

# Les enjeux autour des données de la recherche

*Francis ANDRE  
DIST, CNRS  
francis.andre@cnrs.fr*

# Quid ?

---

- Données de la recherche
  - Produites dans un processus de recherche
    - Générées pour valider une hypothèse
    - Produites en masse
    - Agrégées (observatoires)
    - Coopératives (interdisciplinarité)
  - Réutilisées
    - Données publiques reformatées
  - Financées sur fonds publics
  - Numériques

# Quid ?(2)

---

- OCDE Principles and Guidelines 2006

« factual records (numerical scores, textual records, images and sounds) used as primary sources for scientific research, and that are commonly accepted in the scientific community as necessary to validate research findings. »

- Mais...

*« does not cover the following: laboratory notebooks, preliminary analyses, and drafts of scientific papers, plans for future research, peer reviews, or personal communications with colleagues or physical objects (e.g. laboratory samples, strains of bacteria and test animals such as mice) »*

- Donc

- *Données elles mêmes validées...*

# Quid (3) et fin

---

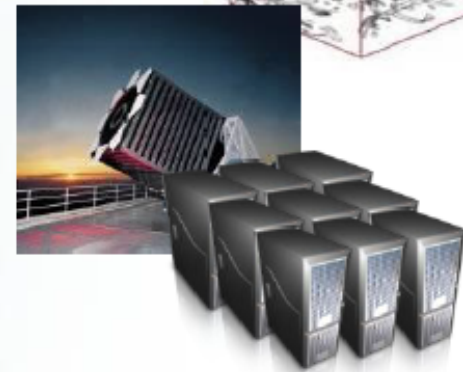
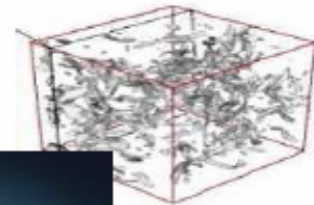
- Rien de nouveau...
  - Des communautés déjà largement organisées : astro, climato, sciences de la terre,...
  - Centres de données ICSU/WDS World data system, ICSU/CODATA, ...
- Si !
  - Numérique
    - Big data et e-science
      - Interdisciplinarité
        - HPC et visualisation
        - ...

# Emergence of a Fourth Research Paradigm

1. Thousand years ago – **Experimental Science**
  - Description of natural phenomena
2. Last few hundred years – **Theoretical Science**
  - Newton's Laws, Maxwell's Equations...
3. Last few decades – **Computational Science**
  - Simulation of complex phenomena
4. Today – **Data-Intensive Science**
  - Scientists overwhelmed with data sets from many different sources
    - Data captured by instruments
    - Data generated by simulations
    - Data generated by sensor networks
  - **eScience is the set of tools and technologies to support data federation and collaboration**
    - For analysis and data mining
    - For data visualization and exploration
    - For scholarly communication and dissemination



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



(With thanks to Jim Gray)

# *The value of data lies in their use*

---

- Partage
  - Open data
  - Collaboration
  - Réutilisation
- Au bénéfice de la recherche elle-même, mais aussi de l'innovation et *in fine* de la société entière
  - Enjeux sociétaux : santé, environnement, énergie,...
  - Commission Communication (COM(2012)401):  
**"Towards better access to scientific information: Boosting the benefits of public investments in research"**

# Quelques repères

---

- Institutions
  - Déclaration de Berlin oct 2003
- Gouvernements
  - OECD Principles and Guidelines Dec 2006
- Chercheurs
  - Principes de Panton Fev 2010
- Commission européenne
  - Recommandations de la Commission Européenne jul 2012

[www.libreaccés.inist.fr](http://www.libreaccés.inist.fr)

[www.donneesdelarecherche.fr](http://www.donneesdelarecherche.fr)

# Accès aux données de recherche

Commission Recommendation C(2012)4890) **on access to and preservation of scientific information**

3. de définir des politiques claires en matière de **diffusion** des données de la recherche financée par des fonds publics et de **libre accès** à ces dernières. Ces politiques

devraient prévoir:

- des **objectifs** et des **indicateurs** concrets permettant de mesurer les progrès accomplis,
- des plans de **mise en œuvre**, incluant la répartition des responsabilités (y compris la concession des **licences** adéquates),
- la **programmation financière** correspondante.



# Accès aux données de recherche

Commission Recommendation C(2012)4890) **on access to and preservation of scientific information**

dans le prolongement de ces politiques, de veiller:

- à ce que les données de la recherche financée par des fonds publics deviennent accessibles, utilisables et réutilisables par le public au moyen **d'infrastructures électroniques**. Il y a lieu de tenir dûment compte des questions relatives, notamment, au respect de la vie privée, aux secrets industriels, à la sûreté nationale, aux intérêts commerciaux légitimes et aux droits de propriété intellectuelle. Les données, le savoir-faire et/ou les informations, quelle que soit leur forme ou leur nature, qui sont détenus par des acteurs privés participant à un partenariat public-privé avant les activités de recherche et qui ont été identifiés comme tels ne sont pas soumis à ce type d'obligation,
- à ce que les ensembles de données soient facilement **identifiables** et puissent être **liés** à d'autres ensembles de données et publications au moyen de mécanismes adaptés, et à ce que des informations supplémentaires soient fournies afin de permettre une évaluation et une utilisation correctes,
- à ce que les organismes chargés de gérer le financement public de la recherche et les établissements universitaires bénéficiaires de **financements publics** contribuent à la mise en œuvre des politiques nationales en mettant en place des mécanismes permettant le **partage de données** de la recherche et le récompensant,
- à ce que les programmes d'études approfondies visant à former des **nouveaux profils professionnels** dans le domaine des technologies de traitement des données soient promus et/ou mis en œuvre;

# Accès aux données de recherche

Commission Recommendation C(2012)4890) **on access to and preservation of scientific information**

4. de renforcer la conservation des informations scientifiques:
- en définissant et en **mettant en œuvre des politiques**, y compris la **répartition des responsabilités** en matière de conservation des informations scientifiques, ainsi que la programmation financière correspondante, afin de garantir la **curation et la conservation à long terme** des résultats de recherche (données de la recherche fondamentale et tous les autres résultats, y compris les publications),
  - en garantissant la mise en place d'un système efficace de **dépôt des informations scientifiques** au format électronique, couvrant les publications d'origine numérique et, le cas échéant, les ensembles de données correspondants,
  - en **conservant** le matériel informatique et les logiciels nécessaires pour lire les informations dans l'avenir, ou en migrant régulièrement les informations vers de nouveaux environnements matériels et logiciels,
  - en créant les conditions permettant aux parties prenantes d'offrir des **services à valeur ajoutée** basés sur la **réutilisation d'informations scientifiques**;

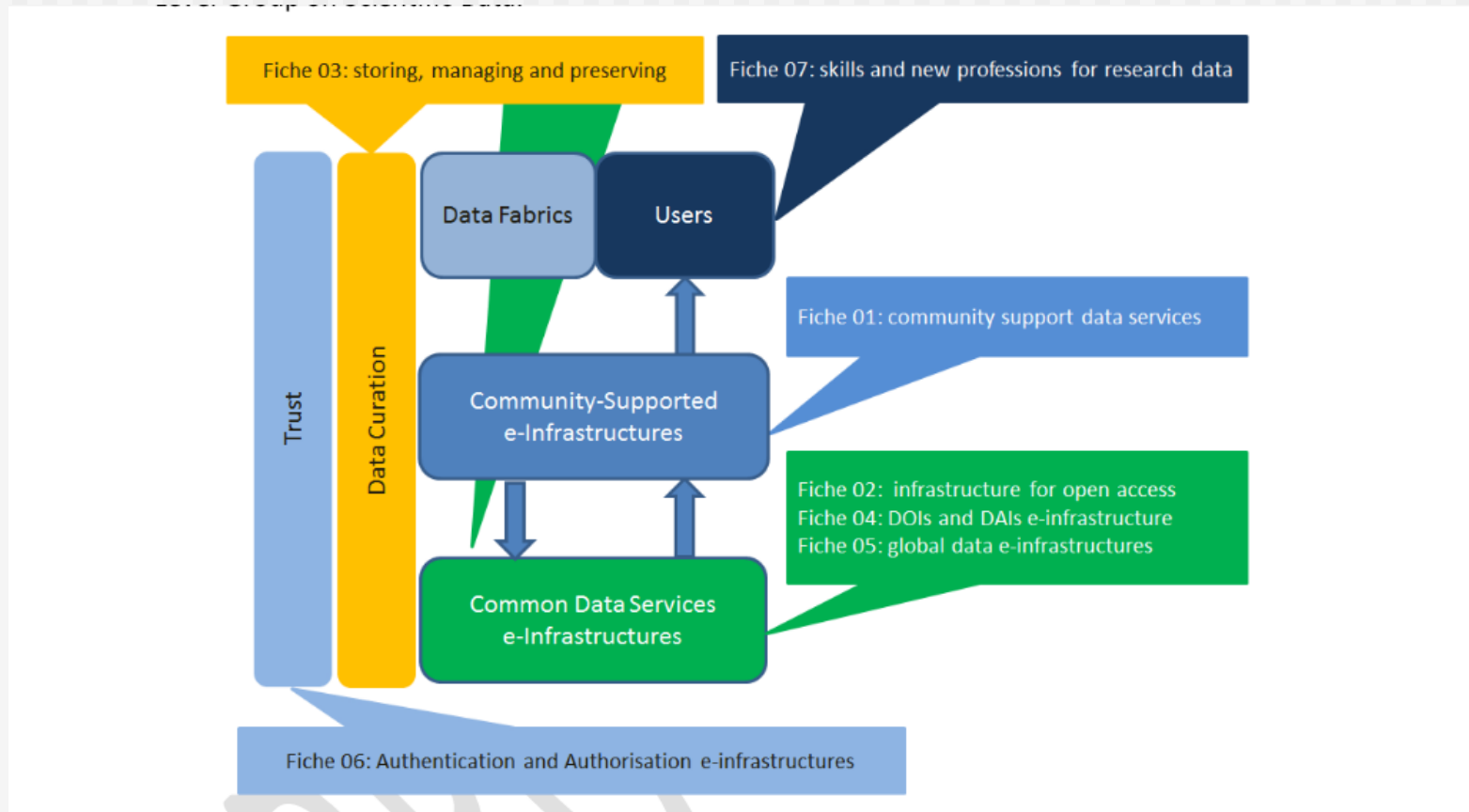
*from Riding the Wave report,  
High-Level Group on Data*



"Our vision is a scientific e-infrastructure that supports **seamless access, use, re-use of data.**"

"With a proper scientific einfrastructure, researchers in different domains can **collaborate** on the same data set, finding new insights. They can **share** a data set easily across the globe, but also protect its integrity and ownership. They can **use, re-use and combine data**, increasing productivity. They can more easily solve today's Grand Challenges, such as climate change and energy supply. Indeed, they can engage in whole\_ **new forms of scientific inquiry**, made possible by the unimaginable power of the e-infrastructure to find correlations, draw inferences and trade ideas and information at a scale we are only beginning to see."

# Horizon 2020 : vers une infrastructure collaborative pour les données de recherche



# Des éléments d'état des lieux

---

- La feuille de route ESFRI : Lifewatch, Anaee, Dariah,...des infrastructures attendues en 2015
- Des infrastructures existantes :
  - Internationales : Elixir, CDS
  - Nationales : Huma-Num, ReNaBi
- Des projets européens : Openaire, Parse/Aparsen, ODE
- Des initiatives CNRS diverses : Mastodons, réseau bdb, Bbees
- Une réflexion institutionnelle avancée à l'Inra

# Des tentatives de coordination

---

- G8+O5 global research infrastructure sub group on Data
- ScienceEurope Data WG
- Liber, Knowledge Exchange
- RDA : Research Data Alliance



# De nouveaux services

- Dryad
- Édition : mégarevues et data journaux



## Geoscience Data Journal



© 2013 John Wiley & Sons Ltd.

Each article is made available under the terms of the [Creative Commons Attribution License](#)



Edited By: Dr Rob Allan, Met Office, UK

Online ISSN: 2049-6060

Associated Title(s): Atmospheric Science Letters, International Journal of Climatology, Meteorological Applications, Quarterly Journal of the Royal Meteorological Society, Weather



# Pour résumer, quelques travaux en cours...

---

- Quelle données partager ?
- Cycle de vie des données
- Liens publications/données
- Sensibilisation des chercheurs
- Plan de gestion de données
- Normes et standards de description, citation, qualité, interopérabilité,...
- Propriété intellectuelle applicable aux données
- Archivage pérenne
- Infrastructures
- Reconnaissance dans le processus d'évaluation
- Gestion des compétences et formations



# Pour conclure

---

- Développer la culture du partage...
- Traiter simultanément organisation (infrastructures et DMP), financement et développement des compétences
- Positions institutionnelles ...coordonnées !
  - Intégré au schéma d'orientation stratégique IST du CNRS
- Régulation à l'échelle nationale (BSN10) et internationale (RDA)

---

Merci de votre attention

*francis.andre@cnrs.fr*

[www.donneesdelarecherche.fr](http://www.donneesdelarecherche.fr)