

Avis 8 - Enjeux éthiques et déontologiques du partage et de la gestion des données issues de la recherche

Séminaire de clôture du chantier
« PARTAGE DES DONNEES »

INRA
22 novembre 2016

Comité consultatif commun
d'éthique pour la recherche agronomique



AVIS
8

SUR les enjeux éthiques et déontologiques
du partage et de la gestion des données
issues de la recherche



SAISINE

Nombreuses considérations

- **juridiques**
 - propriété intellectuelle (chercheur et/ou institution)**
 - protection des données personnelles**
- **meilleure efficacité de leur traitement (bailleurs)**
- **meilleure valorisation économique**
- **« révolution » méthodologique (data vs. hypothesis driven)**

SAISINE (2)

Implications éthiques et déontologiques peu traitées ⁽¹⁾

- doit-on partager toutes les données ?
- partage libre de droits ou soumis à conditions ?
- conséquences pour les acteurs de la recherche ?
 - juniors
 - pays du Sud
 - crowdsourcing
- quelle organisation pour mutualiser les données ?
- comment documenter, tracer, valoriser l'usage des données d'autrui ?

CONTEXTE

- Freedom of Information Act, 1966
- Loi sur l'amélioration des relations avec le public, 1978
- connaissances comme bien commun (\Rightarrow open access)
- transparence de la recherche (open science)
- données comme produit final et matière première
- recherche financée sur fonds publics – OCDE, 2007
- H2020, open access to research data, 2015
- Loi pour une république numérique (Art. 17), 2016
- valorisation économique rapide (G Soros et Budapest initiative)
- crise de la reproductibilité
- crise et mutations de l'édition scientifique

CONTEXTE (2)

L'accès libre aux données est déjà pratiqué

géophysique et sciences de l'environnement

mathématiques

physique (CERN)

astrophysique...

Définitions

les données de la recherche sont des enregistrements factuels sous forme numérique (résultats d'expérience, observations de terrain, résultats d'enquêtes, images...) collectés pour être traités et analysés dans le cadre d'un travail de recherche

la production de données scientifiques valides suppose la conjonction de compétences nécessaires à la construction du système expérimental approprié à leur production, ainsi qu'à leur recueil et leur traitement

les données sont donc le produit d'activités humaines et leur production est frappée d'un caractère éminemment local

les données sur les données ou métadonnées conditionnent leur nouvelle utilisation en décrivant très précisément les modalités précises de leur acquisition

Définitions (2)

l'ouverture des données de la recherche désigne la mise à disposition sur le réseau informatique mondial (internet) d'enregistrement factuels collectés pour être analysés et traités dans le cadre d'un travail de recherche, notamment en vue de leur réutilisation

les données comme nouveaux « actifs »

Méthode

rencontre avec les équipes travaillant sur le chantier numérique

Odile Hologne et Isabelle Blanc (INRA)

Pascal Bonnet (CIRAD)

**revue des travaux initiés dès 2009, notamment le rapport
« gestion et partage des données » du CS INRA (2012)
définissant 11 principes selon 5 axes**

**chantiers lancés en avril 2013 traduisant ces principes en
termes opérationnels**
disciplinaires
juridique
technique
compétences

Méthode (2)

rencontre d'équipes de recherche

INRA

CIRAD (Montpellier)

rencontre avec un éditeur

Elsevier

documentation complémentaire

rapport du CS du CNRS

colloques, notamment de philosophie

présentation du projet d'avis à toutes les personnes sollicitées

Enjeux

Vaste spectre de phénomènes et systèmes biologiques, écologiques, techniques et socio-économiques

Acquisition de jeux de données très variés dont certains sont uniques ou stratégiques

INRA CIRAD : 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, climat, cohortes animales et humaines...
- ❖ **Données extraites du Web, des publications...** : analyse textuelle, fouille et extraction de données externes pour ingénierie des connaissances

Enjeux (2)

Jeux de données différant profondément par leur portée pour la collectivité, en conséquence par leur possibilité de partage

quatre exemples

observatoire du développement rural et politiques agricoles

évolution phénotypique de lignées végétales ou animales

santé sanitaire des cheptels et risques de zoonose

atteinte grave à l'environnement (pollution)

Enjeux (3)

Diversité des partenaires

bailleurs publics

bailleurs privés / entreprises commerciales

partenaires de recherche, notamment pays du Sud

Enjeux (4)

**pour les chercheurs, la double injonction
produire de nouvelles connaissances
tout en partageant leurs données dans un contexte très
compétitif
peut *a priori* paraître paradoxale**

**les chercheurs consultés marquent une certaine ambivalence et
expriment un grand besoin de formation et de conseils**

**cette nouvelle tâche leur demandera un travail
supplémentaire qu'il faut évaluer, guider, organiser et faciliter,
dans la mesure où leur implication sera un facteur décisif du
succès d'une politique de partage des données**

Question éthiques

à l'étape de production

preuve de validité des données : contrôle qualité très strict conçu avec le protocole expérimental

réserve (légitime) des chercheurs vis-à-vis d'un risque de pillage

protection des collaborateurs (jeunes chercheurs ou chercheurs des pays du Sud)

protection des données personnelles et impossibilité de les « anonymiser »

données à caractère dual

Question éthiques (2)

à l'étape de traitement et de mise en ligne

mise en ligne de données incomplètes (par ex. protection des données personnelles)

mise en ligne de données compromettant une filière économique ou un patrimoine collectif

normalisation des présentations dont le format peut « censurer » l'activité des chercheurs

incertitude des chercheurs sur le statut de leurs données

accaparement des données par des sites internet

Question éthiques (3)

à l'étape de la réutilisation

utilisation de logiciels propriétaires de recherche de données et risques de biais

détournement ou accaparement de données publiques pour un usage commercial

traçabilité, citation et valorisation des données d'autrui lors de leur utilisation

utilisation responsable des données d'autrui (domaine de validité, méthodologie appropriée...)

Question éthiques (4)

en résumé

les données de la recherche sont issues de compétences scientifiques et de savoir-faire techniques multiples

elles sont le fruit d'investissements intellectuels et matériels, parfois très lourds et souvent sur des longues durées

les données peuvent alors représenter une valeur propre, représentant un réel patrimoine, susceptible de dépasser les seuls intérêts de l'institut

Recommandations

le premier bénéfice à court terme du partage des données sera d'en améliorer la qualité - il doit donc être encouragé

cependant, le partage des données ne peut s'opérer sans discernement

compte tenu de la diversité et de la complexité des jeux de données produits par les recherches conduites à l'INRA, la décision ne peut être prise qu'au cas par cas

le spectre des décisions s'étendra de la conservation hors ligne à la mise en ligne libre de droits, en passant par la mise en ligne sous conditions d'emploi

ces choix seront fondés sur des considérations scientifiques, mais aussi éthiques, juridiques, administratives, financières et économiques, sans oublier les aspects propres à la conservation des données

Recommandations (2)

1. définir une politique d'établissement

concertée avec les autres institutions (organismes de recherche, universités...) (R2)

préciser les rôles des différents acteurs (R1)

informer et former les chercheurs à la mise en ligne des données et à leur réutilisation responsable (R1)

participer aux instances internationales responsables de l'harmonisation sémantique et de l'élaboration des standards (R3)

Recommandations (3)

2. assurer le succès de cette politique

évaluation précise et allocation des moyens humains et techniques – reconnaissance des nouveaux métiers impliqués dans le contrôle qualité, l'édition, la gestion et la mise en ligne, ainsi que dans la fouille des données (R4)

généralisation du contrôle qualité des données et formation des chercheurs (R5)

reconnaissance et prise en compte dans l'évaluation des chercheurs, des unités et des organismes (R6)

Recommandations (4)

3. créer une expertise collective

création d'une instance spécialisée
rôle d'arbitrage au cas par cas
avec création progressive d'une jurisprudence

rôle de référent - contribue à la politique d'établissement

en cas de partage pour usage commercial, veillera en lien avec l'administration de l'institut, à l'établissement de conditions d'usage précises, définies par un contrat définissant les droits du fournisseur de données ainsi que le mode de leur valorisation (chercheur et collaborateurs)
(R8, R9, R10)

Conclusion

la recherche a pour but de produire de nouvelles connaissances

ces connaissances sont un bien commun

cependant, les données issues de la recherche ne peuvent être partagées sans discernement