L’interview de la donnée

Une méthodologie de découverte et de compréhension d’un jeu de données numériques

Les données ? Une matière froide, peu familière du plus grand nombre. Voire même aride ! Face à un jeu de données, la plupart d’entre nous ne connaît pas les questions pertinentes qu’il convient de se poser. Plutôt qu’une check-list froide et peu motivante, cette courte méthodologie se propose de révéler les principaux aspects d’un jeu de données d’une manière ludique et “sensible” : l’interview de la donnée.

Le principe est simple. Une personne ou une petite équipe joue le rôle d’un jeu de données comme s’il s’agissait d’une personne interviewée par un journaliste/un enquêteur. Ce dernier cherche à connaître le plus de choses d’elle en un minimum de temps.

Les ingrédients sont peu nombreux. Au préalable, il faut un jeu de données, bien sûr. Idéalement un animateur met le jeu en place, donne les consignes et propose une prise de recul à la fin du jeu. On constitue enfin deux petites équipes :

- l’équipe qui va jouer le rôle de la donnée (une à trois-quatre personnes)
- l’équipe qui joue l’interviewer (une à 5-6 personnes)

Le jeu se **déroule de la manière suivante**.

1. L’animateur explique le jeu et son objectif. Il prend quelques exemples de questions qui peuvent être posées.
2. Si besoin la personne/l’équipe qui joue le rôle de la donnée se documente sur celle-ci (15 minutes max). L’équipe d’interviewers peut préparer des questions en parallèle.
3. L’équipe d’interviewers pose ses questions et le jeu de données y répond. Lorsqu’il n’a pas la réponse, on la note quelque part. L’équipe de la donnée peut chercher la réponse à condition de ne pas interrompre l’interview.
4. A la fin de l’interview, l’animateur consulte la liste des questions type (cf. ci-dessous) et mentionne celles qui ont été oubliées. “Vous auriez aussi pu poser les questions suivantes”. La donnée répond à ces questions.

Variante à deux jeux de données : les deux équipes sont de nombre égal et chacune prend en charge un jeu de données. Ainsi la phase de documentation du jeux de données avant interview peut se faire en parallèle par les deux équipes.

Idéalement, la méthodologie ne dure pas plus d’une demi-heure par jeu de données. Elle peut être utilisée dans des contextes assez différents :

- débat : pour mieux comprendre à qui on a affaire,
- incubation de projet, pour prendre du recul sur un jeu de données,
- sensibilisation à la question des données, pour habituer les participants à se poser des questions sur les données,
- publication de données, pour estimer ce qui est important à mettre dans la fiche descriptive des données.

L’interview de la donnée, version beta 2, janv. 2015-juin 2017. Par Charles Nepote (FING), pour la Communauté Infolab. http://infolabs.io/interview-de-la-donnee
Licence Creative commons 3.0 Attribution France.

Questions-types

Qui es-tu ?
Comment t'appelles-tu ? C'est ton nom de tous les jours ? C'est sur tes papiers ? Tu as un surnom ? Quel est ton nom "générique" ?

Si tu avais à te raconter en quelques mots ?

Comment te trouve-t-on ? Où habites-tu habituellement ? Tu as déjà changé de lieu d'habitation ? Tu as plusieurs résidences ?
Tu es de quelle nationalité ? Tu parles quelle(s) langue(s) ?

Comment te décrirais-tu physiquement ? De quelle(s) couleur(s) sont tes cheveux, tes yeux ? Comment est ton visage ...
Est-ce tu aimes changer d'aspect ? Est-ce que parfois tu teinds tes cheveux ? Portes-tu des lentilles ?...
Est-ce que tu mesures des constantes médicales ? Es-tu en bonne santé ? A-t-on des radiographies de toi ?

Quels sont les différents lieux que tu fréquentes ? Est-ce qu'il y a des frontières au-delà desquelles tu ne vas pas ? As-tu un territoire de préférence ?
Comment aimes-tu te déplacer, sur les petites ou les grandes routes ?
Quels sont tes passions, tes sujets d'intérêt ?
Est-ce que tu as un passé sulfureux ? Des choses à cacher ? Es-tu pudique ?
Quelles sont tes valeurs ? Quel est ton code de conduite ?

Qui sont tes parents ? tu as un père ? une mère ? Tu appartiens à quelqu'un ? qui est responsable de toi ? tu as quelle nationalité ?

Quels sont tes amis ? Tu as des frères et soeurs ? Un jumeau qui n'a rien à voir avec ta famille ? Qui fréquentes-tu ? As-tu des amis ? Qui sont-ils ?
As-tu des cousins ? Fais-tu partie d'une tribu ? d'un clan ? d'une famille particulière ?
Est-ce que tu t'es déjà marié ?

Quel âge as-tu ? Tu vas avoir cet aspect pendant combien de temps ? Tu serais plutôt une photo ? à quelle date ? Tu connais ton espérance de vie ?
Est-ce que tu te sens une personne de ton époque ? Est-ce que tu as une bonne mémoire ?
Quels sont tes rythmes d'activités ? As-tu un biorythme particulier ?

Est-ce que n'importe qui peut te voir ? peut te demander des choses ? peut répéter ce que tu lui as dit ?
As-tu déjà participé à un événement : une soirée, une fête ... ?
Es-tu connu, inconnu ? Par qui ? On te désire ? on t'adule ? on te pille ? on te pirate ?
On te vend ? on te revend ? on te redistribue ?
As-tu fait l'objet d'articles dans la presse ?
A quel âge as tu eu ta majorité ?

L'interview de la donnée, version beta 2, janv. 2015-juin 2017. Par Charles Nepote (FING), pour la Communauté Infolab. http://infolabs.io/interview-de-la-donnee
Licence Creative commons 3.0 Attribution France.

Aller plus loin : la fiche descriptive

L'interview de la donnée est un excellent exercice pour **renseigner la fiche descriptive** d'un jeu de données, par exemple au moment de sa publication sur un portail. Il est possible de **faire faire cette fiche aux participants** en conclusion de l'exercice. Souvent elle contient :

- le titre du jeu
- le nom du producteur
- la description du jeu : a minima une dizaine de lignes ; une documentation plus complète est publiée à part lorsqu'il s'agit d'un jeu complexe
- sa date de production et sa fréquence de mise à jour
- sa couverture spatiale
- sa couverture temporelle
- sa langue
- sa licence d'usage
- des mots-clés
- des exemples d'usages

Le dialogue ouvert par l'interview permet de compléter ces informations plus clairement. Un titre trop long pourra être questionné par les interviewer. Un titre trop court les poussera à poser d'autres questions pour le préciser.
Cet exercice agit comme une **maïeutique** très efficace : elle permet littéralement d'**exprimer les savoirs cachés** des producteurs de la donnée. Son caractère collectif augmente encore son efficacité.

Page suivante, nous donnons un exemple de fiche descriptive, tiré du portail national de l'état : http://data.gouv.fr/

Licence Creative commons 3.0 Attribution France.

Base Sirene des entreprises et de leurs établissements (SIREN, SIRET)

Ce jeu de données provient d'un service public certifié

DONNÉE DE RÉFÉRENCE

Ce jeu de données permet d'accéder aux 9 millions d'entreprises et 10 millions d'établissements actifs du répertoire Sirene de l'Insee qui enregistre quotidiennement leur état civil :

- quelle que soit leur forme juridique ;
- quel que soit leur secteur d'activité (industriels, commerçants, artisans, professions libérales, agriculteurs, collectivités territoriales, banques, assurances, associations...) ;
- situés en France métropolitaine, ainsi qu'en Guadeloupe, Martinique, Guyane, La Réunion, Mayotte, Saint-Barthélémy, Saint-Martin et Saint-Pierre-et-Miquelon. Les organismes publics ou privés et les entreprises étrangères qui ont une représentation ou une activité en France y sont également répertoriés.

Le répertoire Sirene est ainsi la principale source exhaustive sur l'ensemble des entreprises et des établissements actifs.

AVERTISSEMENT

La base Sirene contenant des données à caractère personnel, l'Insee attire votre attention sur les obligations légales qui en découlent:

- Le traitement de ces données relève des obligations de déclaration de la Loi 78-17 du 6 janvier 1978 modifiée, dîte Loi CNIL : https://www.cnil.fr/fr/loi-78-17-du-6-janvier-1978-modifiee
- Selon votre usage du jeu de données, il est de votre responsabilité de tenir compte du statut de diffusion le plus récent de chaque personne physique. En effet, l'article A123-96 du code de commerce dispose que : "Toute personne physique peut demander soit directement lors de ses formalités de création ou de modification, soit par lettre adressée au directeur général de l'Institut national de la statistique et des études économiques, que les informations du répertoire la concernant ne puissent être utilisées par des tiers autres que les organismes habilités au titre de l'article R. 123-224 ou les administrations, à des fins de prospection, notamment commerciale."

Fichiers : Trois types de fichiers compactés (format ZIP) sont mis à disposition (1 fichier stock et 2 fichiers de mises à jour mensuelles ou quotidiennes, France entière). Chaque fichier compacté (Format ZIP) contient un fichier de données en format CSV.

Le fichier stock et le fichier des mises à jour mensuelles sont déposés le premier jour de chaque mois avec les données (dessin de fichier L2) au dernier jour du mois précédent. Le fichier des mises à jour mensuelles étant au même dessin que le fichier stock, vous pouvez faire le choix chaque mois, d'annuler et remplacer le fichier stock par le nouveau fichier stock ou de l'actualiser avec le fichier des mises à jour mensuelles. Tous les mouvements relatifs à un établissement au cours du mois sont agrégés, rendant l'information facilement utilisable.

Le fichier des mises à jour quotidiennes est déposé dans la nuit suivant la journée de gestion qu'il concerne (dessin de fichier XL2).

***NB** : Les fichiers sont déposés un peu plus tardivement début janvier et début juillet en raison d'opérations de mise en concordance semestrielle de la base de diffusion.*

Le site de l'Insee www.sirene.fr fournit des informations sur le contenu de ces bases, ainsi que la documentation associée et une Foire Aux Questions pour vous aider, comprenant plus de 50 questions-réponses.

Ressources

ZIP	**Sirene : mise à jour quotidienne du 04 juillet 2017** Dernière modification le mercredi 5 juillet 2017
ZIP	**Sirene : mise à jour quotidienne du 03 juillet 2017** Dernière modification le mercredi 5 juillet 2017

Producteur

Insee
Mesurer pour comprendre

L'Institut national de la statistique et des études économiques (Insee) collecte, produit, analyse et diffuse des informations sur l'économie et la société françaises. Ces ...

SUIVRE

Informations

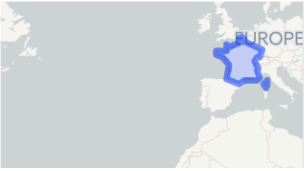
- Donnée de référence
- Licence Ouverte / Open Licence
- Quotidienne
- 27 décembre 2016
- 5 juillet 2017
- 5 juillet 2017
- Autre

ASSOCIATIONS COMPANIES ENTREPRISES ETABLISSEMENTS IMMATRICULATIO... REGISTER REGISTRE SIREN SIRENE SIRET

SUGGÉRER UN MOT-CLÉ

DÉTAILS

Couverture spatiale



L'interview de la donnée, version beta 2, janv. 2015-juin 2017. Par Charles Nepote (FING), pour la Communauté Infolab. http://infolabs.io/interview-de-la-donnee
Licence Creative commons 3.0 Attribution France.