

INRA
SCIENCE & IMPACT

OPEN SCIENCE

CATI CODEX – MONTPELLIER 14 FÉVRIER 2013



Open access

Déclaration de Berlin 2004

promotion du libre accès à l'information scientifique et technique: publications

- Voie « verte » du libre accès (dépôt des publications par les chercheurs),
- Voie « gold » (publication payante dans des revues diffusées gratuitement)



Big data open data

❖ Information
données interprétées

❖ Open data s'intéresse aussi
aux données brutes ou élaborées



Croissance des informations

Calling
International
Rescue:
knowledge lost in
literature and
datalandslide!

Teresa K. ATTWOOD, Douglas B. KELL, Philip
MCDERMOTT, James MARSH, Steve R. PETTIFER and
David THORNE
Biochem. J. (2009) 424, 317–333

doi:10.1042/BJ20091474

A lire avec UTOPIA <http://getutopia.com>

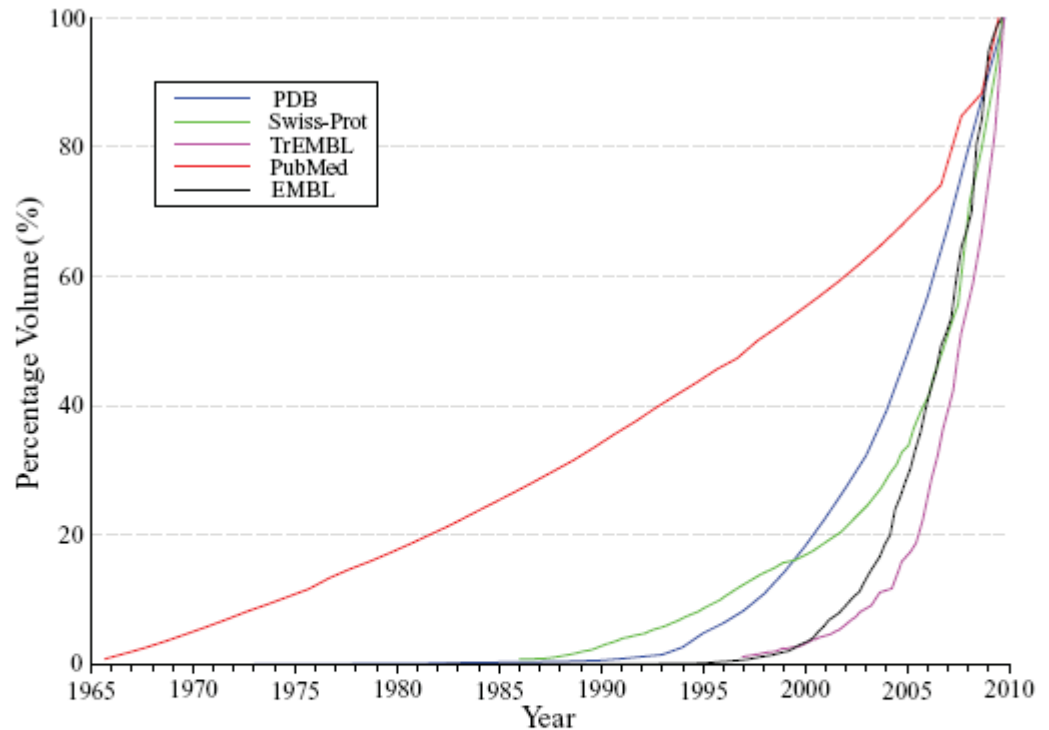


Figure 1 Graphical illustration of the growth of biomedical research publications (red; current total >19 million), alongside the accumulation of research data, including nucleic acid sequences (black; current total ~163 million), computer-annotated protein sequences (magenta; current total 9 million), manually annotated protein sequences (green; current total 500 000) and protein structures (blue; current total 60 000)

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 dans les 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).

Des 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.

Data citation index

WEB OF KNOWLEDGESM



ABOUT | PRODUCTS & TOOLS | BENEFITS & RESOURCES | TRAINING & SUPPORT | NEWS & EVENTS | CONTACT US

Site Search

SEARCH

Products and Tools > Multidisciplinary > Data Citation Index

THE DATA CITATION INDEXSM

CONNECTING THE DATA TO THE RESEARCH IT INFORMS

What is it?

VIEW VIDEO



Data repository: a database or collection comprising data studies, data sets

Data study: description of studies or experiments held in repositories with the associated data which have been used in the data study.

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.

Open data au quotidien



PARISDATA

MAIRIE DE PARIS 



Logements sociaux financés à Paris ▶

22/11/2012

csv

0.533 Mo

téléchargement(s) : 248



Panneaux d'affichage associat

13/11/2012

csv

0.017 Mo

téléchargement(s) : 117



Les références les plus réservées en 20...

25/10/2012

csv

Mo



téléchargement(s) : 230



1000 titres les plus empruntés

25/10/2012

csv

Mo



téléchargement(s) : 1278

[lire les avis\(2 \)](#)



Les références les plus affichées en 20...

25/10/2012

csv

Mo



téléchargement(s) : 122

[lire les avis\(1 \)](#)



1000 références les plus emp

25/10/2012

csv

Mo



téléchargement(s) : 191

[lire les avis\(1 \)](#)

Enjeux du partage de données

Open science

Accessibilité des données dans un objectif de transparence et de preuve

Réutilisation des données dans un objectif de création de valeur et d'accélérateur de découvertes scientifiques et de l'innovation
ex: cartographie du génome humain

Pourquoi partager ses données

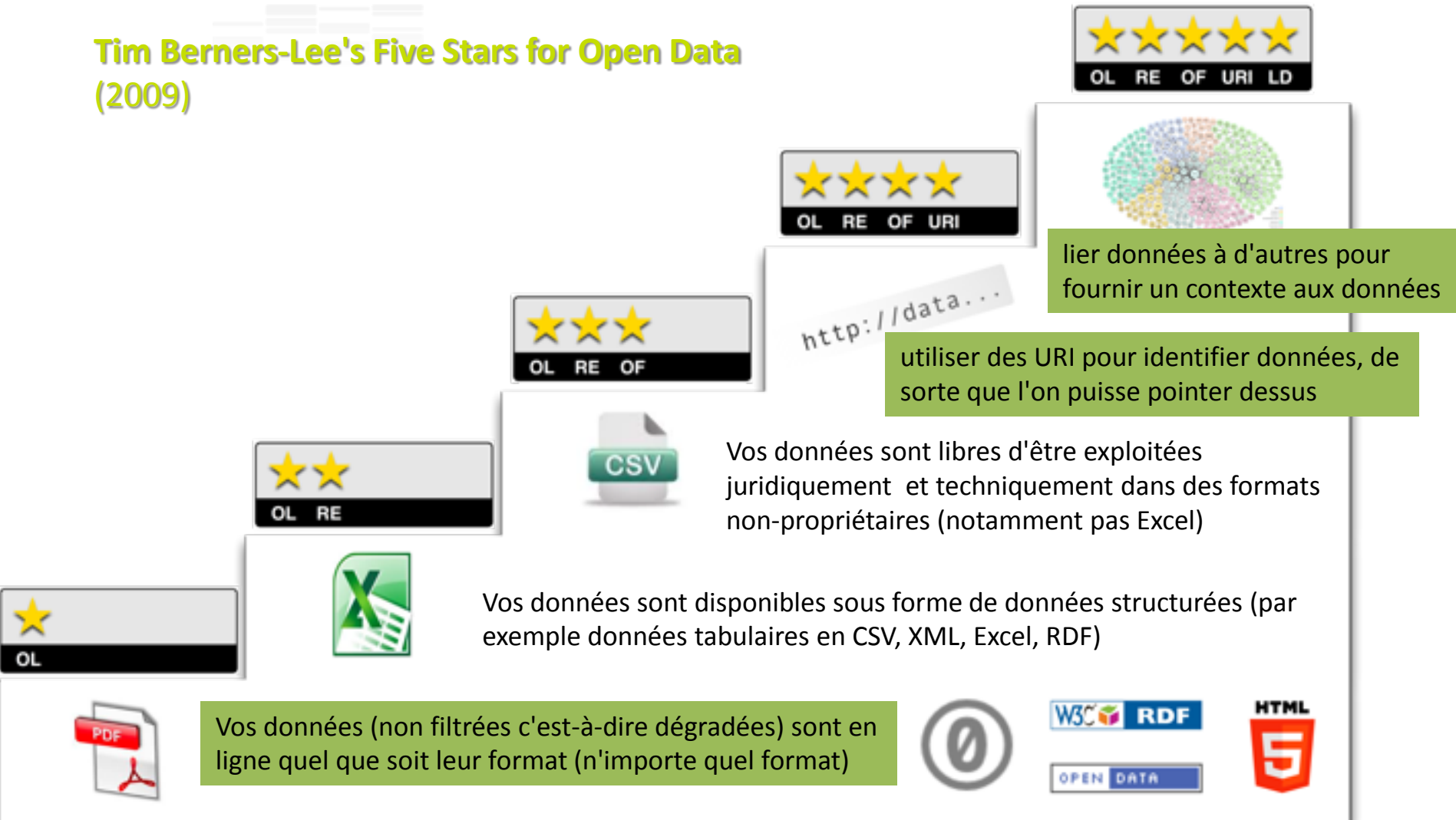
<http://www.data-archive.ac.uk/media/2894/managingsharing.pdf>

- 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



Échelle de qualité des données ouvertes

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



Open science

- Une science ouverte à des scientifiques ou des utilisateurs de tous les pays
- Une science ‘non-concurrentielle’, collaborative entre chercheurs
- Une science ouverte à l’étude, à la critique
- **Partage des publications, des connaissances et des ressources**

La volonté de l'Inra

S'inscrire durablement dans la dynamique de l'open-science

- Démarche engagée avec ProdInra2.0 : archive ouverte des publications
- Contribuer activement aux progrès scientifiques, à l'innovation, à la science participative, à la transparence des résultats de recherche
au travers du partage et de la réutilisation des données de recherche

Démarche Inra

- **Un cadre** : 11 principes en faveur du partage des données de la recherche

- **Une démarche d'accompagnement**

- Stratégie de valorisation par nature de données

Évaluation des ressources Inra : entrepôts et jeux de données

Définition de la stratégie de dépôt : identification des entrepôts externes, définition des bonnes pratiques disciplinaires, analyse des stratégies éditeurs

Préconisations sur le cycle de vie des données par domaine disciplinaire

- Juridique et déontologique
- Des solutions techniques
- Une montée en compétences

Inra dans l'open science

International

- Avril 2013 : G8 partage des données de l'agriculture des pays du G-8 avec des partenaires africains
- Mars 2013 : RDA research data alliance

Europe

- Février 2013 : Science Europe : groupe open data
- Avril 2013 : IGLO : élaborer mesures pour la mise en œuvre du libre accès dans Horizon 2020
- 2013 DG-connect : projet pilote (prospective)

France

- 2013 : BSN 10 : chantier données

Inra

- Février 2013 - Principes en matière de partage de données – document diffusable
- Avril 2013 – lancement chantier « stratégie de valorisation des données de recherche »

Séminaire Inra 15-17 avril 2013

Stratégie de la valorisation des données

15 Avril 2013 : Appropriation des enjeux et des principes

14 h – 14h30	Introduction du séminaire : objectifs / organisation	
14h30 – 15h00	Cadrage politique des enjeux pour l'Inra	Olivier Le Gall
15h-16h	La politique de la commission européenne : open data et infrastructures, H2020, articulation DG-Connect/DG-Agri	Carlos Morais Pires (CE, DG-Connect)
17h-18h	Ateliers N°1 : appropriation des principes : analyse SWOT pour l'Inra	3 ateliers

16 Avril 2013 : Approches thématiques

9-10h	Omiques	EBI
10-11h	L'open data dans le domaine agricole : Enjeux et initiatives internationales (RDA, G8 ...) – projets de la FAO	Johannes Keiser (FAO)
11h-12h	Projet Damaro : The Data Management Rollout at Oxford (DaMaRO) Project is creating a research data management policy for the University and the infrastructure to enable researchers to comply with it	Univ Oxford Prof. James Wilson
	GRDI 2020 (Global research data infrastructures) + Centre de données astronomiques Strasbourg	Françoise Génova
14-15h30	Ateliers N°2 : cahier des charges et identification des groupes	3 ateliers
16h -17h30	Restitutions ateliers	
17h30-18h	Synthèse	

17 Avril 2013 : Dispositif projet

9h	Ateliers N°3 par thèmes : cadrage des sujets à traiter et répartition du travail	3 ateliers
	Restitution	
14h-16h	Synthèse	