



HAL
open science

Le cas du stockage des données en sciences humaines et sociales

Annie Hofstetter

► **To cite this version:**

Annie Hofstetter. Le cas du stockage des données en sciences humaines et sociales. Storage Day 2018, Jan 2018, Paris, France. 11 p. hal-02790596

HAL Id: hal-02790596

<https://hal.inrae.fr/hal-02790596>

Submitted on 5 Jun 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



INRA
SCIENCE & IMPACT



Retours d'expériences : le cas du stockage des données en sciences humaines et sociales

**Annie Hofstetter – INRA Montpellier – Cati CITISES
Storage Day - Paris - 16 janvier 2018**

SOMMAIRE

- Qui ?
- Quoi ?
- Il était une fois...
- Pourquoi ?
- Quand ?
- Comment ?
- Donc !

Qui ?

Sciences Économiques et Sociales



Quoi ?



Ref Alexandre Serres et al U-Rennes2

- ❖ Chiffres
- ❖ Textes
- ❖ Images
- ❖ Etc.

Données sources **utilisées**
= souvent **données contractées**

Données **produites**

Particularités :

données micro économiques
(RICA, entreprises, ménages) =
sensibles, sous contrat
Secret statistique
Clauses d'utilisation

économie expérimentale
BD sujets humains
Informatique et liberté

Il était une fois...

Mission Base de Données SAE2



Avant

- Données achetées
 - ✓ 1 personne = 1 base = dispersion + à la demande = chacun sa base
 - ✓ Plusieurs labos = 1 achat = regroupement et partage du coût mais pas des données = chacun sa donnée...
- Données produites... chacun sa base + chacun sa donnée



2004 création de la Mission BD

Objectifs

- Améliorer le niveau de connaissance des BD utilisées et leur circulation dans le département SAE2
- Maintenir les relations avec les organismes producteurs et diffuseurs de données (INSEE)
- Approfondir les liens avec les partenaires non institutionnels (Centres d'économie rurale, Instituts techniques)

Appui

- Correspondants
- Pages web (pseudo méta données)
- Achat groupé

Réflexion en cours

- Présentation sous forme d'annuaire
- Outil production Cati : Dat@gri
- Newsletter + kit de bonnes pratiques + forum

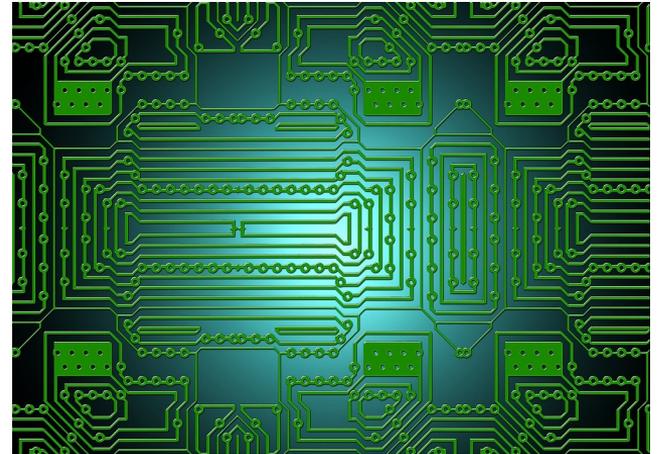
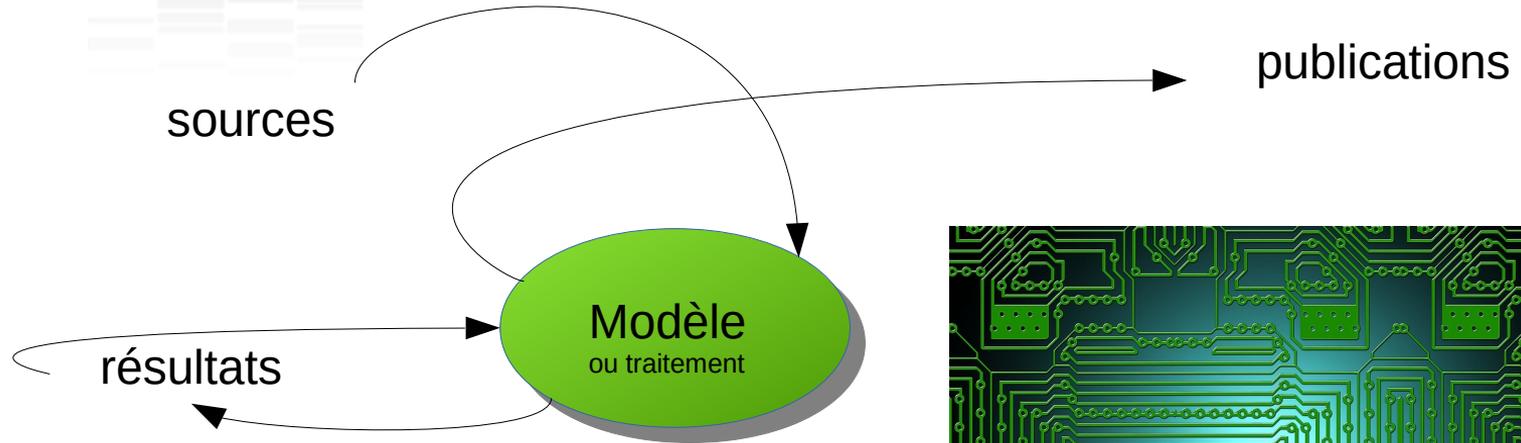
Pourquoi ?

Économie : réduire les coûts...

- ❖ Sauvegarder : et dire que c'est évident...
- ❖ Organiser
- ❖ Archiver
 - Patrimoine
 - Reproductibilité
- ❖ Partager
 - Mutualiser

Quand ?

Ça dépend... ça varie...



Comment ?

On fait comme on peut

Surtout ne pas réinventer l'eau tiède !



- ❖ Système artisanal de gestion de fichiers

Observatoires et plateformes

- ❖ ODR
 - datacenter Toulouse 1T
 - BDNI et BD EFESE 1T
- ❖ PAP 2T
 - Datacenter Paris 1/3
 - Ivry 2/3
- ❖ OQALI
- ❖ Cortext
- ❖ LEEM

- ❖ Système idéal ?

Donc !

Freins	Leviers
Coût // partage Richesse, trésor à exploiter Valorisation intermédiaire Des fonds perdus (ex Grignon Data Base)	S'adapter conservation/accès Reconnaissance institutionnelle Data papers
Données contractuelles	Aspects juridiques « catalogue » Dat@gri
Volume augmente car accès de + en + facile Perte, vol, dégradation Pérennité	Changer méthodes de traitement (travail) Plan de gestion de données Métadonnées Distinguer l'information (quoi stocker) Choisir le bon format
Travail individuel Nomadisme Collaborations extra (ex USA) Pratiques individuelles (ex sauvegarde) Pratiques chronophages (ex doc) Paradoxe entre partage et google-isme	Cibler les priorités Réflexion pour recentrer la recherche (ex métaprogrammes) Éviter le cas par cas Normaliser (≠ méthodes personnelles) Aller vers plus de collectif Formation à la sécurité (info) Diffuser les bonnes pratiques (ex référentiel qualité, recherche reproductible) Démarche FAIR