

Enjeux des “data” en SHS et en sociologie (1/2)

- 1) Conservation pérenne des données (les données ne sont rien sans toutes les connaissances qui ont conduit à les produire : conditions et contextes de production ; méthodes ; échantillon ; choix méthodologiques ; arbitrage(s), etc.) En conséquence, les données doivent être documentées. Cela pose des enjeux techniques (conserver de manière pérenne) et aussi humains (il faut des compétences et du temps-homme pour conduire proprement cet archivage)
- 2) Enjeux réglementaires sur la propriété des données : à qui appartiennent-elles ? Droit à l’oubli ?
- 3) Enjeux sur la mise à disposition et la valorisation des données auprès d’autres chercheurs ; conditions de leur réutilisation (dans la durée)

Enjeux des “data” en SHS et en sociologie (2/2)

4) Enjeux d’anonymisation des données pour leur exploitation primaire puis secondaire

5) Enjeux méthodologiques 1: données de plus en plus volumineuses ; données produites à des fins qui ne sont pas scientifiques mais administratives ou commerciales (ex : enquêtes à partir de données issues de Facebook, base de commentaires du Monde.fr ou de forums Doctissimo.fr, traces d’activités de type *Quantified Self*, etc. - et donc codées de manière malcommode pour le sociologue (ce qui nécessite un gros travail d’ingénierie)

6) Enjeux méthodologiques 2 : aspiration, crawling... nécessitant des compétences inhabituelles pour le sociologue, en termes de collecte notamment, mais aussi d’analyse

Enjeux de formation et de politique scientifique

- On ne peut plus se contenter de sociologues qui “bidouillent” un peu d’informatique ; ni d’informaticiens qui posent quelques questions sociologiques
- Il faut rassembler les gammes de compétences, par la formation de statisticien-informaticien-sociologie/SHS ; par la coopération étroite d’équipes
- Le besoin en interdisciplinarité est donc plus que jamais important :
 - Travail sur la limite de la réutilisation des données (anonymisation, consentements, CNIL...) > enjeux juridiques, philosophiques, éthiques et politiques
 - Les grandes bases de données chiffrées avec l’aide des statisticiens pourraient être encore plus explorées
 - Enrichissement de la recherche grâce à des formations interdisciplinaires

Besoins en infrastructures

- Des ordinateurs performants permettant la rapidité des calculs d'un nombre important de lignes,
- Des serveurs pour stocker et archiver les données,
- De la sécurité informatique pour le partage des données...
- ...et au-delà des infrastructures matérielles il faut aussi des postes permettant le stockage, l'archivage, la pérennisation et la sécurité des données

Merci !

eric.dagiral@parisdescartes.fr

olivier.martin@parisdescartes.fr

melanie.shaiek@parisdescartes.fr