

L'Open data dans tous ses états, open science, RGPD

Jean-Gabriel Ganascia

Équipe ACASA - LIP6 – CNRS (UMR 7606)

Sorbonne Université – Labex OBVIL

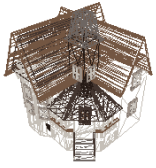
Institut Universitaire de France

Président du comité d'éthique du CNRS

4, place Jussieu, 75252 Paris Cedex 05, FRANCE

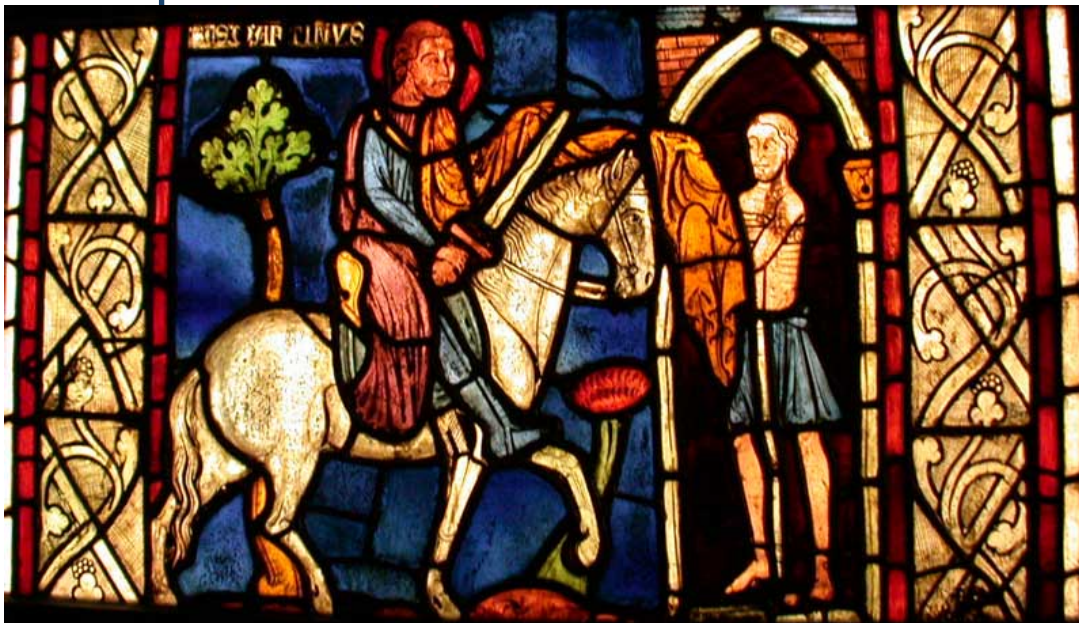
Jean-Gabriel.Ganascia@lip6.fr



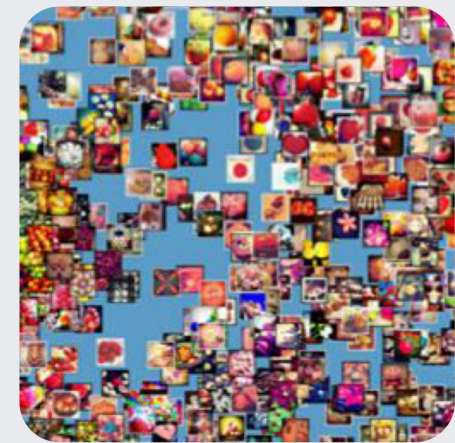


La logique du partage

Du partage matériel,
où l'on perd l'usage de ce
que l'on met en partage...



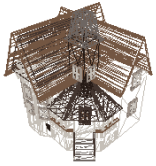
... Au partage numérique
où rien n'est perdu.



Photos Shared
Per Day

2×10^9

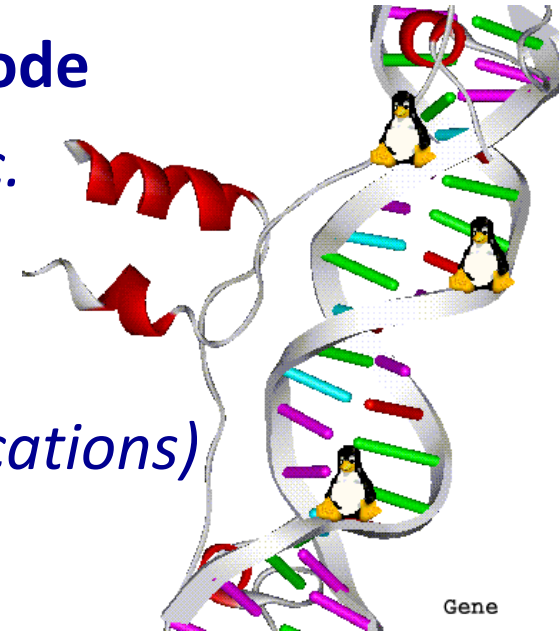
Le partage numérique est « gagnant – gagnant »!



Vertus du partage ...

Le partage ne coûte rien – il est à la mode

- « Data Sharing »: *partage données sc.*
- « Open data »: *données publiques ouvertes*
- « Open access »: *accès ouvert (publications)*
- ...



- Les données sont le **pétrole de l'avenir!**

facebook

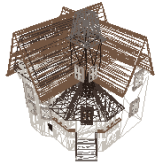
IBM

Google DeepMind



Le partage est-il **équitable?**





Histoire du partage

Partage de données scientifique:

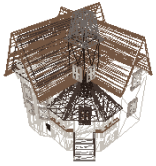
- Origine: biologie, *Bermuda principles*, 1996
- Déclaration de Berlin, 2003 et 2005
- CNRS, 2014

Vient des communautés scientifiques

- Astrophysique, biologie, ...
- Médecine ?? – industrie pharmaceutique?

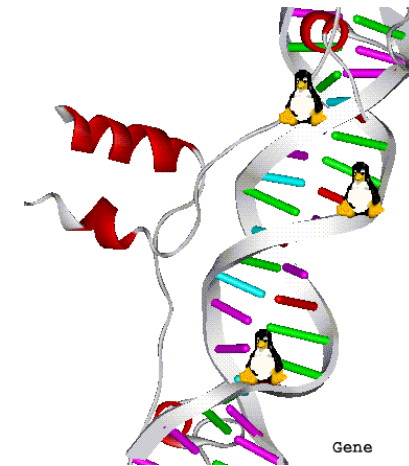
Objectifs

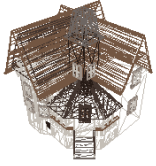
- Mettre la science à la disposition de tous
- Empêcher les fraudes en permettant à tous de reproduire les expériences



Partage: de quoi parle-t-on?

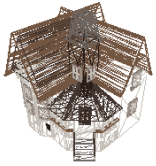
- Nature des données
 - astronomie, génome, images, parole, ...
- Types de données:
 - Données primaires:
 - radio-télescope 7 Po/mn!
 - Données secondaires
 - dérivées des données primaires
 - Filtrées, annotées, enrichies interprétées
 - Métadonnées
 - Structurent, gèrent, facilite l'accès
 - Tracent l'origine des données



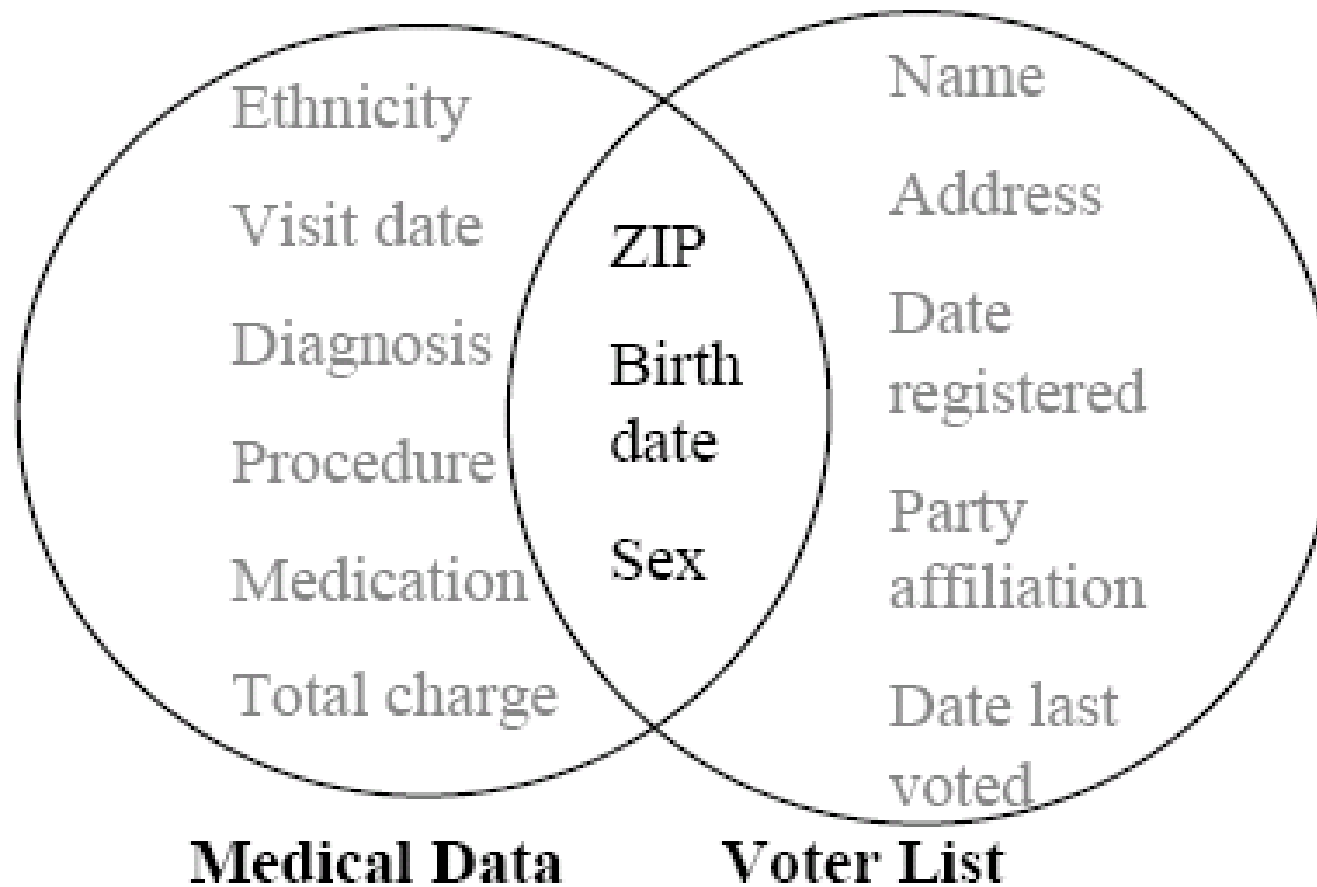


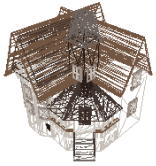
Limites

- **Données sensibles et données à caractère personnel**
 - Médecine, interfaces, images, sciences sociales, etc.
 - Anonymisation: difficultés → croisement



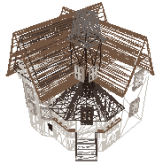
Difficultés de l'anonymisation





Limites

- **Données sensibles et données à caractère personnel**
 - Médecine, interfaces, images, sciences sociales, etc.
 - Anonymisation: difficultés → croisement
 - loi *Informatique et Libertés* du 6 janvier 1978, mod. 2004, RGPD...
 - CNIL: principes de finalité et de proportionnalité (*peu approprié au traitement des grandes masses de données*)
 - **Question:** doit-il y avoir une exception pour des données scientifiques?
- **Limites économiques**



« Open Data » - *Données publiques*

Initiatives gouvernementales

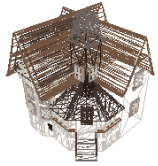
- Europe, Directive sur la réutilisation des informations du secteur public, 2003
- États-Unis, Obama, Gouvernement 2.0, accès publiques aux archives administratives, 2008
- France ETALAB, gère l'*Open data* public sous l'autorité du Premier ministre Circulaire du 17 septembre 2013 - <https://www.etalab.gouv.fr/>

Nature des données

- démographiques, géographiques, météorologiques, économiques, financières, culturelles, touristiques, etc.

But

- Création de gisements éventuellement monayables de données
- Favoriser la transparence des politiques publiques et l'exploitation par les citoyens.

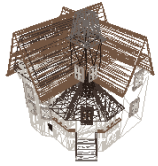


Accès ouvert – « Open Access »

Définition: fournir accès en ligne gratuit à l'« **information scientifique** » - déclaration de Budapest, 2002

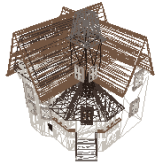
« Information scientifique »:

- a) Articles de recherche, éventuellement évalués par les pairs*
- b) Données de recherche ayant servi aux publications (données primaires et secondaires)*



Accès ouvert aux publications

- **Auto-archivage (« green »)**: les articles évalués sont déposés en ligne **par les auteurs** sur un réceptacle avant ou après publication. *Exemple: HAL, ArXiv*
- **Publication en accès ouvert (« gold »)** les articles sont mis à disposition des lecteurs en accès ouvert (et gratuit). Financement par les institutions ou les auteurs.



Accès ouvert aux données

- Droit d'accéder aux données et de les réutiliser sous un ensemble de conditions spécifiées dans chaque cas.
- Les données sont accessibles et peuvent être exploitées, fouillées, reproduites et diffusées, sans charge financière.

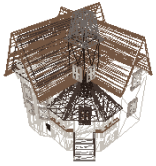
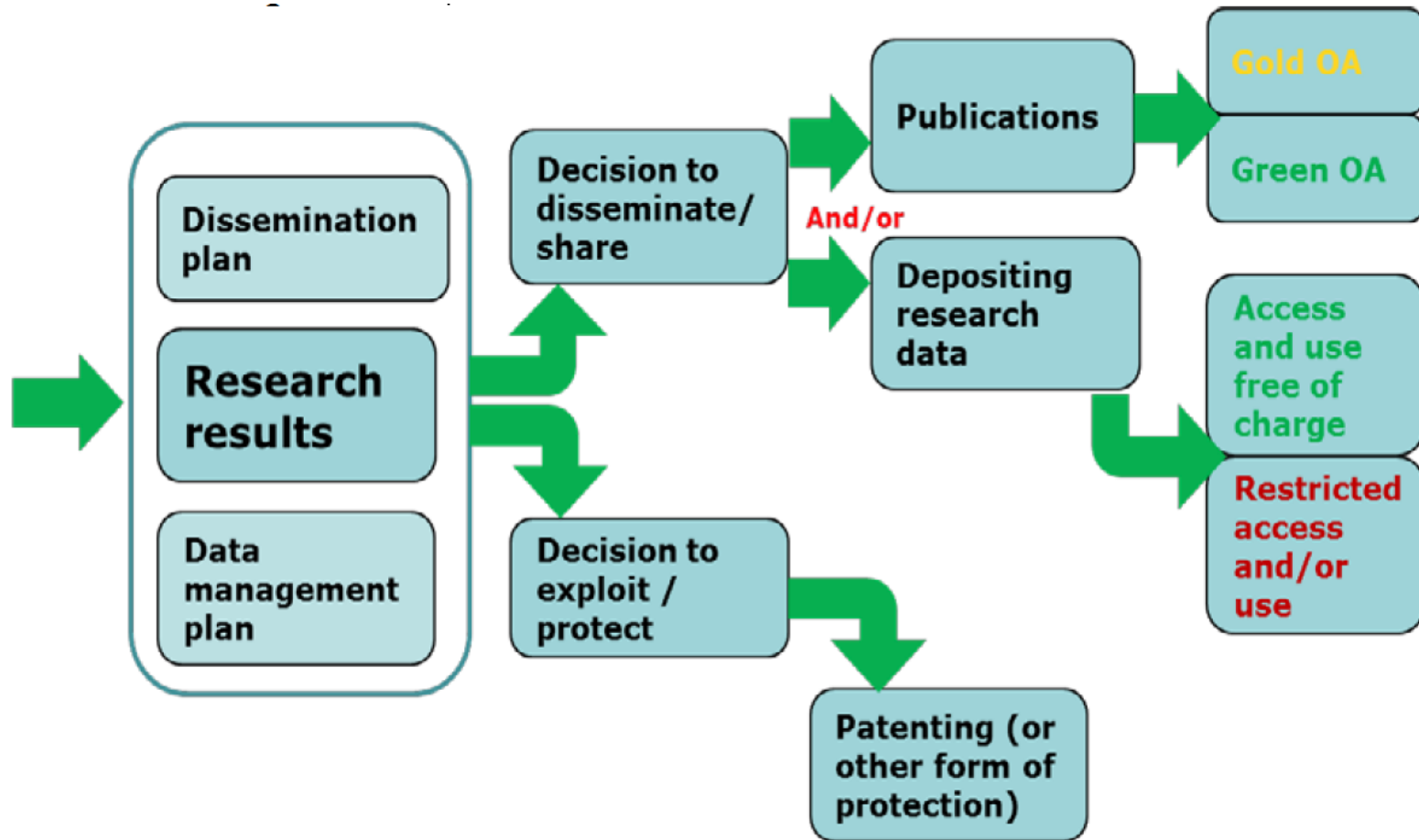
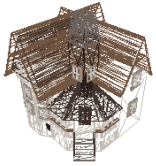


Schéma d'ensemble

R
e
s
e
a
r
c
h



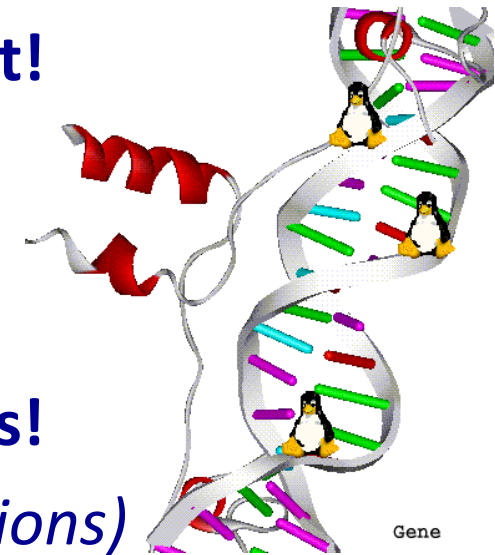


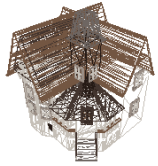
Gratuité du partage ...

Le partage ne coûte rien...

... mais il n'est pas nécessairement gratuit!

- « Data Sharing »: *partage données sc.*
Gratuit!
- « Open data »: *données publiques ouvertes, non nécessairement gratuites!*
- « Open access »: *accès ouvert (publications)*
Gratuit pour le lecteur
Modèle gratuit – green
Modèle payant – gold (auteur payeur)
- **Équité du partage?**





Bibliographie

- *Les enjeux éthiques du partage des données scientifiques, rapport COMETS, 2015 (site COMETS)*
- *Une science ouverte dans une république numérique - Guide stratégique d'application, CNRS, DIST*
- *Déclaration internationale sur le libre accès de Budapest le 14 février 2002, connue sous le sigle BOAI (Budapest Open Access Initiative, <http://www.budapestopenaccessinitiative.org/>)*
- *Déclaration de Berlin sur le libre accès à la connaissance dans les sciences, 2003 puis 2005, <http://openaccess.inist.fr/?Declaration-de-Berlin-sur-le-Libre>*