

1, 2, 3 data, expérimentez !

Ce cahier propose un panorama du monde de l'exploration et de l'expérimentation avec les données. Il montre comment l'expérimentation peut être à la portée de tous, avec des moyens adaptés à toutes les organisations.

Ce cahier est le résultat d'un travail et d'échanges collectifs réalisés tout au long de l'année 2016 au sein de la Campagne Infolab de la FING. Le groupe a été animé par Charles Népote, avec l'étroite collaboration d'Armelle Gilliard et la complicité de Sarah Labelle et Simon Chignard. Le groupe était composé d'Emmanuelle Katz et Axelle Reynier (Groupe UP), Isabelle Micheu et Zouhoura Hadji (Groupe La Poste), Érik Arnaud, Véronique Morault et Sophie Vitrat (MAIF). Le cahier a été rédigé sous la direction de Charles Népote.

Retrouver ce document en ligne :

<http://infolabs.io/xp16>

INTRODUCTION

Expérimenter avec des données, pourquoi faire ?

Les données ce n'est pas nouveau. Voilà des années qu'elles fournissent une des matières premières du fonctionnement et du développement des entreprises. Pourquoi vouloir expérimenter avec les données ?

La première raison tient à l'évolution du paysage même des données. Il y a quelques années encore, les données étaient l'apanage des grandes entreprises ou des grands acteurs publics (État, grandes collectivités...). Aujourd'hui presque tout le monde produit des données (consciemment ou non). Multiplication des sources de données, développement sans précédent des capacités de stockage et de traitement : tout concourt à rendre les données plus nombreuses et plus accessibles. Il est temps de s'approprier cette matière nouvelle et si mouvante qu'elle résiste à un apprentissage classique. L'expérimentation est la clé de sa découverte !

Une seconde raison d'expérimenter avec les données tient dans la nature même de la donnée, un matériau réputé aride, froid, technique. Tout semble indiquer qu'il est réservé aux seuls spécialistes. Quel jeune enfant se rêve adulte à manipuler des colonnes de chiffres, de libellés, de codes ?

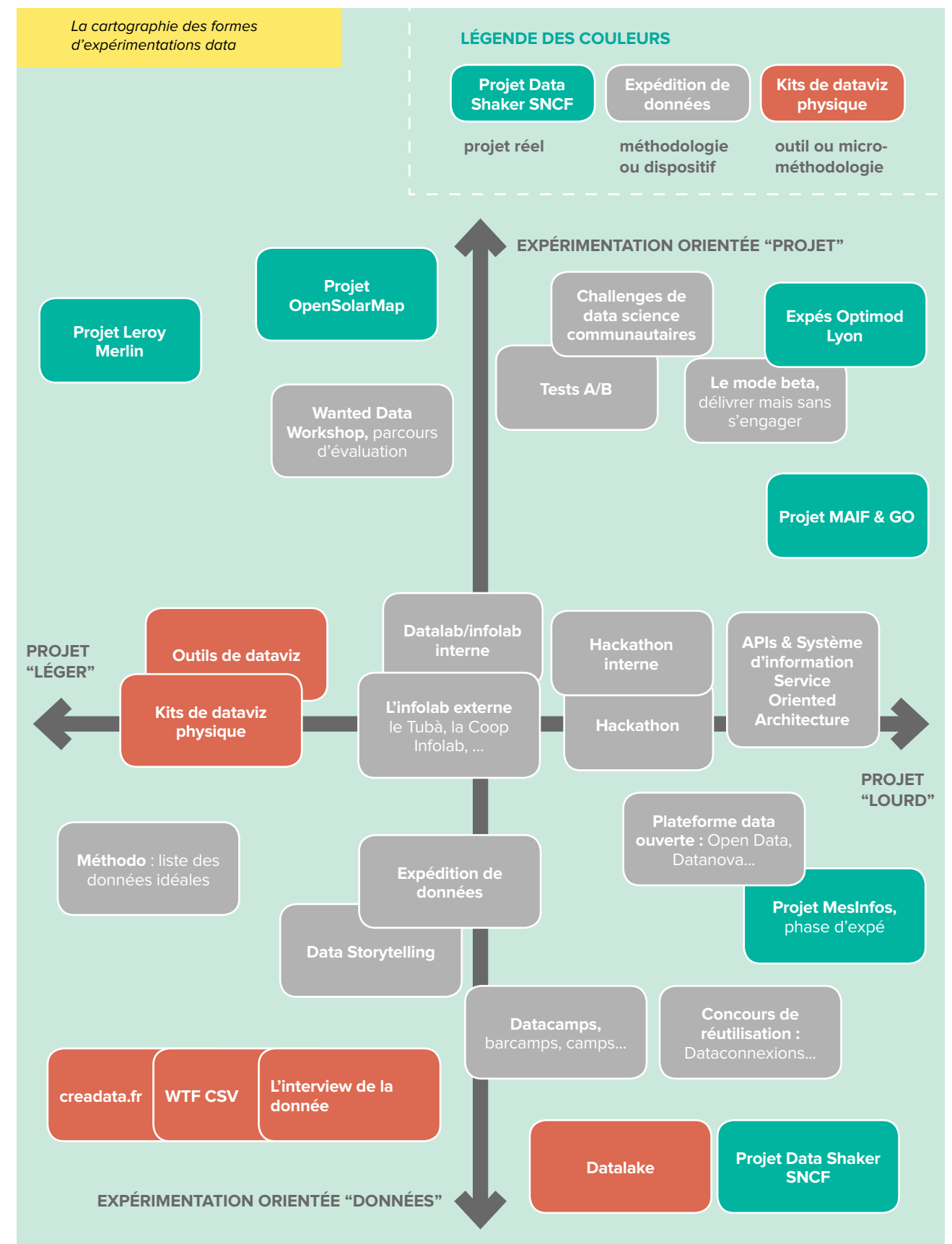
L'expérimentation permet de rendre plus tangible ce matériau, de le triturer, de produire des croisements imprévus et féconds. L'expérimentation dédramatise, suscite la curiosité, peut devenir une véritable aventure d'équipe. Elle peut rendre la donnée attractive voire stimulante.

Enfin, il nous semble qu'en matière de données, rien ne remplace l'expérimentation. Seule la

réalisation concrète permet de saisir les opportunités et les limites des données. Dans ce domaine, il est difficile de séparer la théorie de la pratique des données. L'expérimentation est la clé de l'apprentissage.

Quelles expérimentations au juste ?

Qu'est-ce qu'une expérimentation data ? Où les trouver ? Nous avons repéré plusieurs dizaines d'entre elles et avons souhaité une vision d'ensemble sous la forme d'une cartographie. Deux caractères sont rapidement apparus structurants. Tout d'abord, la cible même de l'expérimentation. D'un côté nous avons observé des expérimentations orientées projets : l'expérimentation sous-tend un besoin clair, avec ou sans données, il peut s'agir par exemple d'accompagner un projet d'innovation de l'entreprise ; l'expérimentation vise à voir comment les données transforment le problème, le besoin ou le projet. A l'inverse certaines expérimentations portent sur les données elles-mêmes, sans projet ou besoin clairement identifié : à l'extrême, il s'agit d'une pure phase d'exploration des données. Le second critère structurant tient à la "taille" de l'expérimentation. Certaines expérimentations ne prennent que quelques heures, avec peu de moyens et des compétences rudimentaires, mise en production comprise ! Quand d'autres expérimentations se déploient sur des mois, embarquant jusqu'à plusieurs dizaines de personnes.



Formats clé en main

Nous présentons ici des formats clé en main pour l'expérimentation. D'ores-et-déjà tous éprouvés, les temps renseignés dans chacune des fiches donnent un ordre de grandeur. Les temps de préparation et de suivi sont dépendant de la résonance que l'on souhaite donner à l'expérimentation.

La liste des données idéales

Aussi appelée *Wanted Data List*. Il s'agit de la liste des données idéales qui seraient utiles ou que l'on pense en rapport avec un sujet donné.

Quoi ?	Atelier créatif : lister les données idéales sur un sujet donné. Voir : http://infolabs.io/liste-donnees-ideales
Qui ?	* Équipes projets, sans compétence particulière, d'univers hétérogène * 5 à 20 participants
Combien de temps ?	Préparation : 30 minutes la première fois Animation : 30-40 minutes environ Suivi : 2 h pour formaliser les résultats.
Qu'est-ce que cela produit ?	* Cartographie de données * Cohésion d'équipe * Fait ressortir les compétences de chacun
Exemple ?	Voir celui qui est présenté dans la fiche méthodo

Le hackathon (et l'après hackathon)

Pour de nombreux acteurs, le hackathon représente un format d'expérimentation clé. De fait, il produit une effervescence et des résultats tangibles en un temps record. Il connaît d'ailleurs un gros succès : des entreprises, des communautés, des territoires, des acteurs publics, des collectifs citoyens, s'en saisissent pour un oui ou pour un non. Les plus gros y ont recours : Axa,

C'est une manière de prendre pied avec un sujet. C'est un exercice simple à mettre en oeuvre, au cours duquel les participants sont souvent très créatifs. Il ne demande aucune compétence sur les données, c'est donc une manière d'embarquer des équipes très hétérogènes sur un sujet "data", en commençant par leur faire réussir un exercice collectif simple.

C'est enfin un exercice très amont dans le processus d'expérimentation ou d'exploration. Il peut être le point de départ d'explorations plus complètes : ces données idéales existent-elles dans la réalité ? où les trouver ? etc.

Orange, Pernod Ricard, la SNCF, la RATP, Vinci, le ministère de l'Environnement, de l'Énergie et de la mer, celui de la Ville, de la Jeunesse et du sport, etc. Même le très sérieux Figaro en fait des gorges chaudes et clame que le hackathon "réinvente l'innovation en entreprise" (rien que ça !). Il est vrai qu'il a de quoi séduire : il crée de la rencontre, plus spontanée, et de l'émulation.

Quoi ?	Atelier créatif : à partir de données fournies pour l'occasion, concourir pour faire germer des idées d'usages innovants
Qui ?	* Équipes projets d'univers hétérogènes ; un développeur par groupe indispensable, graphistes et designers très appréciés * 30 à 100 participants répartis en équipes de 3-7
Combien de temps ?	Préparation : Minimum une semaine Animation : Une journée minimum ; idéalement 2 jours nuit comprise. Suivi : tout dépend de vous !
Qu'est-ce que cela produit ?	* 4-12 maquettes d'usage voire des protos fonctionnels * Crée de la rencontre ; fait naître des équipes * Fait ressortir les compétences de chacun * Effet de communication sur des cercles éloignés
Exemple ?	Le hackathon Vinci Autoroute de mai 2014, a poursuivi un double objectif : en interne sensibiliser les cadres de haut niveaux à cette forme d'interactions avec le grand public. Et en externe repérer des jeunes pousses et les soutenir après le hackathon, notamment en travaillant avec les spécialistes du groupe. Les meilleurs candidats ont exposé au Salon de l'automobile qui a suivi. > http://www.vinci-autoroutes.com/fr/article/vinci-autoroutes-recompense-les-laureats-du-hackathon-dedie-aux-mobilites-du-quotidien

Malgré son succès qui ne se dément (presque) pas, des critiques s'élèvent contre le hackathon².

Le hackathon, c'est un truc vendu par des consultants à de vieilles orgas animé par des idéologues pour faire bosser des esclaves...
—@nodesign

... un écosystème où les développeurs sont toujours considérés comme des unités de production, et non pas comme les designers de solutions à haute valeur ajoutée ...
-- Valentin Squirelo

Un autre type de critique tient à la temporalité de l'exercice : certains parlent d'événement "en cloche" dont la valeur retombe très rapidement. De fait, un très grand nombre de projets meurent

le lendemain du hackathon (la quasi totalité ?). Pour être réussi, un hackathon ne doit pas être pensé comme un événement ponctuel, mais bien comme une étape dans une démarche projet plus large. On peut par exemple imaginer un hackathon comme point de départ ou comme point d'orgue d'une démarche d'ouverture des données. Dans le premier cas, illustré par exemple par l'évènement OpenDAMIR organisé par l'Assurance Maladie en 2014, le hackathon permet de préparer l'ouverture d'un important jeu de données (les dépenses détaillées de santé), de mobiliser les équipes internes au sein de l'administration et de tester in vivo l'intérêt d'un public plus large pour ces données. A l'issue de la journée, les responsables de l'Assurance Maladie ont pu ainsi annoncer que les données

1. Lucie Ronfaut : Comment le "hackathon" réinvente l'innovation en entreprise, Figaro du 31/03/2014. <http://www.lefigaro.fr/secteur/high-tech/2014/03/31/32001-20140331ARTFIG00112-hackathons-les-marathons-de-l-innovation-font-courir-les-geants-de-l-industrie-et-des-services.php>
2. Nous renvoyons notamment à l'excellente synthèse de Paul Richardet (dont sont extraites les citations suivantes) : Hackathon : pipo ou bingo ? Pour une innovation (con) descendante ? <https://medium.com/paul-r/hackathon-pipo-ou-bingo-e0368c17e410#1e21yben4>

prises à disposition des participants le seraient à tous, en open data, quelques semaines plus tard. Dans un autre registre, les hackathons organisés par le Ministère de l'Environnement s'inscrivent dans une démarche plus globale autour de la GreenTech Verte, outre les hackathons on y retrouve un incubateur de startups, des concours de datavisualisation, etc. Le hackathon est ici pensé comme l'un des moyens pour faire interface avec les écosystèmes tiers.

Les concours de réutilisation

Il s'agit d'un format classique qui a beaucoup été utilisé par les collectivités territoriales et

quelques acteurs privés, pour la promotion de leurs données "open data". Le principe en est simple : sur la base de jeux de données publiés, il est proposé d'élire les meilleures idées, maquettes ou prototypes d'usage de ces données. Un règlement clair, une date limite de dépôt, un jury d'experts, une dotation (si possible), un peu de communication et d'animation, et le tour est joué ! Certains ont pu donner des résultats impressionnants comme les concours Open PACA et celui de Rennes Métropole, alignant plus de 40 projets candidats. Certains concours ont une étonnante longévité et régularité comme Dataconnexions d'Étalab qui a connu sa 6e édition en 2016, avec un succès qui ne faiblit pas.

Quoi ?	Concours de réutilisation : à partir de jeux de données, les meilleures réutilisations sont récompensées
Qui ?	* Équipes métier, marketing, prospective, innovation, communication * jusqu'à plusieurs centaines de participants hétérogènes qui s'auto-organisent
Combien de temps ?	Préparation : a minima plusieurs semaines, souvent 3-5 mois Animation : selon le projet retenu. Suivi : Minimum quelques jours pour les dossiers récompensés
Qu'est-ce que cela produit ?	* De la communication sur la culture des données de l'organisation, de l'exposition des données, du repérage de pépites du territoire * Des projets d'usage (jusqu'à plusieurs dizaines) * De la rencontre (participant et organisateurs) * De la notoriété pour les projets lauréats
Exemple ?	> https://www.data.gouv.fr/fr/dataconnexions/ > http://opendata.regionpaca.fr/concours-regional-open-paca.html

Les challenges de data science communautaires en ligne

L'analyse de données complexes est-elle réservée aux data scientists, aux experts et sociétés spécialisées³ ? L'exploration et l'analyse de données peut-elle être plus accessible à tout un chacun ? C'est le pari tenté par plusieurs entreprises privées. Elles proposent aux organisations de

répondre à un problème métier, en organisant un vaste concours auprès d'une communauté de centaines de data scientists. La première d'entre elles, l'américaine Kaggle, est aujourd'hui l'une des plus connues, mais elle possède aujourd'hui plusieurs concurrents comme DrivenData, Tuned IT, Crowd Analytix ou bien le francophone datas-cience.net.

3. <http://www.internetactu.net/2012/02/16/la-democratisation-de-lanalyse-de-donnees/>

Quoi ?	Concours de résolution de problème : à partir d'un problème donné d'une organisation, un concours est proposé à une communauté de centaines de data scientists, sur la base d'un prix financier
Qui ?	* Équipes métier, marketing, innovation ; communauté de data scientists * jusqu'à plusieurs centaines de participants répartis en équipes
Combien de temps ?	Préparation : a minima plusieurs semaines, souvent 3-5 mois Animation : selon le projet retenu. Suivi : minimum plusieurs semaines, souvent plusieurs mois
Qu'est-ce que cela produit ?	* De nombreuses solutions pour un coût fixé d'avance (c'est un concours) * De la rencontre : entre les entreprises et les data scientists ; entre les data scientists entre eux * De la réputation : pour ceux qui remportent le challenge et qui s'en servent ensuite pour trouver ou changer d'emploi
Exemple ?	<i>En 2016, la MAIF a eu recours à datascience.net pour déterminer l'équation ou l'algorithme permettant de calculer les prix d'une couverture d'assurance automobile.</i> > https://www.datascience.net/fr/challenge/26/details

Les datacamps, barcamps et autre "camps"

Les "camps" et autres "barcamps", forment un autre type de format. Événements très hétérogènes, il n'existe pas vraiment de "camp" canonique. La plupart du temps l'événement

dure une journée et c'est l'exploration et la rencontre qui guide les participants : il ne s'agit pas de construire un produit mais plutôt d'explorer collectivement une question. Parfois la frontière avec le hackathon n'est pas si nette. Mais le camp ne récompense personne, il n'est pas tourné vers la compétition.

Quoi ?	Atelier créatif : à partir de données fournies pour l'occasion, se rassembler pour apprendre et faire progresser un sujet
Qui ?	Équipes projets d'univers hétérogènes ; compétence dev possible mais pas indispensable * 30 à 100 participants répartis en équipes
Combien de temps ?	Préparation : 2-5 jours Animation : souvent une journée ; éventuellement un peu moins et, pourquoi pas, jusqu'à 2 à plus de 3 jours Suivi : une journée
Qu'est-ce que cela produit ?	* De la connaissance voire des réalisations (maquette, proto...) ; de la documentation * Rencontres spontanées * Fait ressortir les compétences de chacun
Exemple ?	> https://www.etalab.gouv.fr/openasso-le-10-septembre-decouvrez-donnees-ouvertes-et-associations-participantes > https://www.etalab.gouv.fr/open-data-camp-elections > https://www.etalab.gouv.fr/datacamp-cadelect

Wanted Data Workshop ou “parcours d’évaluation data”

Wanted Data Workshop ou “parcours d’évaluation data”
Ce format d’expérimentation part d’une question simple : comment les données peuvent-elles

transformer un projet existant au sein d’une organisation ? Ils ont une dimension plutôt interne et très opérationnelle, ils permettent de mobiliser en un temps court des profils différents (informatique, marketing, commerce) sur la problématique data.

Quoi ?	Atelier créatif orienté projet : quelles données pour booster mon projet ?
Qui ?	* Équipe projet existante * 5 à 10 participants
Combien de temps ?	Préparation : une journée minimum Animation : une journée idéalement ; éventuellement 3-4 demi-journées réparties sur plusieurs semaines Suivi : une journée minimum
Qu’est-ce que cela produit ?	* De la connaissance commune de l’environnement data du projet ou de l’organisation. * Une priorisation des sources de données utiles, tenant compte de leur accessibilité * Un plan d’action data partagé
Exemple ?	* les Wanted Data Workshop du groupe UP (cf. encadré)

LES WANTED DATA WORKSHOP DU GROUPE UP

Dans le cadre de la démarche Data du groupe Up, un atelier Wanted Data a été déployé auprès de chacune des zones géographiques du groupe afin de :

- accélérer une montée en compétences Groupe, notamment des responsables Marketing
- créer dans les filiales du groupe un réflexe DATA pour identifier les données à ne pas oublier, dès la phase de démarrage des projets
- proposer une méthodologie duplicable afin d’ajouter un volet Data à tous les projets du groupe.

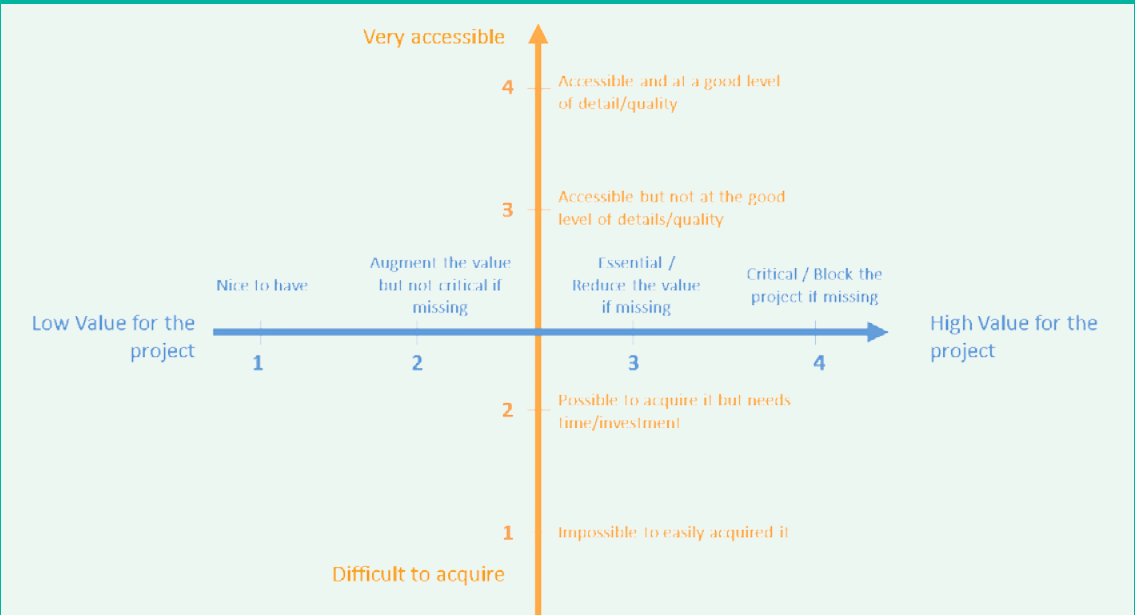
Inspirée par une méthodologie de la Fing, la journée est découpée en 5 ateliers autour d’un projet en phase de démarrage et préalablement sélectionné par le marketing.

1. Use cases definition : permet d’identifier et de prioriser les cas d’usages de données que la filiale souhaite adresser dans le cadre de ce projet. Un cas d’usage est sélectionné avec tous les participants pour la suite de l’atelier.

2. The Wanted Data List identification : permet de lister l’ensemble des données que l’on aimerait idéalement avoir afin de répondre à ce cas d’usage.

3. Data collection : un exercice de recherche et de vérification des données permet de sensibiliser les équipes à la collecte de données et d’identifier les difficultés ou interrogations qu’elles peuvent rencontrer.

4. Data mapping : les données sont positionnées en fonction de leur valeur pour le projet et de leur accessibilité, évaluées dans les 2 phases précédentes.



5. The Data Roadmap : cette dernière phase permet d’identifier les stratégies de sourcing de données à mettre en place donnée par donnée et de prioriser les prochaines actions à mener au sein de l’équipe afin d’assurer la bonne prise en charge du volet Data du projet.



L’atelier permet d’accompagner une équipe projet opérationnelle de 5 à 10 personnes : le responsable marketing de la zone, le/les responsables marketing locaux, le/les responsables Informatiques locaux. Et en fonction du projet/du pays : des experts métiers, des responsables Base de données/Data, juridique...

Pour quels résultats ?

- Un atelier très apprécié qui permet de travailler de façon agile, transversale et collaborative.
- Une méthodologie facile à dupliquer et utiliser sur différents projets et problématiques.
- Un outil de sensibilisation et de mobilisation des équipes : langage commun, intérêt partagé, limites identifiées, etc.
- Une mise en situation sur un projet concret qui apporte immédiatement de la valeur au projet et permet d’identifier des actions concrètes à mettre en place
- Une façon percutante de prendre conscience du capital Data de l’entreprise et des contraintes de qualité et de sourcing

Expéditions de données

L'expédition de données est un format d'animation bien éprouvé. Né vers 2012 sous la houlette de la School of Data (Ecole des données <http://ecolesdesdonnees.org/>), plusieurs dizaines ont été mises en oeuvre aux 4 coins du globe. À par-

tir d'une question simple, l'objectif est d'explorer le plus loin possible le chemin qui conduit à trouver une réponse ou raconter une histoire avec les données : le data pipeline selon la terminologie de la School of data.

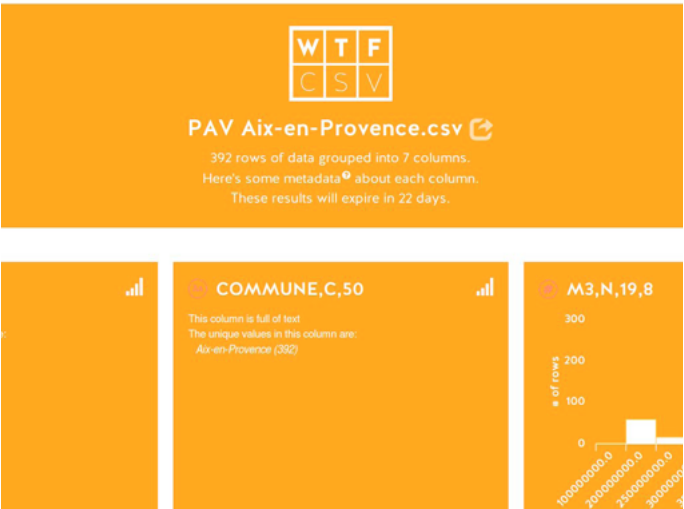
Quoi ?	Atelier créatif partant d'une question posée et explorant au maximum le chemin menant à une réponse
Qui ?	* Tout type de participant * 5 à 60 participants répartis en groupes basés sur la diversité de compétences
Combien de temps ?	Préparation : quelques jours Animation : une journée dans l'idéal ; ½ journée est un peu court Suivi : quelques jours, pour la valorisation en particulier
Qu'est-ce que cela produit ?	* De la connaissance commune sur une question donnée * Achever une ou plusieurs étapes du data pipeline : trouver des données -> obtenir les données -> nettoyer les données -> analyser -> présenter * Fait groupe ou société, favorise le vivre/travailler ensemble
Exemple ?	Le site web de la School of data regorge d'exemples et de documentation sur l'expédition de données : http://schoolofdata.org/category/data-expeditions/ Simon Chignard a documenté l'une de ses expéditions pas à pas : https://donneesouvertes.info/2013/12/03/une-data-expedition-pas-a-pas/

Cette animation a été employée avec tous types de public mais pas dans de grandes organisations. Le data pipeline (graphe ci-dessus) est pourtant un grand classique, probablement utilisé de manière informelle dans de nombreux projets. L'expédition de données permet de mieux cadrer ce processus. C'est une forme d'expérimentation qui se déroule dans le temps, qui est basée sur la convivialité et les bonnes volontés, le but est plus dans le "vivre ensemble" et comprendre les différentes approches des uns et des autres que de faire émerger un projet. Cette forme d'expérimentation exige d'être vigilant sur les dynamiques de groupes et de savoir faire mûrir un projet dans les contraintes de l'exercice.



Des outils ou micro-animations d'aide à l'exploration

L'exploration de données peut être conduite avec quelques outils et méthodologies simples que nous détaillons ici. WTFcsv Littéralement What The Fuck CSV, il permet une première plongée rapide dans un jeu de données. Après avoir ingurgité vos données tabulaires (CSV), cet outil en ligne vous présente une vision synthétique des données. Pour chaque colonne, l'outil propose quelques représentations statistiques simples qui éclairent les données : nombre, somme, moyenne, valeurs les plus fréquentes, etc.



<https://www.databasic.io/en/wtfcsv/>

Creadata : des croisements aléatoires, chez vous aussi

Du hasard peut naître des croisements inédits et féconds. Creadata est un outil en ligne qui génère aléatoirement des idées et des croisements de données (mashups). Limité aux données ouvertes de Loire-Atlantique, il cependant aisé de reproduire son fonctionnement à peu de frais. Listez les données qui vous intéressent, élaborer une fiche pour chaque, mélangez et testez ! Les résultats vous surprendront. Prenons l'exemple ci-contre : l'outil nous propose d'établir un service sur la base des données des manifestations sportives et des voies navigables : l'aspect initialement étrange de la proposition fait rapidement naître quelques idées à explorer : quoi de plus magnifique que de courir le long des voies sur berge... un marathon sur berge est-il envisageable ? Peut-on profiter des voies d'eau pour des cross géants d'un nouveau genre ? Tout un champ de questions s'offre à notre imaginaire.



Créer un service croisant ou confrontant les données des manifestations sportives en Loire-Atlantique avec les données de la longueur des voies navigables concédées au Département de Loire-Atlantique



<http://www.creadata.fr/>

L'interview de la donnée

<http://infolabs.io/interview-de-la-donnee>

Le principe en est simple et sa mise oeuvre immédiate : un participant personifie un jeu de données et se voit poser des questions par les autres participants, sur sa personnalité, sa vie, ses goûts, ses relations, etc. : comment t'appelles-tu ? où habites-tu ? qui sont tes parents ? as-tu des frères et soeurs ? des cousins ? etc. D'une apparence peu sérieuse, cette animation est redoutable pour explorer un jeu de données sous toutes ses dimensions. Pas de vocabulaire technique ! Pas d'informatique. Tout le monde est à même de poser des questions et de décortiquer le potentiel d'un jeu de données. Nous l'avons pratiqué plus d'une dizaine de fois dans des contextes très variés : le succès est toujours au rendez-vous. Il faut simplement oser une première fois.

L'interview de la donnée peut être utilisé dans des contextes variés : au début d'un projet, d'un hackathon ou d'un camp ou complètement improvisé, afin de faire monter un groupe en compétence sur un jeu de données particulier.

Des outils de visualisation comme aide à l'exploration

La visualisation de données, ou "datavisualisation" voire "dataviz", est généralement utilisée à des fins de communication et de reporting. Mais c'est aussi un véritable outil d'exploration de données et de compréhension des phénomènes. Pour des phénomènes simples, la dataviz permet de repérer des tendances sans passer par le calcul (cycles, augmentations/baisses, accidents ou crises, effets de seuil, réseaux, etc.). En amont des projets et/ou des décisions, elle permet de rendre compte de phénomènes complexes et de représenter des volumes de données dont le gigantisme résiste à l'imagination et à l'analyse.

Pour des phénomènes qui possèdent des ensembles de données simples et de taille raisonnable -- comprendre, qui peuvent encore s'ouvrir dans un tableur --, il existe de nombreux outils en ligne ou déployables sur le réseau de l'entre-

prise. Pour ne parler que des plus connus : Tableau Public (<https://www.tableau.com/>), Spotfire (<http://spotfire.tibco.com/>) ou bien encore SAS Visual Analytics (<https://www.sas.com/>). *"Ils rendent la datavisualisation accessible à plus de gens. Les opérationnels peuvent ainsi découvrir et s'appropriier des données déjà sélectionnées pour leur pertinence, et ils vont pouvoir zoomer, filtrer. Mais attention aux risques d'analyse erronée en l'absence d'accompagnement de l'accès aux données, car pour visualiser de la donnée il faut l'avoir comprise ou être guidé par celui qui l'a comprise. Il est important de donner aux utilisateurs opérationnels des données préprogrammées pour qu'ils ne se noient pas"* souligne Véronique Morault, responsable Marketing Stratégique à la MAIF. L'assureur français a déployé Spotfire à l'interne. Mais *"ces logiciels ne sont pas adaptés à une grosse et même une moyenne volumétrie. L'intérêt d'exploration interactive est limitée par les volumes"* tempère Véronique Morault.

Pour des explorations plus poussées, d'autres outils et des connaissances statistiques avancées sont nécessaires. *"On ne datavisualise pas des données nombreuses avant de les avoir analysées et d'avoir compris les principaux liens entre elles. En effet, datavisualiser peut mettre en avant des liens apparents et conduire à de fausses conclusions (un lien qui n'est pas causal, qui est dû à d'autres variables absentes de la visualisation)."* souligne encore Véronique Morault. *"On commence donc par chercher des liens très forts via l'analyse de données (algorithmes...) et on modélise ces liens. Puis on regarde ce qu'on n'a pas pu expliquer : la visualisation est alors utile pour compléter la compréhension et la connaissance. Par exemple, la visualisation cartographique peut montrer qu'un phénomène est dans le cœur des villes et absent des quartiers périphériques. L'analyse des réseaux est souvent visualisée pour schématiser les liens, car on n'a pas d'autre moyen de le faire".*

Pour de la visualisation de données poussées, la MAIF utilise R et communique ces dataviz sur l'intranet.

Soulignons que l'exploration de données massives est un domaine de recherche qui progresse beaucoup, comme en témoigne par exemple l'activité prolifique de l'équipe AVIZ d'INRIA -- <http://aviz.fr/>.

Les kits de dataviz physique

Les dataviz tangibles, également appelées physicalisation de données, gagnent en intérêt pour des usages spécifiques : modélisation rapide, usage par des non-spécialistes, systèmes de comptage collaboratifs, visualisation sensible des données, etc. Nous renvoyons au site <http://dataphys.org/> qui est une référence sur le sujet. Ces outils peuvent avoir une vraie valeur dans le cadre d'expérimentations, notamment si vous mêlez des non-spécialistes, du grand public, etc.

Il s'agit d'un domaine qui n'a pas encore véritablement touché les organisations. Cette situation pourrait changer avec la naissance de kits de dataviz physique. Le designer colombien Jose Duarte, a créé un kit prêt à l'emploi disponible sur <http://www.handmadevisuals.com/>. En France plusieurs designers ou bricoleurs y travaillent : Sylvia Fredriksson, Fréquence Écoles et le CRIJ Rhône-Alpes, Samuel Huron, Loïc Haÿ, etc.

Portails data, entrepôts, datalake : supports d'expérimentation data ?

Les entrepôts de données ont connu d'importantes transformations ces dernières années et n'ont plus rien à voir avec les "datawarehouse" d'antan. Ils ont suivi deux directions utiles à l'exploration de données.

De leur côté les datalakes⁴ ont permis de centraliser les données sous une forme "sans struc-

ture", à l'inverse des bases de données relationnelles ou des datawarehouse qui possèdent des structures très verticalisées. Ainsi la structure des données est créée au moment de l'analyse, permettant des modalités d'exploration non contraintes.

Les portails de données, comme OpenDataSoft, se sont fixés quant à eux de rendre la donnée accessible et réutilisable par le plus grand nombre. Et de permettre de promouvoir les données de



⁴. Pour aller plus loin, on pourra consulter l'article complet de Philippe Nieuwbourg : http://www.decideo.fr/Le-concept-de-data-lake-lac-de-donnees-explication-de-texte_a6976.html

l'organisation, de manière à les valoriser, tisser des liens, voire contribuer à créer un véritable écosystème. L'entreprise s'y retrouve parce qu'elle acquiert une position incontournable dans les nouvelles chaînes de valeur. L'acteur public y réalise la transparence qu'on attend de lui, et suscite également des innovations qu'il n'a pas le temps ni les compétence d'investiguer.

Dans un cas comme dans l'autre, datalake et portails sont donc une manière de stimuler l'exploration et les expérimentations data.

La Poste a mis en oeuvre ces deux outils : un datalake interne mais aussi un portail de ses données "open data", <https://datanova.laposte.fr/>.

D'autres ressources utiles pour l'exploration de données

Nous passons ici en revue quelques ressources complémentaires, identifiées par le groupe de travail mais qui n'ont pas fait l'objet d'investigations détaillées.

Du côté des interfaces et des services en ligne, les **tests A/B** et le **mode "beta"** (voire "alpha") ont toujours le vent en poupe. Le premier permet de tester les comportements des utilisateurs en faisant varier un élément d'une page web. Une des clés de réussite de Google serait sa capacité à avoir su industrialiser massivement les tests A/B de manière à optimiser au maximum ses interfaces et la conversion de ses "clients". Le second, le mode "beta" consiste à lancer quantité de services en alertant les utilisateurs sur l'aspect non terminé voire provisoire du service. On se souvient de tous les services fermés très tôt par Google, dont le "mode beta" est presque une marque de fabrique.

De retour au système d'information, le déploiement d'**APIs** et, plus globalement, d'une **architecture orientée service (SOA)**, facilite et stimule énormément les expérimentations : les données ou les services, ainsi devenus disponibles et documentés pour le plus grand nombre, peuvent faire l'objet d'explorations et d'expérimentations.



Des dispositifs d'expérimentations continus : labs, datalabs, infolabs, etc.

Depuis quelques mois, quelques années pour les plus avancés, certains acteurs ont fait le choix de **dispositifs dédiés à l'expérimentation aux données** : les labs, infolabs ou autres datalabs. Ils prennent leur place aussi bien dans les grandes organisations que chez des acteurs modestes portés sur l'animation d'un territoire ou d'un public particulier. Ils ont des formes très diverses mais partagent souvent des valeurs et missions communes. Trois missions peuvent être identifiées :

- l'incubation de projet et/ou d'expérimentation : cette mission est une constante ; on reconnaît là leur parenté affirmée avec les autres labs, comme par exemple les fablabs.
- la sensibilisation et la formation : la démarche de sensibilisation est nécessaire à leur existence même (à quoi servent-ils ?) ; la formation n'est pas toujours assumée, d'aucun préférant parfois un certain entre soi pour expérimenter tranquillement
- le débat : c'est la mission la plus marginale ; elle ne concerne aujourd'hui que les infolabs territoriaux : leur vocation d'accueil du public les y dirige assez naturellement.

Ils partagent également des valeurs, qui sont souvent la marque de fabrique de l'esprit "lab". La charte des infolabs résume ainsi les **4 valeurs de l'Infolab** :

1. **L'Infolab est un espace d'accueil ouvert.**
Il recherche la diversité et la complémentarité des publics, des talents, et des envies.
2. **L'Infolab est un espace accessible.**



Il n'est pas nécessaire d'avoir fait de longues études pour comprendre et y partager des choses.

3. **L'Infolab privilégie le collectif et la collaboration.**
Le croisement est à la source des idées les plus neuves et les plus enrichissantes.
4. **L'Infolab, si possible, est un lieu d'accueil continu.**
A tout moment chacun peut participer aux animations en cours et réaliser ses projets.

QUATRE À CINQ PROTO-INFOLABS PUBLICS EN FRANCE ?

Si les fablabs sont nés aux États-Unis, les infolabs sont-ils en train de naître en France ? Nous pouvons nous enorgueillir de plusieurs projets aboutis ou en cours de gestation.

Le Tubà -- <http://www.tuba-lyon.com/> -- , situé à Lyon, probablement le plus ancien infolab, affiche deux ans au compteur. Animé par un partenariat public-privé (Grand Lyon, Région et plusieurs grands comptes) il se propose *d'aider au développement de services innovants tout en privilégiant la participation des citoyens à cette dynamique de création et d'innovation.*

La Coop Infolab -- <http://www.la-coop.net/> -- , à Grenoble, mixe espace de co-working et infolab. Cette initiative privée est portée par la Scop La Péniche, agitateur du territoire depuis de longues années. *"Outils l'innovation socio-économique avec les données numériques"* résume leur programme.

Les Infolabs des CRIJ Rhône-Alpes et Poitou-Charente, participent aux renouvellement des Centres Régionaux d'Information Jeunesse. Ils inventent une culture

des données pour les moins de 25 ans et *“propose(nt) aux jeunes une nouvelle façon d'aborder l'information. Longtemps producteur de données, le CRIJ est désormais avant tout un lieu d'éducation à l'information”*.

D'autres infolabs sortent lentement de terre comme le Donut à Marseille (spécialisé dans la biodiversité), n'Infolab à la Cantine de Brest, etc. Le site de la communauté des infolabs en donnent une première cartographie sommaire : <http://infolabs.io/content/v0-alpha-de-la-cartographie-des-infolabs>

Les organisations connaissent un développement rapide des datalabs ou infolabs. On en dénombre chez AXA, Airbus, GrDF, Pernod Ricard, etc., mais aussi chez des grands acteurs publics comme le département de la Gironde, le Grand Poitiers, la CNIL ou encore la Région PACA. L'expérimentation est leur ADN. Leur principale qualité est cette temporalité, cette régularité qui manque aux hackathons et autres dispositifs de stimulation de l'expérimentation.

LES DATALAB DANS LA PRESSE ET SUR LA TOILE

Les datalabs communiquent peu leur activité à l'extérieur. Heureusement, certains font parler d'eux et nous permettent de comprendre un peu ce qui s'y joue : expérimentations, cellule big data, gouvernance des données pour certains, Nous proposons ici quelques articles pour aller plus loin :

> https://frama.link/datalab_Pernod_Ricard
> https://frama.link/datalab_Airbus
> https://frama.link/datalab_GRDF_avr16
> https://frama.link/datalab_GRDF_jan16

Aller plus loin :

Les rencontres Dataliteracy conference (dataliteracyconference.net) dont la première édition a eu lieu à Aix-en-Provence en septembre 2016 a réuni plus de 180 personnes passionnées par la culture des données. Un nouveau rendez-vous est prévu en 2017. Nous espérons vous y voir nombreux !

Bonnes pratiques de l'expérimentation

Tous les dispositifs, toutes les méthodologies, tous les formats d'animation ne sauraient épuiser la matière de l'expérimentation. Une demi-douzaine d'interviews ainsi qu'une veille régulière du sujet sur la toile, nous ont montré l'extraordinaire diversité des modes d'expérimentation. Notre groupe de travail a souhaité en explorer les bonnes pratiques à travers un format original mais efficace. Les 10 commandements de l'expérimentation data sont présentés et détaillés dans cette partie.

Les 10 commandements de l'expérimentation data

1. À apprendre, pas forcément à réussir, tu chercheras
2. Des hypothèses tu formuleras ou l'exploration tu assumeras
3. Tes réussites comme tes échecs tu partageras
4. À la diversité des profils et des contributeurs tu veilleras
5. Frugalité, agilité, simplicité tu chériras
6. Un accès aux données tu obtiendras
7. De l'intérêt des données tu ne préjugeras pas
8. Dans des univers nouveaux, les données tu chercheras
9. Face aux données, un esprit critique tu garderas
10. À ces principes tout le monde adhérera

Sous une forme synthétique et un peu décalée, ces commandements, expliqués plus bas,

représentent 10 bonnes pratiques d'une expérimentation data. Ils peuvent s'adresser à tous les acteurs de l'expérimentation data, sous diverses formes :

- affichés en grand dans la salle où se déroule l'expérimentation
- en annexe des documents présentant une expérimentation, avec une vocation quasi-contractuelle : que le sponsor, le commanditaire comprennent et acceptent bien ces principes
- et bien sûr dans l'infolab/datalab ou les bureaux de l'équipe de datascience de l'organisation.

En amont ou en début d'expérimentation, ils peuvent être utilement communiqués aux principaux intervenants : sponsors, DSI, experts data, métiers, etc. Il permettent de partager une culture commune de l'expérimentation.

En cours d'expérimentation et en fonction du contexte, il est utile de rappeler certains d'entre eux, par exemple en introduction de réunion ou lors d'un point d'étape.

Ces 10 commandements sont le fruit d'un atelier "Expérimentation Data" du 13 octobre 2016⁵ où ils ont été retenus par les participants comme les 10 plus importants.

Les 10 commandements ont fait l'objet d'un travail graphique et sont disponibles sous 3 formes :

- un poster, au format PDF, prêt à être imprimé (en A4, A3 ou plus) : <http://infolabs.io/cd-xp-poster>
- un fichier image, prêt à être intégrée dans une présentation ou un document : <http://infolabs.io/cd-xp-image>
- un document A4 qui détaille la démarche et ses résultats

5. L'atelier et la rédaction du document, animés par Simon Chignard, ont vu la participation d'Emmanuelle Katz (Groupe Up), Érik Arnaud (MAIF), Hugues Roumezin (MAIF) et Charles Nepote (FING).



Les 10 commandements expliqués

1. À apprendre, pas forcément à réussir, tu cherches

Échouer est un sujet tabou en entreprise et par définition les résultats d'une expérimentation ne sont pas donnés d'avance. À travers ce commandement nous réaffirmons clairement que le but d'une expérimentation n'est pas de réussir quelque chose mais d'apprendre, y compris d'un échec patent. L'apprentissage doit guider l'expérimentation, ce qui implique d'autres commandements comme le partage.

2. Des hypothèses tu formuleras ou l'exploration tu assumeras

Une fois admis qu'on cherche à apprendre plutôt qu'à réussir, il reste à expliquer comment. Pour expérimenter, il est possible de formuler des hypothèses et chercher à les vérifier ou les infirmer. Il est aussi possible d'assumer une exploration sans hypothèse préconçue. Le croisement de données au hasard peut constituer un véritable objectif d'expérimentation.

3. Tes réussites comme tes échecs tu partageras

Une expérimentation n'a d'intérêt que si elle est partagée : pour expliquer les échecs et les ré-orientations, pour montrer les gains, pour expliciter ce qu'il faudrait faire pour une industrialisation, pour faire naître la prochaine expérimentation.

4. À la diversité des profils et des contributeurs tu veilleras

L'expérimentation a souvent lieu en petit groupe. Pour autant, il est important d'y agréger des profils hétérogènes : un "pro" de la data, un expert métier, un communicant, un facilitateur, etc., sont des compétences à embarquer. De la diversité naît la richesse.

5. Frugalité, agilité, simplicité tu chériras

Un projet long, (trop) planifié et complexe, est-ce encore une expérimentation ? On privilégiera dans une expérimentation des valeurs de rapidité, frugalité, agilité, simplicité : échouer fréquemment et rapidement, mais à un faible coût et une capacité de rebond permanente.

6. Un accès aux données tu obtiendras

L'expérimentation est ce moment particulier où l'on peut sortir des cadres, assouplir voire s'affranchir de certaines règles ou de postures rigides. C'est ce moment où l'on peut légitimement obtenir un accès direct et sans frein à toutes données utiles. Revendiquez-le ! sans négliger un véritable effort de pédagogie et de restitution sur votre expérimentation.

7. De l'intérêt des données tu ne préjugeras pas

La phase d'expérimentation doit permettre de valider ou invalider des hypothèses ou d'explorer le potentiel d'utilisation des données. Un état d'esprit ouvert et sans a priori est nécessaire. Une donnée qui pourrait vous apparaître sans intérêt pourra se révéler très utile, dans un autre contexte ou pour une autre personne. Il faut garder une certaine humilité face aux données.

8. Dans des univers nouveaux, les données tu chercheras

Expérimenter c'est aussi l'occasion d'aller au-delà des sources habituelles de données que nous utilisons, le plus souvent les systèmes d'information internes. Il faut aller voir du côté des données ouvertes (open data), des courtiers en données (data brokers), ou des données que pourraient nous fournir des partenaires.

9. Face aux données, un esprit critique tu garderas

Même si l'expérimentation se doit d'être menée de façon légère, elle doit pointer du doigt les questions qu'il sera nécessaire de se poser si elle devait être poussée plus loin vers une phase d'industrialisation :

- qualité et accessibilité des données à une échelle plus grande et dans la durée,
- propriété et droits d'utilisation
- contraintes réglementaires...

10. À ces principes tout le monde adhérera

Il est essentiel que toutes les parties prenantes partagent ces principes, cela inclut notamment le sponsor de l'expérimentation et les managers de l'organisation. Celui qui participe, qui organise, qui soutient, qui commandite une expérimentation les accepte. On peut les rappeler aux participants par exemple en les affichant dans la salle où se déroule l'expérimentation.

Conclusion

Explorer, expérimenter avec les données est donc à la portée de tous. Plus encore, l'expérimentation doit être l'affaire de tous : les données ne font pas appel qu'à des savoirs techniques et chaque collaborateur de l'organisation peut apporter ses savoirs-faire métier, sa créativité, voir sa sensibilité, essentielles en la matière.

De nombreuses pistes d'expérimentation proposées sont à la portée de toutes les organisations. Certaines ont des coûts dérisoires et des résultats pourtant immédiats. Dans tous les cas, chacune d'entre elle contribue à développer une véritable culture de la data. Elles développent le nécessaire contact charnel avec ce matériau d'apparence si difficile.

CAHIER #2

1, 2, 3 data, expérimentez !

<http://infolabs.io/xp16>



CC License Creative Commons BY 3.0

Ce cahier est une production de la Campagne Infolab de la Fing



avec le soutien des organisations participantes



La Fing a le soutien de ses grands partenaires

