Explorer associations et transformations à travers des datascapes

"There is thus no breach in humanistic epistemology. Whether knowledge be taken as ideally perfected, or only as true enough to pass muster for practice, it is hung on one continuous scheme." William James

Paul Girard, Mathieu Jacomy et Christophe Leclercq
Sciences Po, médialab, Paris, France

Présentation : http://medialab.github.io/publications/datascape@scienceXXL

Le développement des technologies numériques engendre une longue et profonde mutation de notre rapport à la connaissance. Quelle que soit l'étiquette utilisée pour désigner ce phénomène, de nombreuses personnes s'interrogent sur les évolutions de nos pratiques académiques (Lazer et al 2009, Ollion & Bollaert 2016).

Explorer les parties pour construire les touts

Latour et al avancent en 2012 que de nouvelles façons de représenter et surtout de naviguer dans les données permettraient de revenir sur le rapport entre tout et partie, débat aussi vieux que la sociologie elle-même (Latour 2012). D'après ces auteurs, chaque tout n'est qu'une façon particulière de voir les parties, un trait commun qui rassemble. Un trait parmi d'autres. Le tout est une prise tellement utile, tellement signifiante qu'on en oublie qu'il cache une réduction d'une foule de particularités. Or ce tout peut aujourd'hui être décomposé ou plutôt recomposé dynamiquement dans de nouveaux moyens d'exploration des données appelés *Datascapes* - autrement dit paysages de données. Reprenant ces intentions, nous avons depuis 2012 conçu des outils d'analyse exploratoire de données (Tukey 1977) qui permettent de multiplier les perspectives sur un même objet.

Les mondes de l'art de Experiments in Art and Technology (E.A.T.)

La première enquête utilisant l'approche Datascape s'intéresse aux collaborations artistes-ingénieurs en étudiant les archives de l'association Experiments in Art and Technology (Leclercq et Girard 2013, Leclercq 2016). Des informations sur ces activités sont extraites des documents d'archives par le chercheur, pour être inscrites dans une base de données. La vie des œuvres, projets et acteurs de l'association sont alors mises en forme dans une série de visualisations de données interactives dont l'exploration révèle un monde de l'art particulier (Becker 1984) - monde composé d'acteurs définis par leur participation à des œuvres, celles-ci n'étant rien d'autre qu'une liste d'acteurs ayant participé à leur vie. La réversibilité des "entités" est exploitée par l'exploration hypertextuelle. En explorant, le chercheur suit peu à peu le fil des associations pour découvrir des agencements, singularités et similitudes. C'est dans la dynamique de l'exploration qu'on peut saisir la complexité des agencements de parties déployant leur particularité.

Les chaînes de référence

La reconfiguration par le mouvement de cette *"métonymie très économe en science comme en politique par laquelle une minuscule partie permet de saisir l'immense tout"* doit être complétée par une autre dynamique : celle qui permet de circuler le long de la chaîne de transformation de la référence (Latour 1993). Cette chaîne de transformation réduit peu à peu la complexité, la variété du réel. Le chercheur, par sa pratique, isole des *traits pertinents* d'un terrain, d'une observation pour les amplifier et les rendre opérationnalisables dans son

discours, pour répondre à sa question. Cette chaîne doit demeurer réversible, afin que chaque étape s'imbrique dans la suivante, transportant et transformant la réalité jusqu'au lecteur tout en permettant à ceux qui doutent, y compris lui-même, de se *référer* aux étapes précédentes. La chaîne de référence est construite *par approximations successives,* par continuités et ruptures *entre chose et signe*, entre matière et forme.

Décomposer la chaîne de fabrication de la loi

Le site La fabrique de la loi propose de remonter la chaîne de transformation qui aboutit aux textes de loi produits par le parlement français en décrivant 210[1] des 415 textes de lois adoptés depuis 2010. On peut y visualiser la durée de gestation entre dépôt, interruption, Assemblée nationale, Sénat, commission mixte paritaire et promulgation. En choisissant un texte, on découvre les versions intermédiaires générées par chaque étape de la fabrication du texte. Choisir une étape révèle non plus seulement le résultat mais la longue liste des amendements proposés puis les débats qu'ils ont engendrés. Ce Datascape déploie la complexité de la procédure, dépliant étape par étape chaque loi en autant de débats, propositions et ajustements qui la constituent.

Explorer le commerce aux XVIIIème et XIXème siècles

Deux projets d'histoire de l'économie RICardo (Dedinger & Girard 2017) et TOFLIT18[2] proposent de réinterpréter les archives des statistiques du commerce, au XIXème siècle bilatéral et mondial d'une part, en France par produits au XVIIIème d'autre part. L'exploration visuelle et interactive de telles données demande de contextualiser toute agrégation en révélant hétérogénéité et données manquantes (Girard et al 2016). Dans TOFLIT18, l'hétérogénéité des dénominations de marchandises sur un siècle à l'échelle d'un pays comme la France aboutit à une liste de plus de cinquante mille variantes. Pour permettre des analyses quantitatives, nous avons conçu et développé un système de classification de ces dénominations qui est progressif (différents niveaux d'agrégation), concurrent (plusieurs agrégations thématiques différentes à un même niveau) et dynamique (toute visualisation/agrégation commence par le choix d'une classification) (Girard & Plique 2016). L'hétérogénéité devient alors une richesse qui peut être exploitée en fonction de sa question en utilisant ou en créant une classification. En cas de doute, chaque classification peut être décomposée pas à pas jusqu'aux sources.

Diversité des données

Chacun de ces outils a été conçu et développé avec et pour une équipe de recherche. S'ils partagent la mise en pratique des mêmes concepts méthodologiques, ce sont des solutions techniques et des modèles de données dédiés à des intentions spécifiques (Bowker 2000). Leur originalité et leur intérêt ne reposent pas sur les algorithmes, mais bien dans l'hybridation qu'ils proposent entre techniques quantitatives (système de base de données, agrégation visuelle, etc.) et expériences qualitatives (nettoyage et classement manuels des données, exploration et interprétation des formes, etc.). Cette combinaison est rendue possible par l'abaissement du coût de développement informatique qui permet de créer des instruments de gestion et d'exploration visuelle et interactive de données, dédiés à une recherche interdisciplinaire mêlant sciences humaines et sociales, ingénierie informatique, design d'interaction et de l'information.

---

[1] extraites dans la douleur par le collectif Regards Citoyens faute d'open data officielles
[2] accès au datascape http://toflit18.medialab.sciences-po.fr login:toflit18 pass:douelle (confidentiel)

Bibliographie

Becker, Howard Saul. 1984. *Art Worlds*. University of California Press.

Bowker, G. C. 2000. « Biodiversity Datadiversity ». *Social Studies of Science* 30 (5): 643‑83. doi:10.1177/030631200030005001.

Dedinger, Béatrice, et Paul Girard. 2017. « Exploring trade globalization in the long run : the RICardo project ». *Historical Methods*. http://ricardo.medialab.Sciences-po.fr/Dedinger_Girard_RICardo_HistoricalMethods.pdf.

Girard, Paul, Béatrice Dedinger, Donato Ricci, Benjamin Ooghe-Tabanou, Mathieu Jacomy, Guillaume Plique, et Grégory Tible. 2016. « RICardo Project : Exploring XIX Century International Trade ». In . Kraków, Poland. http://ricardo.medialab.sciences-po.fr/Girardetal_RICardo_dh2016_fr.pdf. http://medialab.github.io/ricardo/

Girard, Paul, et Guillaume Plique. 2016. « Des archives du commerce à des données quantifiables, une longue chaîne de transformation des données ». présenté à ANF MATE-SHS, Collecter et produire des données pour la recherche en SHS, Fréjus, novembre 16. http://medialab.github.io/toflit18/ANFmateSHS.

James, William. 1905. *The Essence of Humanism*. The Journal of Philosophy, Psychology and Scientific Methods. http://archive.org/details/jstor-2012206.

Latour, Bruno. 1993. « Le topofil de Boa-Vista. La référence scientifique: montage photophilosophique ». *Raisons pratiques* 4: 187‑216.

Latour, Bruno, Pablo Jensen, Tommaso Venturini, Sébastian Grauwin, et Dominique Boullier. 2012. « 'The Whole Is Always Smaller than Its Parts' - a Digital Test of Gabriel Tardes' Monads ». *The British Journal of Sociology* 63 (4): 590‑615. doi:10.1111/j.1468-4446.2012.01428.x.

Lazer, David, Alex Pentland, Lada Adamic, Sinan Aral, Albert-Laszlo Barabasi, Devon Brewer, Nicholas Christakis, et al. 2009. « Social Science. Computational Social Science ». *Science (New York, N.Y.)* 323 (5915): 721‑23. doi:10.1126/science.1167742.

Leclercq, Christophe. 2016. « Le projet E.A.T. Datascape ou ce que le design numérique peut faire à l'histoire sociale de l'art ». In *à paraître*. www.collecta.fr.

Leclercq, Christophe, et Paul Girard. 2013. « The Experiments in Art and Technology Datascape ». In *Collections électroniques de l'INHA. Actes de Colloques et Livres En Ligne de l'Institut National D'histoire de L'art*. INHA. http://inha.revues.org/4926.

Ollion, Étienne, et Julien Boelaert. 2015. « Au-delà des big data ». *Sociologie*, n° N°3, vol. 6 (novembre). https://sociologie.revues.org/2613.

Tukey, John Wilder. 1977. *Exploratory Data Analysis*. Addison-Wesley Publishing Company.