

Soft data. Essai d'une nouvelle définition des données numériques pour les sciences sociales

Marta Severo, Geriico, Université de Lille 3

Les soft data, appelées « soft » en opposition aux « hard data » traditionnelles produites par les institutions statistiques, peuvent être définies de façon très générale comme des informations disponibles sur Internet, non contrôlées par une administration publique. Elles sont un effet de l'usage des médias électroniques. En effet, toutes les interactions qui traversent ces médias laissent des traces numériques qui peuvent être facilement enregistrées, massivement stockées, puis récupérées et analysées. Ainsi, les médias numériques offrent de nouvelles bases de données énormes qui peuvent être utilisées pour améliorer l'analyse des phénomènes sociaux et, par conséquent, le processus de prise de décision qui leur est lié.

Les traces numériques ne sont pas seulement produites de façon automatique par les technologies numériques, mais aujourd'hui, nous avons aussi de grandes quantités de données provenant de nouveaux fournisseurs de données tels que des membres de réseaux sociaux en ligne et des utilisateurs des plateformes de partage de contenu. Ces nouveaux types de données issues du Web 2.0 s'offrent au décideur public et au chercheur comme une source originale et riche d'informations sur les phénomènes sociaux.

Cette communication sera organisée en deux parties. Dans une première partie, les défis méthodologiques que l'usage de ces données soulève seront présentés. Un court état de l'art permettra de montrer comment les sciences sociales sont passées d'une attitude très optimiste vers ces nouvelles données du web à une posture bien plus prudente vers leur usage dans des études scientifiques. Attention particulière sera donnée à la naissance et développement d'une nouvelle série de méthodes, dites méthodes numériques, considérées plus adaptées pour traiter ces nouvelles sources d'information présentes sur la Toile.

Dans une deuxième partie, la définition de soft data sera approfondie. Après avoir marqué la différence entre cette étiquette et d'autres étiquettes comme 'big data' et 'open data', nous déclinons les caractéristiques de ces données qui montrent l'imperfection des définitions déjà existantes et justifient la nécessité d'une nouvelle définition.