



# Open Science – Open Data

Enjeux et stratégie pour l'Inra



Séminaire méthodes et outils pour l'Open data



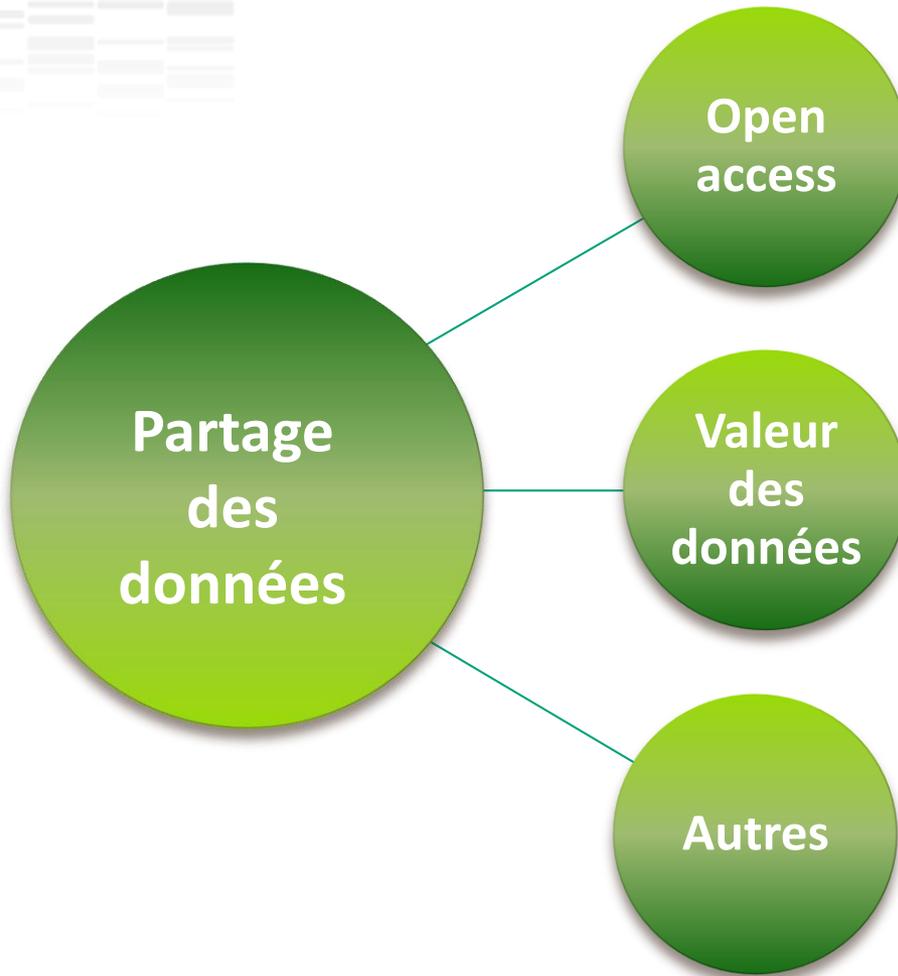
**\_01**

# Contexte international

Des prises de position multiples

# De l'Open Access à l'Open Data

Open science



- Déclaration de Berlin 2003 : Soutien institutionnel au libre accès
- OCDE 2007 : Recherche financée sur fonds publics
- Déclaration de Panton en 2010
- GODAN 2013
- BOUCHOUT 2014



# \_02

## Leviers et catalyseurs...

Acteurs de la recherche

## Science Europe Working Group on Research Data, 21 January 2014, Brussels

European  
Commission

### OA to publications mandate in H2020

Each beneficiary must ensure OA to all peer-reviewed scientific publications relating to its results:

- **Deposit** a machine-readable copy of the published version or final peer-reviewed manuscript accepted for publication in a repository of the researchers choice (possibly OpenAIRE compliant)
- **Ensure OA** on publication or at the latest within 6 months (12 for SSH)
- **Aim to deposit** at the same time **the research data needed to validate the results ("underlying data")**
- Ensure **OA to the bibliographic metadata** that identify the deposited publication, via the repository

L'open access est surtout affirmé pour les publications mais le lien entre publication et données est sollicité et encouragé ....



### OA to publications mandate in H2020

#### Routes towards OA:

- OA publishing and self-archiving considered valid and complementary routes
- Deposit into a repository also in the case of OA publishing

#### Costs for OA publishing:

- Eligibility of OA publishing costs during the grant (as in FP7)
- Piloting a mechanism for open access publishing after the end of the grant agreement (call EINFRA-2-2014 – eInfrastructure on open access)

#### Licensing:

- Encouragement to authors to retain their copyright and grant adequate licences to publishers (e.g. Creative Commons)

# Infrastructure européenne

## Open access, Open data

❖ Publications : OpenAire



❖ Données : OpenAire +



ZENODO is a repository service that enables researchers, scientists, projects and institutions to share and showcase multidisciplinary research results (data and publications) that are not part of existing institutional or subject-based repositories.

# Code de la recherche modifié par la loi ESR de juillet 2013

## Article L112-1

Modifié par [LOI n°2013-660 du 22 juillet 2013 - art. 16](#)

La recherche publique a pour objectifs :

- a) Le développement et le progrès de la recherche dans tous les domaines de la connaissance ;
- b) La valorisation des résultats de la recherche au service de la société, qui s'appuie sur l'innovation et le transfert de technologie ;
- c) Le partage et la diffusion des connaissances scientifiques en donnant priorité aux formats libres d'accès ;
- c bis) Le développement d'une capacité d'expertise et d'appui aux associations et fondations, reconnues d'utilité publique, et aux politiques publiques menées pour répondre aux défis sociétaux, aux besoins sociaux, économiques et du développement durable ;
- d) La formation à la recherche et par la recherche ;
- e) L'organisation de l'accès libre aux données scientifiques.

Les établissements publics de recherche et les établissements d'enseignement supérieur favorisent le développement des travaux de coopération avec les associations et fondations, reconnues d'utilité publique. Ils participent à la promotion de la recherche participative et au développement des capacités d'innovation technologique et sociale de la Nation. Ces coopérations s'exercent dans le respect de l'indépendance des chercheurs et, en l'absence de clauses contraires, dans un but non lucratif. Les travaux de recherche menés dans le cadre de ces coopérations sont, en l'absence de clauses contraires, rendus publics et accessibles.

# Research Data Alliance



- ❖ Lancée en mars 2013 par Commission européenne, NSF, Australie
- ❖ Vise à accélérer et faciliter le partage et l'échange des données scientifiques
- ❖ Contribution INRA à la création de 2 groupes

## Agricultural Data Interoperability IG



Status: Recognised & Endorsed

The Agricultural Data Interest Group is a domain oriented interest group to work on all issues related to data important for the development of global agriculture. The interest group aims to represent all stakeholders producing, managing, aggregating, sharing and consuming data for agricultural research and innovation.

## Wheat Data Interoperability WG



The Wheat Data Interoperability Working Group aims to provide a common framework for describing, representing linking and publishing Wheat data with respect to open standards.

Co chair : E. Dzalé (IST Inra), R. Fulss (CYMMIT)



# Du côté de l'édition...

## ❖ Note aux auteurs – revues classiques

The logo for the journal Science, featuring the word "Science" in white serif font 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.

## ❖ De nouvelles revues (Data Journals...)

The logo for GigaScience, featuring the text "(GIGA)<sup>n</sup> SCIENCE" in a stylized font with a green leaf-like graphic element.

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.

## ❖ De nouveaux entrepôts de données

The logo for Dryad, featuring a stylized green tree with roots and branches inside a circle, surrounded by a grid of green dots, with the word "DRYAD" in green capital letters below.

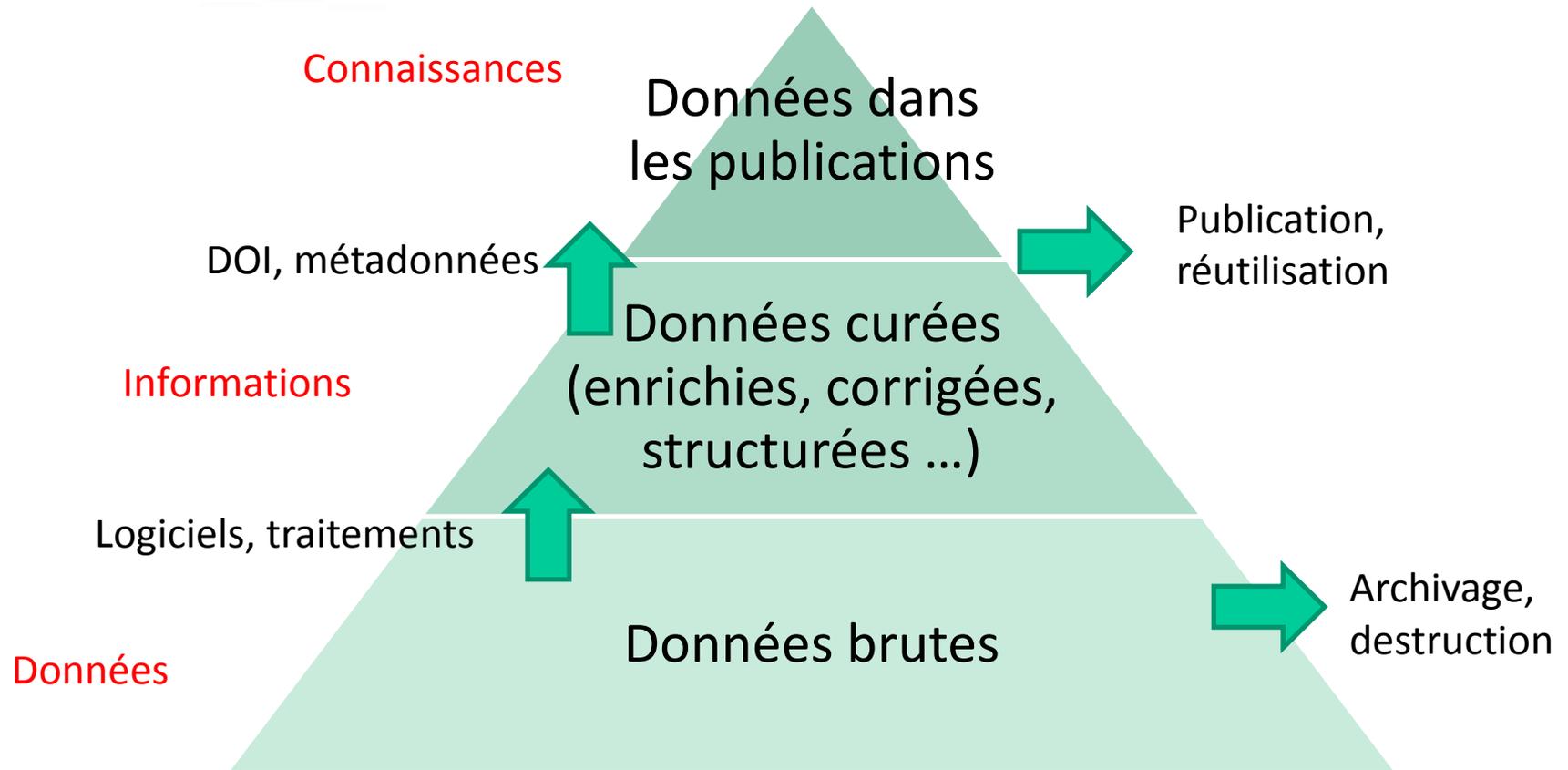
**Dryad is an international repository of data** underlying peer-reviewed articles in the basic and applied biosciences. Dryad enables scientists to validate published findings, explore new analysis methodologies, repurpose data for research questions unanticipated by the original authors, and perform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





<http://fr.slideshare.net/brianhole/from-open-access-to-open-data>

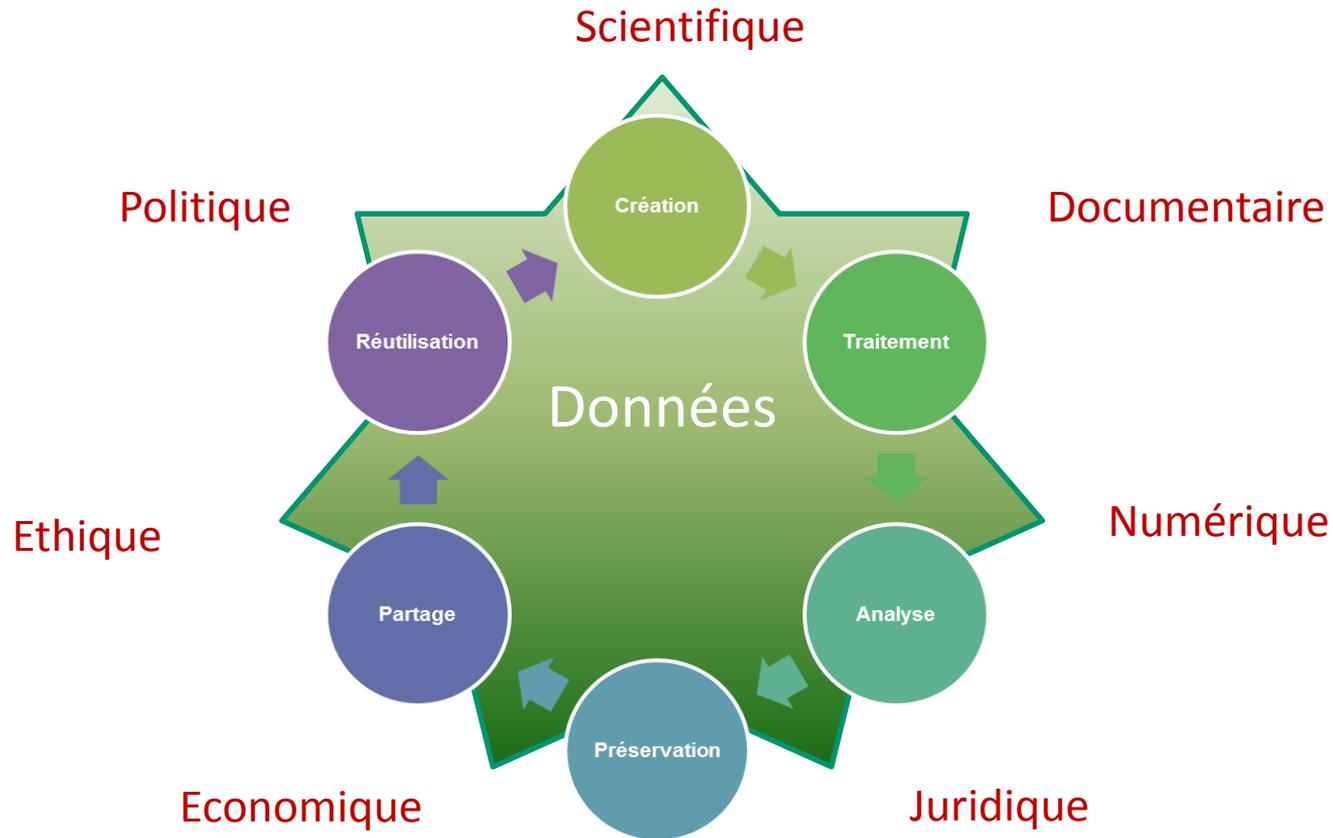
# Les états de la donnée



# Patrimoine de données très diversifié

- ❖ **Données de génétique et génomique**: plantes, arbres, champignons, animaux et micro-organismes...
- ❖ **Données d'expérimentation et observation** : écologie, climat, paysage, sol, plante, arbre, animal, biodiversité, système de culture, socio-écologiques, ...
- ❖ **Données de modélisation** : environnement, paysage, eau, plante, forêt, animal, socio-écologie et systèmes, géographique...
- ❖ **Données d'enquêtes et de cohortes** : Alimentation, consommateurs, industries agro-alimentaires, politiques publiques, exploitation agricole, commerce, cohortes animales et humaines...
- ❖ **Données extraites du Web, des publications...** : analyse textuelle, fouille et extraction de données externes pour ingénierie des connaissances

# Les données, un objet complexe -> mobiliser différents acteurs et compétences





# \_04

## Mise en œuvre du partage des données à l'Inra

# Partage des données à l'Inra : Chronologie

2011-2012 : le CS instruit la question

- Groupe de travail piloté par D. Pontier
- 9 recommandations pour l'Inra (rapport Juin 2012)

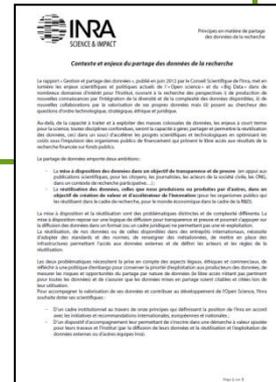


Rapport CS

<http://prodinra.inra.fr/record/206746>

Fin 2012 : élaboration d'une politique

- 11 principes pour mieux gérer et partager les données de la recherche



Note de cadrage :

[Principes en matière de partage des données de la recherche](#)

Avril 2013 : lancement des chantiers de mise en œuvre

- 3 chantiers disciplinaires / familles de données
- Chantier juridique
- Chantier technique (outils et méthodes)
- Chantier compétences

Chantier

« Data Partage »

# Chantiers : planning et livrables

Avril 2013...

...Hiver 2013

... Automne 2014  
08-09/10

Séminaire  
de  
lancement  
Avril 2013

Groupes de travail  
Données  
Méthodes et outils  
Juridique  
IST

Séminaire  
1<sup>ère</sup>  
restitution  
Dec 2013

Groupes de travail  
Nouveaux livrables  
Ebauche /  
Offre de services

## Définir les recommandations, méthodes, outils...

- **Axe données** : omiques et génétiques, environnement-observation-expérimentation, données d'enquêtes et de cohortes
- **Axe méthodes et outils** : format de données, ontologies, repérage, visualisation, qualité, extraction, interprétation des données...
- **Juridique** : propriété intellectuelle de nos données, se doter des cadres juridiques
- **Offre de services** : plan de gestion des données, standards de description (métadonnées), citation des données, entrepôts de données interopérables, identifiants numériques DOI...



05

# Objectifs de ce séminaire

Méthodes et outils pour l'open data

# Séminaire méthodes et outils pour l'open-data décembre 2014

- ❖ Pluri-organismes entre des équipes math-info
- ❖ Favoriser les collaborations entre équipes et consolider une communauté composite naissante
- ❖ Partager des méthodes pour la capitalisation et la pérennisation des données
- ❖ Identifier les futurs outils autour du partage de données
- ❖ Identifier des fronts de sciences
- ❖ Identifier les compétences à développer et leurs moyens de diffusion (tutoriels, formations...)

# Trame et livrables

## ❖ Programme séminaire 17 – 19 décembre Montpellier:

- ✓ Session Introduction au Web de données
- ✓ Session Fouille de textes et d'images et datamining
- ✓ Session Sémantisation, réutilisation et visualisation des données
- ✓ Session Linked open data

## ❖ Premières briques pour une école chercheurs / ingénieurs en 2015

- Faire évoluer les compétences d'un collectif ingénieurs/chercheurs « math-info » par l'acquisition de connaissances et de savoir-faire dans le domaine du partage de données
- Acquérir des compétences dans des méthodologies spécifiques comme la mise en œuvre et la construction de modèles de métadonnées