

“Quels outils/méthodes/animations autour des données existe-t-il déjà ? Ou comment construire un guide des médiations”

La Campagne Infolab a été lancée le jeudi 11 avril à l'occasion de la première rencontre du groupe de travail national, portant sur le thème “ Culture de la donnée et médiations”. Acteurs associatifs (Libertic, Open Street Map, collectifs citoyens Open Data Tours, Poitiers, Les Petits-Débrouillards, Décider Ensemble, La Fonderie...), membres de collectivités territoriales (Région PACA, Communauté urbaine de Bordeaux, CG Gironde, CG Hauts-de-Seine, CG Manche, Mairies de Paris et de Rennes...) et grands comptes (Alcatel Lucent, La Poste, La Caisse des dépôts, Gemalto...) étaient rassemblés pour entamer une réflexion sur la culture de la donnée et ses différentes formes de “médiation”. Retour sur cette journée...



La dernière présentation de la journée, menée par Amandine Brugière (FING), portait sur l'analyse de la veille qu'a initiée, depuis 2 mois, l'équipe INFOLAB (voir sur [Diigo le fil de veille](#), ainsi qu'ici le [tableau partagé](#) où sont référencés plus d'une centaine de « dispositifs de médiation », qu'ils soient formats d'animation, outils, méthodes...L'objectif n'étant pas de capter l'exhaustivité mais la diversité des exemples).

Type de médiation 1) Outils-logiciels, 2) Ressources documentaires, 3) Animation, 4) Serious Game, 5) Plateforme (?), 6) Méthodologies (?), 7) Offre de service, 8) Experts	Type d'actions réaliser	Nom	Lien de référence	Porté par qui ?	Pour quoi faire ?	Comment ?	Avec quelles données ?	Pour quel public ?	Langue	Niveau de difficulté / prise en main	Rmq
	Convertir, relier, nettoyer des données	Digikam			photos à usage via tagging	photos et vidéos via tag, reconnaissance faciale			Multilingue	Débutant	Outil open source
0		Mr.Data Converter	http://shancar		Transformez des données Excel ou CSV en format "web-friendly" : XML, JSON, HTML, SQL et pleins d'autres.				Anglais	Débutants	
0		Data WRangler	http://vis.stan	Stanford	Application web interactive pour classer, formater et optimiser l'analyse de données	On importe des données. L'outil les repropose au sein d'une application où l'on peut trier, croiser, analyser, etc.			Anglais	Débutants	
0		Datalift	http://datalift	Consortium Datalift (Atos, INRIA, Mondaca, IGN, INSEE, Eurecom, Fing)	Plateforme de conversion et de croisement de données			Tous publics	Français et anglais	Avancé	
0		Open Refine	http://openref	Communauté OpenRefine				Tous publics	Anglais	Intermédiaire	Appelé auparavant Gooq
0		Langage R	http://www.r-project.org/	Communauté du langage R	Langage de manipulation de données	Programmation	Toutes données		Anglais (des ressources en français)	Avancé	
		Sherpa Roméo (similaire à Grokkr)	http://www.sh	Bill Ferrell & CEO	"Application pour smartphone qui détermine le meilleur moment et lieu pour vous afficher de l'information."	Grâce au "machine learning", l'application comprends le contenu de mails, calendrier, etc. Elle classe ensuite les données puis en extrait des informations clefs. Elle finit par créer des zones géographiques appliquées aux mots clefs ; une fois dans la zone, Sherpa nous délivre les informations relatives à ma position (par exemple : informations sur mon vol).					
	Anonymiser des données	DataExplorer	http://explorer	OKFN	Voir, nettoyer et partager des données				Anglais	Intermédiaire	
		Pseudonymisation					Elle permet la suppression des champs considérés comme des identifiants potentiels.				

Il s'agissait de livrer une première photographie de ce qui est directement accessible sur le web : Si on cherche aujourd'hui à exploiter des données, quelles soient ouvertes, fermées, personnelles ou publiques, quels sont les outils, les méthodes, les formats d'animation documentés et accessibles sur le web ? Qu'est-ce qui se trouve « à portée de main » ?

Première réunion du groupe national d'Infolab. from La Fing

Cette analyse de la veille s'est structurée autour :

- du type de médiation développée : Est-ce que l'on est face à des outils, des méthodologies, des formats d'animation, des services, des ressources documentaires, des serious game, des plateformes...?

- du type d'actions possibles : Qu'est-ce que cela permet de faire ? Sensibiliser, explorer, imaginer, partager, analyser...? Les catégories qui ont été retenues ici ne sont pas des catégories « stabilisées ». Le but de la présentation de ce work-in-progress était de faire réagir les participants. Ainsi n'hésitez pas à compléter le tableau, à le commenter, à partager les ressources identifiées, à nous contacter pour en discuter (infolab@fing.org). Ce travail s'enrichira au fil des rencontres, grâce à vos contributions.



Sensibiliser aux données

La sensibilisation aux enjeux de la donnée reste, visiblement, un domaine à investir : peu d'exemples émergent, relativement aux autres catégories d'actions.

Ceux trouvés montraient cependant une grande diversité de formes. Cela allait des ateliers rennais de [sensibilisation des enfants à la manipulation et la représentation de données](#), au serious game sur [les données ouvertes du budget de l'Etat](#), ou encore [l'enquête Orange-lab-OWNI-FING](#) « Où habitez-vous vraiment ? », sorte de quizz – à base de données ouvertes - testant les connaissances territoriales des habitants. Les plateformes [Openstreetmap](#) ou [OpenfoodFacts](#) (dé)montrant facilement la force du crowdsourcing (la production collaborative de données).

Un exemple particulièrement pédagogique – mais s'adressant tout de même à un public d'initiés – est la [School of Data](#) de l'association OKFN : Une des seules initiatives proposant des ressources documentaires et méthodologiques.

On le voit, ces formes de médiation développent chacune des pédagogies très différentes, face à la complexité des données. Elles n'hésitent pas à investir le « ludique » et s'adressent à tout type de public.



Imaginer et produire des usages de la donnée

Avoir des données à disposition est nécessaire mais pas suffisant. Encore faut-il avoir des idées de (ré)utilisation des données, et souvent faut-il même stimuler les réutilisateurs potentiels.

Dans ce champ-là, les exemples trouvés relèvent essentiellement de l'animation : Concours ou appel à projet, hackathon, [dataconnexion](#), Data-apéro ou [Data-Tuesday](#). On est dans le champ de l'incitation, de la stimulation, de la recherche de sérendipité, ou de l'innovation ouverte.

3 caractéristiques communes apparaissent :

- Ces formats d'animation s'adressent à des publics relativement avertis : Innovateurs, start-uppeurs, designers, développeurs...

- Ce sont majoritairement des exercices collectifs : la créativité découle-t-elle du croisement de regard ?
- Ils sont d'une durée relative.
- Ils s'adaptent à des données ouvertes comme « propriétaires ».

A noter toutefois un outil de « sérendipité » : [Creadata](#), développé par Bastien Kerspern proposant des idées de mash-up aléatoires entre les jeux de données mis à disposition par la ville de Nantes.



Cuisiner la donnée

Une fois que les données et les idées sont là, s'ouvre alors un champ d'actions relativement techniques : Récupérer les données ([EasyOpenData](#)) les stocker, les nettoyer ([Open Refine](#), [DataWrangler](#)), les anonymiser ([norme ISO de pseudonymisation](#)), les croiser ([DataLift](#)), etc. Ce champ-là foisonne d'outils destinés à des publics très spécialistes.

Beaucoup d'entre eux émergent du champ des données personnelles : du côté du VRM Vendor Relationship Manager ([My Personal Vault](#)), du quantified self – la mesure de soi ([BodyMedia](#)), ou des pratiques de développement durable ([Green Button Data](#)), pour ne citer que ces exemples-là. La relation aux données peut même être médiée par des objets ([Withings](#)).



Partager des données

Quant on cherche à partager des données brutes, ou à communiquer sur leur analyse, plusieurs solutions s'offrent à nous : La plus connue d'entre elle est la datavisualisation. Énormément de solutions techniques existent en la matière : Voir la très riche boîte à outils d'[ExpoViz](#) (made in [La Fonderie](#)). Ces dataviz vont du graphique / diagramme classiques ([Google Image Chart](#)) à des représentations plus sophistiquées ([Treemap](#), [graphe du web](#), [cartographie](#)). Tous ces outils nécessitent un haut niveau de compétences techniques.

D'autres outils permettent de partager des données avec des tiers, en maîtrisant, autant que possible, les termes de l'échange. Là encore le champ des données personnelles est riche ([Mine!Project](#), [Personal.com](#)). Dans le champ des données de santé, on trouve des plateformes de partage collectif de données ([Curetogether](#), [Asthmapolis](#)).



Analysier

Dans le champ de l'analyse des données, ce qu'on identifie comme outils ([Precog](#), [Continuity](#), [StatMix](#), [Coheris Liberty](#), [Google Prediction API](#)), plateformes de services ([DataPublica](#), [Qunb](#), [Captain Dash](#)) offrent la possibilité de construire des modèles décisionnels, voire des modèles prédictifs, via des interfaces finales, des tableaux de bord. La complexité des opérations sous-jacentes liées au traitement des données n'est pas explicitée : Comment s'opèrent les corrélations entre jeux de données, quels sont les considérations, les choix qui prévalent aux datavisualisations, aux tableaux de bord? Comment se font les modélisations, les simulations ? Où se trouve la part de décision humaine dans les process ? Les pondérations dépendent-elles essentiellement des profils techniques ?

A l'opposé, dans le champ des sciences humaines, émergent des méthodologies – et outils – permettant de mettre les données en débat : la méthode développée par l'ANACT / AlterCarto autour de la [plateforme ATTLAS](#) montrait comment les acteurs pouvaient monter en compétences et opérer leur propre corrélation de données. La méthode de [cartographie des controverses](#), développée par Bruno Latour, va aussi dans ce sens. Face à des phénomènes sociaux (scientifiques, techniques) non stabilisés, les cartographies des controverses mettent en visibilité différents points de vue qui s'opposent, afin de favoriser l'objectivité dans les prises de décision qui peuvent découler.

Et maintenant, que faisons-nous ?

Ce travail de recensement et d'analyse des formes de médiation existantes (qu'elles soient humaines, sous forme d'outils et de méthodologies, de ressources documentaires ou de services...) ne fait que commencer. Il se poursuivra jusqu'à la rentrée 2013.

Livrer ce premier niveau de lecture était instructif rien qu'au regard :

- de la nature des ressources disponibles : plus « outils » que ressources documentaires ou méthodologiques ou ludiques ;
- du niveau requis de compétences pour leur utilisation : du geek au développeur professionnel, majoritairement.

Les données sont encore largement affaire de « spécialistes »... On le savait. Quasiment tout reste à faire en matière de « démocratisation des données » et à tout niveau : sensibilisation, information, formation, accompagnement, services.

Heureusement, la campagne Infolab ne fait que commencer !