

# Descriptif des axes et des ateliers

## Axe 1 : « Des nouvelles sources de données, pour quoi faire ? »

Données du web, open data, données géolocalisées (carroyées), bibliométriques, ou encore relationnelles, les nouvelles sources de données émergent en nombre et rapidement depuis quelques années. La tentation est donc grande de s'emparer de ces nouvelles sources et de les exploiter, et pour cause : il s'agit d'un matériau d'une grande richesse. Pour séduisant que cela peut sembler au premier abord, il faut néanmoins se montrer prudent et revenir à la fois sur les enjeux, mais également sur les conditions favorables à un traitement rigoureux et scientifique de ces objets, en lien avec une question de recherche. La question se pose par exemple de l'origine des informations récupérables, de leur niveau de couverture, de la précision/fiabilité des métadonnées disponibles, etc.

Cette session se propose de questionner l'emploi et de revenir sur les conditions d'utilisation de cette masse de données nouvelles.

Exemples (non exhaustifs) :

- Données du web & Big Data ;
- Les données relationnelles ;
- Analyse qualitative de blog, tweet, ou forum ;
- Etc.

## Axe 2 : « Collecter des données oui, mais pour les exploiter : comment les préparer en amont ? »

Qu'ils soient qualitatifs ou quantitatifs, les données brutes et les corpus ne sont que rarement exploitables en l'état : il convient de les transformer de telle sorte qu'ils coïncident avec un objectif de recherche préalablement formulé (codages, construction d'indicateurs). Cette étape requiert à la fois le recours à des routines et de l'inventivité. Quand les données n'existent pas (encore), leur construction peut être pensée en amont, mais peut également être aménagée au fur et à mesure des analyses et de l'avancée du terrain.

Cette séquence présentera quelques cas concrets de préparation des données collectées à des fins d'exploitation (à partir de l'expérience de plusieurs intervenants qui présenteront leur manière de faire en s'appuyant sur des projets de recherche précis).

Exemples (non exhaustifs) :

- Réflexion sur les métadonnées préalables au recueil des données ;
- Journal de terrain, carnet d'observation... Comment intégrer à l'analyse les outils classiques du terrain qualitatif ? ;
- Les étapes préalables à la réalisation d'un questionnaire ;
- Dans quelle mesure les outils conditionnent le formatage des données ? ...

### **Axe 3 : « Le renouveau des méthodes et outils d'analyse : mais quelles données faut-il ? Éloignements et proximités des pratiques disciplinaires »**

Cet axe traite de l'adéquation des données aux nouvelles méthodes et outils d'analyse. Il s'agira ici de présenter quelques-unes de ces nouvelles méthodes avec pour objectif d'en préciser les contraintes en termes de collecte et de préparation des données.

Exemples (non exhaustifs) :

- Enquêtes prosopographiques et collecte de trajectoires / données biographiques ou rétrospectives ;
- Observations ;
- Focus groups ;
- Analyse de contenu assistée par ordinateur (CAQDAS : Sonal, NVivo, Atlas Ti...);
- La combinaison/articulation méthodes qualitatives \* méthodes quantitatives ;
- La prise en compte du spatial (données localisées, carroyage) ;
- Etc.

## Axe 4 : « Produire des données, est-ce incontournable ? de l'intérêt de l'analyse secondaire »

Produire des données est extrêmement coûteux (tant en temps qu'en argent). De nombreuses enquêtes déjà réalisées peuvent être réutilisées pour étudier un thème, même en partie différent de celui pour lequel l'enquête a été produite. Un vaste mouvement international d'archivage et de mise à disposition des données (quantitatives, mais aussi plus récemment qualitatives) permet de réutiliser des enquêtes. Mais selon quelles modalités accéder à ces données ? Quelle qualité attendre de ces données ? Quelles difficultés spécifiques à la réutilisation doit-on s'attendre à devoir surmonter ?

Exemples (non exhaustifs) :

- Ré-analyse d'enquête qualitative ou analyse secondaire d'enquête quantitative et/ou problèmes spécifiques ;
- La question de la mise à disposition des données et corpus ;
- Etc.

## Axe 5 : « La production de données en SHS : préoccupations transversales »

La production de données en SHS inclut plusieurs étapes incontournables que le 5<sup>e</sup> axe de l'ANF propose de porter à la discussion. Il s'agira d'aborder les principales difficultés rencontrées relatives aux aspects juridiques de la collecte (accès aux documents administratifs, secret statistique, secret fiscal, évolution des lois et des techniques, données commerciales, Web of science...), mais également de sensibiliser les stagiaires à penser l'archivage et la documentation des bases de données dès la phase de la collecte (métadonnées, etc).

Exemples (non exhaustifs) :

- Déontologie des enquêtes de terrain en sciences sociales (problématiques des comparaisons internationales...);
- Les problèmes d'accès aux données et les plans de gestion de données ;
- Les licences *creative common* ;
- Les aspects juridiques des enquêtes ;
- Les problèmes posés par le dépôt des données (question de l'anonymat des enquêtés des enquêtes qualitatives...);

- Etc.

## Axe 6 : « se poser la question de “quelle est la preuve par la donnée ?” pour monter en généralité et élargir le débat »

Durant toute l'ANF, l'objectif aura été d'améliorer le recueil des données avec comme idée sous-jacente que la qualité de la donnée garantit la solidité de la preuve scientifique. Mais la donnée seule suffit-elle pour faire preuve ? Quelle est la place du traitement des données dans l'établissement de la preuve ? Peut-on réaliser des travaux intéressants avec des données de qualité moyenne ou comment en faire une force ? Comment tenir compte d'une qualité perfectible de ses données dans l'analyse ? Mais également, dans quelle mesure nos méthodes de production des données nous préservent-elles de la méconduite scientifique (et comment pouvons-nous aider les chercheurs à l'éviter ?) ?

Exemples (non exhaustifs) :

- Les méconduites scientifiques (données partiellement utilisées, résultats uniquement dans le sens des hypothèses, données et scripts pas suffisamment bien archivés pour permettre leur reproduction, etc.) : comment éviter le mal-travail ? ;
- Des interventions sur qualité des données / qualité du traitement / qualité du résultat ;
- Etc.

## Ateliers-découverte de logiciels sur la collecte et l'analyse des données quali et quanti.

Deux temps seront consacrés à la présentation de logiciels : ces ateliers ont pour but soit de présenter des mises en œuvre de méthodes originales (partant du principe que la méthode est connue), soit de proposer un regard particulier sur les logiciels libres en déclinant les avantages et les inconvénients de la migration vers le libre (par exemple : présentation comparées sur un même jeu de données entre Alcest/IraMuteQ, Spad/Factomoiner...).