

Pourquoi et comment documenter ses données ?

Rencontre annuelle du réseau MATE-SHS
26 novembre 2015

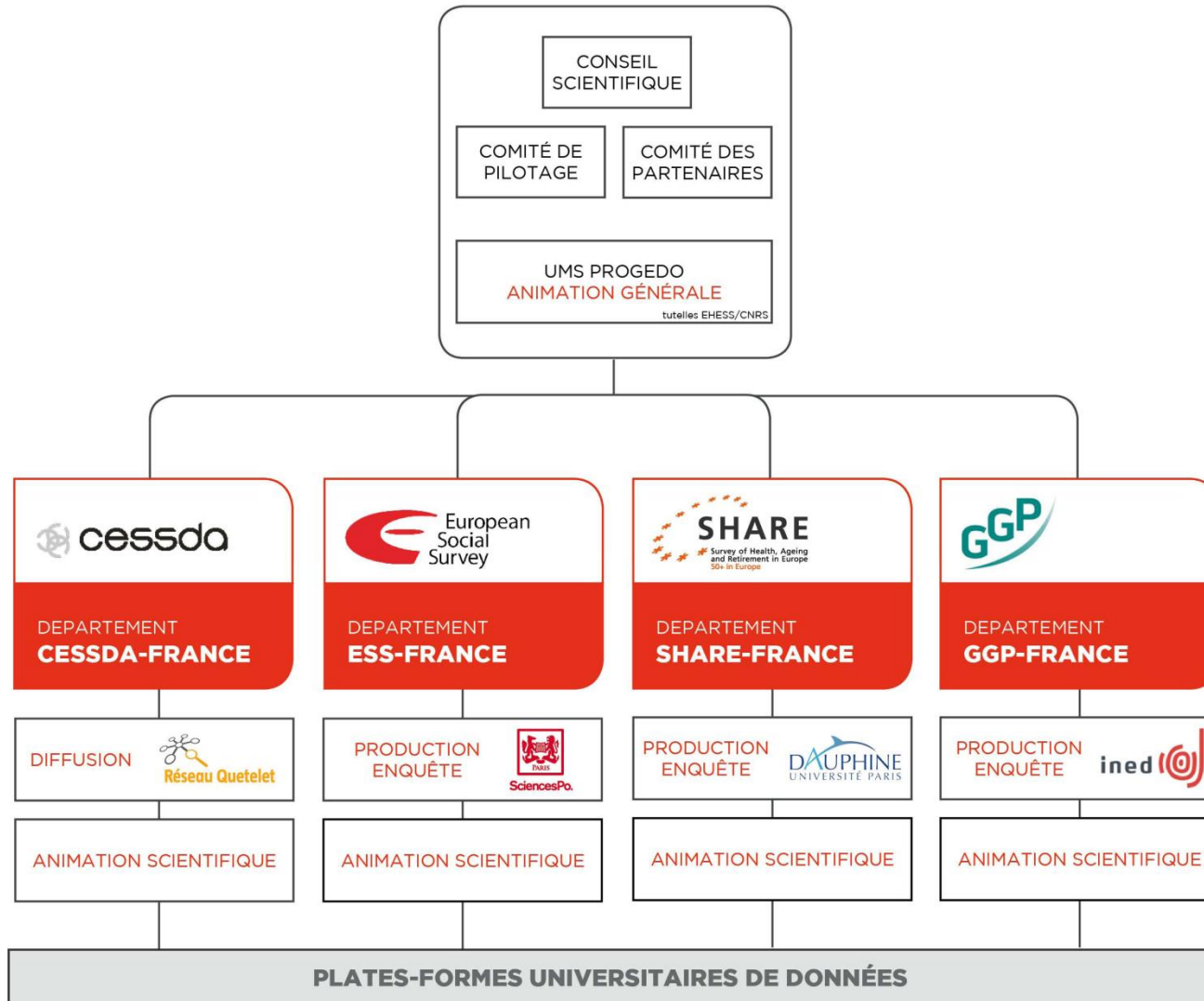
Raphaëlle Fleureux – CMH-ADISP

Guillaume Garcia – CDSP

Quelques éléments de contexte



DATA
INFRASTRUCTURE



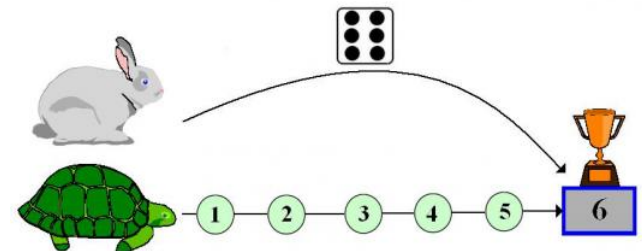
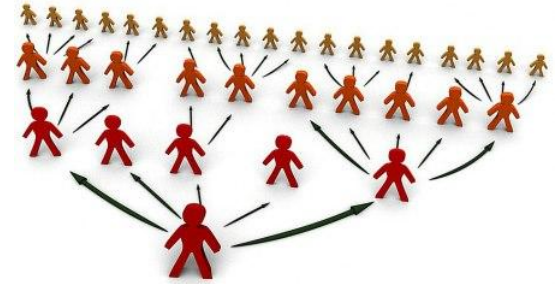
Quelques éléments de contexte

Le **Réseau Quetelet** coordonne les activités de documentation et de diffusion des **données françaises** en sciences humaines et sociales provenant de la recherche, de la statistique publique et d'organismes divers à travers quatre unités partenaires :

- **Adisp-CMH** : diffusion d'enquêtes et bases de données produites par la statistique publique
- **CASD** : diffusion de données individuelles très détaillées dans des conditions de sécurité élevées, après accréditation par le comité du secret statistique.
- **CDSP** : diffusion d'enquêtes socio-politiques de la recherche publique & des résultats électoraux
- **Service des enquêtes de l'INED** : diffusion d'enquêtes socio-démographiques produites par l'INED

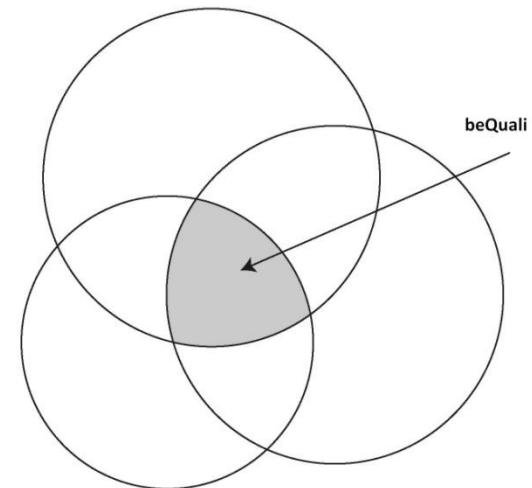
Pourquoi documenter des données ?

- Faciliter la citation et la réutilisation des données
- Optimiser le ou les financements publics investis dans l'enquête en valorisant les données et en assurant leur pérennité
- Une documentation pensée et réalisée dès le départ signifie moins de travail *a posteriori* pour la mise à disposition des données, et ce pour tout le monde...



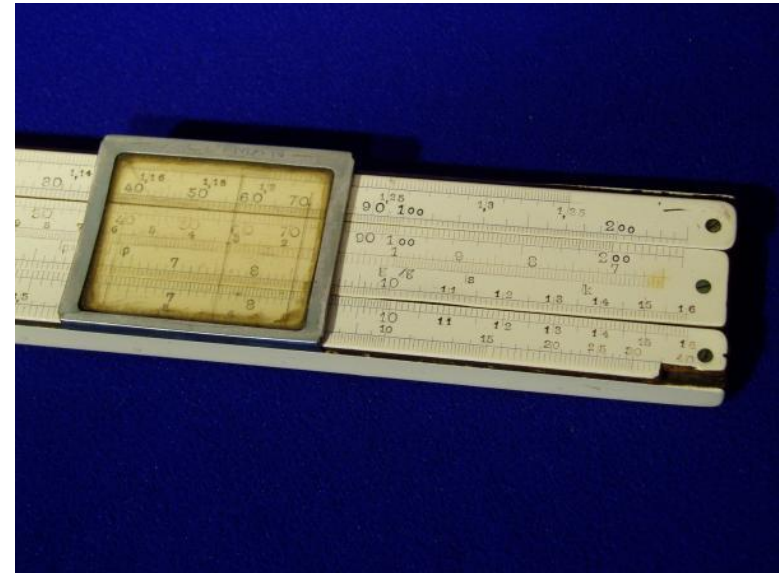
La documentation des enquêtes qualitatives

- Une définition large du périmètre des archives d'une enquête, qui dépasse le corpus de données *stricto sensu*
- A côté des matériaux d'enquête, 2 séries de documents à réunir :
 - ✓ Des archives personnelles : correspondance, journaux de bord, documents intermédiaires d'analyse, etc.
 - ✓ Des archives administratives de la recherche : comptes-rendus de réunion, ordres de mission, etc.



La documentation des enquêtes quantitatives

- Documents concernant la réalisation de l'enquête :
 - Questionnaire(s)
 - Conditions de la production des données : méthodologie, instructions aux enquêteurs, contexte
- Documents permettant d'utiliser le(s) fichier(s) de données :
 - Dictionnaire des codes comportant la signification de l'ensemble des variables
 - Informations sur les redressements / appariements / imputations effectués, pondérations, etc.



Ce qui manque généralement le plus

- Pour les données quantitatives :
 - conditions dans lesquelles l'enquête a été réalisée
 - traitements réalisés sur les fichiers de données pour les apurer, pour les imputations, redressements, pondérations, variables créées...
- Pour les données qualitatives :
 - Informer les chercheurs sur la nécessité d'écrire en amont le plus de renseignements possibles sur le processus d'enquête.
Par ex : interactions avec chaque interviewé, contexte propre à chaque entretien, manière de construire le corpus...
 - Tableau synthétique des matériaux (entretiens notamment) avec les pseudos, dates, lieux, versions...



Difficultés à anticiper

- Les conditions de rediffusion des données, notamment dans le cadre d'enquêtes collectives, ne sont que rarement définies au préalable dans les projets de recherche
- La documentation n'est souvent pas pensée / financée à l'origine
- Anonymisation : opération coûteuse, préalable à la diffusion, impliquant largement les diffuseurs
- Sentiment répandu de défiance vis-à-vis de l'institution et d'un potentiel contrôle scientifique
- Sentiment de propriété sur les données

Et maintenant, qu'est-ce qu'on fait?

- Avoir un discours rassurant sur l'utilité, les (bonnes) raisons de la documentation et de l'archivage : utiliser les bons arguments qui ne laissent pas les chercheurs penser qu'ils peuvent être mis en danger ou non accompagnés dans le processus de mise à disposition
- Des situations contrastées selon les générations ? Selon les disciplines ? Selon les orientations méthodologiques ?
- Une sensibilisation nécessaire, dès les formations doctorales
- La nécessité d'avoir des échanges et un discours commun du monde de l'IST sur le sujet



Pour plus d'informations :

- Adisp :

<http://www.cmh.ens.fr/greco/adisp.php>

- CDSP :

<http://cdsp.sciences-po.fr>

- beQuali

<http://bequali.fr/fr>