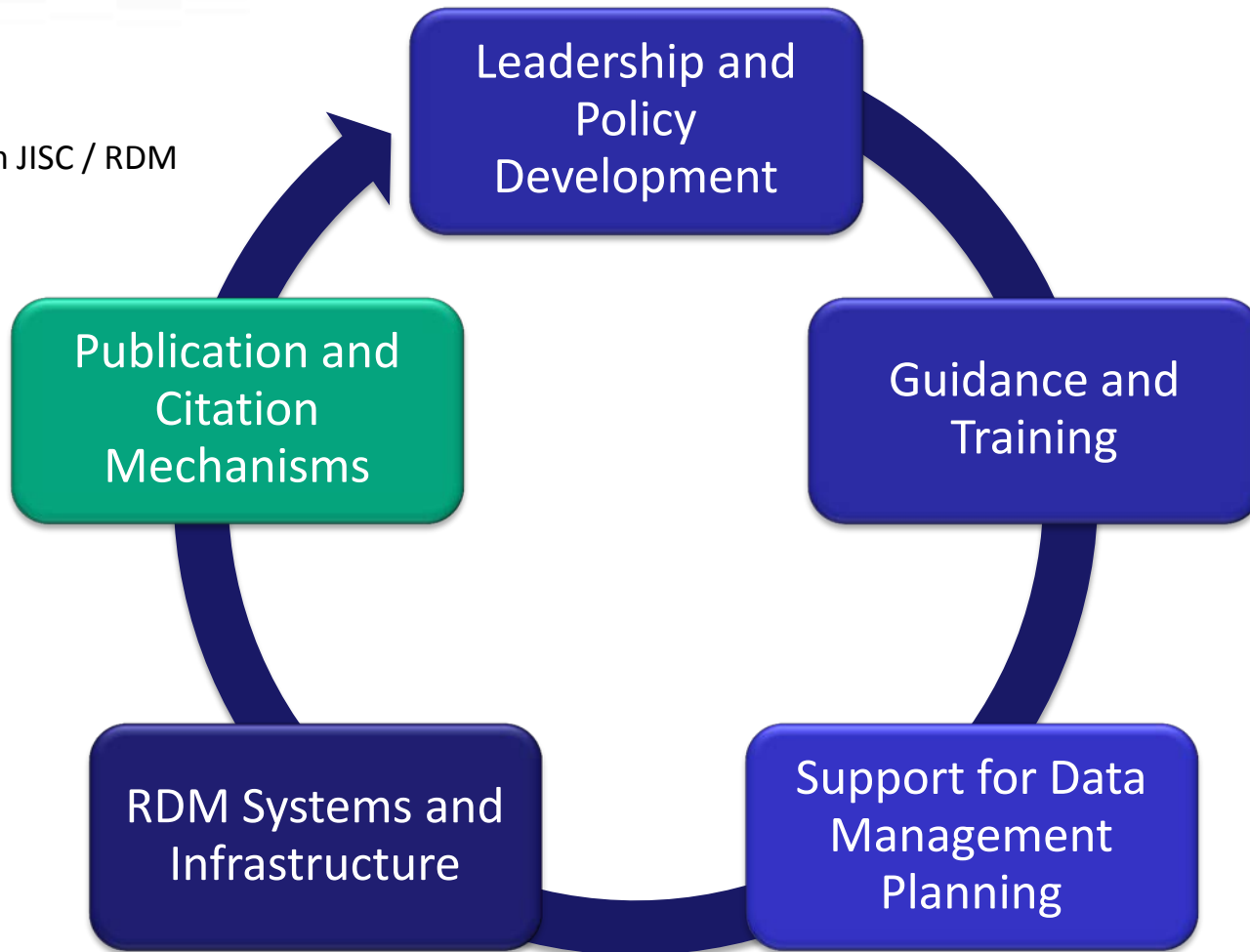
# Différents domaines

Source : S. Hodson JISC / RDM  
Data life cycle



# Contribution possible de l'IST

Source : S. Hodson JISC / RDM



# Domaines de compétences de l'IST

	Données brutes	Données traitées	Données de publication	Publications
Choisir les Métadonnées	DC ? Autres modèles ? <u>MIBBI</u> ?		DC	Maitrise des modèles de description : DC et autres liés aux publi
Décrire l'objet numérique	Non	Non	Oui, En référence à la publication	oui
Maintenir les Référentiels	?			oui
Plateformes logiciel	Ex : CKAN, Figshare Dspace, eprint			Dspace, Eprint Autres ...
Exemples	<a href="http://www.ckan.org">www.ckan.org</a>	<a href="http://www.figshare.com">www.figshare.com</a> <a href="http://www.pan-gaea.de">http://www.pan-gaea.de</a>	<a href="http://datadryad.org/">http://datadryad.org/</a>	<a href="http://www.prodinra.inra.fr">www.prodinra.inra.fr</a> <a href="https://w3.avignon.inra.fr/dspace/">https://w3.avignon.inra.fr/dspace/</a>

# Compétences/connaissances à avoir

## ❖ Traditionnelles :

- ✓ compréhension des Data journals
- ✓ connaissance des entrepôts externes
- ✓ Mécanismes de citation et d'évaluation
- ✓ Connaissances juridiques (CC, PI, BdD)
- ✓ Maintenance de référentiels

## ❖ Technologiques

- ✓ Identifiants et métadonnées
- ✓ Interopérabilité (OAI-PMH et autres)
- ✓ Linked Open Data, Ontologies et standards associés

# Résumé : Apport de l'IST

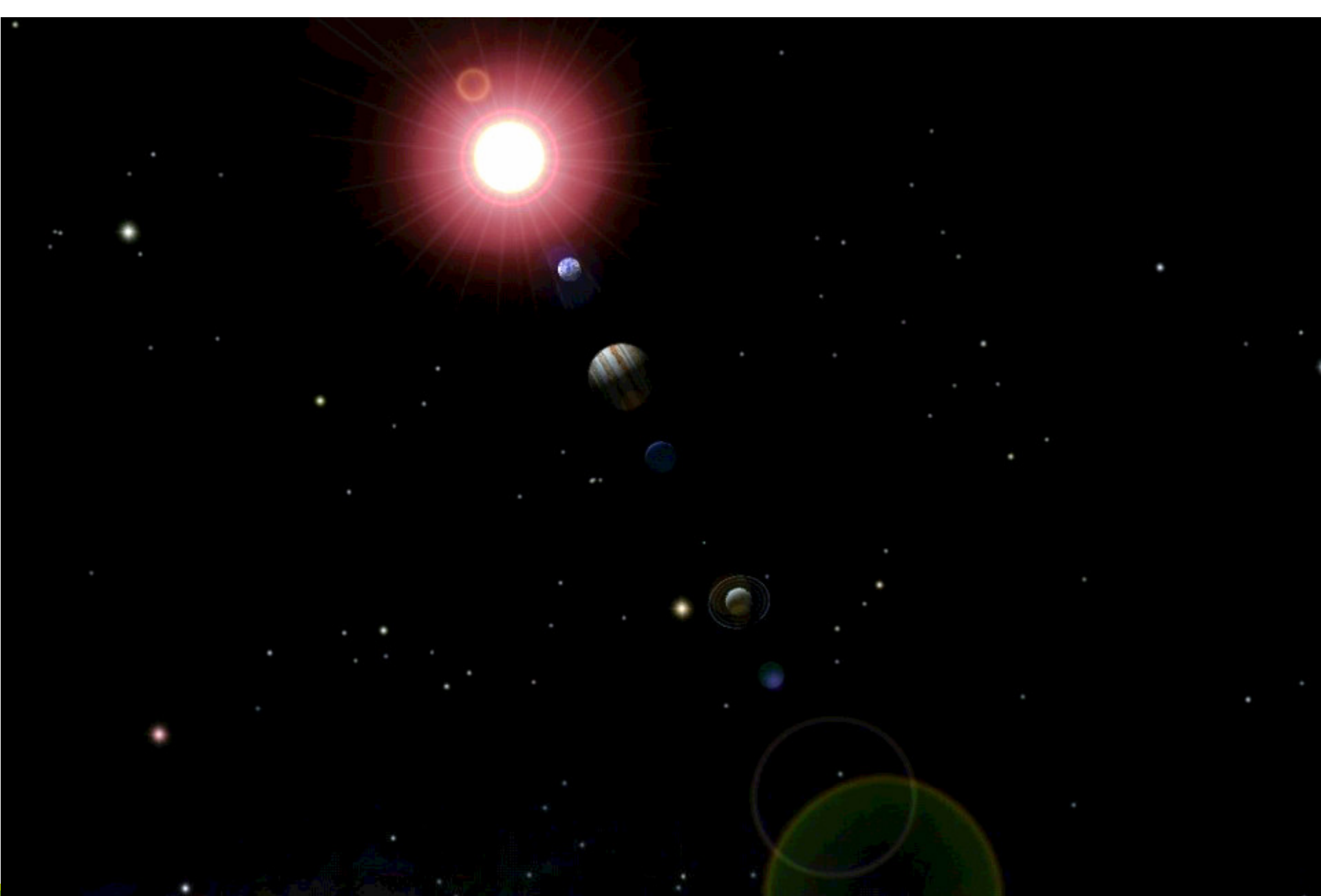
- ❖ Contribuer à l'émergence d'une politique institutionnelle
- ❖ Recenser et valoriser les sources de données
- ❖ Aider à définir les stratégies de dépôt
- ❖ Gestion des DOI et métadonnées associées
- ❖ Sensibiliser les scientifiques à la valorisation des données
- ❖ Etre le relais des bonnes pratiques



\_05

... A L'Inra

Politique de partage des données



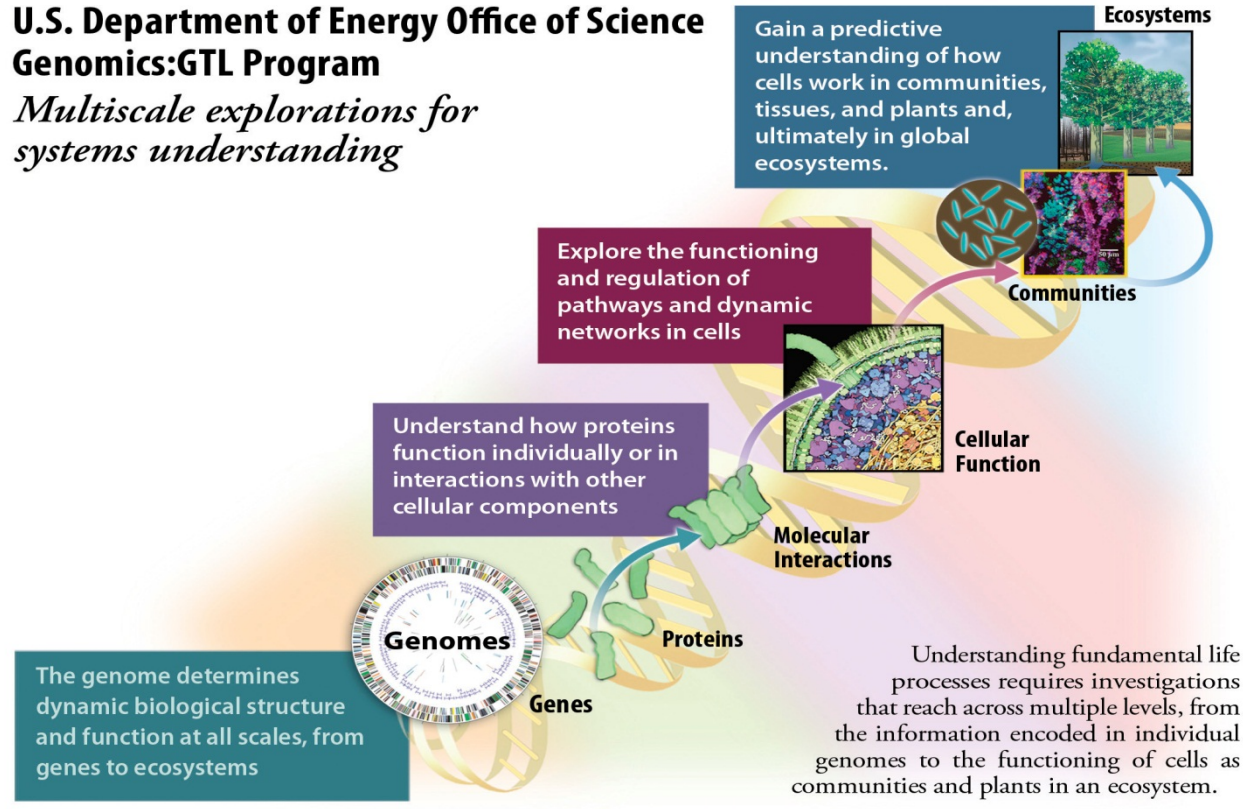


# Nous vivons une révolution technologique

## ❖ Une vision (presque) sans limite des possibles!

Rapprochement des thématiques dans un objectif de vision globale  
=> Vers la biologie systémique

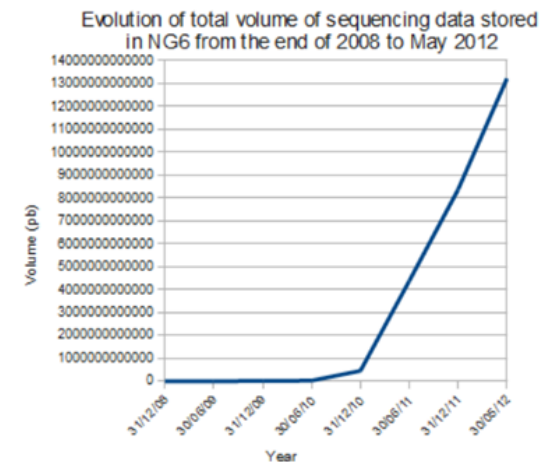
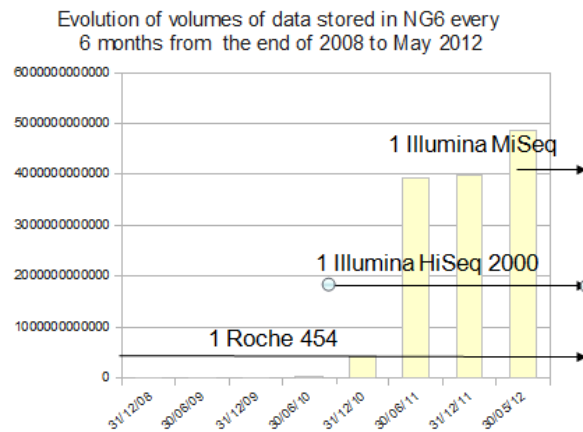
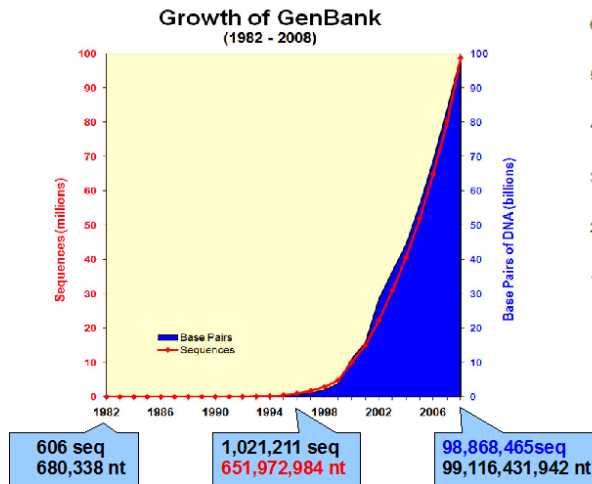
**U.S. Department of Energy Office of Science  
Genomics:GTL Program**  
*Multiscale explorations for  
systems understanding*



# Sept. 2009: Demande au conseil scientifique « Gestion et partage des données »



- ❖ Dans le domaine de la biologie/bioinformatique
- ❖ Dans un contexte de production massive des données



# 2011-2012: le CS instruit la question « Gestion et partage des données »



❖ Groupe de travail piloté par D. Pontier

❖ Objectifs

- ✓ Regard scientifique sur les spécificités thématiques en matière d'acquisition de données, de besoins en matière de gestion et de partage à l'Inra
- ✓ Regard sur les changements induits par les mouvements du Big Data et de l'Open Data
- ✓ Analyse de quelques cadres nationaux et internationaux existants

# 2011-2012: le CS instruit la question « Gestion et partage des données »

## ❖ Conclusions (juin 2012): 9 recommandations

- ✓ R1 - Définir la politique de l'établissement et la communiquer
- ✓ R2 - Mettre en place un dispositif d'évaluation des données produites par l'Inra
- ✓ R3 - S'impliquer dans les comités internationaux de standardisation
- ✓ R4 - Développer un portail d'accès à un ensemble de ressources distribuées
- ✓ R5 - Prendre en compte le cycle de vie des données dès l'élaboration des projets de recherche

# 2011-2012: le CS instruit la question « Gestion et partage des données »

- ❖ Conclusions (juin 2012): 9 recommandations
  - ✓ R6 - Définir un cahier des charges pour les plateformes
  - ✓ R7 - Doter l'Inra d'infrastructures dimensionnées pour les stockages et calculs hautes performances
  - ✓ R8 - S'engager dans une politique de gestion des compétences répondant aux besoins en émergence
  - ✓ R9 - Conduire une réflexion inter-organismes
- ❖ Validation des recommandations par le CS en mai 2012

# Comment bâtir une politique ?

- ❖ Démarche « ingénieur » : réfléchir, groupe de travail → aboutir à la politique
- ❖ Démarche stratégique : affirmer la politique → mise en œuvre, groupes de travail

# Elaboration d'une politique

- ❖ 11 principes pour une **politique de partage des données de la recherche**
- ❖ Un projet inscrit au schéma directeur des systèmes d'information
  - ✓ création de data centers

# 15-17 Avril 2013 : Lancement du projet

- ❖ Séminaire avec : scientifiques, IST, Informatique
- ❖ Objectifs : Appropriation des principes et organisation du projet
- ❖ Résultats : création de groupes de travail :
  - ✓ Données par grands domaines
  - ✓ Méthodes et outils
  - ✓ Stratégie : juridique, communication
  - ✓ Compétences



# Conclusion

- ❖ Données : des objets complexes
- ❖ De nombreux enjeux : scientifiques, économiques ....
- ❖ Nécessité d'approches :
  - ✓ Top down : politique
  - ✓ Bottom up : appropriation par les scientifiques
- ❖ Des compétences à consolider pour les scientifiques, informaticiens et IST

# Lectures et sites intéressants



Researche Data Alliance <http://rd-alliance.org/>

Science as an open entreprise <http://royalsociety.org/policy/projects/science-public-enterprise/report/>

JISC <http://www.jisc.ac.uk/whatwedo/programmes/mrd.aspx>

Digital curation center

<http://www.dcc.ac.uk/>

<http://www.dcc.ac.uk/training/rdm-librarians> (formation pour les « Librarian »)

<http://www.dcc.ac.uk/events/idcc13/programme-presentations> Conf #idcc13

Australian national data service <http://www.ands.org.au/>



Merci pour votre attention...

[odile.hologne@versailles.inra.fr](mailto:odile.hologne@versailles.inra.fr)

@Holo\_08

