



THESIS / THÈSE

MASTER EN SCIENCES INFORMATIQUES

Open data et smart-city

regard sur les politiques et processus d'ouverture des données

Camara, Agnès

Award date:
2021

Awarding institution:
Universite de Namur

[Link to publication](#)

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal ?

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.



**UNIVERSITÉ
DE NAMUR**

FACULTÉ
D'INFORMATIQUE

**Open data et smart-city: regard sur les
politiques et processus d'ouverture des
données**

Agnès Fama CAMARA

Remerciements

La réalisation de ce mémoire a été possible grâce au concours de plusieurs personnes à qui je voudrais témoigner toute ma gratitude.

Je voudrais tout d'abord adresser toute ma gratitude à la promotrice ce mémoire, Mme Claire Lobet-Maris, pour sa patience, sa disponibilité et surtout ses judicieux conseils, qui ont contribué à alimenter ma réflexion.

Je souhaite particulièrement remercier la famille Deneufbourg et la famille Diemé-Gillain pour sa précieuse aide à la relecture et à la correction de mon mémoire.

Je désire aussi remercier les professeurs et assistants de l'Université de Namur, qui m'ont fourni les outils nécessaires à la réussite de mes études universitaires.

Je remercie ma famille et mes amis qui ont toujours été là pour moi. Leur soutien inconditionnel et leurs encouragements ont été d'une grande aide.

À tous ces intervenants, je présente mes remerciements, mon respect et ma gratitude.

Résumé

Notre mémoire se situe dans le domaine des données ouvertes mises à disposition du public par des administrations ou des entreprises. Dans un premier temps, nous avons voulu comprendre le trajet qu'effectuent ces données pour être « ouvertes » et disponibles au public. Nous nous sommes donc intéressé à comprendre et formaliser les différentes étapes du processus de transformation des données. Du point de vue des utilisateurs de ces données, la littérature nous apprend qu'il existe une différence entre ouverture des données et leur accessibilité pour une utilisation efficace par le public. Dans cette perspective centrée sur l'utilisateur, nous nous sommes intéressé aux outils permettant de renforcer la capacité des utilisateurs à utiliser ces données dans l'exploration d'un projet ou la résolution d'un problème. Nous avons identifié différents outils et méthodologies qui visent à cette meilleure accessibilité. Parmi ceux-ci, notre mémoire s'est centré sur les outils de visualisation. Ces outils de visualisation sont très nombreux. Par contre, peu d'évaluation systématique existe pour les comparer dans ce qu'ils peuvent apporter pour une utilisation efficace des données ouvertes. C'est donc sur cette question de l'évaluation et de la capacité de ces outils de visualisation à rendre les données plus accessibles aux usagers que s'est focalisé notre mémoire. Après avoir identifié les outils de visualisation les plus souvent cités par la littérature, nous avons développé une grille d'analyse permettant de les explorer et de les comparer. Sur base de tests effectués sur chacun d'eux, nous avons complété la grille d'analyse de manière à pouvoir les comparer et informer les futurs utilisateurs sur les contraintes d'utilisation mais aussi les apports spécifiques de chacun de ces outils.

Mot clés : Données ouvertes, OGD, processus de transformation des données, visualisation, accessibilité

Abstract

Our thesis is in the field of open data made available to the public by administrations or companies. First, we wanted to understand the path that this data takes to be “open” and available to the public. We are therefore interested in understanding and formalizing the different stages of the data transformation process. From the point of view of the users of this data, the literature tells us that there is a difference between openness of data and its accessibility for effective use by the public. From this user-centric perspective, we looked at tools to strengthen the ability of users to use this data in the exploration of a project or the resolution of a problem. We have identified various tools and methodologies aimed at this better accessibility. Among these, our thesis focused on visualization tools. These visualization tools are very numerous. On the other hand, little systematic evaluation exists to compare them in what they can bring for an effective use of open data. It is therefore on this question of evaluation and the capacity of these visualization tools to make data more accessible to users that our thesis has focused. After having identified the visualization tools most often cited in the literature, we have developed an analysis grid allowing them to be explored and compared. On the basis of tests carried out on each of them, we have completed the analysis grid so as to be able to compare them and inform future users about the constraints of use but also the specific contributions of each of these tools.

Keywords: Open data, OGD, data transformation process, visualization, accessibility

Table des matières

Introduction	2
Chapitre 1 Open Data : Contexte et Définition	4
1 Open Data	4
2 Open Gouvernement Data	5
3 Open Government Data : les processus et les écosystèmes	5
4 Les politiques de données ouvertes	6
5 Open Data : Les utilisateurs	7
6 Open data utilisation	8
7 Données ouvertes et données accessibles : le concept d'accessibilité aux données . .	9
Chapitre 2 De la donnée brute à la donnée accessible : le trajet des données	10
1 Objectif du chapitre et méthodologie	10
2 Processus d'ouverture des données	12
3 Analyse des besoins des citoyens en termes de données ouvertes	16
4 Résolution de problèmes autour des données ouvertes	18
Chapitre 3 Exploration des outils de visualisation : Méthodologie et résultat	21
1 Introduction	21
2 Sélection et présentation des outils de visualisation	21
3 Explication des critères en lien avec l'accessibilité et l'utilisabilité	24
Chapitre 4 Analyse et comparaison des outils	31
Conclusion	53
Glossaire	55

Introduction

Le mémoire se situe dans le domaine de l' *open data* ou encore des données ouvertes libérées ou mises à disposition du public par des instances privées, scientifiques ou publiques. Ces données peuvent concerner des domaines aussi variés que les mesures de qualité de l'air, les chiffres détaillés du chômage, de l'accès aux soins de santé, du nombre d'entreprises créées dans un secteur donné, etc. Partout dans le monde, le phénomène de l' *open data* prend de l'ampleur. En témoignent le nombre important d'événements organisés ces dernières années autour des données ouvertes que ce soit par les villes, les gouvernements, les entreprises ou encore les universités. En Wallonie, le baromètre des villes wallonnes de 2020 présenté par Installé [52] de FuturoCité, montre une évolution significative du nombre de villes ayant entamé une politique d' *open data* à travers différentes activités de gouvernance et d'exploitation de leurs données. Dans le domaine public des gouvernements et des administrations, ces politiques et ces pratiques d'ouverture des données visent à rendre plus transparente et plus lisible l'action publique. Mais elles visent aussi et peut-être surtout, par la ré-utilisation et l'exploitation de ces données ouvertes, à la création d'entreprises ou de services innovants axés sur -et parfois portés par- les citoyens. Dans la logique des *smart city*, l'ouverture des données est envisagée comme une modalité de coproduire de nouveaux services à destination des citoyens[26]. Identifiée comme une ressource essentielle au développement de nouveaux services, la mise à disposition des données doit permettre à des acteurs externes de s'en emparer.

C'est ce phénomène que nous allons explorer dans le cadre de notre mémoire. Plus particulièrement, nous allons dans ce mémoire tenter de comprendre ce qui peut rendre une donnée utilisable par les citoyens. En effet, ouvrir les données, les rendre disponibles n'est pas la même chose que les rendre utilisables par tous. C'est autour de cette tension entre ouverture et utilisabilité des données que nous avons voulu construire notre mémoire.

Dans un premier chapitre, nous nous attacherons à contextualiser notre problématique. Nous veillerons d'abord à bien approcher les termes, le contexte, les acteurs mais aussi les enjeux de cette ouverture des données au public. C'est à travers la revue de la littérature à laquelle nous avons procédé pour instruire ce domaine que la question de l'utilisabilité des données par le public s'est imposée comme un enjeu central pour le succès de ces politiques tant au niveau de la transparence démocratique qu'au niveau de l'innovation par le public.

Partant de cette question, nous avons voulu dans un deuxième chapitre comprendre le cheminement ou le trajet parcouru par les données de leur lieu de production « métier » à leur utilisation par le public pour la compréhension de décision publique ou pour la création de services innovants. En effet, pour comprendre ce qui permet à ces données de pouvoir être utilisées par le public, il nous apparaît important de tracer tout le parcours que suivent des données produites dans des cadres métiers pour devenir des données utilisables et utilisées par un plus large public.

Nous avons structuré ce trajet autour de trois grands étapes. La première est celle de l'ouverture des données. Cette étape nous montre que l'ouverture des données n'est pas un simple processus de mise à disposition au public d'un set de données défini. L'ouverture des données est une étape complexe qui se construit à travers différents processus et qui fait émerger aussi de nouveaux rôles dans les administrations. Une fois les données ouvertes, l'étape suivante est celle de l'analyse des besoins. Nous verrons que différentes méthodes et outils permettent de soutenir l'expression des besoins des citoyens et l'identification de sets de données pertinents. Enfin, la troisième étape est celle qui consiste à se servir des données pour leur donner sens ou répondre aux attentes exprimées

en matière d'innovation ou de résolution de problèmes. Ici aussi nous verrons que des méthodes et des outils peuvent soutenir ces dynamiques.

Dans le trajet accompli dans le chapitre 2, il apparaîtra que les outils de visualisation jouent un rôle assez central auprès des utilisateurs pour comprendre les données ouvertes, les utiliser à des fins d'analyse, leur donner sens dans un projet. C'est la raison pour laquelle, toujours en tissant le fil de l'utilisabilité des données, nous avons voulu focaliser la suite de notre recherche sur ces outils de visualisation.

Dans le chapitre 3, nous procéderons à un état de l'art afin d'identifier dans la littérature les outils de visualisation les plus souvent référencés dans les articles traitant de l'accessibilité aux données ouvertes. Pour évaluer ces outils et les comparer, nous construirons toujours dans ce chapitre 3 une grille d'évaluation portant sur différents critères centrés sur les fonctionnalités offertes par ceux-ci en termes de visualisation et de collaboration mais aussi sur leur accessibilité en termes de compétences requises, de prix ou encore de protection des données, pour ne reprendre que quelques critères.

Enfin, le 4ème et dernier chapitre sera consacré à l'évaluation systématique des 10 outils de visualisation retenus. Après les avoir explorés et testés, nous les évaluerons et comparerons sur base de la grille des critères définie au chapitre précédent.

A travers ce parcours, notre premier objectif scientifique est de bien comprendre les conditions qui permettent de passer de l'ouverture de la donnée à son utilisabilité par le public. Notre second objectif scientifique est de procéder à une évaluation comparée des 10 outils de visualisation afin d'explorer les potentialités et l'accessibilité de chacun en matière d'utilisabilité des données.

Chapitre 1

Open Data : Contexte et Définition

Introduction

Dans cette section nous allons poser le contexte de ce mémoire en donnant les définitions. Nous allons d'abord parler de l'*Open data*, pour introduire ensuite l'Open Government Data que nous retrouverons souvent sous l'acronyme OGD et les Écosystèmes OGD. Nous nous intéresserons aussi à la politique que les gouvernements ou administrations publiques mettent en place pour les OGD. Nous nous questionnerons ensuite sur qui sont les utilisateurs de ces données ouvertes et sur les usages qui sont faits de celles-ci. Nous ferons également le point dans ce chapitre sur les conditions d'utilisation de ces données. Enfin nous terminerons ce chapitre par le concept d'accessibilité des données qui permet de penser l'ouverture des données par le prisme de la capacité des utilisateurs à en faire une utilisation efficace.

1 Open Data

Pour Courmont [26], l'*Open data* est un mouvement international, visant à mettre en circulation les données des administrations publiques afin de les rendre accessibles aux citoyens. Produire, utiliser ou faire circuler une donnée ne peut s'envisager sans des systèmes techniques, des individus, des organisations, des cadres juridiques, des motivations stratégiques, des représentations idéologiques, des cultures professionnelles, des procédures, etc.

Un [67] groupe de travail composé de 30 personnes, parmi lesquelles Tim O'Reilly – auquel on doit la définition du Web 2.0 – et Lawrence Lessig – professeur à l'Université de Stanford et promoteur des licences Creative Commons, a posé ainsi les bases de l'Open data dont les 8 principes fondamentaux sont [93] :

1. Les données sont complètes. Toutes les données publiques doivent être rendues disponibles sauf les données pouvant porter atteinte à la vie privée des citoyens ou à la sécurité.
2. Les données sont primaires. Les données doivent être brutes, telles qu'elles ont été collectées à la source, non agrégées, non modifiées.
3. Les données sont tenues à jour. Elles doivent être rendues disponibles aussi vite que possible afin de préserver leur valeur.
4. Les données sont accessibles. Les données sont disponibles au plus large spectre d'utilisateurs pour l'usage le plus large.
5. Les données doivent permettre un traitement automatisé. Elles doivent être structurées et documentées afin de permettre un traitement informatisé.
6. L'accès aux données est non discriminatoire. Elles sont disponibles à tout le monde de façon anonyme ne nécessitant pas d'enregistrement.

7. Les données sont disponibles dans un format non propriétaire. Elles doivent être rendues disponibles au moins dans un format sur lequel aucune entité ne détient le monopole (ex : non PDF, non Excel).
8. Les données sont libres de droits. Elles ne doivent pas être l'objet de droits d'auteurs, marques déposées, brevets, etc.

2 Open Gouvernement Data

L'Open Government Data (OGD) est une philosophie - et de plus en plus un ensemble de politiques - qui promeut la transparence, la responsabilité et la création de valeur en mettant les données du gouvernement à la disposition de tous¹. Les organismes publics produisent et traitent d'énormes quantités de données et d'informations. En rendant leurs ensembles de données disponibles, les institutions publiques deviennent plus transparentes et plus responsables envers les citoyens. En encourageant l'utilisation, la réutilisation et la distribution gratuite des ensembles de données, les gouvernements encouragent la création d'entreprises et des services innovants axés sur -et parfois portés par- les citoyens.

Dans le domaine de l'administration publique, les données ouvertes sont souvent appelées *Open Government Data* (OGD) pour Torchiano et al. [113] : le secteur public est l'un des principaux producteurs et détenteurs d'informations, qui va, par exemple, des cartes aux enregistrements des entreprises.

3 Open Government Data : les processus et les écosystèmes

Des observateurs experts ont qualifié les programmes OGD de « rues à sens unique » qui doivent devenir des « écosystèmes » avec des cycles de rétroaction entre les utilisateurs de données et les fournisseurs Pollock [89]. C'est une voie à sens unique[89] car il n'y a pas de boucle de rétroaction, pas de partage de données vers les éditeurs et pas de partage entre intermédiaires.

Partant de cette notion d'écosystème, Koch et al. [60] nous propose une la classification de la population du réseau en six rôles d'utilisateurs, à savoir les motivateurs, les attracteurs d'attention, les générateurs d'idées, les communicateurs, les maîtres et les utilisateurs passifs. Les rôles d'utilisateurs identifiés diffèrent considérablement dans leurs activités de communication et de commentaires ainsi que dans leur contribution à l'objectif principal de la plateforme sous la forme d'idées soumises. Il [60] nous suggère que les plates-formes de gouvernement ouvert doivent être considérées comme un « écosystème » de rôles d'utilisateur, qui évolue au fil du temps et où le succès dépend de la bonne combinaison des différents rôles.

Dawes et al. [30] ont développé un modèle d'écosystème dans le but d'aider à la planification et à la conception de programmes Open Gouvernement Data. Ils soulignent aussi que le modèle de l'écosystème n'est pas la recette d'un programme réussi - il est un moyen d'évaluer les conditions existantes et de déterminer quels changements dans les politiques, les stratégies, les pratiques, les relations ou les participants pourraient rapprocher le programme OGD d'un gouvernement de la production des avantages souhaités. Ils pensent aussi que le modèle de l'écosystème, qui met l'accent sur la dynamique qui se déroule dans le temps, peut être utile à la fois pour la conception et l'évaluation des programmes de données du gouvernement ouvert. Ils nous proposent un modèle d'écosystèmes OGD où on peut voir :

- les trois groupes de parties prenantes clés(Les Fournisseurs, les Utilisateurs et les Bénéficiaires),
- les influences qui motivent et façonnent les politiques et stratégies d'un gouvernement donné,
- les processus de préparation et de publication des données par les agences administratives,
- l'utilisation des données et les produits de données développés principalement par des utilisateurs extérieurs au gouvernement,
- la rétroaction et la communication entre les éditeurs de données et les utilisateurs de données,

1. <https://www.oecd.org/gov/digital-government/open-government-data.htm>

— et les avantages associés à l'utilisation des données et à l'adoption de nouveaux produits et services créés avec OGD.

Ces facteurs sont tous susceptibles d'entrer en jeu, mais de diverses manières selon les différences de capacité et d'expérience des acteurs concernés.

Zuiderwijk et al. [124] nous proposent un aperçu des différents processus, des outils et des parties prenantes d'un écosystème de données ouvertes. L'écosystème de données ouvertes qu'ils [124] ont développé dans leur article proposent de montrer les interactions à plusieurs niveaux qui se passent entre les parties prenantes, les outils et programmes de données ouvertes.

À partir de cette recherche, ils [124] ont identifié les éléments clés que les écosystèmes de données ouvertes devraient capturer, à savoir :

1. la libération et la publication de données ouvertes sur Internet,
2. la recherche, la découverte, l'évaluation et la visualisation des données et de leurs licences associées,
3. le nettoyage, l'analyse, l'enrichissement, la combinaison pour relier et visualiser les données et
4. interpréter et discuter des données et fournir des commentaires au fournisseur de données et aux autres parties prenantes.
5. des parcours utilisateur montrant les directions d'utilisation des données ouvertes,
6. un système de gestion de la qualité et
7. différents types de métadonnées pour pouvoir connecter les éléments.

Les différents instruments et outils de données ouvertes se complètent (éléments 1 à 4) et les trois éléments supplémentaires (éléments 5 à 7) garantissent l'intégration des composants individuels, permettant ainsi l'interopérabilité.

Pour Zuiderwijk et al. [124] cette revue montre que la recherche sur les écosystèmes de données ouvertes se concentrent uniquement sur les activités liées à la publication ou à l'utilisation des données ouvertes, mais pas sur l'ensemble des activités qui devraient être présents dans un écosystème de données ouvertes.

4 Les politiques de données ouvertes

Les politiques de données ouvertes sont importantes, car leur objectif est d'assurer la transparence à long terme des informations gouvernementales [11] et de contribuer ainsi au droit des citoyens d'accéder aux informations gouvernementales, qui est considéré comme un principe fondamental de la démocratie [6].

Pour Zuiderwijk and Janssen [123] l'objectif est de proposer un cadre de comparaison des politiques de données ouvertes et de pouvoir utiliser ce cadre pour comparer les politiques de données ouvertes de sept organisations gouvernementales néerlandaises. Celui-ci comprend les facteurs de l'environnement et du contexte, le contenu des politiques, les indicateurs de performance et les valeurs publiques (Voir figure 1.1). Cette analyse leur a montré des opportunités pour les organisations publiques d'apprendre les unes des autres [123].

Pour Courmont [26] la politique d'ouverture des données n'a pas une seule origine linéaire, elle est le résultat d'une constellation de valeurs, de visions et de pratiques plurielles qui conduisent différents acteurs à parler d'ouverture de données de manière contingente. Les données sont techniquement, institutionnellement et cognitivement attachées aux politiques sectorielles. Selon les secteurs, différents objectifs sont assignés à la donnée. Pour chacune de ces problématiques, les acteurs associés et les modes d'attachement ne seront pas les mêmes. Il semble donc illusoire de construire une politique globale et unique de la donnée autonome des secteurs d'action publique.

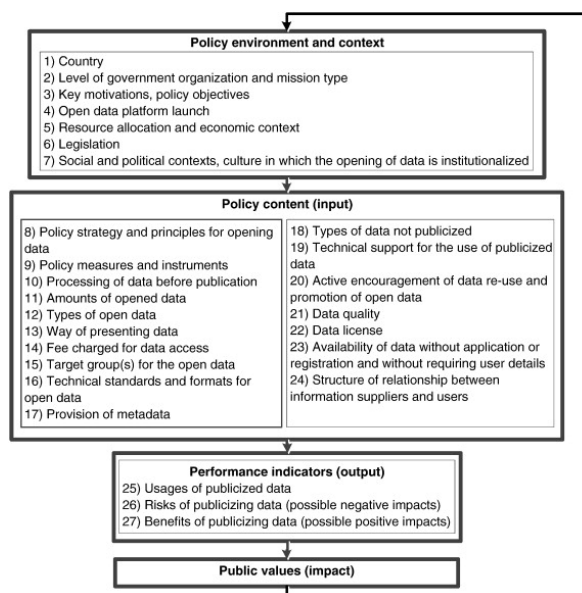


FIGURE 1.1 – Un cadre pour comparer les politiques de données ouvertes à différents niveaux de gouvernement. [123]

5 Open Data : Les utilisateurs

Safarov et al. [103] ont trouvé une distinction entre les utilisateurs directs - ceux qui utilisent eux-mêmes les données publiques ouvertes - et les utilisateurs indirects - ceux qui utilisent des données qui ont été traitées par des intermédiaires. En plus de ceux qui ont des compétences professionnelles et des connaissances techniques, les utilisateurs incluent également ceux qui peuvent utiliser des ensembles de données simples ou qui consomment les effets de l'utilisation des OGD.

Citoyens Parycek et al. [86] identifient les citoyens comme les principales parties prenantes (avec les entreprises, les chercheurs et les journalistes). Ils les considèrent comme les principaux bénéficiaires de l'utilisation des OGD en raison d'une participation accrue, ce qui permet aux citoyens d'avoir un meilleur aperçu des activités gouvernementales.

Pour Graves and Hendler [47], le citoyen est un terme qui représente la grande majorité de la population qui peut éventuellement consommer des données via une application, une visualisation ou un rapport, mais elle n'est pas directement impliquée dans les OGD.

Entreprises Très peu d'études se concentrent uniquement sur les entreprises en tant qu'utilisateurs d'OGD[103]. Malgré la difficulté pratique d'observer l'utilisation des OGD par les entreprises (soit parce que ce n'est pas l'une des activités principales de l'entreprise, soit parce qu'elle est cachée sous des secrets commerciaux), le développement rapide des technologies de données, telles que l'exploration de données et l'analyse de données, a créé des chances prometteuses pour la recherche d'entreprise en tant qu'utilisateur d'OGD.

Développeurs Les développeurs de données ouvertes jouent un rôle important pour encourager l'adoption de politiques OGD et révéler de plus en plus d'ensembles de données[27].

Les développeurs travaillent généralement avec les OGD pendant leur temps libre, créant des applications, des mashups ou adaptant les données à leurs besoins[47].

Selon Desouza and Bhagwatwar [34], la majorité des projets basés sur les OGD sont établis par des développeurs en tant que start-ups.

Journalistes et chercheurs Les chercheurs et les journalistes désignent des personnes hors du gouvernement qui travaillent dans des journaux ou des organisations non gouvernementales (ONG) et qui sont intéressées à utiliser OGD pour leur travail[47]. Par exemple, de nombreuses ONG utilisent des données gouvernementales pour réaliser des études, rédiger des rapports et faire du lobbying pour promouvoir leurs programmes[47].

6 Open data utilisation

Safarov et al. [103] ont abouti à un cadre d'utilisation des OGD, composé de quatre catégories génériques (utilisateurs, effets, types et conditions) avec une variété de sous-catégories[103].

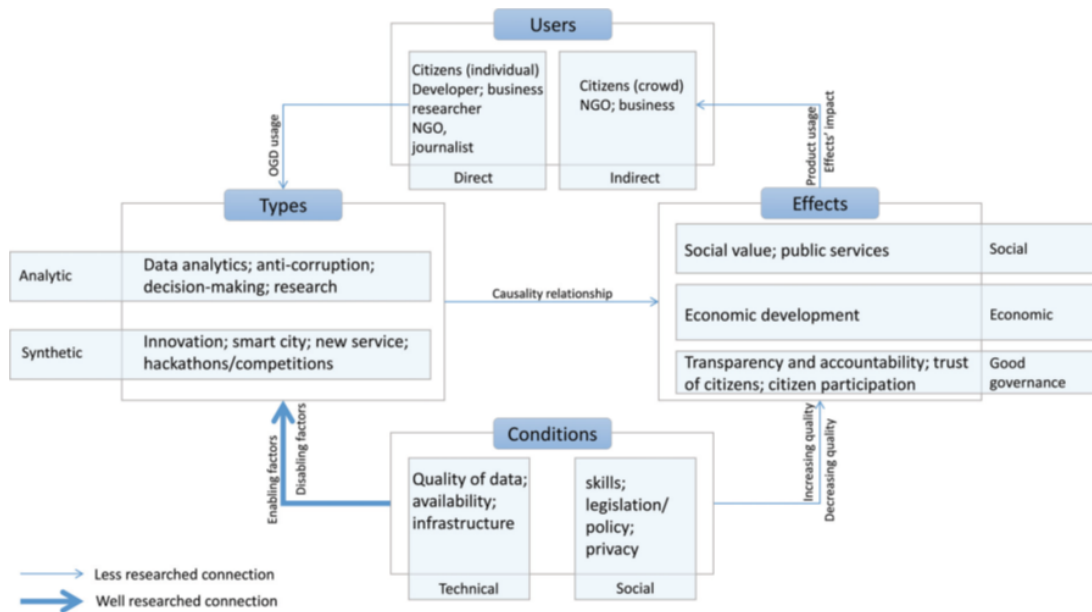


FIGURE 1.2 – framework utilization OGD [103]

Nous allons structurer notre exposé sur base du modèle d'utilisation de Safarov et al. [103]. Nous aborderons les catégories importantes en nous concentrant sur les intermédiaires et sur les différents outils qui existent.

6.1 Les conditions d'utilisation des Open Data

Les conditions d'utilisation des OGD sont séparées en deux catégories[103] : techniques et sociales. Les conditions techniques sont des caractéristiques telles que la qualité des données, leur disponibilité et l'infrastructure permettant de les rendre disponibles. Les conditions sociales sont de nature institutionnelle (législation, politique, etc.), mais elles renvoient également aux compétences des utilisateurs.

Janssen et al. [54] ont réalisé des entretiens avec des membres du public pour recueillir leurs opinions et leurs expériences avec les données ouvertes. Leurs résultats suggèrent un certain nombre d'obstacles qui empêchent actuellement les citoyens de convertir des données ouvertes en quelque chose qu'ils peuvent utiliser. Parmi ceux-ci, le principal est que les citoyens n'ont pas les connaissances nécessaires pour utiliser ou donner un sens aux données.

6.2 Types d'utilisation

Le type d'utilisation couvre les différentes façons dont les praticiens emploient les OGD. Par exemple, les données ouvertes peuvent être utilisées comme outil de recherche [64], dans les hackathons[65] [57] ou dans l'analyse de données[61][68].

Safarov et al. [103] distinguent deux types d'utilisation : l'utilisation analytique et l'utilisation synthétique. A partir de la définition de Moodysson et al. [70], les auteurs distinguent comme suit les différents types d'utilisation :

- -L'utilisation analytique fait référence à l'utilisation des OGD qui explique des caractéristiques spécifiques ou résout des problèmes particuliers, tels que des problèmes publics, commerciaux ou gouvernementaux en mettant en œuvre un ensemble spécifique d'algorithmes pour analyser des ensembles de données publiques spécifiques.
- -L'utilisation synthétique fait référence à l'utilisation d'OGD pour développer des outils et des appareils qui résolvent des problèmes fonctionnels, tels que la fourniture de meilleurs services.

7 Données ouvertes et données accessibles : le concept d'accessibilité aux données

L'intention générale est de rendre les données locales, régionales et nationales (et en particulier les données acquises publiquement) disponibles sous une forme permettant une manipulation directe à l'aide d'outils logiciels, comme par exemple, à des fins de tabulation croisée, de visualisation, de cartographie, etc[48].

Rendre les données disponibles n'est pas la même chose que les rendre utilisables par tout le monde[96].

La suggestion implicite dans la plupart des discussions sur les « données ouvertes » est que « tout le monde » a le potentiel d'utiliser les données[48]. Cependant, « tout le monde » n'a pas accès à l'infrastructure numérique, au matériel ou aux logiciels, ou aux ressources/compétences financières ou éducatives qui permettraient une utilisation efficace des données ou de toute autre ressource numérique[48].

Lorsque les citoyens n'ont pas les infrastructures ou les compétences nécessaires pour accéder aux données ou leur donner un sens et les utiliser pour une action efficace, il est difficile de prétendre qu'il s'agit de « données ouvertes »[8].

Bon nombre des solutions et outils existants, bien qu'ils visent à rendre les données ouvertes plus accessibles aux communautés, sont souvent liés à des ensembles de données statiques, et ne sont pas disponibles gratuitement ou ne disposent pas des fonctionnalités nécessaires pour aider les citoyens à comprendre et à trouver de la valeur dans les données[96].

Gurstein [48] suggère que nous adoptions le concept « d'utilisation efficace » pour distinguer l'accès de l'accessibilité et de l'utilisation. Dans ce cas, nous devons distinguer les « données ouvertes » des « données accessibles ». Le premier renvoie à la définition classique des données gouvernementales ouvertes², tandis que le second est celui qui rendrait véritablement les données ouvertes et accessibles, car le simple fait d'avoir accès à ces différents ensembles de données brutes ne signifie pas un meilleur engagement civique ni même la participation des citoyens[29]. Une approche « d'utilisation efficace » des données ouvertes serait donc une approche qui garantirait que des opportunités et des ressources pour traduire ces données ouvertes en résultats utiles seraient disponibles (et adaptées) pour le plus large éventail possible d'utilisateurs[48].

2. <http://opengovdata.org>

Chapitre 2

De la donnée brute à la donnée accessible : le trajet des données

1 Objectif du chapitre et méthodologie

Dans ce chapitre nous allons décrire les différents processus de transformation des données en se basant sur le trajet des données ouvertes en données accessibles.

Pour illustrer ce trajet, nous allons partir de la pyramide de Bellinger et al. [10] présenté par Burnay [18]. Cette pyramide (figure 2.1) comprend 4 couches décrites ci-dessous :

- Donnée : cette couche est issue des outils opérationnels (population, urbanisme, finance...) et représente la donnée unitaire, *Open Data*
- Information : est issue de l'agrégation et du recoupement de diverses données, afin d'en tirer des interprétations. Ce sont des données organisées ou structurées. Par exemple des visualisations, des articles de blog, des infographies et des rapports écrits.
- Connaissance : cette couche est tirée de l'exploitation et de la visualisation de l'information par un utilisateur pour permettre un processus de décision basé sur l'informatique.
- Sagesse : cette couche concerne le résultat de la décision, par exemple la satisfaction des citoyens. Il s'agit d'une couche qu'on ne peut pas automatiser.

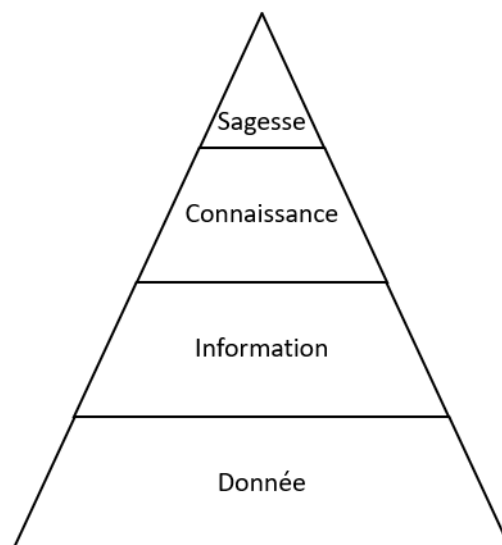


FIGURE 2.1 – Pyramide De la Donnée vers la Connaissance

Nous partons de la couche donnée en bas pour aller de couche en couche, d'abord vers la couche information puis vers la couche connaissance et éventuellement la couche sagesse.

Nous tenterons de répondre aux questions suivantes : quel est le trajet de la donnée entre les producteurs de données, par exemple les gouvernements, et les consommateurs de données, par exemple les citoyens. Quels sont les processus mis en œuvre d'un accès citoyen en matière de données publiques ouvertes? Quels sont les différentes étapes de ce processus? Comment fait-on pour prendre en compte les besoins des citoyens? Comment fait-on pour identifier les données utiles pour les citoyens? Comment exploite-t-on ces données?

Chacune de ces étapes est nourrie par un certain nombre de dispositifs ou d'outils.

La revue a été menée en sélectionnant les éléments pertinents relatifs aux questions de recherche, ceux-ci ont ensuite été regroupés en un concept plus général/de haut niveau.

Sur la base des questions de recherche, nous avons mené des recherches et consulté des experts dans le domaine afin d'obtenir une liste d'articles pertinents[103, 56, 96, 9, 47]. Ces derniers ont ensuite servi de base à la revue systématique afin de trouver d'abord, par effet boule de neige, les références citées dans ces articles et puis les termes de recherche qui répondraient le mieux à nos questions de recherche.

A partir de la lecture de ces articles nous avons essayé de recomposer les étapes des processus et d'accrocher différentes thématiques à ces étapes.

Nous avons réussi à identifier les différentes étapes liées au trajet de la donnée pour devenir une donnée ouverte. Mais nous avons aussi pu identifier les moyens pour y arriver.

Parmi ces grandes étapes nous avons pu identifier : le processus d'ouverture des données avec la préparation, la gouvernance et la publication des données. Ensuite, l'analyse des besoins des citoyens en termes de données ouvertes avec l'analyse collaborative et le storytelling (narration) ainsi que la visualisation. Et pour finir nous verrons l'exploitation des données finalisées avec les hackathons, serious game (jeux sérieux). Nous y ajoutons, à nouveau, la visualisation qui intervient dans cette étape également en plus de celle de l'analyse des besoins. Ces différentes étapes seront détaillées dans les prochaines sections.

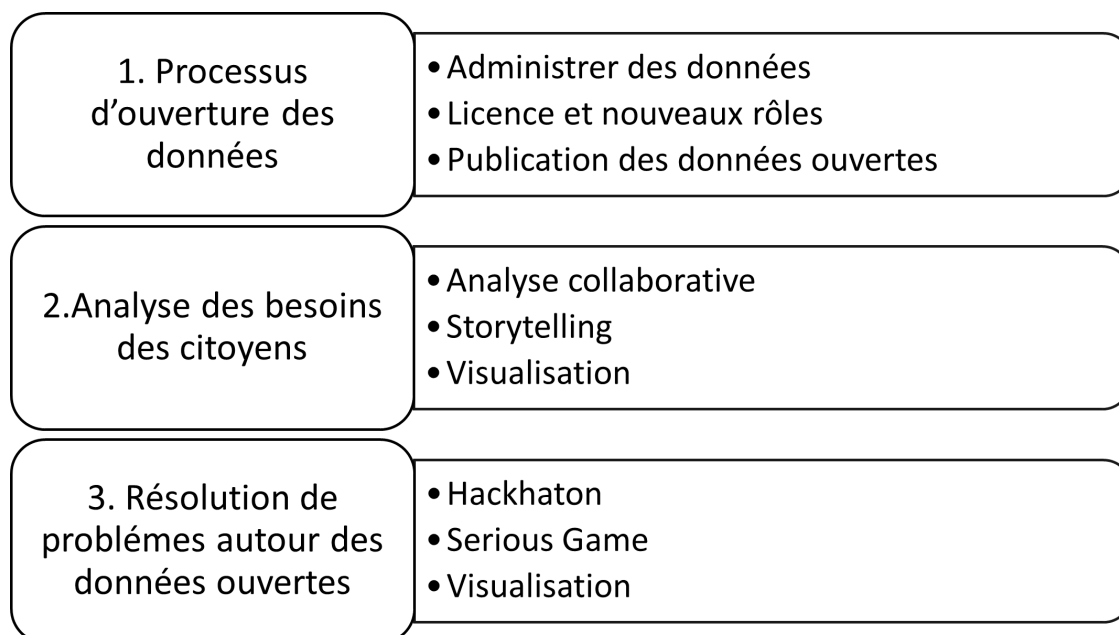


FIGURE 2.2 – Processus Open data

2 Processus d'ouverture des données

Cette section explique le contexte et la littérature derrière les différentes manières de rendre disponibles et utilisables des données ouvertes.

Les données des administrations publiques ne sont pas disponibles en l'état, prêtes à être libérées, il s'agit avant tout d'un processus de transformation de la donnée en donnée ouverte[31].

Nous essayons de répondre aux questions, comment font les administrations pour identifier les données à ouvrir, quelles sont les techniques et les différentes étapes pour réaliser cette ouverture des données, et par quels processus de transformation passent les données pour devenir des données ouvertes.

Nous allons donner un aperçu du processus d'ouverture des données avec, dans un premier temps, l'administration des données, c'est à dire du processus de transformation des données métier en données « brutes » puis en données ouvertes. C'est dans ce processus d'ouverture des données qu'interviennent les licences des données ouvertes qui stipulent si ces données pourront être copiées, diffusées et réutilisées librement. De nouveaux rôles aussi ont été créés dans les administrations pour encadrer et gérer ce processus qui peut être complexe.

Dans un second temps, le processus de publication des données est analysé. Ce processus conduit à la disponibilité des données ouvertes par leur publication sur différentes plateformes afin de les rendre utilisables par les citoyens.

2.1 Administrer la donnée

L'Open data a fait émerger la donnée comme une entité à administrer[25]. La problématisation de cette politique a implicitement cadré la donnée comme une entité homogène et autonome pouvant aisément circuler et être gouvernée de manière transverse [25].

Administrer la donnée requiert de connaître le patrimoine de données du territoire. Cela nécessite de « détacher » les données des différents silos métiers et organisationnels auxquels elles sont attachées pour les gérer de manière transversale [25].

Les données ne sont ni brutes ni autonomes[32] [88]. Elles sont toujours étroitement associées à des infrastructures informationnelles [36] composées d'un ensemble de personnes, de pratiques, de technologies, d'institutions qui les produisent, les maintiennent et les utilisent. Pour permettre leur diffusion et leur utilisation dans un environnement autre, il est nécessaire de délier l'ensemble de ces attachements afin de rendre la donnée autonome de ce cadre initial.

En outre, les systèmes d'information professionnels, souvent propriétaires et rigides, ont rarement été conçus pour extraire aisément les données qu'ils contiennent, nécessitant alors des développements informatiques qui peuvent s'avérer coûteux[25].

Denis and Goëta [31] insistent sur trois étapes cruciales de l'instauration des données en données ouvertes. L'identification, l'extraction, et la « brutification » constituant une part du travail sociotechnique dont les données font l'objet.

Il est nécessaire d'identifier les données, de déterminer leur statut juridique et leur degré de sensibilité, de nettoyer la donnée afin de la « brutifier »[32], de renseigner des métadonnées et de mettre en place des chaînes d'extraction et de traitement des données[87]. La mise en circulation des données nécessite donc le détachement de leur environnement initial. La donnée est modifiée tout au long de la chaîne de diffusion pour acquérir des propriétés de diffusabilité et se stabiliser comme une entité diffusable.

L'ensemble de ces opérations contribue à la fabrication d'une donnée « brute », détachée de son environnement initial pour pouvoir être utilisée par d'autres acteurs[25].

A partir de l'article de Denis and Goëta [31], nous avons reconstitué les différents processus sous la forme de schéma(Voir figure 2.3).

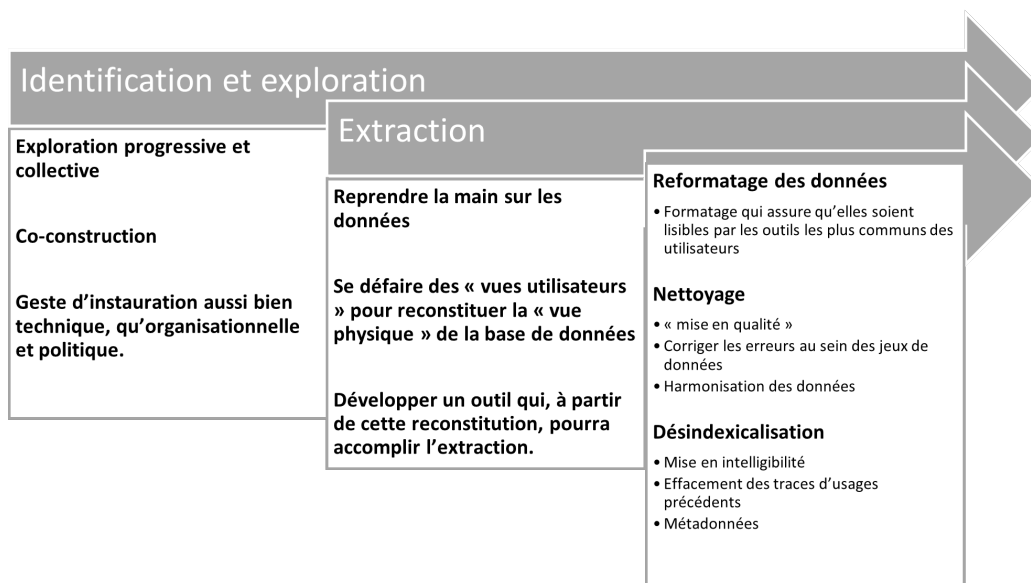


FIGURE 2.3 – Administrer la donnée

Pour Denis and Goëta [31] il existe deux grandes directions possibles pour prendre en considération le travail de fabrication des données brutes. Le travail d'ouverture peut être mené soit a posteriori sur les données métier ou soit ex-ante lors de leur production. A posteriori : Une fois ce travail éprouvé, puis reconnu, c'est-à-dire une fois que l'on assume que l'ouverture des données a un coût, qu'elle représente même un investissement, on peut l'assumer comme une série d'opérations à mener a posteriori sur les données métier, il faut alors inventer des postes et redéfinir des rôles au sein de l'organisation. Ex-Ante : On peut au contraire chercher à intégrer ce travail en amont, en transformant la nature même des données sur les sites de leur production et dans leurs premiers usages.

La différence entre les deux directions ne tient pas tant à la part organisationnelle de la fabrique des données brutes (elle est présente à chaque fois) qu'à la définition sous-jacente de ce que l'on entend par données. Dans le premier cas, la multiplicité des données et la nécessité d'en faire coexister des versions différentes au sein de l'institution sont assumées. Dans le second, le caractère générique des données — leur aspect « brut » — est considéré comme un bien en soi, sur lequel il faut aligner les idiosyncrasies professionnelles.

2.2 Nouveaux rôles

La mise à disposition des données ne relève pas de la livraison de données déjà existantes qu'il s'agirait de restituer, mais d'un travail méticuleux de façonnage de données que l'on s'engage à donner[31].

Dans leur étude de cas, Courmont [25] nous parlent des nouveaux rôles créés au sein de la métropole de Lyon, pour gérer ce processus d'ouverture des données.

L'Administrateur/trice général(e) des données ou *Chief Data Officer (CDO)* joue un rôle majeur dans la gouvernance des données de l'institution. Il doit connaître les données produites sur le territoire, s'assurer de leur qualité, encourager leur circulation, définir les données cruciales à produire et maintenir, expérimenter de nouvelles méthodes basées sur l'usage des données, élaborer et défendre une politique métropolitaine de la donnée pour laquelle il devra rendre des comptes annuellement[25].

Le/la scientifique des données ou *Data Scientist* a des missions visant avant tout à piloter les évolutions de la plateforme Open data en intégrant notamment des technologies de gestion

de données de type Big data. Ce poste représente ainsi le volet technique de l'administration des données. En tant que chef de produit de la plateforme de données, il a pour mission d'assurer la collecte, l'organisation et la valorisation des données provenant de systèmes d'informations multiples et hétérogènes [25].

2.3 Licence

L'ouverture des données implique aussi d'y ajouter une licence fixant les conditions dans lesquelles ces données pourront être copiées, diffusées, réutilisées librement.

Afin de pouvoir mettre les données à disposition de chacun, plusieurs licences existent et sont déjà prêtes à être utilisées, parmi lesquelles on peut citer les licences de Creative Commons[1], de l'Open Knowledge Foundation[2, 83] ou encore celle d'Etalab[39]. Toutefois, dans certaines conditions, certains acteurs peuvent se mettre à élaborer leur propre licence si les licences existantes ne les satisfont pas ou tout simplement s'ils préfèrent garder la maîtrise totale sur la libération de leurs données[67].

Par exemple, la Ville de Bruxelles[74] utilise la licence ouverte conçue par Etalab[39](France).

2.4 Publication des données ouvertes

L'ouverture des données s'effectue par la publication des données sur des sites internet, des portails ou via des API.

Les administrations publiques ont le choix de publier leurs données ouvertes sur les portails publics d'*Open Data* mais aussi sur leurs propres portails, via des solutions clés en main ou de leur propre initiative.

La popularité des données ouvertes et la disponibilité de différentes sources de données ont stimulé le développement d'une variété de plateformes pour utiliser ces ensembles de données[96].

Nous avons construit cette section sur la base d'une exploration de la littérature pour expliquer les différents outils de publication des données ouvertes, nous reprenons aussi des exemples d'outils recommandés par des blogs, livres et articles. Nous commençons d'abord par nous poser les questions suivantes : Où peut-on trouver des données ouvertes ? Quelles sont les différentes solutions disponibles pour accéder aux données ouvertes ? Comment font les administrations pour mettre à disposition les jeux de données ouvertes ?

Nous allons présenter les plateformes de solution de données ouvertes qui permettent une implémentation de portail de données ouvertes, ces solutions clé en mains sont utilisées notamment par les villes, entreprises privées, associations pour mettre à disposition les données. Et ensuite nous allons présenter quelques exemples de portails publics en Europe et en Belgique.

2.4.1 Solutions de plateformes Open Data

De nombreux portails de données ouvertes fonctionnent sur base de plateformes qui offrent des solutions de gestion des données ouvertes. Ces plateformes incluent des fonctionnalités telles que la gestion des métadonnées, des visualisations de base, des API, la gestion des utilisateurs, la publication et le stockage des jeux de données (*datasets*).

La plupart des portails de données ouvertes sont implémentés sur la base de plateformes bien connues, parmi lesquelles nous pouvons citer Socrata [120] et CKAN[21], mais aussi l'Open Government Data Initiative de Microsoft[55], OpenDataSoft[69] ou des solutions personnalisées.

Opendatasoft propose à ses clients des solutions de partage de données de pointe qui permettent aux organisations de publier, gérer, combiner, analyser, visualiser et partager facilement des données dans une variété de formats sur une seule plateforme[84]. Les ensembles de données Opendatasoft sont accessibles par les développeurs via une API HTTP REST. Cette solution est choisie par de très nombreux services publics en Europe et dans le reste du monde.

Par exemple en Wallonie, la solution Opendatasoft a également été déployée par la Ville de Liège[79], ainsi que la Ville de Namur[80].

Open Government Data Initiative de Microsoft(OGDI) est un « kit de démarrage » qui peut être utilisé pour publier des données sur le Web et les rendre universellement accessibles via l'API RESTful et les formats ouverts[23].

Comprehensive Knowledge Archive Network (CKAN) est un outil de création de sites Web de données ouvertes[20]. C'est un système de gestion de données puissant qui rend les données accessibles - en fournissant des outils pour rationaliser la publication, le partage, la recherche et l'utilisation des données. Il aide à gérer et à publier des collections de données. Il est utilisé par les gouvernements nationaux et locaux, les instituts de recherche et d'autres organisations qui collectent de nombreuses données. Une fois les données publiées, les utilisateurs peuvent utiliser ses fonctionnalités de recherche pour parcourir et trouver les données, ainsi que les prévisualiser à l'aide de cartes, de graphiques et de tableaux. CKAN est un logiciel open source, avec une communauté active de contributeurs qui développent et maintiennent sa technologie de base.

Exemple : le portail U.S. Government's open data[81] se base sur CKAN.

Socrata est une solution SaaS basée sur le cloud qui permet aux organisations gouvernementales de mettre leurs données en ligne, de prendre des décisions basées sur les données, de fonctionner plus efficacement et de partager des informations avec les citoyens[106, 92].

Exemple : la ville d'Edmonton[76] au Canada utilise Socrata pour partager ces données.

2.4.2 Portails Publics

Un portail Open Data est un espace de présentation des données publiques ouvertes, il permet l'accès aux données ouvertes sous différents formats, y compris via des APIs afin de pouvoir facilement incorporer les données dans une application Web ou mobile.

Il permet aussi de faire une recherche sur les différents jeux de données disponibles ; Certains permettent aussi une présentation visuelle des données, sous forme de graphique ou de carte ; Ils disposent aussi tous d'un outil permettant l'extraction ou le transfert des données. Selon le type de données, elles peuvent être téléchargées sous forme de fichier au format CSV, JSON, XLS, GeoJSON, Shapefile. Les données peuvent également être utilisées via une API. Ces portails permettent aussi de soumettre ses propres ensembles de données pour les rendre disponibles à tous, dans ce cas et en fonction des plateformes, les procédures sont différentes.

2.4.3 Exemples de portails publics

Le gouvernement local de Chicago(USA) offre un portail[75] pour diffuser des données ouvertes en ligne[59]. La politique du portail permet d'accéder à une grande variété de données et autorise l'exploration et l'interprétation des informations disponibles, au moyen de technologies, développées pour extraire de nouvelles informations et connaissances liées à la ville[9].

En Belgique il existe plusieurs portails pour trouver les données ouvertes en fonction des régions ou localités mais aussi des plus globaux avec, entre autres, le portail[73] fédéral belge qui reprend les données au niveau fédéral.

En Wallonie, dans le portail[82], on y trouve l'ensemble des jeux de données publiés par les portails indépendants de certaines entités publiques Région-Fédération, comme les villes de Liège et Namur.

Il existe aussi un portail en région flamande[78].

La ville de Bruxelles a aussi son portail Open Data[74] pour permettre au citoyen d'y accéder. Datastore.brussels est la plateforme régionale[28] pour l'ouverture et le partage de données et services en Région de Bruxelles-Capitale. Cette plateforme a pour but de répondre aux besoins

tant des utilisateurs que des producteurs de données, de manière à leur permettre d'ouvrir et de réutiliser en toute liberté les données et services disponibles en Région de Bruxelles-Capitale.

Le Portail européen de données recueille les métadonnées des informations du secteur public disponibles sur les portails de données publics des différents pays européens[77]. Il comprend également des informations concernant la publication de données et les avantages offerts par leur réutilisation. Le portail fournit : un catalogue standardisé, facilitant l'accès aux données ouvertes de l'UE ; une liste d'applications et d'outils en ligne réutilisant ces données ; un éditeur de recherches au moyen du point de terminaison SPARQL ; un accès au moyen de l'API REST ; des conseils sur la manière d'utiliser au mieux le site.

3 Analyse des besoins des citoyens en termes de données ouvertes

La disponibilité des ensembles de données n'est pas suffisante pour que les citoyens puissent en faire usage. Pour cela il est nécessaire que ceux-ci comprennent les données ouvertes mises à leur disposition afin d'en faire bon usage.

Une bonne réutilisation des données nécessite de créer un contexte pour les utilisateurs, ce qui pourrait découler d'une compréhension de la façon dont les données ont été créées et seront utilisées[29]. Cela suggère une relation complexe entre ceux qui créent des données ouvertes, ceux qui utilisent des données ouvertes et les technologies qui prennent en charge l'accès aux données ouvertes[37].

Dans cette partie, nous parlerons des moyens existants pour faire une analyse des besoins de citoyens en termes de données ouvertes. Nous tenterons de répondre aux questions : Maintenant que les données ouvertes sont disponibles, comment font les citoyens, les organisations pour les utiliser ? Comment les citoyens peuvent-ils dans un premier temps, comprendre les données publiées et identifier les données pertinentes par rapport aux questions qui se posent ?

En se basant sur la littérature, nous avons pu mettre en évidence l'analyse collaborative, le storytelling et aussi la visualisation qui constituent des moyens de faire une analyse des besoins des citoyens en termes de données ouvertes(voir figure 2.4). La visualisation sera abordée dans la section suivante étant donné qu'elle est utilisée non seulement pour l'analyse des besoins des utilisateurs mais aussi pour la résolution de problèmes autour des données ouvertes.

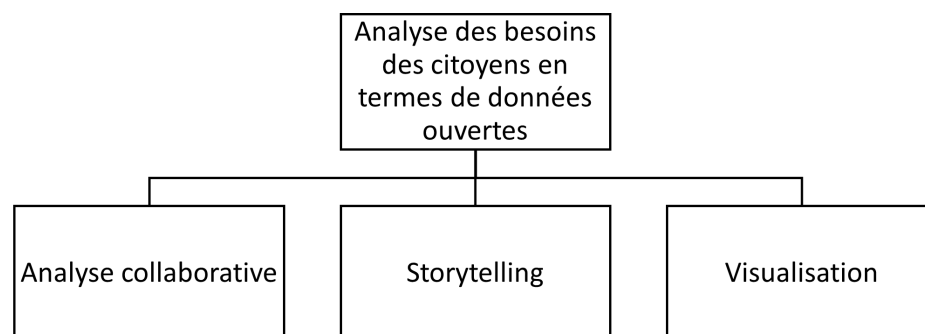


FIGURE 2.4 – Moyens utilisés pour l'analyse des besoins

3.1 Analyse collaborative

Dans leur article, Erete et al. [37] encouragent l'interprétation et la consommation de données grâce à l'analyse collaborative.

La recherche sur l'analyse collaborative suggère des avantages à établir des partenariats avec d'autres pour comprendre les ensembles de données ouvertes[49] [85] [118];

L'analyse collaborative, dans laquelle les consommateurs de données (experts et non-experts) travaillent ensemble pour analyser, interpréter et visualiser les données [29] [43] [49] [116], a retenu l'attention. La production de ces analyses nécessite une communication entre les utilisateurs pour faciliter la prise de décision en « distribuant des informations entre les membres du groupe »[49].

Erete et al. [37] suggèrent que non seulement les experts sont bénéfiques pour les utilisateurs[5, 4], mais que la relation entre les deux peut être mutuellement bénéfique. Ainsi, les technologies qui prennent en charge l'analyse collaborative devraient soutenir l'établissement de relations et la symbiose de partage des connaissances entre les utilisateurs et les experts.

3.1.1 Analyse visuelle collaborative et interactive

Les systèmes d'analyse visuelle collaborative basés sur le Web facilitent la collaboration en permettant aux analystes de lier des commentaires textuels et des annotations graphiques à des vues ou des états spécifiques d'une visualisation interactive [118]. Ces systèmes se sont principalement concentrés sur l'utilisation de commentaires pour partager des questions et des observations dans l'analyse exploratoire, tout en ignorant les tâches analytiques plus complexes telles que la collecte de preuves, l'organisation des résultats, la pondération des alternatives et la synthèse des résultats.

Les outils d'analyse visuelle collaborative des données[37] aident les utilisateurs à consommer plus facilement les données en mettant en évidence des informations et des résultats importants[85].

Pour Clark and Brennan [22] sur un terrain d'entente, la compréhension partagée nécessaire pour une communication réussie implique que les collaborateurs sont plus efficaces lorsqu'ils peuvent se référer à un environnement visuel partagé pour fonder les actions et les commentaires de chacun[42].

Willett et al. [118] examinent comment un petit vocabulaire fixe de balises (question, hypothèse, à faire) et de liens (preuves pour, contre preuve) peut aider les analystes à collecter et organiser de nouvelles preuves, à identifier les conclusions importantes faites par d'autres et à synthétiser leurs résultats.

Jansen et al. [53] définissent les avantages des données physicalisées par rapport aux autres formes de représentation : tirer parti des compétences d'exploration perceptuelle, rendre les données plus accessibles et plus engageantes pour une population plus large, et faciliter l'apprentissage et la compréhension. Ils proposent de penser à la physicalisation des données comme un domaine de recherche qui examine comment les représentations physiques des données assistées par ordinateur (c'est-à-dire les physicalisations) peuvent soutenir la cognition, la communication, l'apprentissage, la résolution de problèmes et la prise de décision [53].

Sur la base de ces idées, un certain nombre d'interfaces de tables tangibles ont été développées pour permettre l'analyse physicalisée de grandes quantités de données. Un exemple est inFORM, qui utilise des barres programmables pour donner aux données une forme physique [40].

Ces interfaces encouragent la discussion de groupe et la collaboration, mais présentent certains inconvénients en ce qu'elles sont volumineuses (et donc difficiles à déplacer) et coûteuses[119].

Il existe un certain nombre d'études visant à rendre l'engagement avec les données plus interactif. Parmi ces études nous pouvons épingler :

VoxBox un système de questionnaire physique. Il s'agit d'un exemple de dispositif de collecte de données tangible avec une série de leviers et de boutons qui le rendent plus attrayant et engageant que les questionnaires papier conventionnels. De même le projet Sens-us[45] dans lequel la

soumission des réponses traditionnelles au questionnaire du recensement britannique a été rendue physique et amusante[44].

Datacatcher vise à susciter la curiosité sur les données grâce à un appareil portatif qui peut être transporté et qui permet d’afficher des données pertinentes à l’emplacement physique actuel des utilisateurs, renforçant ainsi leur compréhension de la relation entre les données et l’environnement[41].

Data-in-Place a cherché à étudier le sens et la signification contextuelle des données avec une combinaison de technologies numériques dans les maisons des résidents et dans les espaces partagés sur une rue[62, 111] .

CommentSpace introduit un modèle général dans lequel les analystes peuvent lier les commentaires à plusieurs états de visualisation et les situer non pas dans une, mais dans de nombreuses discussions thématiques[85].

3.2 La narration - Storytelling

La narration (*data storytelling ou storytelling*) est l’art d’utiliser les données pour raconter une histoire ce qui permet aux informations les plus importantes d’apparaître plus clairement.

Sur la base de l’étude de Boje [12] sur les organisations utilisant la narration, Erete et al. [37] définissent la narration comme « comment les gens [ou les organisations] interprètent des histoires pour donner un sens aux événements, introduire le changement et obtenir un avantage politique ». Les organisations à but non lucratif (OSBL) utilisent la narration pour rendre les données consommables pour leur public cible, qui peut inclure des parties prenantes internes et externes, des fonctionnaires, des bailleurs de fonds, des investisseurs financiers, des résidents.

De nombreux travaux ont été menés sur les organisations utilisant la narration comme mécanisme de changement organisationnel en obtenant le soutien des parties prenantes, en partageant les faits et les connaissances et en co-créant des récits qui éclairent l’orientation future des organisations[33].

Pour que la narration se produise, Otjacques et al. [85] expliquent que les visualisations ont besoin d’une identité. Cela aide les gens à les reconnaître, à les localiser et à évaluer s’ils veulent les commenter. Ils[85] ont nommé «décoration» le processus consistant à donner une identité dans le réseau à une visualisation.

4 Résolution de problèmes autour des données ouvertes

Dans cette partie nous allons aborder la question de l’exploitation des données ouvertes et la résolution de problèmes autour des données ouvertes.

Nous allons présenter des moyens d’exploiter les données ouvertes à travers des outils et des technologies qui permettent de faire des animations autour d’un problème lié aux données ouvertes. Nous présenterons les serious games, les hackatons et les visualisations comme moyens d’exploiter les données ouvertes.

4.1 Serious game

Le terme Serious game ou « jeux sérieux » fait référence à des jeux conçus pour faire plus que simplement divertir [109]. Zyda [125] propose la définition suivante : les jeux sérieux sont « un concours mental, joué avec un ordinateur conformément à des règles spécifiques qui utilise le divertissement pour promouvoir la formation, l’éducation, la santé, la politique publique et les objectifs de communication stratégique du gouvernement ou de l’entreprise ». Deux caractéristiques clés des jeux sérieux sont leurs qualités éducatives et immersives[90].

Les principaux atouts des applications de jeux sérieux peuvent être généralisés comme appartenant aux domaines de la communication, de l’expression visuelle de l’information, des mécanismes de collaboration, de l’interactivité et du divertissement [7, 104].

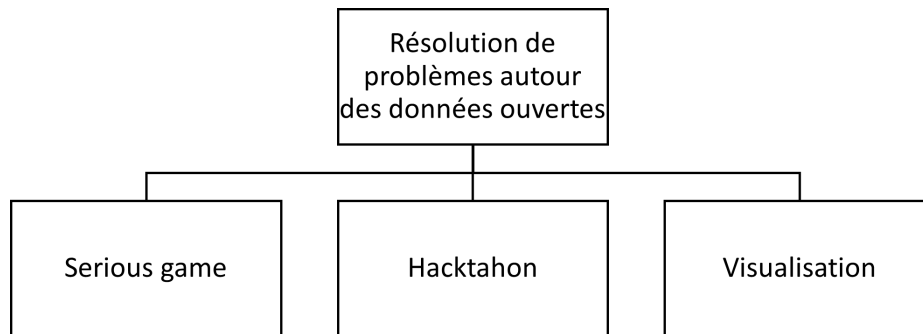


FIGURE 2.5 – Résolution de problèmes autour des données ouvertes

Il existe une longue tradition dans la recherche participative de conception et la co-conception autour des jeux participatifs comme moyen d’impliquer les citoyens et de faciliter la collaboration[15, 38], et cette approche est courante dans les contextes de la société civile, tel que les consultations d’urbanisme[90].

@stake est un exemple de technologie civique pour structurer le dialogue, un *serious game* conçu pour renforcer la capacité de délibération par le jeu de rôle et faciliter des interactions plus délibératives dans les consultations associées aux exercices de budgétisation participative[46].

Datascape est un jeu conçu pour aider les citoyens à mieux comprendre la relation entre les données et l’environnement dont elles sont issues. Le Datascape est initialement présenté comme un jeu destiné à être utilisé avec des communautés qui commencent à réfléchir à l’exploration des données ouvertes[119]. Il peut s’agir de communautés essayant d’utiliser des données pour identifier ou vérifier des problèmes, par ex. avec des services locaux ou des citoyens qui créent de nouvelles applications à partir des données pour servir un objectif particulier[119].

4.2 Hacktaton

Purwanto et al. [95] le définissent comme suit : Un hackathon de données ouvertes fait référence à un concours d’idéation hors ligne/en face-à-face parrainé par des agences gouvernementales dans un endroit centralisé qui rassemble des citoyens d’horizons différents (par exemple, des programmeurs, des concepteurs, d’autres) pour travailler de manière intensive en collaboration en petites équipes pour une courte durée (par exemple, 12 heures, 24 heures, 2 jours) pour créer des artefacts (par exemple, maquettes, conception, prototypes, applications) à l’aide des OGD. Ce type d’engagement prend généralement la forme d’événements hackathon et vise à offrir une valeur économique[57]. En règle générale, à la fin du concours, chaque équipe présente l’idée finale devant des jurys et des sponsors, et une équipe gagnante gagne un prix (par exemple, de l’argent, un investissement).

Pour Purwanto et al. [95], un hackathon de données ouvertes, les organisateurs et les sponsors fournissent presque toutes les ressources et le soutien nécessaires aux équipes pour travailler efficacement [24, 107, 16], y compris des services de restauration, des sacs de couchage/zone, des installations confortables (appareil de jeu, salle de sport), Internet connexion, électricité (câbles)...

La fourniture d’un soutien technique de la part de fournisseurs de données ouvertes ou d’organisa-teurs ou de sponsors d’événements est également courante. Ces équipements sont intentionnellement fournis pour soutenir le développement du groupe lors de l’événement hackathon.

Ce modèle d'innovation est aujourd'hui de plus en plus adopté par les pouvoirs publics afin de favoriser une nouvelle ère de collaboration, de gouvernance ouverte et d'engagement citoyen.

Exemple : Hackathon Citizens Of Wallonia est un hackathon dédié aux Smart Cities[19] .

4.3 Visualisation de données

Dans la mesure où les outils de visualisation seront au coeur de notre analyse dans le chapitre 3, nous nous contentons dans cette partie d'introduire de manière sommaire la visualisation des données.

La visualisation des données ou *Data Visualisation* ou *Dataviz* consiste à représenter de façon visuelle les données, de manière à ce qu'elles soient compréhensibles par ceux qui les manient ou les consultent.

L'un des principaux objectifs de la visualisation des données est d'aider à interpréter les informations. Dans une visualisation réussie, un mélange efficace de contenu, de contexte, de construction et de conception est attendu [121].

Graves and Hendler [47] proposent d'utiliser les visualisations comme moyen de consommer, partager et interagir avec les données. Les outils de visualisation de données permettent à quiconque d'organiser et de présenter les informations de manière intuitive.

Une visualisation est créée en faisant correspondre un ensemble de données avec un composant de visualisation[115].

La visualisation des données peut revêtir des formes diverses : graphiques, camemberts, cartographies, représentations chronologiques, nuages de mots, bulles, radars interactifs, animations vidéo, représentations interactives, etc. afin de permettre de rendre les données plus lisibles.

Pour Viegas et al. [115] non seulement les visualisations interactives sont un moyen clé de communication dans un monde riche en données, mais les rapports préliminaires suggèrent que les visualisations ont potentiellement un effet catalytique sur la narration [114] et l'analyse collective des données[117].

Des recherches antérieures ont mis en évidence plusieurs éléments à prendre en compte pour visualiser les données ouvertes[85] : (1) la technologie doit avoir une faible barrière à l'entrée ; (2) la « décoration » peut aider à soutenir la narration autour des visualisations. Otjacques et al. [85] désignent par « décoration » le processus consistant à donner une identité dans le réseau à une visualisation. (3) garder des traces de l'historique est nécessaire ; (4) une représentation visuelle des différences peut aider à gérer une collection de graphiques similaires ; (5) les éléments sociaux influencent l'interprétation des graphiques. En effet, Hullman et al. [50] ont montré avec une expérience que la présentation d'informations sociales (c'est-à-dire les réactions antérieures d'autres personnes à propos du même graphique) influence le jugement de quelqu'un qui voit un graphique.

Chapitre 3

Exploration des outils de visualisation : Méthodologie et résultat

1 Introduction

Parmi les ressources existantes qui transforment les données ouvertes en informations (voir figure 2.1) plus parlantes pour les utilisateurs finaux, il y a entre autres, les outils de visualisation. Les outils de visualisation de données permettent aux utilisateurs d'organiser et de présenter les données de manière plus visuelle.

Nous allons voir dans cette partie en quoi les outils de visualisations permettent une utilisation efficace des données ouvertes, quels sont les critères en lien avec l'utilisation efficace, quels sont les différents outils qui existent et peuvent-ils être utilisés par tout le monde.

Ce chapitre se déroulera comme suit : Dans un premier temps, nous expliquerons la manière dont nous avons sélectionné les différents outils de visualisation qui seront analysés dans ce mémoire. Nous en ferons, ensuite, une brève présentation. Dans un second temps, nous expliquerons la grille méthodologique des critères que nous avons élaborée pour évaluer l'accessibilité et l'utilisabilité des outils de visualisation retenus.

2 Sélection et présentation des outils de visualisation

Il existe un grand nombre d'outils qui permettent aux utilisateurs de créer des visualisations à partir de données.

Ils sont très nombreux et trop nombreux pour pouvoir faire l'objet d'une évaluation détaillée dans le cadre de ce mémoire. C'est la raison pour laquelle nous avons sélectionné un sous-ensemble pertinent, sur la base d'une méthodologie de sélection définie dans la section suivante.

2.1 Méthode de sélection : état de l'art et citation dans les articles

Pour identifier ces dispositifs et ces outils nous avons retenu les outils qui apparaissent le plus pertinents par rapport à notre problématique d'utilisation efficace et effective des données et qui ont été cités le plus souvent dans la littérature existante.

Nous sommes partis de l'article de Graves and Hendler [47], pour trouver les outils de base. Nous avons ensuite étoffé cette première liste en consultant les références citées par l'article de départ et d'autres articles mentionnés dans ces derniers. Sur cette base, nous avons pu stabiliser une liste

de 10 outils répondant à la fois aux critères de citation dans les articles consultés et d'intérêt pour notre problématique d'utilisation efficace des données.

De nombreuses applications axées sur la *Business Intelligence* (BI) et l'analyse de données permettent aux utilisateurs de créer des visualisations basées sur des données[47]. A ceux-ci identifiés dans la littérature, nous avons ajouté les outils de informatique décisionnelle ou *Business Intelligence* mentionné par Burnay [18]. Ces outils sont : Tableau, Microsoft Power Bi et Qlik.

Pour organiser leur présentation, nous avons regroupé ces outils en fonction de la manière d'y accéder et de leur utilisation. Nous avons tout d'abord la suite Office avec Microsoft Excel. Puis il y a les outils disponibles sur le Web avec Infogram, TIBCO Spotfire et RAW Graphs. Puis les logiciels de Business Intelligence avec Tableau, Microsoft Power Bi et Qlik Sense. Il existe aussi un langage de programmation qui peut être utilisé notamment pour créer des visualisations : R. Nous avons aussi inclus dans les outils de visualisation, des APIs et bibliothèques tels que Processing.JS ou p5.js, Protovis et D3.js.

Parmi les outils les plus cités dans la littérature, il y a en 2 qui ne sont plus maintenus depuis quelques années, Protovis et Processing.JS. Ces outils seront mentionnés dans la description pour information mais ne seront pas évalués dans le chapitre suivant. Nous testerons les nouvelles versions de ces outils, à savoir P5.js. et D3.js. .

Nous allons donc présenter et analyser les outils suivants :

- Microsoft Excel
- Infogram
- TIBCO Spotfire
- Rawgraphs
- Tableau Public
- Microsoft Power Bi
- Qlik Sense
- R langage de programmation
- Processing.JS/p5.JS
- Protovis/D3.JS

Ces outils seront présentés brièvement dans la section suivante.

2.2 Brève présentation - description des outils retenus

Microsoft Excel est un logiciel tableur de la suite bureautique Microsoft Office qui permet d'utiliser des données tabulaires pour créer différents types de graphiques[47].

Infogram est un outil Web qui permet aux utilisateurs de créer des visualisations basées sur les données qu'ils ont téléchargées[51]. Il fournit plusieurs modèles de graphiques (par exemple, camembert, graphiques, graphiques en courbes, graphiques à bulles) [47]. Infogram est disponible ici : <https://infogram.com/>

TIBCO Spotfire est un logiciel de visualisation, qui fournit une interface simple et utilisable pour la création de visualisation de données, de tableaux de bord[69]. Elle permet également de faire des analyses sur les données. La plupart du travail dans Spotfire peut être effectué par glisser-déposer et il existe aussi une multitude de fonctions intelligentes qui rendent les tâches de visualisation plus accessibles. Il possède des fonctionnalités avancées qui lui permettent de s'exécuter rapidement avec de grands ensembles de données[112].

RAW Graphs est un Framework de visualisation de données open source[100]. RAW Graphs vise à fournir un lien manquant entre les applications de tableur (par exemple Microsoft Excel, Apple Numbers, OpenRefine) et les éditeurs de graphiques vectoriels (par exemple Adobe Illustrator, Inkscape, Sketch)[66].

Tableau est une suite logicielle de Business Intelligence qui permet aux utilisateurs de générer des visualisations[63, 3].

Tableau est composé de Tableau Desktop, Tableau Server, Tableau Online, Tableau Prep, Tableau Public mais seuls les outils concernés par la visualisation seront étudiés ici, c'est à dire Tableau Desktop et Tableau Public. Tableau Desktop est une application à télécharger, elle permet de créer des visualisations, soit avec Tableau Desktop Professional Edition, soit avec la version gratuite Tableau Desktop Public Edition. Tableau Public est un service gratuit qui permet à quiconque de publier des visualisations de données interactives sur le Web. Tableau Public est accessible via un navigateur.

Microsoft Power BI est une solution BI d'analyse de données de Microsoft. Il permet de créer des visualisations de données personnalisées et interactives mais aussi des rapports et tableaux de bord.

Qlik Sense est une plateforme d'analyse de données qui permet de faire des personnalisations complètes grâce à ses API ouvertes. Qlik Sense est une solution BI conçue pour ceux qui souhaitent créer rapidement des visualisations et explorer leurs données en profondeur.

R Langage de programmation est un langage de programmation et un environnement logiciel libre qui permet de faire de l'analyse des données et des graphiques. R[99] fournit une grande variété de techniques statistiques (modélisation linéaire et non linéaire, tests statistiques classiques, analyse de séries chronologiques, classification, clustering,...) et graphiques, et est hautement extensible.

Ce genre d'outils fournit des mécanismes puissants pour utiliser des données provenant d'un nombre varié de sources et de formats, mais ils nécessitent également une certaine expertise technique de leurs utilisateurs[47].

Processing.JS est le portage du langage de programmation Processing en JavaScript. Il fournit une série de primitives de bas niveau (points, lignes, ellipses) qui sont appliquées dans l'élément canvas disponible en HTML5 [47]. Processing[101] utilise du code écrit dans le langage de programmation Processing (similaire à C++) pour générer des visualisations 2D et 3D.

Avec le développement de p5js et les avancées de l'API dans Processing, Processing.js qui était en mode maintenance depuis plusieurs années maintenant a été archivé en décembre 2018. Cet outil ne sera donc pas analysé dans le cadre de ce document.

P5.JS est une bibliothèque JavaScript pour le codage créatif, avec pour objectif de rendre le codage accessible et inclusif pour tout le monde. p5.js dispose d'un ensemble complet de fonctionnalités pour le dessin mais aussi pour la vidéo, la webcam et le son. Il dispose aussi d'un Éditeur interactif en ligne.

Protovis fournit des primitives aux éléments visuels de base (lignes, rectangles) appelés marques, qui sont associés à des éléments de données[47]. Ainsi, les données peuvent être représentées sous forme de visualisations en composant plusieurs marques qui représentent un ensemble de données. Sur le site de l'éditeur[94], on peut y lire que Protovis n'est plus sous développement actif et ils recommandent de passer à D3.js. Il ne sera donc pas évalué dans la partie suivante.

D3.js [14, 47] est un Framework pour les visualisations et manipulations de données dans les documents Web dérivés de Protovis[13]. D3(Data-Driven Documents ou D3.js) est une bibliothèque JavaScript qui permet aux développeurs de créer des visualisations à l'aide de graphiques vectoriels adaptables (SVG) [97]. Nous allons aussi voir la plateforme Observable qui est un bloc-notes JavaScript interactif pour travailler avec D3.js et d'autres bibliothèques graphiques. La plateforme Observable contient aussi une nombreuse documentation liée à la nouvelle version de D3; On est automatiquement redirigé vers cette plateforme lorsqu'on veut regarder certains tutoriels.

3 Explication des critères en lien avec l’accessibilité et l’utilisabilité

Pour définir le cadre méthodologique de notre évaluation des outils de visualisation, nous avons cherché dans la littérature des critères ou des grilles qui permettraient de soutenir cette analyse. Étant donnée la perspective générale de notre mémoire qui entend travailler la question de l’accessibilité et l’utilisabilité des données ouvertes, il nous est apparu important de rester dans cette perspective pour l’évaluation des outils de visualisation. C’est la raison pour laquelle, nous avons choisi de nous appuyer sur les critères définis par Gurstein [48] comme particulièrement critiques pour un usage efficace des données ouvertes. Rappelons que cet auteur veut, à travers ces critères, souligner qu’une accessibilité de données ouvertes ne signifie pas automatiquement une utilisabilité de celles-ci par le grand public. Dans notre recherche d’une grille de critères permettant d’évaluer les outils de visualisation, la grille de critères proposée par Gurstein [48] et qui vaut pour tout le processus d’usage des données ouvertes nous est apparu comme une bonne base, particulièrement appropriée pour cette évaluation centrée sur les utilisateurs que nous souhaitons mener sur les outils de visualisation.

Dans son article, Gurstein [48] propose un modèle à 7 couches pour une « utilisation efficace » [48]. Les couches comprennent : (1) Internet, (2) ordinateurs et logiciels, (3) compétences informatiques / logiciels, (4) contenu et formatage, (5) interprétation / création de sens, (6) soutien et collaboration (*Advocacy*), (7) gouvernance.

Partant de cette grille de base, nous avons déployé chacune des 7 couches en sous-critères plus opérationnels afin de les adapter à notre propos, à savoir l’évaluation des outils de visualisation et, par ailleurs, pour structurer et étoffer notre analyse.

3.1 Internet

La première couche est Internet car la majeure partie des outils ont besoin de disposer d’un **accès Internet** pour y accéder, les télécharger ou pour les utiliser en ligne.

Les aspects d’Internet comme l’abordabilité de l’accès Internet, la disponibilité d’une bande passante suffisante et l’accessibilité du réseau ne seront pas développés dans ce travail. En effet, d’une part nous n’avons pu tester ces outils que dans les conditions singulières de notre propre infrastructure Internet à haut débit et d’autre part, en limitant notre réflexion à la Belgique, nous supposons que ces conditions sont remplies pour la plupart des citoyens soit en mode privé, soit via des points d’accès publics.

3.2 Ordinateurs et logiciels

Ce critère est en lien avec la manière d’**accéder à l’outil** de visualisation mais aussi l’équipement nécessaire pour utiliser les outils.

C’est un critère qui joue sur l’accessibilité car, pour utiliser ces outils de visualisation, il est nécessaire d’avoir un ordinateur à disposition et pour certains outils il faut être capable d’installer les logiciels sur cet ordinateur. Nous considérons que les outils les plus accessibles sont ceux qui peuvent être utilisés directement grâce à un navigateur sans configuration spéciale, téléchargement et installation qui nécessitent d’avoir des ressources disponibles sur l’ordinateur.

Parmi les outils que nous avons testé, il y a des applications de bureau à télécharger et à installer, des outils accessibles via un navigateur Web, des bibliothèques et APIs de visualisation.

3.3 Compétences informatiques/ logiciel

Pour pouvoir utiliser les OGDs (dans notre cas, les outils de visualisation), des compétences techniques et des connaissances sur les données sont nécessaires, telles que des connaissances statistiques ou en programmation[103]. La maîtrise de l’informatique est considérée comme une condition plus

importante que les ressources financières et autres pour établir une innovation en utilisant les OGD[108].

Pour pouvoir faire des visualisations, il est nécessaire de comprendre le jeu de données, mais aussi de comprendre les différents types de représentations qui existent (tableaux, graphes, arbres, interactives...) pour pouvoir choisir la bonne représentation visuelle correspondant à la qualité des données disponibles et au sens du résultat que l'on veut obtenir. C'est ce que nous regroupons dans le critère **compétences en statistiques**.

Parmi les compétences à avoir pour utiliser les outils de visualisation, nous avons aussi retenu la **connaissance d'au moins un langage de programmation**, qui est nécessaire pour manipuler certains outils notamment les bibliothèques et les API. C'est un critère qui joue sur l'accessibilité car en fonction des compétences informatiques et logicielle, l'utilisateur pourrait être limité dans l'utilisation de certains outils, voir ne pas savoir les utiliser. Le fait que l'utilisateur n'ait pas à connaître un langage de programmation pour utiliser l'outil est un critère qui rend l'outil plus accessible.

Nous avons inclus aussi dans ces critères, la **difficulté de prise en main** de l'outil que nous avons évalué lors des tests. Elle inclut les difficultés d'installation et d'utilisation. Sur cette base, les outils seront classés des plus intuitifs et simples jusqu'aux plus complexes à utiliser et à comprendre.

Il est à noter que nous sommes bien conscients de la subjectivité de l'évaluation pour la difficulté de prise en main, c'est un regard posé par une étudiante en informatique et qui pratique l'informatique au quotidien. Il faut dès lors en tenir compte dans la lecture des évaluations qui seront faites au chapitre suivant.

3.4 Contenu et formatage

A ce niveau nous parlerons des **données sources** ou *Datasets* qui sont utilisés pour créer des visualisations. Ces données peuvent provenir de différentes sources, comme des fichiers, des bases de données,...

On peut créer des visualisations sur ces outils et dispositifs en utilisant des **fichiers texte**. Parmi les fichiers que l'on peut utiliser, il y a ceux de formats ouverts, par exemple des fichiers CSV, SVG ou JSON.

Les formats ouverts favorisent l'interopérabilité c'est à dire qu'il n'est pas nécessaire d'avoir un logiciel spécial pour les utiliser ce qui rend ces outils plus accessibles.

La plupart des outils autorisent l'utilisation aussi de fichiers de format fermés ou propriétaires, par exemple les XLS.

Les fichiers ne sont pas les seules sources de données possibles, il y a aussi l'import **via une base de données** Oracle ou MySQL par exemple.

Il y a aussi **le lien d'intégration** ou une URL qui pointe vers une source de données en ligne pour permettre de créer des visualisations interactives par exemple.

Les données peuvent provenir **d'une source de données en ligne ou un serveur**, comme Google Drive ou Dropbox. Ces types de source de données permettent aussi d'avoir une information en temps réel ou interactive mais aussi de voir rapidement les changements dans les données.

Ces fonctionnalités comme l'utilisation de base de données, de serveurs, de liens d'intégration nécessitent un public averti dans ce domaine, c'est-à-dire qu'il faut que l'utilisateur ait un minimum de connaissance dans les moyens d'accéder et de récupérer des informations sur une base de données ou sur un serveur.

Pour construire une visualisation de données, les données peuvent provenir d'une **source unique** c'est-à-dire un fichier ou une URL par exemple.

Les données peuvent aussi provenir de **sources multiples**, par exemple plusieurs fichiers qui permettent de créer des visualisations complexes ou plus complètes combinant plusieurs sources

d'informations. Les outils qui permettent de combiner plusieurs sources de données donnent une accessibilité supplémentaire dans le sens où l'utilisateur peut utiliser directement ses différentes sources dans le même outil.

Lorsque l'outil autorise uniquement une source unique, l'utilisateur va devoir faire une action manuelle sur les sources pour les mettre dans la même source ou le même fichier avant de pouvoir les utiliser.

La question de la **réutilisation de données** pour créer des visualisations revient souvent dans la littérature. Il s'agit de savoir si on peut réutiliser des données sources importées par soi-même ou par quelqu'un d'autre dans l'outil pour construire une visualisation. La réutilisation facilite l'accessibilité d'un outil car elle permet à l'utilisateur de ne pas devoir aller rechercher le fichier source et de directement pouvoir l'utiliser.

3.5 Interprétation / création de sens

L'un des principaux objectifs de la visualisation des données est d'aider à interpréter les informations [9]. Dans une visualisation réussie, un mélange efficace de contenu, de contexte, de construction et de conception est attendu[121].

Par rapport à l'interprétation ou à la création de sens, nous nous intéresserons principalement aux fonctionnalités que permettent les outils testés.

Tout d'abord, quels sont les différents **types de visualisation** qu'il est possible d'effectuer au niveau de l'outil.

Il y a les **visualisations prédéfinies**, c'est-à-dire avec un modèle disponible ou des exemples à modifier. Les visualisations prédéfinies rendent l'outil accessible aux utilisateurs qui ont peu de connaissances dans la création de visuel et de la programmation, mais facilitent également la création de visualisations.

Il y a d'autre part des outils qui permettent à l'utilisateur de **programmer ou de personnaliser** ses propres visualisations. Par exemple, les différentes bibliothèques et API offrent une grande flexibilité dans la création de différents types de représentations visuelles. L'utilisation de telles bibliothèques nécessite généralement une certaine connaissance de la programmation de la part de l'utilisateur.

Les **fonctionnalités pour l'exploitation statistique des données** représentent une option en plus dans la démarche de création de visualisations car elles permettent de regrouper ou d'affiner, ou même de mettre en avant certaines valeurs.

Parmi les fonctionnalités statistiques on retrouve l'agrégation de données qui est un regroupement de valeurs correspondantes et l'accumulation de valeurs qui permet d'ajouter chaque valeur suivante de la série à la valeur précédente de la même série. On retrouve aussi des fonctions statistiques comme la corrélation, la régression, les clusters...

Les **Fonctionnalités pour la visualisation des données** représentent les différents types de visualisation possibles. Pour faire ce classement et cette description nous nous appuyons sur le livre [58] qui contient une explication des différentes représentations visuelle possibles. Les diverses représentations visuelles que peuvent permettre les outils testés, en voici quelques unes classés en différentes grandes catégories ci-dessous :

Graphiques univariés ou *Univariate Graphs*, Les graphiques univariés tracent la distribution des données à partir d'une seule variable. La variable peut être catégorielle (par exemple race, sexe) ou quantitative (par exemple âge, poids)[58]. Les illustrations graphiques les plus fréquemment utilisées pour les données univariées sont :

- Diagramme à bandes ou *Bar chart*
- Graphique à secteurs ou *Pie chart*
- Carte d'arbre ou *Tree map*
- Histogramme

- Tracé de densité du noyau ou *kernel density plot*
- Tracé de points ou *dot plot*

Graphiques bivariés ou *Bivariate Graphs* affichent la relation entre deux variables. Le type de graphique dépendra du niveau de mesure des variables (catégoriel ou quantitatif)[58]. Les illustrations graphiques les plus fréquemment utilisées pour les données bivariés sont :

- Graphique à barres empilées ou *Stacked bar chart*
- Graphique à barres groupées *Grouped bar chart*
- Graphique à barres segmentées *Segmented bar chart*
- Graphique en mosaïque ou *Mosaic plots* ; ces graphiques offrent une alternative aux graphiques à barres empilées pour afficher la relation entre les variables catégorielles.
- Nuage de points ou *Scatterplot* : il permet un affichage simple de deux variables quantitatives, chaque variable étant représentée sur un axe.
- Graphique linéaire ou *Line plot* : cet affichage est utilisé lorsque l'une des deux variables représente le temps. Il permet d'afficher l'évolution dans le temps d'une variable.
- Graphiques à barres utilisant des statistiques récapitulatives ou *bar charts using summary statistics*
- Tracés de densité de noyau groupés ou *grouped kernel density plots* : On peut comparer des groupes sur une variable numérique en superposant des graphiques de densité de noyau dans un seul graphe.
- Boîtes à moustaches côte à côte ou *side-by-side box plots* ;
- Parcelles de violon côte à côte ou *side-by-side violin plots* : Les parcelles de violon sont similaires aux parcelles de densité de noyau, mais sont en miroir et tournées de 90 °.
- Parcelles de ligne de crête ou *ridgeline plots or joyplot* : Un graphique de ligne de crête affiche la distribution d'une variable quantitative pour plusieurs groupes. Ils sont similaires aux graphiques de densité de noyau avec des facettes verticales, mais prennent moins de place.
- Graphiques de Cleveland ou *Cleveland plots* : ils permettent de comparer une statistique numérique pour un grand nombre de groupes.
- Nuages de points violon *Beeswarm plots or violin scatter plots* : ils sont similaires aux nuages de points gigue, et affichent également la densité des données à chaque point (d'une manière similaire à un tracé de violon).

Graphiques multivariés ou *Multivariate Graphs* affichent les relations entre trois variables ou plus. Il existe deux méthodes courantes pour gérer plusieurs variables : le regroupement et la mise en facettes[58]. Le regroupement permet de tracer plusieurs variables dans un seul graphique, en utilisant des caractéristiques visuelles telles que la couleur, la forme et la taille. Facette ou *faceting*, permet de partitionner un graphique en une matrice de panneaux. Chaque panneau montre une partie des données.

Cartes ou *Maps* Dessiner une carte en arrière plan est la première étape de toute analyse géospatiale.

- Cartes de densité de points *Dot density maps* : utilisent des points sur une carte pour explorer les relations spatiales.
- Cartes choroplèthes ou *Choropleth maps* : utilisent la couleur ou l'ombrage sur des zones prédéfinies pour indiquer les valeurs moyennes d'une variable numérique dans cette zone.

Graphiques dépendant du temps Un graphique peut être un outil puissant pour afficher les changements au fil du temps.

- Série chronologique *Time series* : est un ensemble de valeurs quantitatives obtenues à des moments successifs. Les intervalles entre les points temporels (par exemple, heures, jours, semaines, mois ou années) sont généralement égaux.
- Graphiques d'haltères ou *Dummbell charts* : sont utiles pour afficher le changement entre deux points dans le temps pour plusieurs groupes ou observations.
- Graphique de pente *Slope graphs* : peut être utilisé lorsqu'il y a plusieurs groupes et plusieurs points dans le temps.

- Graphique en aires simple ou *Area Charts* : est un graphique linéaire, avec un remplissage de la ligne à l'axe des x.

Modèles statistique ou *Statistical Models* : décrit la relation entre une ou plusieurs variables explicatives et une ou plusieurs variables de réponse.

- Graphiques de corrélation ou *Correlation plots* aident à visualiser les relations par paires entre un ensemble de variables quantitatives en affichant leurs corrélations à l'aide de la couleur ou de l'ombrage.
- Régression linéaire *Linear Regression* permet d'explorer la relation entre une variable de réponse quantitative et une variable explicative tandis que d'autres variables sont maintenues constantes.
- Régression logistique ou *Logistic regression* : peut être utilisée pour explorer la relation entre une variable de réponse binaire et une variable explicative tandis que d'autres variables sont maintenues constantes.
- Graphique de survie *Survival plots or Kaplan-Meier Curve* peut être utilisé pour illustrer la probabilité qu'un individu survive jusqu'au temps t inclus.
- Graphiques en mosaïque *Mosaic plots* : peuvent afficher la relation entre les variables catégorielles à l'aide de rectangles dont les zones représentent la proportion d'observations pour une combinaison donnée de niveaux. La couleur des carreaux peut également indiquer la relation de degré entre les variables.

Autre Graphiques

- Nuage de points 3D *3-D Scatterplot*
- Biplot ou *Biplots* : est un graphe spécialisé qui tente de représenter la relation entre les observations, entre les variables et entre les observations et les variables, dans un espace de faible dimension (généralement à deux).
- Graphique à bulles *Bubble charts* est essentiellement un nuage de points où la taille en points est proportionnelle aux valeurs d'une troisième variable quantitative.
- Diagramme de flux *Flow diagrams* représente un ensemble de relations dynamiques. On retrouve les diagrammes de Sankey ou *Sankey diagrams* et Les diagrammes alluviaux ou *Alluvial diagrams* qui sont un sous-ensemble des diagrammes de Sankey.
- Carte thermique ou *Heatmaps* : affiche un ensemble de données à l'aide de tuiles colorées pour chaque valeur de variable dans chaque observation.
- Graphique radar ou *Radar charts / spider chart / star chart* : affiche un ou plusieurs groupes ou observations sur trois variables quantitatives ou plus.
- Matrice de nuage de points ou *Scatterplot matrix* : est un ensemble de nuages de points organisés sous forme de grille. Il est similaire à un graphique de corrélation, mais au lieu d'afficher des corrélations, il affiche les données sous-jacentes.
- Graphique en cascade ou *Waterfall charts* illustre l'effet cumulatif d'une séquence de valeurs positives et négatives.
- Nuage de mots *Word clouds / tag cloud* : est essentiellement une infographie qui indique la fréquence des mots dans une collection de texte.
- Tableau, Un tableau synoptique est la représentation d'informations sous forme de colonnes et de lignes [110].

Le choix graphique est important dans la représentation de la donnée, Il permet de constater visuellement le phénomène que les données permettent de décrire. Pour bien choisir le graphique et la modalité statistique, des communautés d'entraide comme «From Data to Viz» propose une classification des types de graphiques basée sur le format des données d'entrée. Cette classification se présente sous la forme d'un arbre de décision menant à un ensemble de visualisations potentiellement appropriées pour représenter votre jeu de données.

Au plus l'outil propose des choix de représentations visuelles différents, au plus l'utilisateur pourra donner sens aux données. C'est donc un facteur qui renforce l'accessibilité et l'utilisabilité des données même si l'accessibilité à des outils statistiques de visualisation plus sophistiqués peut constituer une barrière pour certains profils d'utilisateurs non experts.

Parmi les critères, nous avons retenu aussi la possibilité d'**éditer des visuels**, d'y ajouter des annotations, des commentaires, mais aussi de pouvoir les enrichir avec des métadonnées.

Yuille and Macdonald [122] expliquent que les visualisations ont besoin d'une identité. Cela aide les gens à les reconnaître, à les localiser et à évaluer s'ils veulent les commenter. Ils ont nommé « décoration » le processus consistant à donner une telle identité dans le réseau à une visualisation[122]. Parmi ces processus de décoration, il y a la possibilité d'ajouter des annotations automatiques (fermées) ou des annotations manuelles (ouvertes).

La possibilité de **réutilisation des visualisations** c'est à dire de créer des visualisations dérivées à partir des visualisations existantes est une option importante car elle permet de ne pas recommencer entièrement un visuel mais de faire des adaptations, des modifications pour adapter ce visuel aux besoins de l'utilisateur et permet ainsi d'éviter à l'utilisateur de recommencer à zéro.

Les outils permettant la déconstruction, la modification et la création d'une nouvelle version d'une visualisation pourraient être utiles aux parties prenantes[47].

Un processus en plusieurs étapes est souvent nécessaire pour concevoir progressivement des visualisations efficaces. Scheidegger et al. [105] ont mis en évidence les avantages de garder trace de cet historique.

Pour permettre une réutilisation de visualisations existantes, l'outil doit enregistrer et encoder des métadonnées sur la façon dont la visualisation a été construite, y compris quelles sources ont été utilisées et comment elles ont été traitées[85].

Une fois la représentation visuelle créée, vient l'étape de l'**exportation des visuels** qui permet à l'utilisateur de pouvoir sauvegarder ou utiliser le visuel créé.

En fonction de la méthode et de l'outil utilisé il existe différents formats pour exporter les visuels créés.

Plus le format d'exportation est ouvert, plus l'utilisabilité de l'outil est grande, car pour rappel les formats ouverts ne nécessitent pas l'obtention de logiciels particuliers.

Les formats d'exportation des visuels de base sont les formats d'image, avec PNG ou JPEG. Certains outils permettent d'exporter les visualisations sous format SVG qui permet de faire en sorte que les visualisations puissent être modifiées avec des applications graphiques vectorielles pour des améliorations supplémentaires. L'export peut aussi se faire sous format PDF, c'est aussi le cas pour les rapports.

Certains outils permettent d'exporter au format JSON qui est un format de données textuelles dérivé de la notation des objets du langage JavaScript. Il permet de représenter de l'information structurée comme le permet XML par exemple.

Le partage et la publication représentent des éléments importants de l'accessibilité, en fonction de ce que l'outil permet, l'utilisateur peut se retrouver limité et ne pas pouvoir voir ses visuels créés hors de l'outil en question.

Concernant les **fonctionnalités de partage et de publication** des visuels créés, certains outils proposent la mise à disposition d'un lien unique(URL) pour le partage et qui permet aussi d'intégrer la visualisation directement sur un site ou un blog.

Le partage sur les réseaux sociaux est aussi une fonctionnalité proposée par certains outils, qui permettent de créer des visuels adaptés au format de ces réseaux sociaux et qui fournissent un bouton de partage en direct. Pour Graves and Hendler [47], il est nécessaire de simplifier l'accès aux différentes méthodes de partage que les parties prenantes utilisent souvent pour partager des informations.

Plus il y a de méthodes de partage et de publication, plus l'outil sera accessible car il aura une large variété de diffusion possible des résultats.

La **création de rapports et de tableau de bord** constitue aussi une des fonctionnalités proposées par la plupart des outils de visualisation. Les rapports et les tableaux de bord permettent

de transmettre et de partager l'information d'une certaine manière.

Un **tableau de bord** ou *Dashboard* est un écran unique comportant les indicateurs les plus importants et qui permet d'obtenir une vision globale de l'activité. Les **rapports** sont des documents qui représentent les données d'une manière structurées à destination d'une audience spécifique.

3.6 Soutien et collaboration

Dans cette catégorie nous aborderons les différentes ressources individuelles ou communautaires de soutien et d'aide disponibles pour l'utilisation ou la création de visuels. Parmi les **ressources individuelles**, certains outils mettent à disposition des informations sous la forme de vidéos, de tutoriels, de foires aux questions ou , qui ne se limite pas seulement à un simple mode d'emploi mais fournit des informations concernant l'utilisation avec des astuces et des exemples détaillés. Ces informations sont en général disponibles sur les plateformes d'outils, certains proposant également l'envoi de bulletins d'information ou *newsletters*.

Les **ressources collectives** représentent les communautés en ligne, les forums de discussion qui favorisent l'entraide en permettant d'obtenir plus d'informations personnalisées ou des réponses à des questions directes. Plus il y a de ressources individuelles ou collectives disponibles pour un outil, plus l'utilisateur pourra en cas de besoin trouver la solution à son problème, ce qui rend l'outil plus accessible et utilisable pour les utilisateurs.

La **collaboration** est une part importante de l'utilisabilité des outils. Doit-on travailler seul sur ces outils ou existe-il une possibilité de faire du travail collaboratif - en mode synchrone ou asynchrone ?

Certains outils prennent en charge la **collaboration synchrone** où des personnes réparties à différents endroits voient et interagissent simultanément avec les mêmes visualisations.

La **collaboration asynchrone** quant à elle, ne permet pas un travail en simultané, et se doit de garder une trace de l'historique d'exploration afin que les autres parties prenantes puissent modifier ces schémas pour créer une version modifiée de la visualisation.

Au niveau collaboration, un outil accessible serait un outil qui permette de faire de la collaboration avec d'autres utilisateurs ou de travailler en équipe.

3.7 Gouvernance

Le **prix** de l'outil utilisé doit aussi être pris en compte dans les critères d'accessibilité car tous les utilisateurs ne peuvent pas se permettre d'utiliser des produits payants. La gratuité de l'outil est un facteur d'accessibilité car il permet que l'outil puisse être utilisé par tous les utilisateurs.

Parmi les outils testés, il existe des outils gratuits, payants, des outils avec une partie gratuite et une partie payante, et des outils Open Source avec le code source mis à disposition. Dans les outils payants, certains proposent un essai gratuit limité dans le temps pour tester les fonctionnalités.

Concernant le prix, un outil accessible est un outil où l'utilisateur ne doit pas payer pour accéder à l'outil, nous mettons donc dans cette catégorie les outils gratuits et les outils open source.

La **protection des données** est importante aussi dans le cadre d'une politique de gestion des données, qui définit si oui ou non les données peuvent être partagées ou doivent être protégées. Nous nous poserons la question de savoir si les outils testés permettent de mettre en place une manière de protéger les données et leurs résultats. Certains outils offrent une protection des données et des résultats dans le sens où tout est privé ou stocké en local. Pour d'autres outils, les résultats sont directement publiés en ligne et donc accessibles à tous les utilisateurs. Les outils les plus accessibles seront les outils qui permettent à l'utilisateur de choisir de mettre une visibilité publique ou non sur ses visualisations.

La **langue** de l'outil et du support disponible est aussi un critère en lien avec l'accessibilité. Les outils les plus accessibles seront les outils dont l'éditeur et le support proposé sont disponibles dans plusieurs langues avec au moins l'anglais ; le français et le néerlandais étant des plus.

Chapitre 4

Analyse et comparaison des outils

Introduction

Dans ce chapitre, nous allons évaluer chacun des outils de visualisation identifiés. Pour ce faire, nous avons travaillé sur chaque outil individuellement, nous les avons explorés, testés et évalués avec les critères sélectionnés précédemment.

Cette analyse repose sur des tests que nous avons réalisés nous-même, et lorsque le test n'a pas pu être réalisé nous le signalons dans le texte. Tout n'était pas testable, dans le sens où il y a des fonctionnalités payantes ou d'autres qui étaient difficiles d'accès. Dans ces cas-là nous nous sommes repliés sur les informations mises à disposition sur le site des éditeurs mais aussi dans les FAQ. Pour les tests des bibliothèques JavaScript et du langage R, n'ayant pas beaucoup de connaissance dans ces langages, nous nous sommes basés sur la documentation.

Nous avons réalisé des tests sur la base de jeux de données disponibles pour Bruxelles¹ ou pour la région wallonne². Pour chaque outil, un ou plusieurs fichiers ont été importés pour tester les fonctionnalités disponibles dans chaque outil.

Il est important de noter que cette évaluation se fait sur la base de notre prise en main de ces différents outils et des tests que nous avons pu réaliser sur chacun d'entre eux. La lecture de cette analyse doit donc se faire à la lumière des compétences informatiques de l'évaluatrice. Il convient dès lors de souligner que l'analyse et la comparaison des outils de visualisation proposés ci-dessous sont exploratoires, elles pourraient dans la suite de ce travail être enrichies, complétées, consolidées par des évaluations faites par des utilisateurs n'ayant pas le même background que la rédactrice de ce mémoire.

Microsoft Excel

Internet, ordinateurs et logiciels C'est une application de bureau à télécharger même si actuellement la version Office 365 permet un accès via navigateur. Il est nécessaire de créer un compte pour pouvoir l'utiliser.

Compétences informatiques / logiciels Il n'est pas nécessaire d'avoir des connaissances techniques significatives en programmation informatique pour utiliser cet outil. En revanche, il est nécessaire d'avoir des compétences de base en statistiques descriptives.

La prise en main de l'outil est assez simple lorsqu'on a l'habitude d'utiliser les outils de Microsoft, Excel reste assez intuitif pour faire des visualisations simples.

1. <https://opendata.bruxelles.be/pages/home/>

2. <https://www.odwb.be/pages/home/>

Contenu et formatage Les données sources peuvent être importées à partir d'un fichier texte ou CSV, XML, JSON, à partir du Web (URL), à partir d'un tableau ou à partir d'une base de données.

Excel permet de fusionner ou de concaténer différentes sources en une seule, mais la création des graphiques se fait avec une source unique. Les données sources qui ont été importées peuvent être modifiées directement dans l'outil pour ensuite créer une visualisation.

Interprétation / création de sens Les graphiques sont prédéfinis dans Excel et en fonction des données, l'outil peut même suggérer le type de graphique approprié. Ils nécessitent généralement la connaissance de formules Excel pour les réaliser. Excel propose toute une liste de formules statistiques prédéfinies que l'utilisateur peut appliquer à ces données. Les différents types de graphiques disponibles dans Excel sont : histogramme, courbe, secteurs, barres, aires, nuage de points, carte choroplèthe, boursier, surface, radar, compartimentage, cascade, entonnoir, graphique combiné. Il existe d'autres modèles ou compléments qui sont téléchargeables sur le site de l'éditeur.

L'utilisateur peut ajouter un complément d'informations à ses graphiques notamment des commentaires ou des étiquettes. Les visualisations créées avec Excel peuvent être réutilisées et modifiées dans Excel ou tout autre outil qui prend en charge les classeurs Excel. Le partage et la publication se fait au format classeur Excel ou PDF/XPS. Il y a aussi la possibilité de publier sur Microsoft Power BI.

Soutien et collaboration Il existe beaucoup de ressources fournies par l'éditeur, dans le logiciel avec une rubrique d'aide et FAQ, mais aussi des vidéos. Il est possible de faire de la collaboration en temps réel dans la version en ligne de Microsoft Excel (disponible uniquement avec un abonnement).

Gouvernance La suite Microsoft 365 qui contient Microsoft Excel, est une suite logicielle payante, à partir de 69,00 €/an pour les particuliers, et propose aussi un essai gratuit de 30 jours pour essayer le logiciel. Il existe aussi d'autres plans tarifaires pour les familles, l'éducation et les entreprises.

Les données importées dans Excel sont privées, on peut choisir lors de la publication de les partager avec d'autres sous forme d'un lien vers le document sauvegardé dans OneDrive.

Excel est disponible en anglais, français, néerlandais mais aussi en plusieurs autres langues.

Exemple Voici un exemple de visualisation (Voir figure : 4.1) créé à partir d'Excel et du fichier de la population bruxelloise trouvé sur le site Open Data de Bruxelles[91]. Cette visualisation a été exportée en PNG.

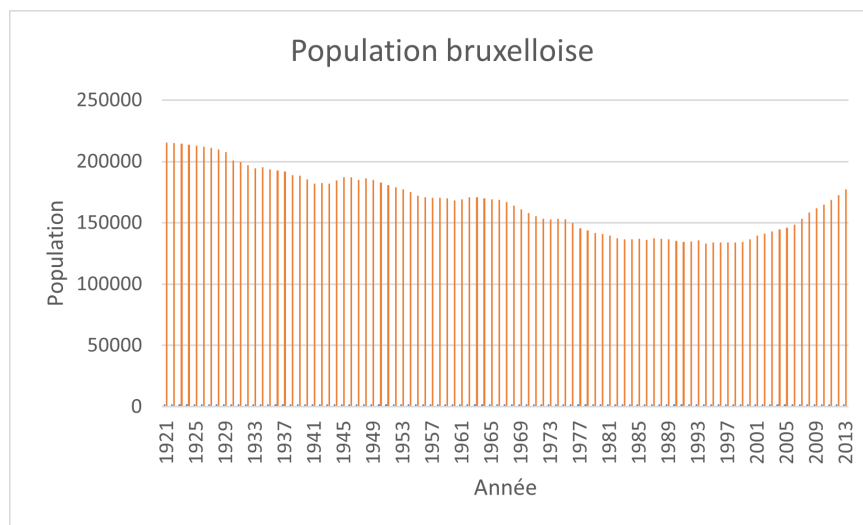


FIGURE 4.1 – Visualisation créer à partir de Excel

Infogram

Internet, Ordinateurs et logiciels Infogram [51] est un outil Web qui permet aux utilisateurs de créer des visualisations basées sur les données. Il est nécessaire de créer un compte pour pouvoir l'utiliser.

Compétences informatiques/logiciels Une compétence en statistiques est nécessaire pour utiliser cet outil de visualisation. C'est un outil accessible dans le sens où il est disponible via un navigateur et il est assez intuitif. Nous avons trouvé la prise en main de l'outil assez intuitive dans le sens où l'outil est dans le navigateur, il n'y a rien à installer. L'interface graphique de l'outil indique clairement les options possibles.

Contenu et formatage Les données sources peuvent être importées d'une source locale mais aussi de Google Drive, Dropbox ou de bases de données (MySQL, PostgreSQL, Amazon Redshift, Oracle, Microsoft SQL). Les fichiers textes pris en charge sont des fichiers XLS/XLSX, CSV, Google Sheets et JSON.

Lors de la création de visuels, l'utilisateur a le choix de n'avoir qu'une source de données ou d'en choisir plusieurs pour créer des visualisations plus complexes. Il peut aussi modifier son fichier source dans la plateforme.

Infogram ne permet pas aux utilisateurs de réutiliser les ensembles de données qu'ils ont téléchargés. Les utilisateurs sont obligés de télécharger à nouveau les données lorsqu'ils souhaitent créer une nouvelle visualisation.

Interprétation/création de sens Cet outil permet de faire des visualisations prédéfinies, c'est à dire qu'on dispose d'un modèle pour chaque visuel possible.

Concernant les fonctionnalités pour l'exploitation statistique des données, il en existe des basiques avec la version gratuite à savoir l'agrégation et l'accumulation.

Pour les fonctionnalités de visualisation des données, Infogram permet de créer différents types de graphiques notamment les graphiques en barres, histogramme et courbes, radial, secteurs, aires, nuage de points, bulles, illustration, tableau, nuage de mots, carte géographique, diagramme de faits et chiffres, minuteur, pyramide, entonnoir et financier. Il permet aussi de faire des tableaux de bord, des infographies, des mises en page Web et mobiles réactives, pour créer un contenu dynamique qui s'ajuste en fonction des dimensions d'affichage des différents écrans, avec des cartes, des tableaux et des rapports interactifs.

Dans ses versions payantes, il y a la possibilité d'ajouter des commentaires et de faire des discussions. Il est aussi possible de faire des annotations manuelles et automatiques des visuels avec l'ajout de textes, d'images, de valeurs, pour aider le lecteur à mieux comprendre la visualisation. La réutilisation des visualisations est possible dans le cadre d'un travail d'équipe par exemple. Par contre dans la version gratuite, les seules visualisations que l'on peut réutiliser sont les siennes ou les exemples fournis par la plateforme.

Pour les utilisateurs de la version payante, il existe un historique des versions qui permet d'accéder et de restaurer des itérations antérieures des projets.

Dans la version payante, les visuels peuvent être exportés, sous forme d'image (PNG, JPEG) ou Vector Graphics (SVG), PDF, Gif animé ou vidéo MP4. Avec la version d'essai il n'est pas possible d'exporter le résultat des visuels créés.

Infogram fournit une URL vers le visuel créé mais aussi des liens pour faire une intégration sur un site Web ou blog. Il facilite aussi le partage sur les principaux réseaux sociaux (LinkedIn, Facebook, Instagram, Twitter, Pinterest et YouTube) avec un bouton de partage en direct et en permettant de créer du contenu adapté à ces réseaux sociaux, par exemple avec la taille de l'image.

Soutien et collaboration Concernant le support, la plateforme dispose d'une page d'assistance avec des tutoriels étape par étape, des exemples, des modèles. Elle offre aussi à ses utilisateurs un blog avec des astuces, des témoignages et des exemples d'utilisation. Elle propose aussi la possibilité de faire un chat ou d'envoyer un mail à leur équipe d'assistance pour avoir une réponse personnalisée.

Infogram, dans sa version payante, offre la possibilité de créer des équipes pour travailler ensemble en temps réel. Il existe des versions spécialement dédiées aux entreprises pour faciliter le travail en équipe.

Gouvernance Il existe différentes options de prix pour Infogram, il y a une version gratuite et des versions payantes (à partir de 25\$ par mois) avec plus d'options, notamment l'export des visuels, la collaboration en temps réel ou une gestion de la vie privée.

Dans la version d'essai, le profil de l'utilisateur est public par défaut. C'est à dire que n'importe quel utilisateur avec le lien vers les visualisations ou le profil y a accès. Dans la version payante, il y a la possibilité de choisir si on veut rendre le résultat ainsi que le profil publics ou privés.

Le site de l'éditeur ainsi que sa documentation est proposé en anglais et en partie en français.

Exemple Voici un exemple de visualisation créé avec cet outil. Il s'agit d'une visualisation de la population bruxelloise et des déclarations de mariages, naissances et divorces par année faite avec un fichier disponible sur le site Open Data de Bruxelles. Dans sa version gratuite, Infogram ne permet pas de télécharger le résultat, voici l'URL vers la visualisation créée :

<https://infogram.com/line-chart-1h0r6rpzegw1w2e?live>

TIBCO Spotfire

Internet, ordinateurs et logiciels TIBCO Spotfire est un logiciel à télécharger localement sur son ordinateur, il dispose également d'une version en ligne. Il est nécessaire de créer un compte pour pouvoir l'utiliser. On peut utiliser l'application téléchargée en ligne ou hors ligne avec options limitées.

Compétences informatiques/logiciels Une compétence en statistiques est nécessaire pour utiliser ce logiciel de visualisation. Pour pouvoir utiliser les fonctionnalités avancées, il est nécessaire de maîtriser au moins un langage de programmation comme le Python ou R. Nous avons trouvé TIBCO Spotfire un peu plus complexe à utiliser. En effet, une fois installé, il faut un petit temps pour comprendre comment naviguer dans l'outil et trouver où se situent toutes les options disponibles. Cette difficulté concerne aussi la partie pour utiliser une bibliothèque pour personnaliser les visuels.

Contenu et formatage Les sources de données que propose le logiciel sont multiples, il y a la possibilité d'ajouter des sources qui proviennent de fichier (XLS, CSV, TXT, ...) mais aussi de se connecter à une source de données comme une base de données (Oracle , MySQL, ...) ou un serveur externe (Amazon, Microsoft SharePoint Online...).

Interprétation/création de sens Il existe une multitude de visualisations prédéfinies utilisables dans le logiciel. Le logiciel intègre aussi une partie développement Python et TERR permettant à l'utilisateur averti de personnaliser ses créations. TERR ou TIBCO Enterprise Runtime for R permet de développer avec le langage open source R.

Il existe plusieurs fonctionnalités pour l'exploitation statistique des données, notamment pour l'analyse des données, comme la possibilité de faire des modélisations, des partitionnements de données, des relations entre les données, des champs calculés ...

Avec Tibco Spotfire, l'utilisateur peut utiliser des visualisations prédéfinies comme : table, tableau croisé, histogramme, diagramme en cascade, courbe de tendance, graphique combiné, camembert,

nuage de points et nuage de points 3D, cartographie, TreeMap, carte de chaleur, diagramme KPI ou indicateurs de performance clés, graphique de coordonnées, box plot ou diagramme en boîte...

Des annotations comme des étiquettes ou des conversations peuvent être ajoutées aux visuels créés.

La réutilisation de visualisation déjà créée est possible avec ce logiciel. Les visuels sont sauvegardés en un fichier complet qui stocke la source et le résultat visuel, ce qui permet de pouvoir l'ouvrir et adapter la source des données et les visuels.

Concernant l'export de visuels créés, il n'y a pas beaucoup de formats disponibles, il y a notamment le format image (PNG), PDF, Power Point et le format propriétaire du logiciel. Il est aussi possible d'exporter la source de données au format XLS ou CSV.

Concernant les fonctionnalités de publication et de partage, il y a -selon le site de l'éditeur- la possibilité de faire un partage sur les réseaux sociaux notamment sur Twitter. La bibliothèque de Spotfire fournit une URL pour accéder directement à la visualisation créée mais ne fournit pas un lien d'intégration direct vers la visualisation.

L'outil offre aussi la possibilité de créer des rapports en PDF et des tableaux de bord.

Soutien et collaboration Il existe plusieurs ressources individuelles proposées sur le site de l'éditeur sous forme de documentation, vidéos pas à pas, FAQ, Wiki. Il propose aussi des ressources collectives avec Tibco Community qui est un espace d'échange et d'entraide, un blog et un centre de support et aussi des formations et des certifications pour pouvoir bien utiliser l'outil.

Le site de l'éditeur indique qu'il y a la possibilité de collaborer en temps réel sur le logiciel via la création et la gestion d'une équipe.

Gouvernance Il existe plusieurs plans tarifaires pour cet outil, il y a une version d'essai gratuite avec des options limitées comme un nombre d'exports gratuits et des versions payantes avec plus d'options comme l'accès à l'API, l'intégration des résultats, la collaboration. En fonction du rôle de l'utilisateur et ce qu'il veut pouvoir faire dans le logiciel, les prix s'échelonnent entre 25 et 125 \$ par mois.

Dans la version d'essai, le profil de l'utilisateur est public par défaut. C'est à dire que n'importe quel utilisateur avec le lien vers les visualisations ou le profil a accès aux fichiers sources. Dans la version payante, il y a la possibilité de choisir si on veut rendre le résultat public ou privé.

Le logiciel ainsi que le site de l'éditeur et la documentation sont disponibles en plusieurs langues notamment l'anglais et le français.

Exemple Voici un exemple de visualisation créée avec Tibco Spotfire (Voir figure : 4.2). Il s'agit d'une carte des bulles à verre de la région wallonne faite avec un fichier disponible sur le site OpenData de la région wallonne[17].

RAW Graphs

Internet, ordinateurs et logiciels RAW Graphs est un outil en ligne pour créer des visualisations vectorielles personnalisées. C'est aussi une application Open source créé à partir de la bibliothèque D3.js. Étant donné que D3.js est testé plus tard dans ce document, l'analyse ici se concentrera principalement sur l'éditeur en ligne de RAW Graphs. Il n'est pas nécessaire de créer un compte pour utiliser cet outil en ligne.

Compétences informatiques/logiciels Dans sa version disponible sur le Web, cet outil est accessible aux personnes qui ont un minimum de compétences en statistiques, pour savoir quel graphique utiliser ou comment mapper les colonnes avec les axes par exemple.

Pour aller plus loin et créer ses propres visualisations ou sa propre instance de l'application, la connaissance la bibliothèque D3.js et du langage de programmation JavaScript est nécessaire.

Bulles à verre Région Wallonne

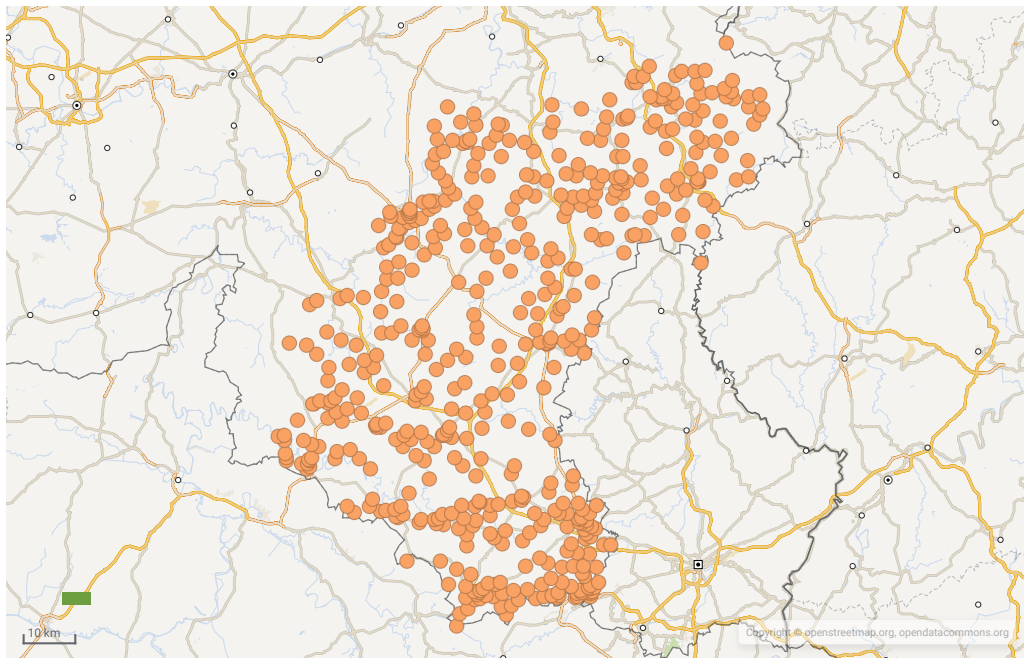


FIGURE 4.2 – Visualisation Spotfire : Bulles à verre Region Wallone

La prise en main de cet outil n'est pas évidente, lors de nos tests il y a eu plusieurs soucis par rapport à l'import de fichier sans message d'erreur clair (fichier trop long ou trop de colonnes). A noter que l'étape de mapping des axes et des colonnes des graphiques se fait manuellement.

Contenu et formatage Les formats de fichier acceptés pour l'import de données sont les fichiers TSV, CSV, DSV et JSON. Le fichier peut être importé directement ou seulement copié dans l'éditeur disponible.

On peut aussi saisir une adresse Web (URL) pointant vers les données par exemple, un fichier Dropbox, via une API, ...

Dans sa version disponible sur le navigateur, une seule source de données est possible pour la création de visualisation.

Il n'y a pas de possibilité de réutilisation de sources étant donné que rien n'est stocké dans l'outil Web. On peut juste modifier les données sources qu'on a téléchargées.

Interprétation/création de sens Pour la création de visuel, il existe plusieurs modèles disponibles et dans le cas où ce ne serait pas suffisant, il y a la possibilité de créer ses propres modèles ou visuels à partir de la bibliothèque Open source.

Concernant les fonctionnalités pour l'exploitation statistique des données, il en existe quelques-unes dans l'éditeur en ligne comme la somme, la moyenne, l'empilement, ... La bibliothèque Open source quand elle offre plus de fonctionnalités statistiques à programmer.

Pour la visualisation des données, il existe plusieurs modèles de graphiques comme Treemap, graphique à bulles, emballage de cercle, dendrogramme, dendrogramme circulaire, diagramme alluvial (diagramme de type Fineo), nuage de points, coordonnées parallèles... Il y a une étape manuelle de mapping pour mettre en correspondance le contenu de la source de données et le graphique.

Il n'y a pas de possibilité de créer des tableaux de bord ou des rapports, ni de créer un contenu interactif avec la version en ligne.

Il y a très peu de fonctionnalités d'édition disponibles dans l'application Web. Il y a la possibilité d'ajouter des annotations automatiques, par rapport à la légende par exemple.

Etant donné que l'application Web ne stocke rien, il n'y a pas de fonctionnalités de réutilisation des visualisations ni même une sauvegarde de l'historique des modifications et des métadonnées.

L'outil permet d'exporter des visualisations sous forme d'images vectorielles (SVG) ou PNG. Si on a les connaissances nécessaires, il est possible d'exporter la visualisation au format vectoriel (SVG) et de la modifier avec n'importe quel éditeur de graphiques vectoriels.

Dans sa version Web il n'existe pas d'option de partage direct. On peut juste télécharger localement les visuels créés.

Contrairement aux autres outils de visualisation disponibles sur le Web, cette application ne fournit pas une URL ou un lien direct vers la visualisation qui pourrait faciliter l'intégration.

Soutien et collaboration Rawgraph propose des ressources individuelles d'aide comme des tutoriels, FAQ et une *newsletter*. Il fournit également de la documentation avec un guide du développeur et une référence API. Rawgraph met à disposition des ressources collectives avec la possibilité d'envoyer à leur équipe de support un message ou de poser des questions sur le forum Google Groupes.

L'application Web Rawgraph ne propose pas d'option de collaboration entre utilisateurs.

Gouvernance C'est une application Web Open source gratuite.

Les données insérées dans l'application Web ne sont traitées que par le navigateur, il n'y a pas de stockage côté serveur.

L'outil ne propose que l'anglais pour l'application Web et pour la documentation.

Exemple Voici un exemple de visualisation créée avec cet outil (Voir figure : 4.3). Il s'agit d'une visualisation de la population bruxelloise de 2000 à 2014 faite avec un fichier disponible sur le site Open Data de Bruxelles[91] mais modifié pour avoir moins de lignes.

Tableau Public

Internet, ordinateurs et logiciels Tableau Desktop est une application de bureau à télécharger, elle permet de créer des visualisations. Tableau Public est un service gratuit accessible via un navigateur qui permet à l'utilisateur de publier des visualisations de données sur le Web. Il est nécessaire de créer un compte pour pouvoir utiliser l'application ou la plateforme en ligne.

Compétences informatiques/logiciels Pour créer des visualisations avec Tableau il n'est pas nécessaire d'avoir des compétences particulières en informatique telle que la connaissance d'un langage de programmation. Il est toutefois nécessaire d'avoir des compétences en statistique et informatique plus approfondie pour utiliser toutes les options proposées par l'outil mais aussi dans le cas de sources de données complexes.

Concernant la prise en main, Tableau n'est pas un outil évident à comprendre et à utiliser. Il faut un peu de temps pour comprendre comment bien utiliser cet outil, trouver comment créer une visualisation, les options qui existent, comprendre les liens entre la partie en ligne et l'application de bureau.

Contenu et formatage Il existe une multitude de sources de données possible. Tout d'abord il y a les sources de type fichier (Excel, Texte, JSON, Microsoft Access, PDF), des fichiers de données spatiales (GeoJSON, KML...), des fichiers statistiques (SAV ,RDA, RDATA...) et classeur

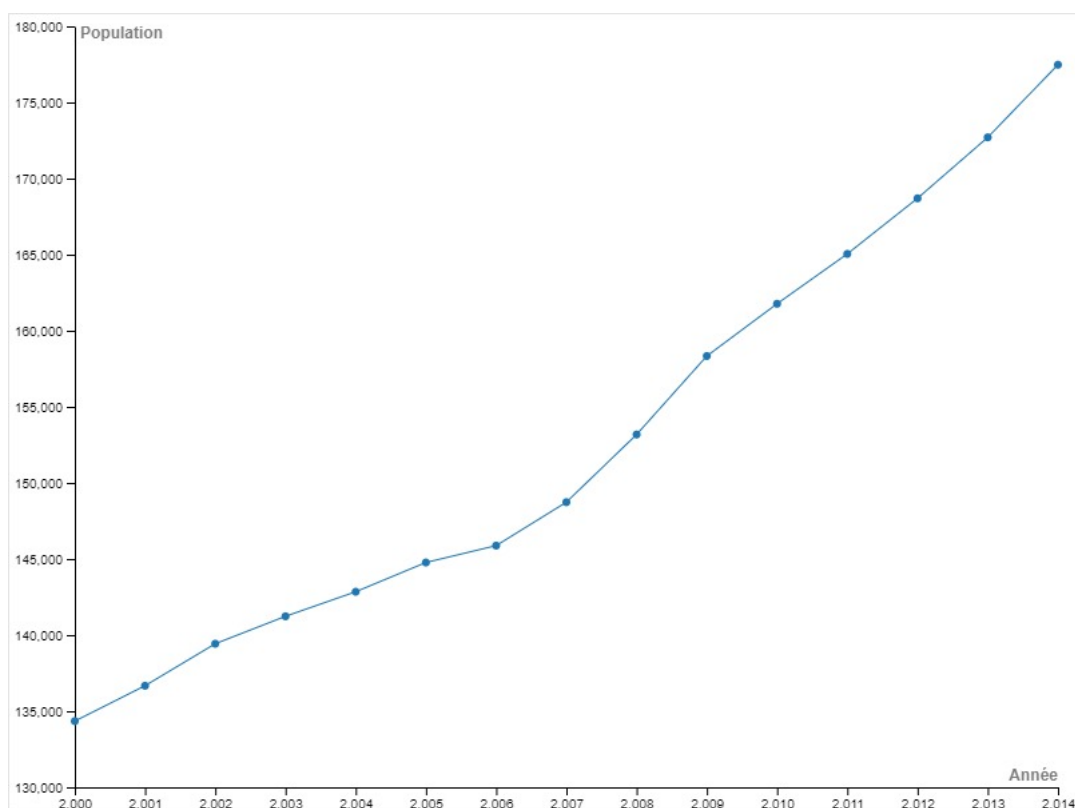


FIGURE 4.3 – Visualisation Rawgraph : Évolution population bruxelloise de 2000 à 2014

Tableau. Il y a aussi la possibilité d’avoir des sources de données en ligne sur un serveur : Microsoft SQL Server ou d’une base de données MySQL, Oracle ...

Plusieurs sources de données sont possibles pour créer des visualisations plus complexes.

La réutilisation des données sources est permise dans l’application Tableau. Les données sources sont stockées au format classeur Tableau et il suffit de les récupérer sur Public Tableau pour pouvoir les réutiliser.

Interprétation/création de sens Les visualisations sont prédéfinies dans Tableau, on peut choisir parmi une liste de visuels. Tableau propose également des fonctionnalités puissantes pour l’exploration des données, telles que le tri, le filtrage, le regroupement et le pivotement des données.

Concernant les fonctionnalités pour la visualisation des données, Tableau permet de créer différents types de graphiques notamment les graphiques en barres, histogramme, radial, secteurs, nuage de points, bulles, tableau, nuage de mots, carte de chaleur et géographique, diagramme de Gant, ...

Il est possible d’ajouter des annotations aux visuels, avec des annotations automatiques et manuelles avec l’ajout d’étiquettes, de textes, de légendes... Il est possible de réutiliser des visualisations, tout comme pour les données sources, les visualisations sont stockées au format classeur Tableau. Avec Tableau Public ou Gallery on a accès à des visualisations Tableau d’autres utilisateurs, on peut les télécharger ou les modifier en ligne.

Il existe différentes manières d’exporter et d’utiliser les visuels créés avec Tableau. La visualisation peut être exportée au format image (PNG), Power Point ou PDF.

Il est possible d’exporter au format classeur la visualisation et/ou la source de données pour l’enregistrer en local.

Dans la plateforme de bureau seule la publication sur Tableau Public est permis. A partir de Tableau

Public on peut avoir un lien pour l'intégration ou une URL. Il existe une option de partage sur les réseaux sociaux dans Tableau Online.

L'outil permet aussi de faire des tableaux de bord interactifs, mais aussi de créer des histoires. Dans Tableau, une histoire est une séquence de visualisations qui fonctionnent ensemble pour transmettre des informations.

Soutien et collaboration Comme support proposé il y a les vidéos didactiques, une FAQ, un blog. Tableau met aussi en place une communauté en ligne d'utilisateurs avec un forum des utilisateurs de Tableau. Il propose aussi des formations mais aussi des formations guidées en ligne (e-learning) aux utilisateurs qui le souhaitent. Ces formations ne sont pas toutes gratuites.

Il est possible de travailler en équipe avec la version payante de Tableau.

Gouvernance Il existe différents plans tarifaires pour chaque type d'utilisateur. Tableau propose un essai gratuit de quelques jours pour tester l'application. Concernant les versions payantes (à partir de 70 \$) , il en existe pour les utilisateurs individuels, pour les équipes et entreprises et une avec l'option analytique embarquée. Il n'est pas possible d'acheter juste Tableau Desktop. Pour faire des visualisations, il est donc nécessaire d'acheter toute la suite. Les étudiants et les enseignants du supérieur disposent d'une licence gratuite pendant 1 an.

Dès qu'un classeur est publié sur Tableau Public, la visualisation est accessible à tous sur Internet à moins que l'utilisateur dispose d'un compte professionnel, dans ce cas il peut gérer la confidentialité de ses visualisations.

Le site internet ainsi que les différents logiciels de la suite Tableau sont disponibles entièrement en plusieurs langues notamment en anglais et en français. Le support est majoritairement disponible en anglais.

Exemple Voici un exemple de tableau de bord assez simple créé avec Tableau Public(voir figure 4.4). Il a été fait sur base du fichier Open Data des agences bancaires de Bruxelles disponible sur le site Open Data de Bruxelles[35].

Exemple de lien vers ce tableau de bord contenant une visualisation interactive :

https://public.tableau.com/views/Classeur2_16117462405340/Tableaubord1?:language=fr&:display_count=y&publish=yes&:origin=viz_share_link

Microsoft Power BI

Internet, ordinateurs et logiciels Power BI se compose de plusieurs éléments qui fonctionnent tous ensemble : une application de bureau Windows appelée Power BI Desktop, un service SaaS (Software as a Service) en ligne appelé service Power BI et des applications mobiles Power BI pour les appareils Windows, iOS et Android. L'outil inclut une API et la possibilité d'ajouter des bibliothèques pour développer ses propres modèles. Il est nécessaire de créer un compte utilisateur pour pouvoir l'installer et l'utiliser.

Compétences informatiques/logiciels Pour utiliser cet outil, pour créer une visualisation simple, il n'est pas nécessaire d'avoir des connaissances techniques significatives en programmation informatique. C'est un outil assez puissant de business Intelligence qui contient beaucoup d'options, il est nécessaire d'avoir des connaissances en statistiques. Pour aller plus loin et créer ses propres modèles de visualisation, il est nécessaire de connaître au moins un langage de programmation.

Lorsqu'on a déjà travaillé avec la suite Microsoft Office, on a l'avantage de connaître la charte graphique et cela offre des facilités dans la prise en main de l'outil.

Agence

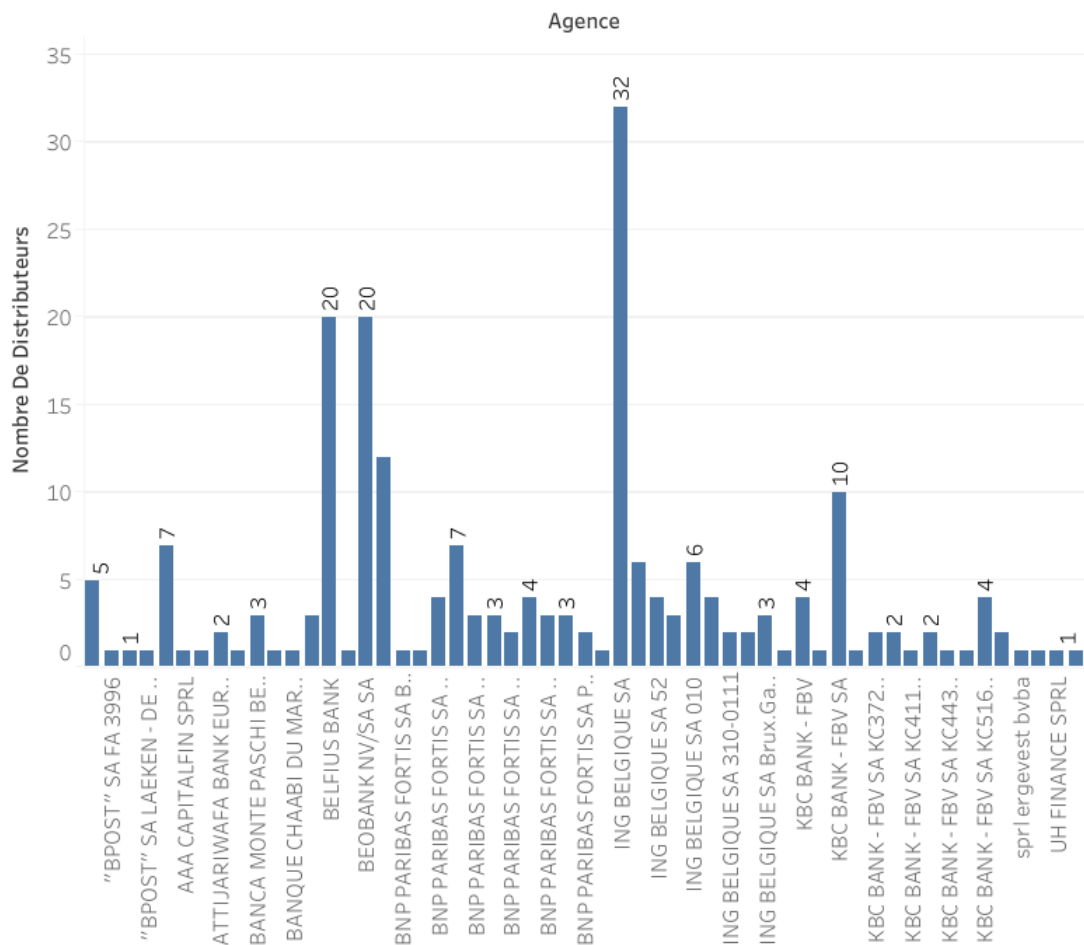


FIGURE 4.4 – Visualisation Tableau : agences bancaires de Bruxelles

Contenu et formatage Les données sources peuvent être importées à partir d'un fichier texte ou CSV, XML, JSON, à partir du Web (URL), à partir d'un tableau ou à partir d'une base de données.

Il est possible d'avoir plusieurs sources de données dans l'outil. Un écran permet de les ajuster ou changer les liens entre les sources. Les données sources peuvent être modifiées et réutilisées dans l'outil et dans la plateforme en ligne.

Interprétation/création de sens Les graphiques disponibles sont prédéfinis dans l'outil. Il existe des modèles ou compléments qui sont téléchargeables sur le site de l'éditeur ou mis à disposition par la communauté en ligne. Pour développer ses propres visuels Power BI, il est possible d'installer des bibliothèques supplémentaires, telle qu'une bibliothèque D3.js, React, R, Python...

Pour les fonctionnalités pour l'exploitation statistique des données, il existe différentes options notamment l'utilisation des formules comme pour Excel.

Les différents types de graphiques disponibles dans Power BI sont : histogramme, courbe, secteurs, barres, aires, nuage de points, carte, boursier, surface, radar, compartimentage, cascade, entonnoir, graphique combiné.

L'utilisateur peut ajouter un complément d'informations à ses graphiques notamment des com-

mentaires, des étiquettes...

Les visualisations créées avec Power Bi peuvent être réutilisées et modifiées dans Power Bi en ligne ou dans l'application de bureau. L'export des visuels se fait au format PowerPoint, PDF ou au format Power Bi. Les données sources peuvent aussi être exportées en Excel .

Il y a la possibilité de publier le résultat en ligne sur Power BI pour les utilisateurs qui ont une licence pro, pour ensuite le rendre public ou le partager grâce à un lien d'accès ou un QR code. Power bi fournit aussi un lien pour l'intégration des résultats.

Le partage en direct sur les réseaux sociaux n'est pas prévu.

Il est possible de créer des rapports et des tableaux de bord interactifs dans l'outil.

Soutien et Collaboration Il existe de multiples ressources fournis par l'éditeur, dans le logiciel avec une rubrique d'aide et FAQ, de vidéo, et aussi sur la plateforme en ligne. Il existe aussi un espace communauté avec un forum, un blog, des webinaires, des événements...

Grâce à son application en ligne, il est possible de créer des espaces de travail pour la collaboration ou le travail d'équipe.

Gouvernance Le téléchargement est gratuit mais pour activer certaines options telles que la collaboration ou le partage, il est nécessaire d'avoir une licence valide qui, elle, est payante.

Par défaut les données sources ainsi que les résultats sont privés. Avec une licence, on peut gérer les droits d'accès des autres utilisateurs pour qu'ils puissent modifier les visuels ou juste les regarder.

L'outil est disponible en plusieurs langues dont l'anglais et le français. Le support en ligne et les ressources collectives sont majoritairement disponibles en anglais.

Exemple Voici un exemple de rapport PDF créé avec cet outil (Voir figure : 4.5). Il s'agit d'une visualisation de la population bruxelloise et des déclarations de mariages,naissances et divorces par année faite avec un fichier disponible sur le site Open Data de Bruxelles.

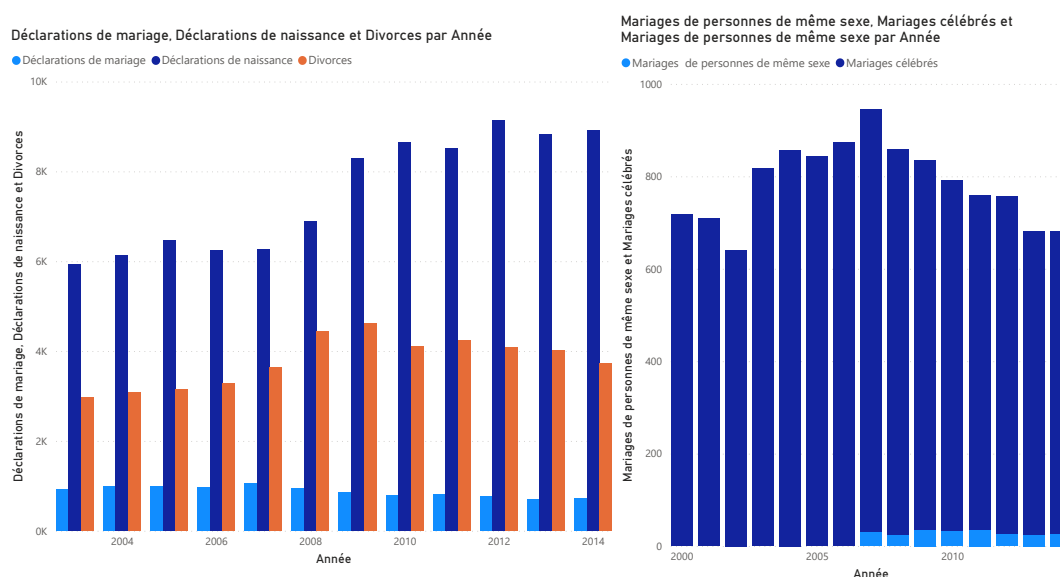


FIGURE 4.5 – Visualisation Power Bi : démographie bruxelles

Qlik Sense

Internet, ordinateurs et logiciels C'est un outil de visualisation disponible en ligne. Il existe aussi une application de bureau mais elle n'est pas disponible dans la version d'essai gratuite. Il est nécessaire de créer un compte pour pouvoir l'utiliser.

Compétences informatiques/logiciels C'est un outil de *Business Intelligence* qui permet de créer des visualisations mais aussi de faire des analyses sur les données. Une compétence en statistiques est nécessaire pour utiliser cet outil de visualisation.

C'est un outil accessible dans le sens où il est disponible via un navigateur et il est assez intuitif pour créer une visualisation simple. Il fournit même des suggestions en fonction des données, mais pour aller plus loin il est nécessaire de chercher un peu ou de chercher dans la documentation.

Contenu et formatage La source de données peut provenir de fichiers CSV ou XLS, de bases de données, de Services Cloud. Il y a aussi la possibilité d'ajouter des données manuellement sous la forme d'un tableau.

Pour créer des visualisations, on peut combiner plusieurs sources de données.

Les données sources peuvent être modifiées manuellement ou on peut y appliquer des formules. La réutilisation de données sources est aussi possible avec les données qu'on a importées ou mises à disposition par l'équipe.

Interprétation/création de sens Il existe dans l'outil toute une liste de visualisations prédéfinies. Grâce aux extensions et à l'API, les utilisateurs qui ont des connaissances en programmation peuvent créer des visualisations personnalisées.

L'outil propose des fonctionnalités pour l'exploitation statistique des données, notamment les fonctions de base comme la moyenne. Il offre également la possibilité de créer des formules à appliquer sur les données sources.

Cet outil propose les visuels ci-dessous : Diagramme à barres, diagramme à puces combiné, conteneur, histogramme, KPI, diagramme en courbes, carte, diagramme Mekko, diagramme à secteurs, tableau croisé dynamique, nuage de points, tableau, carte en arbre, diagramme en cascade,... L'outil permet d'ajouter des annotations aux visuels, notamment des bookmarks, des labels, des tags et autres commentaires.

L'outil permet aussi de dupliquer des visualisations. Il permet aussi de réutiliser des visualisations déjà existantes ou partagées par d'autres membres de l'équipe.

Le résultat de la visualisation peut être téléchargé uniquement sous format PDF.

L'outil propose un lien direct vers la visualisation à partager ou à intégrer dans une application. Une fois le résultat publié il n'est pas possible de le modifier, l'utilisateur peut le dupliquer pour le modifier ou le remettre en accès privé.

Soutien et collaboration Parmi les ressources individuelles proposées, il y a des documents et des vidéos. Il existe une ligne de support dédiée aux entreprises. Il y a une communauté en ligne, un blog, des questions et discussions, des formations et certifications, et des Webinaires.

Concernant la collaboration, l'outil propose de créer des équipes qui peuvent ensuite collaborer sur une visualisation. Dans la console de management de l'outil, l'utilisateur peut gérer les autres utilisateurs, les espaces, les événements qui ont eu lieu c'est à dire les modifications faites par les utilisateurs.

Gouvernance Il existe différents plans tarifaires (à partir de 30 \$ par mois) avec des options et des fonctionnalités différentes pour utiliser Qlik Sense et les autres modules de Qlik. C'est un outil en ligne payant qui propose un essai gratuit de 30 jours. Il y a aussi un programme académique qui permet aux étudiants et aux professeurs d'utiliser l'outil gratuitement.

L'outil offre une protection immédiate des données et la possibilité pour l'utilisateur de les publier pour qu'elle soit accessible à tout le monde, ou pas.

L'outil est disponible en plusieurs langues dont l'anglais et le français. Le support en ligne et les ressources collectives sont majoritairement disponibles en anglais.

Exemple Voici un exemple (Voir figure : 4.6) de visualisation créé avec cet outil. Il s'agit d'un rapport PDF avec plusieurs graphiques de la population de Namur faite avec un fichier disponible sur le site Open Data de Wallonie[71].

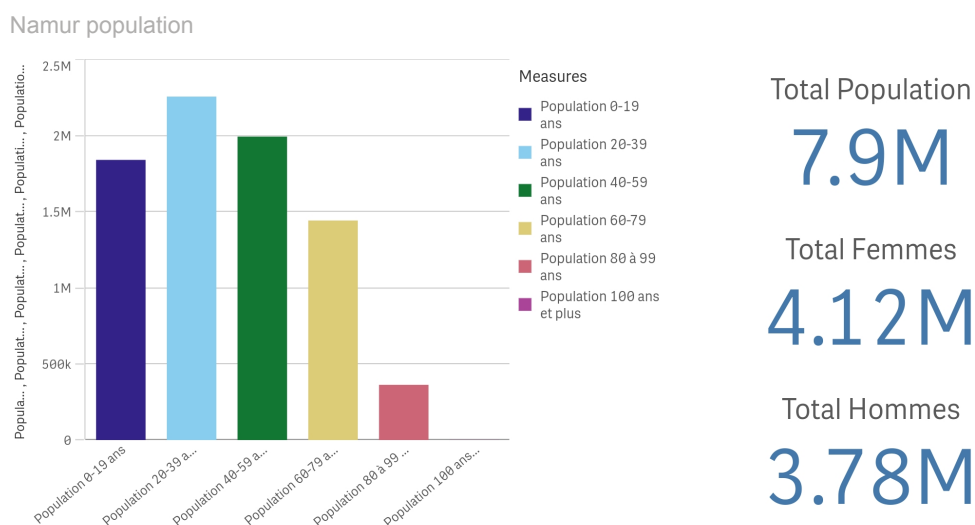


FIGURE 4.6 – Visualisation avec Qlik Sense : Population

R Langage de programmation

Internet, ordinateurs et logiciels Pour utiliser R, il faut le télécharger sur CRAN. Le téléchargement comprend les différentes bibliothèques et un éditeur exécutable Windows GUI.

Compétences informatiques/logiciels R est un langage de programmation accompagné de plusieurs bibliothèques. Il est nécessaire d'avoir des connaissances dans ce langage et en informatique pour pouvoir l'utiliser.

Lorsqu'on n'a pas de connaissance dans le langage R, il est nécessaire de prendre le temps d'apprendre ce langage pour pouvoir faire des visualisations. Étant donné que c'est un langage que nous ne connaissons pas, nous nous sommes donc basés sur la documentation[58] pour faire l'analyse de cet outil.

Contenu et formatage R peut importer des données de presque toutes les sources, y compris des fichiers texte, des feuilles de calcul Excel, des packages statistiques et des systèmes de gestion de bases de données.

Les données sources peuvent provenir de plusieurs sources différentes.

Il est possible de modifier ou d'appliquer des calculs sur les données sources importées.

Interprétation/création de sens Il existe des bibliothèques R de visualisation qui fournissent un ensemble de graphiques prédéfinis, mais il est toujours possible de développer ses propres visualisations.

De base R est un langage de programmation et un logiciel destiné aux statistiques, il existe une multitude de fonctionnalités statistiques (modélisation linéaire et non linéaire, tests statistiques classiques, analyse de séries chronologiques, classification, clustering,...) et des bibliothèques pour l'exploitation statistique des données.

Dans les bibliothèques de visualisation R, il existe toutes sortes de graphiques, notamment : Les graphiques univariés : histogramme, diagramme à secteurs, carte arborescente, histogramme, diagramme de densité du noyau ou diagramme à points. Les graphiques bivariés avec graphiques à barres empilés, groupés ou segmentés, graphique en mosaïque... Et pour finir, il y a les graphiques dépendant du temps, cartes géographiques, nuage de points tridimensionnel, tracé biographiques, diagramme à bulles, carte de chaleur, cartes radar, matrice de nuage de points, graphique en cascade, nuages de mots.

Il est possible d'ajouter des annotations aux visualisations sous la forme d'étiquette, de texte ou de formes.

Si les visualisations ont été enregistrées au format SVG ou PDF, l'utilisateur pourra les modifier ou les réutiliser dans l'application ou dans un éditeur de format vectoriel.

Les visualisations peuvent être exportées au format : PDF, JPEG, TIFF, PNG, SVG et WMF.

Il existe dans les packages de R des fonctionnalités qui permettent de poster le résultat d'une visualisation sur le Web.

Il n'y a pas de fonctionnalité de partage au niveau de l'éditeur par défaut de R, et donc pas de partage en direct sur les réseaux sociaux.

Il n'y a pas de fonctionnalité à proprement parler pour des tableaux de bord, mais il est possible d'en simuler un, et de créer des visualisations interactives. Pour la création des rapports, il faut faire preuve de créativité et ajouter les éléments nécessaires aux rapports.

Soutien et collaboration Sur le site de R, on peut y trouver plusieurs sortes de ressources individuelles comme une rubrique d'aide sur le site de l'éditeur, une FAQ, des tutoriels, des manuels et des livres.

Il existe plusieurs ressources collectives avec les communautés en ligne liées à R. Il y a StackOverflow, R-bloggers[98] et aussi des forums.

Il n'est pas possible de faire de la collaboration avec l'éditeur par défaut. Mais il existe d'autres logiciels R qui permettent de créer des équipes et de collaborer, comme RStudio[102] par exemple.

Gouvernance R est un outil disponible gratuitement au téléchargement.

R s'exécute en local, les données sources utilisées pour travailler avec R restent stockées en local et ne sont accessibles qu'aux utilisateurs de la machine.

Le site de R[99] ainsi que l'ensemble de la documentation existante est disponible en anglais. L'éditeur par défaut (R GUI) est quant à lui disponible aussi en Français.

P5.js

Internet, ordinateurs et logiciels P5.js est une bibliothèque JavaScript qui inclut un éditeur en ligne³ et une bibliothèque Open source.

3. <https://editor.p5js.org/>

Compétences informatiques/logiciels Il est possible d'utiliser l'éditeur en ligne p5.js sans avoir de connaissances en programmation, par exemple en utilisant les exemples fournis. Mais pour faire des visualisations personnalisées ou plus complexes, il vaut mieux avoir des connaissances en un langage de programmation comme JavaScript.

Lorsqu'on n'a pas de connaissance en JavaScript ou en informatique cela peut prendre un peu de temps pour bien comprendre le langage. Mais l'interface graphique avec l'éditeur permet en direct de voir l'évolution de la construction de notre visualisation ou directement d'avoir des messages d'erreur.

Contenu et formatage Il existe des fonctionnalités pour charger des fichiers de données au format txt, CSV, JSON et XML.

Pour créer une visualisation on peut utiliser plusieurs sources de données, il est nécessaire de gérer le mapping manuellement. Il est possible de modifier les données sources qu'on a importées dans l'application.

Interprétation/création de sens On peut faire des visualisations prédéfinies avec p5, mais aussi entièrement les personnaliser grâce aux différentes bibliothèques.

Les bibliothèques p5.js ainsi que les bibliothèques JavaScript fournissent un ensemble de primitives pour l'exploitation statistique des données.

Pour la visualisation des données, avec p5.js et en combinaison avec d'autres bibliothèques JavaScript, on peut faire de nombreux visuels comme : des graphiques, des animations, la 2D et la 3D, ainsi que l'intégration de sons sur les graphiques.

Les visuels créés peuvent être édités en y ajoutant des annotations, des étiquettes et du texte.

Il y a la possibilité d'envoyer des liens vers le dossier de travail, comprenant le code source et les fichiers sources d'une visualisation, pour permettre à d'autres de la réutiliser.

Les bibliothèques p5 fournissent des fonctions qui permettent de sauver le résultat sous la forme d'image (JPG,PNG,SVG) et JSON. Les tableaux peuvent être sauvés en HTML, TXT, CSV ou TSV.

L'éditeur p5 fournit un lien pour partager et embarquer le résultat de la visualisation.

Soutien et collaboration p5 fournit de multiples supports pour avoir plus d'informations, comme la documentation, des livres, des tutoriels et des exemples. Parmi les ressources collectives, il y a la communauté en ligne, des forums de discussions, un Wiki.

L'éditeur en ligne ne permet pas de faire de la collaboration ou de créer des équipes.

Gouvernance p5.js est une bibliothèque gratuite et Open source.

La bibliothèque p5.js ainsi que son éditeur gardent les fichiers importés en local sur la machine de l'utilisateur. Mais il y a des possibilités de partager ses fichiers ou résultats avec d'autres.

Le site de l'éditeur propose plusieurs langues dont l'anglais, c'est également le cas pour l'éditeur et la plupart de la documentation.

Exemple Voici un exemple de la fenêtre de création d'une visualisation à partir de l'éditeur p5.js (Voir figure : 4.7). Étant donné nos connaissances limitées en JavaScript, cette visualisation a été faite à partir d'un exemple de code disponible sur gitHub⁴ qui montre le nombre d'occurrence des lettres de l'alphabet dans un fichier.

Le lien vers la visualisation : <https://editor.p5js.org/agcamara/present/q1ZH8dITn>

4. <http://sepans.github.io/p5d3/bar-chart>

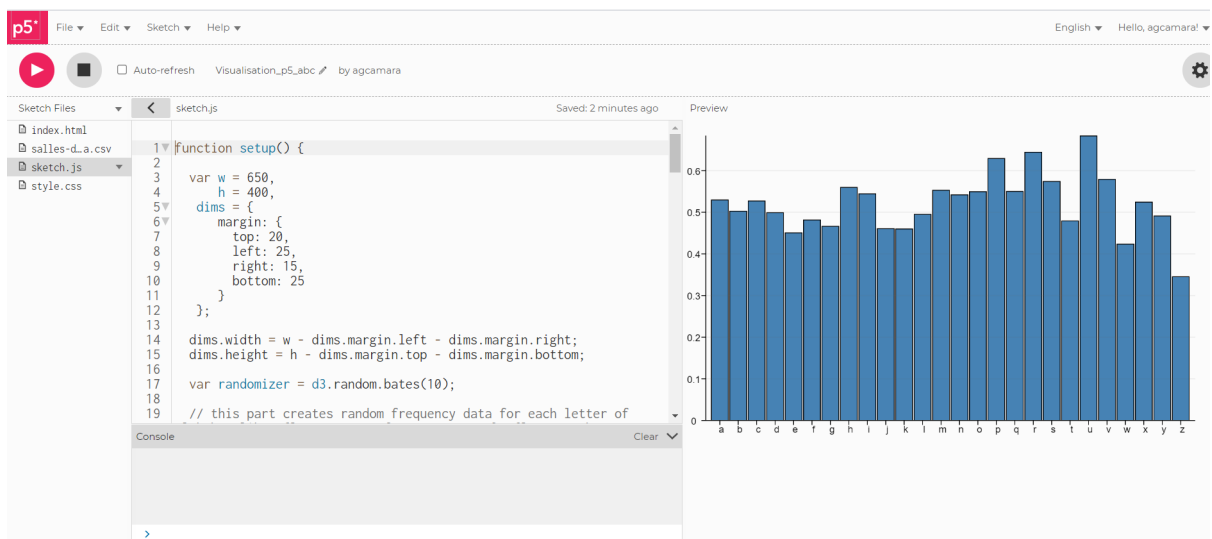


FIGURE 4.7 – Visualisation P5 : occurrence lettre alphabet

D3.js

Internet, ordinateurs et logiciels D3.js est une bibliothèque JavaScript permettant de visualiser des données à l'aide de standards Web. D3.js propose aussi d'utiliser Observable, un éditeur en ligne[72].

Compétences informatiques/logiciels Pour pouvoir utiliser D3, il est nécessaire d'avoir un minimum de connaissances en développement avec le langage de programmation JavaScript.

Comme pour les autres bibliothèques Javascript, avoir des connaissances dans ce langage de programmation est une nécessité pour pouvoir utiliser l'outil entièrement. Il est assez difficile à prendre en main lorsqu'on n'a pas l'habitude des langages de programmation. Nous nous baserons sur la documentation [14, 72] pour compléter cette analyse.

Contenu et formatage Il existe plusieurs sources de données possibles comme : intégrer des données dans le bloc-notes sous forme de code, des fichiers (CSV, SQLite...), un accès via API - interrogée à partir d'un serveur distant, et un accès via un client de base de données. Avec des compétences en JavaScript ou grâce aux ressources et exemples disponibles, on peut utiliser n'importe quelle source de données et même plusieurs sources si nécessaire.

Les données sources peuvent aussi être modifiées en local par l'utilisateur mais pas dans l'éditeur par défaut.

Interprétation/création de sens Étant donné que c'est une bibliothèque JavaScript, il y a des visualisations prédéfinies qui nous sont fournies en exemple, mais il y a aussi la possibilité de créer ses propres visuels.

Il existe des bibliothèques JavaScript puissantes pour l'exploitation statistiques des données.

La galerie D3 propose plusieurs modèles de graphiques parmi lesquels : TreeMap, regroupement d'arêtes hiérarchiques, diagramme de Sankey, contours de densité, graphe dirigé par la force, projections cartographiques. Il y a également la possibilité de créer des visualisations interactives.

D3 fournit de nombreuses fonctions et fabriques de fonctions réutilisables intégrées, telles que des primitives graphiques pour les graphiques en aires, en courbes et en secteurs[14].

Selon l'éditeur, il est possible d'ajouter des annotations aux visuels tels que les étiquettes, légendes, axes, titres, guides... qui permet de lier plus d'informations à la visualisation.

L'export des visualisations peut se faire grâce aux différentes bibliothèques dans des formats image (PNG, SVG...), PDF ... Avec la plateforme Observable, l'utilisateur a accès à un bouton pour partager ses publications sur les réseaux sociaux ou avec une URL. Dans sa version en local, il est aussi possible d'utiliser des fonctionnalités pour partager ses visualisations. La plateforme en ligne Observable proposée sur le site en lui-même ne permet pas de créer de rapports ou de tableaux de bord interactifs comme dans les outils en ligne vus précédemment.

Avec des connaissances d'au moins un langage de programmation, il est possible de créer des tableaux de bord ainsi que des rapports.

Soutien et collaboration Il existe différentes ressources communautaires pour aider les utilisateurs à communiquer entre eux et avoir du support, comme sur StackOverflow avec le tag [d3.js], Google Group, Slack et Gitter. Il existe aussi de la documentation, mais aussi des pas à pas et des exemples. Au niveau du GitHub de D3, on y mentionne d'essayer la plateforme Observable[72] pour la visualisation, elle contient notamment plusieurs ressources individuelles et collectives.

Pour la collaboration, D3 propose d'utiliser la plateforme Observable qui est un environnement permettant le travail collaboratif en temps réel[72].

Gouvernance D3 est une bibliothèque Open source gratuite.

Les données sont protégées, avec l'éditeur en ligne ou celui par défaut, car tout est stocké en local et on peut choisir de créer des règles de visibilité pour tout.

L'outil ainsi que les ressources mis à disposition sont disponibles totalement en anglais mais il existe aussi des traductions non officielles.

Exemple Voici un exemple de visualisation basique créée à partir de l'éditeur Observable de D3 (Voir figure : 4.8). Elle a été faite avec un fichier de la population de Bruxelles sur le site Open Data de Bruxelles[91]. Cette visualisation est aussi disponible à cette adresse : <https://observablehq.com/@agcamara/visualisation>

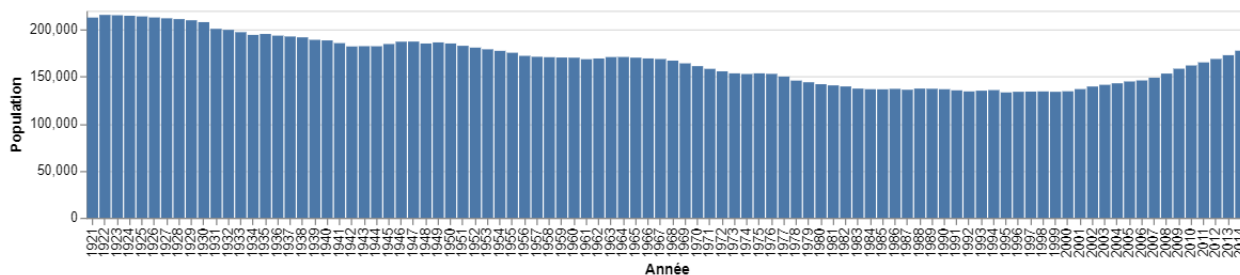


FIGURE 4.8 – Visualisation D3 avec éditeur Observable : population de Bruxelles

Comparaison

Dans cette partie, nous allons comparer les différents outils analysés précédemment sur base des critères définis dans notre méthodologie. Nous les avons regroupé dans des tableaux et nous allons les comparer afin d'évaluer leur accessibilité pour les utilisateurs.

Internet, ordinateurs et logiciels

Pour comparer cette partie, nous allons utiliser le système de comparaisons ci-dessous :

- En vert : Disponible
- En jaune : Disponible moyennant utilisation d'un complément ou d'un éditeur supplémentaire

— En rouge : Pas possible/Pas disponible

Outils	Excel	Infogram	Spotfire	Rawgraphs	Tableau	Power Bi	Qlick	R language	p5.js	D3.js
Application de bureau	Green	Red	Green	Red	Green	Green	Green	Red	Green	Red
Accès via navigateur	Green	Green	Green	Green	Green	Green	Green	Yellow	Green	Yellow
Bibliothèques de visualisation et API	Red	Red	Red	Red	Red	Yellow	Red	Red	Red	Red
Langage de programmation	Red	Red	Red	Red	Red	Red	Red	Green	Red	Red

TABLE 4.1 – Infrastructure d'accès

Dans cette catégorie nous considérons que les outils les plus accessibles sont les outils qui sont accessibles directement via navigateur. Les outils qui proposent aussi plusieurs méthodes d'accès permettent plus de flexibilité.

Compétences informatiques/logiciels

Pour tous les outils il est nécessaire d'avoir des compétences statistiques qui portent à la fois sur les possibilités de traitement statistiques en fonction des données disponibles et sur le sens des différents calculs statistiques disponibles et leurs mises en graphiques. Les compétences informatiques requises varient, quant à elles, d'un outil à l'autre.

Pour comparer cette partie, nous allons utiliser le système de comparaison ci dessous :

- En vert : Ne nécessite pas de compétences/Pas de difficulté pour la prise en main
- En jaune : Un minimum de connaissance est nécessaire/Quelques difficultés pour la prise en main
- En rouge : Compétences requises/Prise en main difficile

Outils	Excel	Infogram	Spotfire	Rawgraphs	Tableau	Power Bi	Qlick	R language	p5.js	D3.js
compétences en statistiques	Red	Red	Red	Red	Red	Red	Red	Red	Red	Red
compétences en informatique	Green	Green	Yellow	Yellow	Green	Yellow	Yellow	Red	Red	Red
Difficulté de prise en main	Green	Green	Yellow	Yellow	Green	Yellow	Yellow	Red	Red	Red

TABLE 4.2 – Compétences requises

Les outils qui nécessitent le moins de compétences sont les plus accessibles, on a par exemple Excel et Infogram qui combine aussi une facilité de prise en main .

Contenu et formatage

Pour les fonctionnalités de contenu et formatage, nous les avons regroupées dans le tableau ci-dessous. En fonction de ce que l'utilisateur veut utiliser comme données, certains outils seront mieux adaptés que d'autres.

Pour comparer la partie contenu et formatage, nous allons utiliser le système de comparaison ci-dessous :

- En vert : Disponible
- En jaune : Disponible avec difficulté/programmation nécessaire
- En rouge : Pas disponible

Name of Tools	Excel	Infogram	Spotfire	Rawgraphs	Tableau	Power Bi	Qlick	R language	p5.js	D3.js
Fichier Texte Format -Ouvert et Fermé	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green
Base de donnée en ligne/ serveur / URL	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green
Fichier JSON	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green
Collection de données : source unique	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green
Collection de données : plusieurs sources	Green	Red	Green	Red	Green	Green	Green	Green	Green	Green
Réutilisation/modification des sources	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green

TABLE 4.3 – Données source

Les données sources des outils peuvent provenir de différents types de source. En vert on peut voir que la plupart des outils testés permettent d'avoir des sources variées. 2 outils parmi les outils testés, Infogram et Rawgraph, n'acceptent pas de sources multiples de données et ne permettent

pas de faire des réutilisations ce qui peut être un frein pour faire certaines actions. Rawgraph quant à lui ne se connecte pas à la base de données dans sa version éditeur en ligne, mais il y a moyen de le programmer.

Interprétation / création de sens

Pour les options d'interprétation et création de sens, nous les avons regroupées dans le tableau ci-dessous. En fonction de ce que l'utilisateur veut faire, certains outils seront mieux adaptés que d'autres.

Pour faire la comparaison des différents outils nous allons utiliser le système de couleur ci-dessous :

- En vert : Disponible
- En jaune : Disponible avec installation /programmation nécessaire
- En rouge : Pas possible/Pas disponible

Name of Tools	Excel	Infogram	Spotfire	Rawgraphs	Tableau	Power Bi	Qlick	R language	p5.js	D3.js
Types de visualisation										
pré-définie										
programmable										
Fonctionnalités pour l'exploitation statistique des données										
de base : somme, moyenne,...										
corrélation,agregation,...										
régression ,cluster,...										
Fonctionnalités pour la visualisation des données										
Graphiques univariés										
Graphiques bivariés										
Graphiques multivariés										
Géospatiale, Carte,...										
Graphiques dépendant du temps										
Tableau										
Autres Graphiques										
Fonction d'édition des visuels créés										
Annotations automatiques (fermées)										
Annotations manuelles (ouvertes)										
Réutilisation des visualisations										
Exportation des visuels										
Sous forme d'image (PNG, JPG)										
Sous forme d'image Vector (SVG)										
Data Model (JSON)										
Print Vector / PDF										
fonctionnalités de partage et publication des visualisations										
Lien pour intégration/Embed SVG Code										
URL										
direct sur les réseaux sociaux.										
Création de rapports										
Création de dashboard										

TABLE 4.4 – Interprétation/création de sens

Les outils les plus accessibles sont les outils qui proposent des visualisations prédéfinies, ce qui permet aux utilisateurs de pouvoir créer plus facilement des visualisations. On remarque que tous les outils de visualisations testés proposent des visualisations prédéfinies directement dans l'outil ou via des modèles ou exemples.

Les outils testés proposent à chaque fois au moins quelques fonctionnalités pour l'exploitation statistique des données. Les outils les plus accessibles seront ceux pour lesquels il existe plusieurs fonctionnalités permettant à l'utilisateur d'avoir plus de choix.

Tous les outils testés proposent des fonctionnalités de visualisation des données notamment les graphiques univariés, bivariés... Excel propose des visualisations simples mais pour les autres types de graphique il est nécessaire de télécharger un complément. Pour le langage R, et les bibliothèques p5 et D3, il est nécessaire de les programmer.

Les fonctionnalités d'édition des visuels créés, comme l'ajout des annotations fermées, sont disponibles dans tous les outils testés, directement ou avec un peu de programmation.

La possibilité de réutilisation des visuels créés permet de rendre l'outil plus accessible car cela permet aux utilisateurs de ne pas devoir les recréer de zéro.

Il existe plusieurs fonctionnalités d'exportation des visuels, les plus répandues dans les outils sont les formats images ou PDF. Plus l'outil propose de formats différents, plus on aura de possibilités d'avoir le format qui correspond à l'utilisation qu'on veut en faire.

Les fonctionnalités de partage et de publication des visualisations ne sont pas disponibles directement dans tous les outils testés, ce qui les rend moins accessibles.

Pour les outils liés à des bibliothèques Javascript ou R, p5.js et D3.js, toutes ces fonctionnalités sont disponibles mais il faut les programmer pour pouvoir les utiliser.

Soutien et collaboration

On peut voir que tous les outils analysés prévoient comme support des ressources individuelles et collectives. Pour la collaboration, certains outils permettent une collaboration synchrone alors que d'autres pas.

Nous allons utiliser le système de comparaison ci-dessous :

- En vert : Disponible
- En rouge : Pas disponible

Name of Tools	Excel	Infogram	Spotfire	Rawgraphs	Tableau	Power Bi	Qlick	R language	p5.js	D3.js
Support proposé										
Ressource individuelle										
Ressource collective										
Collaboration										
Travail collaboratif - synchrone										
Travail collaboratif - asynchrone										

TABLE 4.5 – Soutien et Collaboration

Tous les outils testés disposent de ressources individuelles et collectives qui permettent aux utilisateurs d'avoir les ressources nécessaires pour utiliser l'outil et d'avoir de l'aide quand ils en ont besoin.

Concernant la collaboration synchrone, elle n'est pas disponible dans tous les outils testés. C'est le cas des bibliothèques et des langages de programmation qui ne disposent pas d'une interface permettant ce genre de collaboration.

Gouvernance

Pour la gouvernance, nous allons utiliser le système de comparaisons ci-dessous en nous centrant sur deux critères importants : le prix et la protection des données. La langue d'utilisation et de documentation de l'outil n'est pas prise en compte ici, tous les outils évalués offrant au minimum une interface de travail et de documentation en anglais. Nous les indiquons directement dans le tableau.

- En vert : les outils entièrement gratuits / Protection immédiate des données
- En jaune : les outils possédant une version d'essai gratuite / Protection des données soumises à des conditions
- En rouge : les outils payants / Pas de protection des données

Name of Tools	Excel	Infogram	Spotfire	Rawgraphs	Tableau	Power Bi	Qlick	R language	p5.js	D3.js
Prix										
protection des données										
Langue de l'outils	FR, en, nl	fr,en	fr,en	en	fr,en	fr,en	fr,en	en	en	en

TABLE 4.6 – Gouvernance

Concernant le prix, les outils les plus accessibles sont les outils gratuits ou Open source.

La plupart des outils testés offrent une protection des données immédiate. Les outils les plus accessibles sont ceux qui permettent aux utilisateurs de choisir la visibilité de leurs visualisations, ou qui offrent une protection directe.

La langue de l'outil constitue aussi un point important, ici nous considérons les outils qui sont en anglais (EN) et français (FR) et/ou néerlandais (NL) comme les plus accessibles.

Conclusion comparaison

Outils	Excel	Infogr.am	Spotfire	Rawgraphs	Tableau	Power Bi	Qlick	R language	p5.js	D3.js
Application de bureau	Green	Red	Green	Red	Green	Green	Green	Red	Green	Red
Accès via navigateur	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green
Bibliothèques de visualisation et API	Red	Red	Green	Green	Red	Yellow	Green	Red	Green	Green
Langage de programmation	Red	Red	Red	Red	Red	Red	Red	Red	Red	Red
Compétences										
compétences en statistiques	Red	Red	Red	Red	Red	Red	Red	Red	Red	Red
compétences en informatique	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green
Difficulté de prise en main	Green	Yellow	Yellow	Yellow	Yellow	Yellow	Yellow	Yellow	Yellow	Yellow
Donnée source										
Fichier Texte Format -Ouvert et Fermé	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green
Base de donnée en ligne/ serveur / URL	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green
Fichier JSON	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green
Collection de données : source unique	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green
Collection de données : plusieurs sources	Green	Red	Green	Red	Green	Green	Green	Green	Green	Green
Réutilisation/modification des sources	Green	Red	Green	Red	Green	Green	Green	Green	Green	Green
Types de visualisation										
pré-définie	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green
programmable	Red	Red	Green	Yellow	Red	Yellow	Green	Yellow	Yellow	Yellow
Fonctionnalités pour l'exploitation statistique des données										
de base : somme, moyenne,...	Green	Red	Green	Green	Green	Green	Green	Green	Yellow	Yellow
corrélation,agregation,...	Green	Red	Green	Green	Green	Green	Green	Green	Yellow	Yellow
régression ,cluster,...	Green	Red	Green	Red	Green	Green	Green	Green	Yellow	Yellow
Fonctionnalités pour la visualisation des données										
Graphiques univariés	Green	Green	Green	Green	Green	Green	Green	Green	Yellow	Yellow
Graphiques bivariés	Green	Green	Green	Green	Green	Green	Green	Green	Yellow	Yellow
Graphiques multivariés	Green	Green	Green	Green	Green	Green	Green	Green	Yellow	Yellow
Géospatiale, Carte,...	Green	Green	Green	Green	Green	Green	Green	Green	Yellow	Yellow
Graphiques dépendant du temps	Green	Green	Green	Green	Green	Green	Green	Green	Yellow	Yellow
Tableau	Green	Green	Green	Green	Green	Green	Green	Green	Yellow	Yellow
Autres Graphiques	Yellow	Green	Green	Green	Green	Green	Green	Green	Yellow	Yellow
Fonction d'édition des visuels créés										
Annotations automatiques (fermées)	Green	Green	Green	Green	Green	Green	Green	Green	Yellow	Yellow
Annotations manuelles (ouvertes)	Green	Green	Green	Red	Green	Green	Green	Green	Yellow	Yellow
Réutilisation des visualisations	Green	Yellow	Green	Red	Green	Green	Green	Green	Yellow	Yellow
Exportation des visuels	Green	Green	Green	Green	Green	Green	Green	Green	Yellow	Yellow
Sous forme d'image (PNG, JPG)	Green	Green	Green	Green	Green	Red	Green	Green	Yellow	Yellow
Sous forme d'image Vector (SVG)	Red	Red	Red	Red	Red	Red	Red	Red	Yellow	Yellow
Data Model (JSON)	Red	Red	Red	Red	Red	Red	Red	Red	Yellow	Yellow
Print Vector / PDF	Red	Red	Red	Red	Red	Red	Red	Red	Yellow	Yellow
fonctionnalités de partage et publication des visualisations										
Lien pour intégration URL	Red	Green	Red	Red	Green	Green	Red	Green	Yellow	Yellow
direct sur les réseaux sociaux,....	Red	Green	Red	Red	Green	Green	Red	Green	Yellow	Yellow
Création de rapports	Red	Green	Red	Red	Green	Green	Red	Green	Yellow	Yellow
Création de dashboard	Yellow	Green	Red	Red	Green	Green	Red	Green	Yellow	Yellow
Support proposé										
Ressource individuelle	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green
Ressource collective	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green
Collaboration										
Travail collaboratif - synchrone	Red	Green	Red	Red	Green	Green	Green	Green	Red	Red
Travail collaboratif - asynchrone	Red	Green	Red	Red	Green	Green	Green	Green	Red	Red
Prix	Yellow	Yellow	Yellow	Yellow	Yellow	Yellow	Yellow	Yellow	Yellow	Yellow
protection des données	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green
Langue de l'outils	FR, en, nl	fr,en	fr,en	en	fr,en	fr,en	fr,en	en	en	en

TABLE 4.7 – Récapitulatif accessibilité

Cette analyse comparative (Voir aussi tableau complet de comparaison 4.7) montre qu'il est très difficile de faire des recommandations générales quant à l'accessibilité des outils de visualisation. L'intérêt de l'analyse et la comparaison n'en est pas moins important. En effet, les analyses et les tableaux de comparaison ont comme vocation d'informer un futur utilisateur dans la contingence contextuelle qui est la sienne sur les différents outils de visualisation qui s'offrent à lui. Ceux-ci sont, nous l'avons dit, nombreux et parfois complexes à comprendre et à prendre en main. Le travail que nous avons mené permet déjà de simplifier la démarche d'un futur utilisateur en identifiant les outils les plus souvent référencés et en les analysant à l'aide d'une même grille d'analyse. Les tableaux et analyses de comparaison qui en ressortent sont destinés à être lus en contexte, c'est à

dire sur base des compétences de l'utilisateur, du projet qui est le sien, des données dont il dispose, du caractère collectif ou individuel de sa démarche, du budget à sa disposition... Ce n'est qu'avec ce contexte en tête que les analyses et tableaux de comparaison peuvent servir d'aide à la décision pour un futur utilisateur. En effet, en fonction des différents critères comme l'utilisation qu'on veut en faire, le niveau de compétence, le budget, certains outils seront plus accessibles que d'autres.

Pour pouvoir faire de la collaboration, les outils les plus accessibles sont les outils comme Infogram et Spotfire mais aussi les outils BI comme Tableau, Power Bi et Qlick Sens car ils permettent de créer des équipes et de les gérer au sein de l'outil. Mais là où Infogram ou Tableau, par exemple, ne requièrent pas de compétence informatique, il n'en n'est pas de même pour les autres ce qui peut rendre leur accessibilité plus délicate pour un groupe de citoyens, sans background informatique, souhaitant collaborer ensemble autour de jeux de données de leur quartier.

Pour le critère de réutilisabilité, les outils qui permettent cette fonctionnalité seront très utiles pour permettre aux utilisateurs de modifier ou d'adapter plus facilement des visualisations. Les outils comme Tableau, Power Bi et Qlick Sens mais aussi Excel et Spotfire permettent plus d'options concernant la réutilisabilité des visualisations, mais à nouveau les compétences nécessaires peuvent varier d'un outil à l'autre tant pour créer des visualisations personnalisées que pour adapter ou modifier des visualisations existantes.

Par rapport au prix, il existe très peu d'outils gratuits et disponibles en ligne, ici nous pouvons citer Raw Graphs qui est entièrement gratuit mais n'offre pas beaucoup d'options en termes de collaboration ou de réutilisation.

Infogram est sans doute l'outil qui ressort comme le meilleur candidat pour pouvoir faire des visualisations simples pour un travail citoyen individuel ou en groupe sans avoir de compétences en informatique. En effet, ils offre beaucoup de fonctionnalités pour la création de visualisations et est simple d'utilisation, ne nécessitant pas non plus d'installation de logiciel, il permet aussi aux utilisateurs de collaborer sur des visualisations.

Pour un profil de citoyen seul qui souhaite mener un travail ou une réflexion individuelle sur des données ouvertes, et qui n'a pas de compétences en langage de programmation, Excel et Infogram sont de bon candidat de part leur simplicité d'accès et les fonctionnalités qu'ils permettent. Raw-graph peut aussi être utilisé par ce profil mais il est moins facile d'accès.

Avec un profil de citoyen individuel, mené par un informaticien par exemple, nécessitant du coup des compétences en informatique et en langage de programmation, les outils basés sur des bibliothèques JavaScript D3 et p5, ou R permettent plus de souplesse et de fonctionnalités statistiques et graphiques.

Pour un profil de citoyens en groupe avec des en statistique et quelques compétences en informatique qui souhaitent collaborer ensemble ou créer des équipes, Spotfire ou les outils de Business Intelligence (Tableau, Power Bi ou Qlick) fournissent des mécanismes puissants, d'une part pour explorer les données via des visualisations, des tableaux de bord ou des rapports, et d'autre part pour faire des analyses de données ou du *storytelling*.

On peut voir qu'il est nécessaire de combiner plusieurs critères pour trouver l'outil le plus accessible pour un contexte d'usage particulier. Le tableau complet (voir Tableau en annexe :4.7) permet d'avoir une vue globale et d'aider à la décision.

Conclusion

L'ambition de ce mémoire était de mieux comprendre ce qui permet l'utilisation par le public de données produites dans des cadres métiers, qu'ils soient liés aux administrations, aux associations ou aux entreprises privées.

Dans un premier chapitre du mémoire, nous avons pris le temps de bien définir les termes, le contexte et les enjeux de ce qu'on qualifie aujourd'hui d'OGDs. De manière intéressante, cette partie définitionnelle a fait surgir la question de l'accessibilité de ces données ou plus particulièrement la question de la capacité du public en les utiliser de manière efficace. Ainsi, nous avons pu poser qu'ouverture de la donnée ne signifie pas nécessairement son accessibilité par le grand public.

Pour comprendre ce qui permet à ces données de pouvoir être utilisées par le public dans des dossiers variés que des projets de quartier ou de consommation énergétique, il fallait explorer tout le parcours que suivent des données produites dans des cadres métiers pour devenir des données utilisables et utilisées par un plus large public. Nous nous sommes d'abord intéressé aux différents acteurs et processus qui interviennent dans l'ouverture et ensuite l'utilisation de ces données par le public. C'est l'objet du chapitre 2. Ce chapitre commence par la première étape du processus, à savoir celle de l'ouverture de la donnée. En effet, nous nous sommes vite rendu compte qu'il ne suffit pas de libérer la donnée, il faut en construire l'ouverture et pour cela des processus et des rôles nouveaux sont nécessaires. S'agissant des processus, produire des données ouvertes est le résultat d'un assez long cheminement qui comporte plusieurs étapes. Il s'agit de la préparation, la gouvernance et la publication des données mais aussi du choix d'une licence qui permet de dire ce qu'on peut faire avec les données. Pour accompagner et administrer ce processus, des nouveaux rôles ont vu le jour avec l'administrateur/trice général(e) des données ou *Chief Data Officer (CDO)* et le/la scientifique des données ou *Data Scientist*.

Une fois les données ouvertes, il convient encore de pouvoir identifier les données nécessaires, utiles et utilisables par des citoyens ou un public confronté à un problème ou souhaitant mener un projet. Nous avons qualifié cette deuxième étape de processus d'analyse des besoins. La littérature nous en parle comme d'un processus collaboratif qui rassemble le plus souvent des experts et des citoyens dans des dynamiques qui visent à faire s'exprimer les besoins ou attentes des citoyens et identifier les sets de données qui pourraient y répondre. Nous avons vu que différentes méthodes et techniques peuvent outiller ce processus comme l'analyse collaborative, le *storytelling* et la visualisation.

Enfin, la troisième étape est celle qui consiste à se servir et donner du sens à des données pour répondre aux besoins ou attentes identifiés. Ici aussi différentes méthodes et techniques existent pour soutenir ce processus comme les hackathons et *serious games*, mais aussi la visualisation. La visualisation est à la fois utilisée lors de l'analyse des besoins des citoyens mais aussi pour l'exploitation des données finalisées. A travers ce chapitre, nous avons vu tout le long cheminement qui ponctue le trajet des données depuis leur libération par leur entité de départ jusqu'à leur utilisation par les citoyens à des fins publiques ou associatives. Dans ce cheminement, il nous est aussi apparu que repérer des données utiles, les comprendre, les utiliser à des fins d'analyse, leur donner sens dans un projet est une démarche difficile qui nécessite que des outils puissent aider les citoyens ou le public dans la prise en main de ces données, leur lisibilité et leur utilisation. Parmi ces outils, nous avons choisi sur base d'un état de l'art de nous intéresser plus particulièrement aux outils de visualisation dont nous avons vu qu'ils interviennent dans les 2 dernières étapes du cheminement décrit dans le chapitre précédent.

Nous consacrons le chapitre 3 à une revue de littérature sur les outils de visualisation existants afin de comprendre de quoi ils sont faits et ce que sont leurs fonctionnalités. Restant alignée sur notre objectif d'accessibilité, c'est à travers ce regard aussi que nous avons choisi d'évaluer 10 outils de visualisation parmi les plus souvent référencés par la littérature. En effet, l'objectif de notre mémoire est bien celui de renforcer la capacité des citoyens à servir efficacement des données ouvertes. Si les outils de visualisation peuvent y concourir, ils doivent eux aussi être accessibles ou en d'autres termes ne pas ajouter par leur complexité ou leur prix une nouvelle barrière pour les citoyens. C'est la raison pour laquelle, nous avons pour les explorer, les évaluer et les comparer entre eux développé une grille d'évaluation portant sur différents critères centrés utilisateurs.

Le chapitre 4 propose une évaluation de chacun des 10 outils de visualisation identifiés sur base de la grille commune des critères définie dans le chapitre précédent. Nous les avons analysés, testés et comparés.

Lors des tests nous avons été limités par le prix de certains outils payants qui ne nous ont pas permis de tester correctement certaines fonctionnalités. Nous avons aussi été limités par le manque de connaissance des langages de programmation JavaScript et R, mais aussi en statistiques pour construire des visualisations plus parlantes. Tout ceci a joué sur notre travail mais joue aussi sur l'accessibilité de ces outils. Au terme de cette évaluation, nous avons dressé un tableau qui devrait permettre à chacun, en fonction du contexte de compétence, de données disponibles, de modalités de travail, de budget... d'être aidé dans sa décision sur le choix de l'outil qui correspond le mieux à sa situation et ses attentes.

Ce travail reste un travail de mémoire, c'est-à-dire exploratoire. Nous y voyons deux prolongements possibles. Le premier concerne la trajectoire de la donnée de sa production métier à son ouverture publique. Ce trajet est fascinant. Quelques études de cas permettent de se rendre compte de la difficulté du processus mais aussi de son côté encore fort artisanal dans certaines administrations. Un travail méthodologique pour aider les administrations dans ce processus pourrait être proposé sous la forme d'un roadmap de questions à se poser et de réponses apportées par des administrations ayant ouvert avec succès leurs données. Le second concerne l'évaluation des outils de visualisation. Afin de valider les évaluations réalisées sur les différents outils de visualisation, il nous paraîtrait intéressant qu'un travail de mémoire ultérieur puisse organiser des tests d'accessibilité au départ de profils en compétences informatiques et statistiques autres que celui de la rédactrice du présent mémoire.

Glossaire

API En informatique, une interface de programmation d'application ou interface de programmation applicative est un ensemble normalisé de classes, de méthodes, de fonctions et de constantes qui sert de façade par laquelle un logiciel offre des services à d'autres logiciels. Wikipédia. 14, 15

API RESTful Une API REST (également appelée API RESTful) est une interface de programmation d'application (API ou API web) qui respecte les contraintes du style d'architecture REST et permet d'interagir avec les services web RESTful.. 15

BI Business Intelligence ou L'informatique décisionnelle est l'informatique à l'usage des décideurs et des dirigeants d'entreprises. Elle désigne les moyens, les outils et les méthodes qui permettent de collecter, consolider, modéliser et restituer les données, matérielles ou immatérielles, d'une entreprise en vue d'offrir une aide à la décision et de permettre à un décideur d'avoir une vue d'ensemble de l'activité traitée. Wikipédia. 22, 23, 52

big data les mégadonnées ou les données massives, désigne les ressources d'informations dont les caractéristiques en termes de volume, de vélocité et de variété imposent l'utilisation de technologies et de méthodes analytiques particulières pour générer de la valeur, et qui dépassent en général les capacités d'une seule et unique machine et nécessitent des traitements parallélisés. Wikipédia. 14

CRAN Le Comprehensive R Archive Network (CRAN) est un site web où l'on peut trouver et télécharger le matériel concernant R.[99]. 43

CSV Comma-separated values, connu sous le sigle CSV, est un format texte ouvert représentant des données tabulaires sous forme de valeurs séparées par des virgules. Ce format n'a jamais vraiment fait l'objet d'une spécification formelle. Wikipédia. 15, 25, 32, 33, 36, 45

Dropbox est un service de stockage et de partage de copies de fichiers locaux en ligne proposé par Dropbox, Inc., entreprise localisée à San Francisco, en Californie. Dropbox fournit des logiciels client ainsi qu'une interface web. Wikipédia. 36

DSV Le Delimiter-Separated Values est un format informatique qui stocke les tableaux en deux dimensions en séparant les valeurs de chaque rangée par un ou plusieurs délimiteurs. Wikipédia. 36

FAQ Frequently Asked Questions - Rubrique (d'un site Internet...) qui regroupe les réponses aux questions les plus fréquemment posées.. 30-32, 35, 37, 39, 41, 44

Framework En programmation informatique, un framework désigne un ensemble cohérent de composants logiciels structurels, qui sert à créer les fondations ainsi que les grandes lignes de tout ou d'une partie d'un logiciel. Un framework se distingue d'une simple bibliothèque logicielle principalement par : Wikipédia. 22, 23

GeoJSON est un format ouvert d'encodage d'ensemble de données géospatiales simples utilisant la norme JSON. Wikipédia. 15, 37

Gif Le Graphics Interchange Format (littéralement « format d'échange d'images »), plus connu sous l'acronyme GIF, est un format d'image numérique couramment utilisé sur l'Internet. Wikipédia. 33

- GitHub** est un service web d'hébergement et de gestion de développement de logiciels, utilisant le logiciel de gestion de versions Git. Wikipédia. 47
- Gitter** est un système de messagerie instantanée et de salle de discussion open-source pour les développeurs et les utilisateurs de référentiels GitLab et GitHub. Wikipédia (anglais). 47
- Google Group** est un service de groupe de discussion proposé par Google et qui permet d'accéder notamment aux services de messages Usenet. Son archive contient plus d'un milliard de messages consultables depuis 1981. Wikipédia. 47
- GUI** En informatique, une interface graphique (en anglais GUI pour graphical user interface) ou un environnement graphique est un dispositif de dialogue homme-machine. Wikipédia. 43
- HTML** Le HyperText Markup Language, généralement abrégé HTML ou dans sa dernière version HTML5, est le langage de balisage conçu pour représenter les pages web. Ce langage permet : d'écrire de l'hypertexte, d'où son nom, de structurer sémantiquement la page, de mettre en forme le contenu, de créer des formulaires de saisie, (Wikipédia). 45
- JavaScript** est un langage de programmation de scripts principalement employé dans les pages web interactives et à ce titre est une partie essentielle des applications web. Avec les technologies HTML et CSS, JavaScript est parfois considéré comme l'une des technologies cœur du World Wide Web. Wikipédia. 23, 29
- JPEG** Joint Photographic Experts Group est une norme qui définit le format d'enregistrement et l'algorithme de décodage pour une représentation numérique compressée d'une image fixe. Wikipédia. 29, 33, 44
- JSON** JavaScript Object Notation est un format de données textuelles dérivé de la notation des objets du langage JavaScript. Il permet de représenter de l'information structurée comme le permet XML par exemple. Wikipédia. 15, 25, 29, 32, 33, 36
- KML** Keyhole Markup Language que l'on peut traduire par « langage à base de balises géolocales », est un langage fondé sur le formalisme XML et destiné à la gestion de l'affichage de données géospatiales dans les logiciels de SIG. Depuis 2008, le format est normalisé par l'Open Geospatial Consortium. Wikipédia. 37
- MP4** MPEG-4 Part 14, ou MP4, également appelé ISO/CEI 14496-14, est une partie de la norme MPEG-4 spécifiant un format conteneur pour encapsuler des données de type multimédia. L'extension de nom de fichier généralement associée à ce format est « .mp4 ». Wikipédia. 33
- Open Source** La désignation open source, ou code source ouvert, s'applique aux logiciels (et s'étend maintenant aux œuvres de l'esprit) dont la licence respecte les critères précisément établis par l'Open Source Initiative, c'est-à-dire les possibilités de libre redistribution, d'accès au code source et de création de travaux dérivés. Mis à la disposition du grand public, ce code source est généralement le résultat d'une collaboration entre programmeurs. Wikipédia. 30
- PDF** Le Portable Document Format, communément abrégé en PDF, est un langage de description de page présenté par la société Adobe Systems en 1992 et qui est devenu une norme ISO en 2008. La spécificité du PDF est de préserver la mise en page d'un document — polices de caractère, images, objets graphiques, etc. — telle qu'elle a été définie par son auteur, et cela quels que soient le logiciel, le système d'exploitation et l'ordinateur utilisés pour l'imprimer ou le visualiser. Wikipédia. 33, 44
- PNG** Le Portable Network Graphics est un format ouvert d'images numériques. Wikipédia. 29, 33, 44
- RDA** Le format RData (généralement avec l'extension .rdata ou .rda) est un format conçu pour être utilisé avec R. 37

- RDATA** Le format RData (généralement avec l'extension `.rdata` ou `.rda`) est un format conçu pour être utilisé avec R. 37
- SaaS** Le software as a service ou logiciel en tant que service, est un modèle d'exploitation commerciale des logiciels dans lequel ceux-ci sont installés sur des serveurs distants plutôt que sur la machine de l'utilisateur. Wikipédia. 15
- SAV** Saved Game -Les fichiers SAV peuvent être générés par différents jeux et programmes.. 37
- Shapefile** ou « fichier de formes » est un format de fichier pour les systèmes d'informations géographiques. Wikipédia. 15
- Slack** est une plateforme de communication collaborative propriétaire ainsi qu'un logiciel de gestion de projets créé par Stewart Butterfield, Eric Costello, Cal Henderson et Serguei Mourachov en août 2013 et officiellement lancée en février 2014. L'entreprise appartient au groupe Salesforce depuis décembre 2020. Wikipédia. 47
- SPARQL** est un langage de requête et un protocole qui permet de rechercher, d'ajouter, de modifier ou de supprimer des données RDF disponibles à travers Internet. Son nom est un acronyme récursif qui signifie SPARQL Protocol and RDF Query Language. Wikipédia. 16
- StackOverflow** est un site web proposant des questions et réponses sur un large choix de thèmes concernant la programmation informatique. Il fait partie du réseau de sites Stack Exchange. Wikipédia. 44, 47
- SVG** Le Scalable Vector Graphics, ou SVG, est un format de données ASCII conçu pour décrire des ensembles de graphiques vectoriels et basé sur XML. Ce format inspiré directement du VML et du PGML est spécifié par le World Wide Web Consortium.(Wikipédia). 23, 25, 29, 33, 44
- TIFF** Le Tag Image File Format généralement abrégé TIFF est un format de fichier pour image numérique. Adobe en est le dépositaire et le propriétaire initial. Plus exactement, il s'agit d'un format de conteneur, à la manière de avi ou zip, c'est-à-dire pouvant contenir des données de formats arbitraires. Wikipédia. 44
- TSV** Tab-separated values est un format texte ouvert représentant des données tabulaires sous forme de « valeurs séparées par des tabulations ». Ce format de fichier peut être notamment utilisé pour exporter ou importer des données d'une base de données ou d'un tableur. Wikipédia. 36, 45
- TXT** un fichier texte ou fichier texte brut ou fichier texte simple est un fichier dont le contenu représente uniquement une suite de caractères ; il utilise nécessairement une forme particulière de codage des caractères qui peut être une variante ou une extension du standard local des États-Unis, l'ASCII.Wikipédia. 45
- URL** Une URL(Uniform Resource Locator), couramment appelée adresse web, est une chaîne de caractères uniforme qui permet d'identifier une ressource du World Wide Web par son emplacement et de préciser le protocole internet pour la récupérer.Wikipédia. 25, 29, 32
- webinaires** est un mot-valise associant les mots web et séminaire, créé pour désigner toutes les formes de réunions interactives de type séminaire faites via internet généralement dans un but de travail collaboratif ou de transmission d'informations pour une audience plus ou moins importante en nombre. Wikipédia. 42
- Wiki** Site web collaboratif, au développement duquel les visiteurs peuvent participer par l'écriture de pages web et la création de liens, et dont la permanence repose sur l'identification des auteurs, l'archivage des versions et le suivi des modifications.Le Robert. 35
- WMF** Windows Metafile est un format d'image numérique sur les systèmes Windows. Wikipédia. 44
- XLS** est une extension de nom de fichier pour tableur au format de Microsoft Excel.Wikipédia. 15, 25, 33

XLSX est une extension de nom de fichier pour tableur au format Office Open XML utilisé par Microsoft Office à partir de la version 2007. Wikipédia. 33

XML L'Extensible Markup Language, généralement appelé XML, « langage de balisage extensible » en français, est un métalangage informatique de balisage générique qui est un sous-ensemble du Standard Generalized Markup Language. Wikipédia. 29, 32

XPS Le XML Paper Specification est un langage de description de page, développé par Microsoft, destiné à concurrencer en partie le PDF pour la description de documents fixes et destinés à l'archivage et à l'impression. Il peut théoriquement être créé par n'importe quelle application. Wikipédia. 32

Bibliographie

- [1] About The Licenses - Creative Commons, . URL <https://creativecommons.org/licenses/>.
- [2] Licenses — Open Data Commons : legal tools for open data, . URL <https://opendatacommons.org/licenses/>.
- [3] Tableau : Business Intelligence and Analytics Software. URL <https://www.tableau.com/home>.
- [4] Mark S. Ackerman and David W. McDonald. Answer garden 2 : Merging organizational memory with collaborative help. In *Proceedings of the 1996 ACM Conference on Computer Supported Cooperative Work, CSCW '96*, page 97–105, New York, NY, USA, 1996. Association for Computing Machinery. ISBN 0897917650. doi : 10.1145/240080.240203. URL <https://doi.org/10.1145/240080.240203>.
- [5] Mark S Ackerman, Juri Dachtera, Volkmar Pipek, and Volker Wulf. Sharing knowledge and expertise : The cscw view of knowledge management. *Computer Supported Cooperative Work (CSCW)*, 22(4-6) :531–573, 2013.
- [6] Kenneth B. Allen. Access to government information. *Government Information Quarterly*, 9(1) :67 – 80, 1992. ISSN 0740-624X. doi : [https://doi.org/10.1016/0740-624X\(92\)90033-I](https://doi.org/10.1016/0740-624X(92)90033-I). URL <http://www.sciencedirect.com/science/article/pii/0740624X9290033I>.
- [7] Eike Falk Anderson, Leigh McLoughlin, Fotis Liarokapis, Christopher Peters, C Petridis, and Sara de Freitas. Serious games in cultural heritage. 2009.
- [8] Mariam Asad, Christopher A. Le Dantec, Becky Nielsen, and Kate Diedrick. Creating a sociotechnical api : Designing city-scale community engagement. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems, CHI '17*, page 2295–2306, New York, NY, USA, 2017. Association for Computing Machinery. ISBN 9781450346559. doi : 10.1145/3025453.3025963. URL <https://doi.org/10.1145/3025453.3025963>.
- [9] Raissa Barcellos, José Viterbo, Leandro Miranda, Flávia Bernardini, Cristiano Maciel, and Daniela Trevisan. Transparency in practice : using visualization to enhance the interpretability of open data. In *Proceedings of the 18th Annual International Conference on Digital Government Research*, pages 139–148, 2017.
- [10] Gene Bellinger, Durval Castro, and Anthony Mills. Data, information, knowledge, and wisdom, 2004.
- [11] John C. Bertot, Paul T. Jaeger, and Justin M. Grimes. Using icts to create a culture of transparency : E-government and social media as openness and anti-corruption tools for societies. *Government Information Quarterly*, 27(3) :264 – 271, 2010. ISSN 0740-624X. doi : <https://doi.org/10.1016/j.giq.2010.03.001>. URL <http://www.sciencedirect.com/science/article/pii/S0740624X10000201>.
- [12] David M Boje. The storytelling organization : A study of story performance in an office-supply firm. *Administrative science quarterly*, pages 106–126, 1991.
- [13] Michael Bostock and Jeffrey Heer. Protovis : A graphical toolkit for visualization. *IEEE transactions on visualization and computer graphics*, 15(6) :1121–1128, 2009.

- [14] Mike Bostock. D3.js - Data-Driven Documents. URL <https://d3js.org/>.
- [15] Eva Brandt and Jörn Messeter. Facilitating collaboration through design games. In *Proceedings of the Eighth Conference on Participatory Design : Artful Integration : Interweaving Media, Materials and Practices - Volume 1*, PDC 04, page 121–131, New York, NY, USA, 2004. Association for Computing Machinery. ISBN 1581138512. doi : 10.1145/1011870.1011885. URL <https://doi.org/10.1145/1011870.1011885>.
- [16] Gerard Briscoe and Catherine Mulligan. Digital innovation : The hackathon phenomenon.(2014). *Google Scholar There is no corresponding record for this reference*, 2014.
- [17] Bulles à verre. Bulles à verre. URL <https://www.odwb.be/explore/dataset/bulles-a-verre/map/>.
- [18] Corentin Burnay. De la Donnée vers la Connaissance : comment la Business Intelligence contribue-t-elle à rendre nos villes plus smart?, March 2021. URL <https://www.youtube.com/watch?v=HKJ3iVsCsMg&t=6499s>.
- [19] Citizens of Wallonia. Home. URL <https://www.citizensofwallonia.be/>.
- [20] CKAN. ckan. URL <https://ckan.org/>.
- [21] CKAN Blog. The Comprehensive Knowledge Archive Network (CKAN) Launched Today, July 2007. URL <https://blog.okfn.org/2007/07/04/the-comprehensive-knowledge-archive-network-ckan-launched-today/>.
- [22] Herbert H Clark and Susan E Brennan. Grounding in communication. 1991.
- [23] codePlex Archive. CodePlex Archive. URL <https://archive.codeplex.com/>.
- [24] Grazia Concilio, Francesco Molinari, and Nicola Morelli. Empowering citizens with open data by urban hackathons. In *2017 Conference for E-Democracy and Open Government (CeDEM)*, pages 125–134. IEEE, 2017.
- [25] Antoine Courmont. Ce que l’open data fait à l’administration municipale. *Rezeaux*, (6) : 77–103, 2019.
- [26] Antoine Courmont. Ce que l’open data fait à l’administration municipale. La fabrique de la politique métropolitaine de la donnée. *Rezeaux*, 218(6) :77–103, 2019. ISSN 9782348054846. doi : 10.3917/res.218.0077. URL <https://www.cairn.info/revue-rezeaux-2019-6-page-77.htm>. Place : Paris Publisher : La Découverte.
- [27] Jocelyn Cranefield, Oliver Robertson, and Gillian Oliver. Value in the mash : Exploring the benefits, barriers and enablers of open data apps. 2014.
- [28] Datastorebrussels. datastore.brussels. URL <https://datastore.brussels/web/>.
- [29] Tim Davies and Mark Frank. ‘there’s no such thing as raw data’ : Exploring the socio-technical life of a government dataset. In *Proceedings of the 5th Annual ACM Web Science Conference*, WebSci ’13, page 75–78, New York, NY, USA, 2013. Association for Computing Machinery. ISBN 9781450318891. doi : 10.1145/2464464.2464472. URL <https://doi.org/10.1145/2464464.2464472>.
- [30] Sharon S Dawes, Lyudmila Vidasova, and Olga Parkhimovich. Planning and designing open government data programs : An ecosystem approach. *Government Information Quarterly*, 33(1) :15–27, 2016.
- [31] Jérôme Denis and Samuel Goëta. La fabrique des données brutes. Le travail en coulisses de l’open data. In *Penser l’écosystème des données. Les enjeux scientifiques et politiques des données numériques*, Paris, France, February 2013. URL <https://halshs.archives-ouvertes.fr/halshs-00990771>.
- [32] Jérôme Denis and Samuel Goëta. Rawification and the careful generation of open government data. *Social Studies of Science*, 47(5) :604 – 629, October 2017. doi : 10.1177/0306312717712473. URL <https://halshs.archives-ouvertes.fr/halshs-01617976>.

- [33] Robert F Dennehy. The springboard : How storytelling ignites action in knowledge-era organizations. *Journal of Organizational Change Management*, 2001.
- [34] Kevin C Desouza and Akshay Bhagwatwar. Citizen apps to solve complex urban problems. *Journal of Urban Technology*, 19(3) :107–136, 2012.
- [35] Distributeurs bancaires. Distributeurs bancaires. URL https://opendata.bruxelles.be/explore/dataset/bruxelles_distributeurs_bancaires/table/.
- [36] Paul N Edwards, Geoffrey C Bowker, Steven J Jackson, and Robin Williams. Introduction : an agenda for infrastructure studies. *Journal of the Association for Information Systems*, 10(5) :6, 2009.
- [37] Sheena Erete, Emily Ryou, Geoff Smith, Khristina Marie Fassett, and Sarah Duda. Storytelling with data : Examining the use of data by non-profit organizations. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing, CSCW '16*, page 1273–1283, New York, NY, USA, 2016. Association for Computing Machinery. ISBN 9781450335928. doi : 10.1145/2818048.2820068. URL <https://doi.org/10.1145/2818048.2820068>.
- [38] Mette Agger Eriksen, Eva Brandt, Tuuli Mattelmäki, and Kirsikka Vaajakallio. Taking design games seriously : Re-connecting situated power relations of people and materials. In *Proceedings of the 13th Participatory Design Conference : Research Papers - Volume 1, PDC '14*, page 101–110, New York, NY, USA, 2014. Association for Computing Machinery. ISBN 9781450322560. doi : 10.1145/2661435.2661447. URL <https://doi.org/10.1145/2661435.2661447>.
- [39] Etalab. Licence Ouverte / Open Licence. URL <https://www.etalab.gouv.fr/licence-ouverte-open-licence>.
- [40] Sean Follmer, Daniel Leithinger, Alex Olwal, Akimitsu Hogge, and Hiroshi Ishii. Inform : Dynamic physical affordances and constraints through shape and object actuation. In *Proceedings of the 26th Annual ACM Symposium on User Interface Software and Technology, UIST '13*, page 417–426, New York, NY, USA, 2013. Association for Computing Machinery. ISBN 9781450322683. doi : 10.1145/2501988.2502032. URL <https://doi.org/10.1145/2501988.2502032>.
- [41] William Gaver, Andy Boucher, Nadine Jarvis, David Cameron, Mark Hauenstein, Sarah Pennington, John Bowers, James Pike, Robin Beitra, and Liliana Ovalle. The datacatcher : Batch deployment and documentation of 130 location-aware, mobile devices that put sociopolitically-relevant big data in people’s hands : Polyphonic interpretation at scale. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems, CHI '16*, page 1597–1607, New York, NY, USA, 2016. Association for Computing Machinery. ISBN 9781450333627. doi : 10.1145/2858036.2858472. URL <https://doi.org/10.1145/2858036.2858472>.
- [42] Darren Gergle, Robert E Kraut, and Susan R Fussell. Language efficiency and visual technology : Minimizing collaborative effort with visual information. *Journal of language and social psychology*, 23(4) :491–517, 2004.
- [43] Jeremy Goecks, Amy Volda, Stephen Volda, and Elizabeth D. Mynatt. Charitable technologies : Opportunities for collaborative computing in nonprofit fundraising. In *Proceedings of the 2008 ACM Conference on Computer Supported Cooperative Work, CSCW '08*, page 689–698, New York, NY, USA, 2008. Association for Computing Machinery. ISBN 9781605580074. doi : 10.1145/1460563.1460669. URL <https://doi.org/10.1145/1460563.1460669>.
- [44] Connie Golsteijn, Sarah Gallacher, Lisa Koeman, Lorna Wall, Sami Andberg, Yvonne Rogers, and Licia Capra. Voxbox : A tangible machine that gathers opinions from the public at events. In *Proceedings of the Ninth International Conference on Tangible, Embedded, and Embodied Interaction, TEI '15*, page 201–208, New York, NY, USA, 2015. Association for

- Computing Machinery. ISBN 9781450333054. doi : 10.1145/2677199.2680588. URL <https://doi.org/10.1145/2677199.2680588>.
- [45] Connie Golsteijn, Sarah Gallacher, Licia Capra, and Yvonne Rogers. Sens-us : Designing innovative civic technology for the public good. In *Proceedings of the 2016 ACM Conference on Designing Interactive Systems*, DIS '16, page 39–49, New York, NY, USA, 2016. Association for Computing Machinery. ISBN 9781450340311. doi : 10.1145/2901790.2901877. URL <https://doi.org/10.1145/2901790.2901877>.
- [46] Eric Gordon, Becky Michelson, and Jason Haas. @stake : A game to facilitate the process of deliberative democracy. In *Proceedings of the 19th ACM Conference on Computer Supported Cooperative Work and Social Computing Companion*, CSCW '16 Companion, page 269–272, New York, NY, USA, 2016. Association for Computing Machinery. ISBN 9781450339506. doi : 10.1145/2818052.2869125. URL <https://doi.org/10.1145/2818052.2869125>.
- [47] Alvaro Graves and James Hendler. Visualization tools for open government data. In *Proceedings of the 14th Annual International Conference on Digital Government Research*, pages 136–145, 2013.
- [48] Michael B Gurstein. Open data : Empowering the empowered or effective data use for everyone? *First Monday*, 2011.
- [49] Jeffrey Heer and Maneesh Agrawala. Design considerations for collaborative visual analytics. *Information visualization*, 7(1) :49–62, 2008. doi : 10.1109/VAST.2007.4389011.
- [50] Jessica Hullman, Eytan Adar, and Priti Shah. The impact of social information on visual judgments. In *Proceedings of the SIGCHI conference on human factors in computing systems*, pages 1461–1470, 2011.
- [51] Infogram. infogram. URL <https://infogram.com/fr/about>.
- [52] Nicolas Installé. Open Data en Wallonie : comment aider les administrations à faire le premier pas?, March 2021. URL <https://www.youtube.com/watch?v=HKJ3iVsCsMg&t=6499s>.
- [53] Yvonne Jansen, Pierre Dragicevic, Petra Isenberg, Jason Alexander, Abhijit Karnik, Johan Kildal, Sriram Subramanian, and Kasper Hornbæk. Opportunities and challenges for data physicalization. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, CHI '15, page 3227–3236, New York, NY, USA, 2015. Association for Computing Machinery. ISBN 9781450331456. doi : 10.1145/2702123.2702180. URL <https://doi.org/10.1145/2702123.2702180>.
- [54] Marijn Janssen, Yannis Charalabidis, and Anneke Zuiderwijk. Benefits, adoption barriers and myths of open data and open government. *Information Systems Management*, 29(4) : 258–268, 2012. doi : 10.1080/10580530.2012.716740. URL <https://doi.org/10.1080/10580530.2012.716740>.
- [55] Zdeněk Jiříček and Francesca Di Massimo. Microsoft open government data initiative (ogdi), eye on earth case study. In *International Symposium on Environmental Software Systems*, pages 26–32. Springer, 2011.
- [56] Ian G Johnson, Aare Puusaar, Jennifer Manuel, and Peter Wright. Neighbourhood data : exploring the role of open data in locally devolved policymaking processes. *Proceedings of the ACM on Human-Computer Interaction*, 2(CSCW) :1–20, 2018.
- [57] Peter Johnson and Pamela Robinson. Civic hackathons : Innovation, procurement, or civic engagement? *Review of policy research*, 31(4) :349–357, 2014.
- [58] Rob Kabacoff. *Data Visualization with R*. URL <https://rkabacoff.github.io/datavis/>.
- [59] Maxat Kassen. A promising phenomenon of open data : A case study of the chicago open data project. *Government Information Quarterly*, 30, 10 2013. doi : 10.1016/j.giq.2013.05.012.

- [60] Giordano Koch, Katja Hutter, Peter Decarli, Dennis Hilgers, and Johann Füller. Identifying participants' roles in open government platforms and its impact on community growth. pages 1900–1910, 01 2013. ISBN 978-1-4673-5933-7. doi : 10.1109/HICSS.2013.254.
- [61] Kenneth Kuhn. Open government data and public transportation. *Journal of Public Transportation*, 14(1) :5, 2011.
- [62] Siân E Lindley, Anja Thieme, Alex S Taylor, Vasilis Vlachokyriakos, Tim Regan, and David Sweeney. Surfacing small worlds through data-in-place. *Computer Supported Cooperative Work (CSCW)*, 26(1-2) :135–163, 2017.
- [63] Jock Mackinlay, Pat Hanrahan, and Chris Stolte. Show me : Automatic presentation for visual analysis. *IEEE transactions on visualization and computer graphics*, 13(6) :1137–1144, 2007.
- [64] Erika G Martin, Natalie Helbig, and Guthrie S Birkhead. Opening health data : what do researchers want? early experiences with new york's open health data platform. *Journal of Public Health Management and Practice*, 21(5) :E1–E7, 2015.
- [65] Ricardo Matheus, Manuella Maia Ribeiro, and José Carlos Vaz. Brazil towards government 2.0 : Strategies for adopting open government data in national and subnational governments. In *Case Studies in e-Government 2.0*, pages 121–138. Springer, 2015.
- [66] Michele Mauri, Tommaso Elli, Giorgio Caviglia, Giorgio Uboldi, and Matteo Azzi. Rawgraphs : A visualisation platform to create open outputs. In *Proceedings of the 12th Biannual Conference on Italian SIGCHI Chapter*, CHIItaly '17, New York, NY, USA, 2017. Association for Computing Machinery. ISBN 9781450352376. doi : 10.1145/3125571.3125585. URL <https://doi.org/10.1145/3125571.3125585>.
- [67] Branislav Meszaros, Sitthida Samath, Sonia Guérin-Hamdi, and Céline Faure. Livre blanc sur les données ouvertes. 2015.
- [68] Petar Milić, Nataša Veljković, and Leonid Stoimenov. Framework for open data mining in e-government. In *Proceedings of the Fifth Balkan Conference in Informatics*, pages 255–258, 2012.
- [69] Hafiz Usman Mohammed. Creating social value through the use of open data, 2015.
- [70] Jerker Moodysson, Lars Coenen, and Bjørn Asheim. Explaining spatial patterns of innovation : Analytical and synthetic modes of knowledge creation in the medicon valley life-science cluster. *Environment and Planning A : Economy and Space*, 40(5) :1040–1056, 2008. doi : 10.1068/a39110. URL <https://doi.org/10.1068/a39110>.
- [71] Namur Population. Namur - Population - 7 - Chiffres et densité de population. URL <https://data.namur.be/explore/dataset/namur-population-7-evolution-indicateurs/api/>.
- [72] Observable. Observable - Make sense of the world with data, together / Observable. URL <https://observablehq.com/>.
- [73] Open data Belgique. Data.gov.be. URL <https://data.gov.be/fr>.
- [74] Open Data Bruxelles. Home — OpenData Bruxelles. URL <https://opendata.bruxelles.be/page/home/>.
- [75] Open data Chicago. City of Chicago | Data Portal | City of Chicago | Data Portal. URL <https://data.cityofchicago.org/>.
- [76] Open data Edmonton. Edmonton - Open Data Portal. URL <https://data.edmonton.ca/>.
- [77] Open data EU. Accueil | Portail européen de données. URL <https://www.europeandataportal.eu/fr>.
- [78] Open data Flandre. Bienvenue | Portail flamand des données ouvertes. URL <https://opendata.vlaanderen.be/>.

- [79] Open data Liege. Accueil — Open Data Liège. URL <https://opendata.liege.be/pages/accueil/>.
- [80] Open data Namur. Accueil — Namur - OpenData. URL <https://data.namur.be/pages/accueil/>.
- [81] Open data US Gov. Data.gov. URL <https://www.data.gov/>.
- [82] Open Data Wallonie. Portail Open Data — ODWB. URL <https://www.odwb.be/pages/home/>.
- [83] Open definition. Conformant Licenses - Open Definition - Defining Open in Open Data, Open Content and Open Knowledge. URL <https://opendefinition.org/licenses/>.
- [84] Opendatasoft. Opendatasoft - Make Your Data Bright. URL <https://www.opendatasoft.com>.
- [85] Benoît Otjacques, Mickaël Stefas, Maël Cornil, and Fernand Feltz. Open data visualization keeping traces of the exploration process. In *Proceedings of the First International Workshop on Open Data, WOD '12*, page 53–60, New York, NY, USA, 2012. Association for Computing Machinery. ISBN 9781450314046. doi : 10.1145/2422604.2422612. URL <https://doi.org/10.1145/2422604.2422612>.
- [86] Peter Parycek, Johann Hocht, and Michael Ginner. Open government data implementation evaluation. *Journal of theoretical and applied electronic commerce research*, 9(2) :80–99, 2014.
- [87] Ana Peraica. Big data, little data, no data : Scholarship in the networked world, 2016.
- [88] Jean-Christophe Plantin. «raw data» is an oxymoron lisa gitelman (dir.), cambridge, mit press, 2013. *Communication langages*, (3) :155–156, 2013.
- [89] R Pollock. Building the (open) data ecosystem.[web log post] retrieved 8 march, 2013 from <http://blog.okfn.org/2011/03/31/building-the-open-data-ecosystem>. Archived by Web-Cite® at <http://www.webcitation.org/6CefTj7HU>, 2011.
- [90] Alenka Poplin. Playful public participation in urban planning : A case study for online serious games. *Computers, environment and urban systems*, 36(3) :195–206, 2012.
- [91] Population bruxelloise. Population bruxelloise. URL <https://opendata.bruxelles.be/explore/dataset/population-bruxelloise/>.
- [92] Portail Socrata. Socrata Portail des données ouvertes. URL <https://data.europa.eu/euodp/fr/node/6541>.
- [93] Principles. The 8 Principles of Open Government Data (OpenGovData.org). URL <https://opengovdata.org/>.
- [94] Protovis. Protovis. URL <https://mbostock.github.io/protovis/>.
- [95] Arie Purwanto, Anneke Zuiderwijk, and Marijn Janssen. Group development stages in open government data engagement initiatives : A comparative case studies analysis. In *International Conference on Electronic Government*, pages 48–59. Springer, 2018.
- [96] Aare Puussaari, Ian G Johnson, Kyle Montague, Philip James, and Peter Wright. Making open data work for civic advocacy. *Proceedings of the ACM on Human-Computer Interaction*, 2(CSCW) :1–20, 2018.
- [97] A. Quint. Scalable vector graphics. *IEEE MultiMedia*, 10(3) :99–102, July 2003. ISSN 1941-0166. doi : 10.1109/MMUL.2003.1218261.
- [98] R bloggers. R-bloggers, April 2021. URL <https://www.r-bloggers.com/>.
- [99] R Project. R : The R Project for Statistical Computing. URL <https://www.r-project.org/>.
- [100] Rawgraph. Home. URL <https://rawgraphs.io/>.
- [101] John Resig. Processing. js.

- [102] RStudio. RStudio | Open source & professional software for data science teams. URL <https://rstudio.com/>.
- [103] Igal Safarov, Albert Meijer, and Stephan Grimmelikhuijsen. Utilization of open government data : A systematic literature review of types, conditions, effects and users. *Information Polity*, 22(1) :1–24, 2017.
- [104] Katie Salen, Katie Salen Tekinbaş, and Eric Zimmerman. *Rules of play : Game design fundamentals*. MIT press, 2004.
- [105] Carlos Scheidegger, Huy Vo, David Koop, Juliana Freire, and Claudio Silva. Querying and creating visualizations by analogy. *IEEE transactions on Visualization and Computer Graphics*, 13(6) :1560–1567, 2007.
- [106] Socrata. Socrata | Data & Insights | Tyler Technologies. URL <https://www.tylertech.com/products/socrata>.
- [107] Porya Mohajer Soltani, Kalevi Pessi, Karin Ahlin, and Ida Werner. Hackathon : A method for digital innovative success : A comparative descriptive study. In *Proceedings of the 8th European Conference on IS Management and Evaluation*, pages 367–373, 2014.
- [108] Iryna Susha, Åke Grönlund, and Marijn Janssen. Driving factors of service innovation using open government data : An exploratory study of entrepreneurs in two countries. *Information polity*, 20(1) :19–34, 2015.
- [109] Tarja Susi, Mikael Johannesson, and Per Backlund. *Serious games : An overview*, 2007.
- [110] Tableau synoptique. Tableau synoptique, September 2020. URL https://fr.wikipedia.org/w/index.php?title=Tableau_synoptique&oldid=174653902. Page Version ID : 174653902.
- [111] Alex S. Taylor, Siân Lindley, Tim Regan, David Sweeney, Vasillis Vlachokyriakos, Lillie Grainger, and Jessica Lingel. Data-in-place : Thinking through the relations between data and community. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems, CHI '15*, page 2863–2872, New York, NY, USA, 2015. Association for Computing Machinery. ISBN 9781450331456. doi : 10.1145/2702123.2702558. URL <https://doi.org/10.1145/2702123.2702558>.
- [112] TIBCO Spotfire. TIBCO Spotfire. URL <https://www.tibco.com/products/tibco-spotfire>.
- [113] M. Torchiano, A. Vetrò, and F. Iuliano. Preserving the benefits of open government data by measuring and improving their quality : An empirical study. In *2017 IEEE 41st Annual Computer Software and Applications Conference (COMPSAC)*, volume 1, pages 144–153, 2017.
- [114] Fernanda B Viégas, Danah Boyd, David H Nguyen, Jeffrey Potter, and Judith Donath. Digital artifacts for remembering and storytelling : Posthistory and social network fragments. In *37th Annual Hawaii International Conference on System Sciences, 2004. Proceedings of the*, pages 10–pp. IEEE, 2004.
- [115] Fernanda B Viegas, Martin Wattenberg, Frank Van Ham, Jesse Kriss, and Matt McKeon. Manyeyes : a site for visualization at internet scale. *IEEE transactions on visualization and computer graphics*, 13(6) :1121–1128, 2007.
- [116] Amy Volda, Ellie Harmon, and Ban Al-Ani. Homebrew databases : Complexities of everyday information management in nonprofit organizations. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '11*, page 915–924, New York, NY, USA, 2011. Association for Computing Machinery. ISBN 9781450302289. doi : 10.1145/1978942.1979078. URL <https://doi.org/10.1145/1978942.1979078>.
- [117] Martin Wattenberg. Baby names, visualization, and social data analysis. In *IEEE Symposium on Information Visualization, 2005. INFOVIS 2005.*, pages 1–7. IEEE, 2005.
- [118] Wesley Willett, Jeffrey Heer, Joseph Hellerstein, and Maneesh Agrawala. Commentspace : Structured support for collaborative visual analysis. In *Proceedings of the SIGCHI Conference*

- on *Human Factors in Computing Systems*, CHI '11, page 3131–3140, New York, NY, USA, 2011. Association for Computing Machinery. ISBN 9781450302289. doi : 10.1145/1978942.1979407. URL <https://doi.org/10.1145/1978942.1979407>.
- [119] Annika Wolff, Matthew Barker, and Marian Petre. Creating a datascape : A game to support communities in using open data. In *Proceedings of the 8th International Conference on Communities and Technologies*, pages 135–138, 2017.
- [120] Andrew Wong, Vicky Liu, William Caelli, and Tony Sahama. An architecture for trustworthy open data services. In *IFIP International Conference on Trust Management*, pages 149–162. Springer, 2015.
- [121] Han-Ming Wu, ShengLi Tzeng, and Chun-houh Chen. Matrix visualization. In *Handbook of data visualization*, pages 681–708. Springer, 2008.
- [122] Jeremy Yuille and Hugh Macdonald. Feature

the social life of visualization. *Interactions*, 17(1) :28–31, January 2010. ISSN 1072-5520. doi : 10.1145/1649475.1649482. URL <https://doi.org/10.1145/1649475.1649482>.
- [123] Anneke Zuiderwijk and Marijn Janssen. Open data policies, their implementation and impact : A framework for comparison. *Government Information Quarterly*, 31(1) :17–29, 2014.
- [124] Anneke Zuiderwijk, Marijn Janssen, and Chris Davis. Innovation with open data : Essential elements of open data ecosystems. *Information Polity*, 19(1, 2) :17–33, 2014.
- [125] Michael Zyda. From visual simulation to virtual reality to games. *Computer*, 38(9) :25–32, 2005.