Institutional Repository - Research Portal
Dépôt Institutionnel - Portail de la Recherche
researchportal.unamur.be
University of Namur

THESIS / THÈSE

MASTER IN COMPUTER SCIENCE

Understanding Citizens' Usability Requirements toward Visualization Features on Open Government Data Portals

BIHIN, Jehan

Award date:
2022

Awarding institution:
University of 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.

Download date: 12. May. 2024

Université de Namur
Faculté d'informatique
Année académique 2021–2022

Understanding Citizens' Usability Requirements toward Visualization Features on Open Government Data Portals

Jehan Bihin

Promoteur : _____ (Signature pour approbation du dépôt - REE art. 40)
Benoît Vanderose

Co-promoteur : Antoine Clarinval

Mémoire présenté en vue de l'obtention du grade de
Master en Sciences Informatiques.

Résumé

Les gouvernements publient de grandes quantités de données et utilisent des portails avec des fonctionnalités de visualisation pour fournir un accès et une consultation facilitée aux citoyens. Malgré cela, les portails open data restent peu utilisés en pratique. De nombreuses raisons peuvent expliquer cela, une potentielle étant l'utilisabilité des fonctionnalités de visualisation. En effet, la littérature scientifique ne fournit que peu d'indications sur comment garantir une bonne utilisabilité de ces fonctionnalités de visualisation pour les utilisateurs finaux. L'objectif de ce mémoire est donc de comprendre les exigences des citoyens concernant les fonctionnalités de visualisation sur les portails OGD et de les implémenter dans un prototype haute-fidélité. Nous avons sélectionné la solution de portail OpenDataSoft en raison de sa position sur le marché et avons mené un test utilisateur avec 5 participants, ainsi qu'une évaluation experte avec 2 experts en utilisabilité pour comprendre les difficultés rencontrées par les utilisateurs et leurs exigences. Le résultat montre une incompréhension des interfaces, un problème de navigation, des fonctionnalités graphiques difficiles à utiliser, une fonctionnalité cartographique inutile, des freins à la lecture des tables de jeux de données, et bien plus encore. Après regroupement et ordonnancement, nous avons défini 8 barrières à résoudre au cours de ce mémoire dans un prototype moyenne-fidélité mais nous avons également implémenté un prototype haute-fidélité résolvant la plupart de ces barrières.

Abstract

Governments publish large amounts of data and use portals with visualization features to provide easy access and consultation to citizens. Despite this, open data portals remain little used in practice. There are many reasons for this, one potential being the usability of visualization features. Indeed, the scientific literature provides only few indications on how to guarantee good usability of these visualization features for end users. The objective of this thesis is therefore citizens' requirements about visualization features on OGD portals, and to implement them into a high-fidelity prototype. We [1] selected the OpenDataSoft portal solution due to its position in the market and conducted a user test with 5 participants, as well as an expert assessment with 2 usability experts to understand the difficulties faced by users and their requirements. The result shows a misunderstanding of the interfaces, a navigation problem, difficult to use graphic functionalities, useless cartographic functionality, barriers to reading dataset tables, and much more. After grouping and ordering, we defined 8 barriers to solve during this thesis in a medium-fidelity prototype but we also implemented a high-fidelity prototype solving most of these barriers.

[1] This document uses the we pronoun to acknowledge the help of the supervisors

Keywords

Open Data, Open Governement Data, OGD, Portal, Usability, Visualization Features, Requirements, Prototype

Acknowledgment

I would like to thank Antoine Clarinval and Benoît Vanderose for supervising this thesis and a special thank you to Antoine for his assistance and guidance.

A big thank you also to Émeline for her support, and to all the friends and participants for your encouragement.

Contents

1 Introduction

Governments publish large amounts of data in various formats, for transparency, accountability reasons and to allow citizens to create new services or make better decisions based on this data. In order for government data to be opened, it needs to be made accessible to everyone through a new solution, under the form of dedicated online platforms called Open Government Data (OGD) portals.

These portals have become widespread, the Data Portal[1] census lists 593 existing open data portals. Overall, these portals look similar, with a welcome page presenting the dataset categories, a search engine, and a datasets tab that shows all the published datasets, allowing users to browse and filter datasets or download them. OGD portals do not limit themselves to giving access to data, but attempt to present it in a more understandable way for the end-user, using for example charts or maps.

Despite that, OGD portals are not much used by the population in practice. The barriers to OGD use are multiple and several studies have identified and listed them. Users can experience impediments such as lack of data or low-quality data, but also a gap between their technical skills and the skills required to use OGD portals and the visualization tools they offer. This is unfortunate, because these visualization tools are there to overcome these barriers.

These features provide several kind of visualizations such as charts, map, 2D or 3D rendering, which should alleviate barriers instead of creating new ones. The issue is that these visualization features have not been evaluated with end-users to ensure that they have a good usability, and the requirements of these users about the visualization features have not been identified.

To answer this gap, we chose an application case, an OGD portal, that is implemented with the OpenDataSoft solution because of its leading position on the OGD portal provider market and because it provides several visualization features and thus fits the research goal. In particular, since this master thesis is done with the University of Namur, the OGD portal of the city of Namur was picked. On this portal, a user test was conducted to evaluate the ease of use and usefulness of the visualization features with 5 citizens, and was supplemented with the review of 2 usability experts.

The results show the **barriers encountered** by participants and experts, such as the problem of multi-layers charts, the use of data tables or maps, or the navigation. From these usability issues and suggestions made by the participants of the user

[1] https://dataportals.org/

test, we designed a new solution in the form of a **medium-fidelity prototype** that answers the problems with the current OGD portal raised by the participants. And to put the cherry on the cake, we implemented a **web high-fidelity prototype** to illustrate part of the solution. These are the 3 contributions of this thesis.

This thesis is structured as follows. Chapter 2 presents the background on Open Government Data, the challenges users have to exploit it and how visualization can be part of the solution. In Chapter 3, we conduct an analysis of the requirements citizens have about visualization features on OGD portals via a user test of an existing OGD portal complemented by the opinion of two usability experts. Chapter 4 presents a medium-fidelity prototype illustrating possible solutions and how it works with the second scenario. Then in Chapter 5 we implement as a web application parts of the features in a high-fidelity prototype. Finally, in Chapter 6, we present leads for improvements to go further with possible solutions. Chapter 7 concludes the thesis with a summary of its contributions.

2 Background

2.1 Open Government Data

Governments, companies and citizens constantly publish large amounts of data that can be unstructures, structured (following global standards or not), or even linked. Examples include geospatial data, topography imagery, road networks and addresses, demographic information, pictures, profiles, and many others. When published for unrestricted use, such data can be qualified as open data (Open Data Handbook, 2015). If the open data is published by a public organization, it consists in Open Government Data (OGD) (Attard et al., 2015). Examples of OGD datasets include, among others, budgets, subsidies, extruded buildings of the cities, planned road works, population density, parking points, pedestrian counter, Walloon heritage sites, etc. The Dutch authorities have published a list of 25 datasets that they consider as the most valuable[1]. For example, traffic intensity data, location of monuments, public grants, and public art are part of this list.

According to the Organisation for Economic Co-operation and Development (OECD) [2] Open Government Data (OGD) is a philosophy - and increasingly a set of policies - that promotes transparency, accountability and value creation by making government data available to all, and is the most widespread source of Open Data since 2009-2011. At this time, a lot of countries decided to join the effort and made plans to build solutions. For exemple, the European Union Open Data Portal established by the Commission Decision (2011/833/EU), or in United States, by an initiative of Barack Obama in 2009 with a memorandum called 'Transparency and Open Government' (Bauer & Kaltenböck, 2011; Obama, 2009).

Still according to OECD[2], by making their datasets available, public institutions become more transparent and accountable to citizens. By encouraging the use, reuse and free distribution of datasets, governments promote business creation and innovative, citizen-centric services.

Open Government Data is found in various format, which are defined on Tim Berners-Lee 5-star model [3] (Figure 2.1). At the first level, data is published on the Internet under an open licence, in unstructured formats such as PDF. At the second level, the format is structured but proprietary like an Excel table for example. At the third level, the data is published online in a structured non-proprietary format such as CSV. Most of the OGD published today are at this level. At the fourth level, URIs are related to the data so that it can be referred to. At the final

[1] https://data.overheid.nl/community/maatschappij/high-value/gemeenten
[2] https://www.oecd.org/gov/digital-government/open-government-data.htm
[3] https://5stardata.info/en/

level, also known as Linked Open Data, the data is linked to other data to provide context. It refers more precisely to a set of best practices for publishing and connecting structured data on the Web (Bizer et al., 2011). These best practices have been adopted by a lot of data providers and conduct to build the Web of Data or also known as Semantic Web.

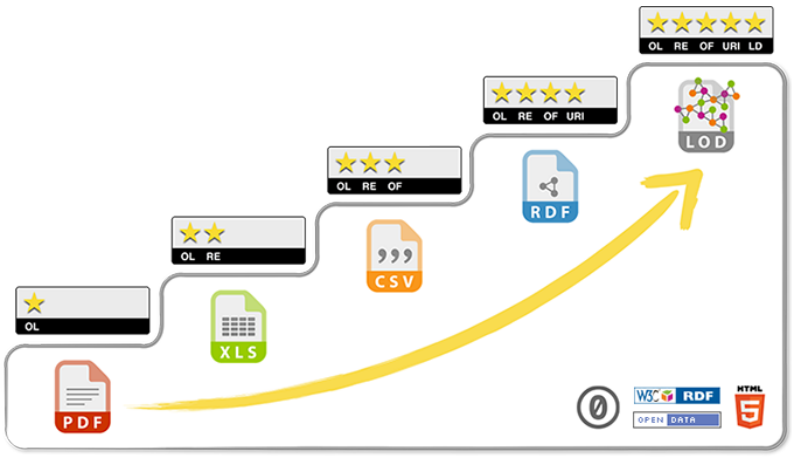


Figure 2.1: 5-star steps

2.2 OGD Portals

In order for government data to be opened, it needs to be made accessible to everyone. Before the introduction of OGD, governments already had communication channels to give information to the population. For example, municipal newsletter have a long history in this regard. Digital examples include social media pages and dedicated platforms to share the decisions taken by the municipal college [4]. However, these are not suited for structured datasets and a new solution was needed, under the form of dedicated online platforms called OGD portals, which allows giving the data back to the citizen or any interested stakeholder, for unrestricted use and redistribution (Kostovski et al., 2012; Kubler et al., 2016).

More concretely, an OGD portal is a website specially designed to present datasets from potentially different sources and present them to a user as a single point of access in a uniform way. [5] [6] [7] Following the context, the portal can offer multiple services such as personalization, dashboards, notifications, search engine, and access to the data in multiple ways, including via an API.

In Belgium, OGD portals are developed internally by governments or purchased from a solution provider, this last option being much more frequently opted for. An example of OGD portal developed internally is the one of the region of Brussels [8] shown in Figure 2.2.

[4] https://www.deliberations.be/
[5] https://en.wikipedia.org/wiki/Web_portal
[6] https://www.techopedia.com/definition/17352/web-portal
[7] https://www.liferay.com/fr/resources/l/web-portal
[8] https://datastore.brussels/web/

Figure 2.2: Welcome page - OGD custom portal of Bruxelles

On the solution provider side, OpenDataSoft has a large lead. Most OGD portals rely on their solution, such as the portal of the Wallonia-Brussels Federation, the portal of city of Liège and the portal of the city of Namur shown in Figure 2.3.

Figure 2.3: Welcome page - OGD portal of Liège

All OGD portals look similar, with a welcome page presenting the dataset categories and a search engine (Figure 3.2) and a datasets tab that shows all the published datasets and allow users to browse and filter datasets (Figure 2.3). Each dataset has a dedicated page that includes the following elements:

1. An information tab showing metadata such as the publisher, the number of downloads and the last update

2. A data tab showing the dataset in a user-friendly tabular form
3. An analysis and/or map tab that shows interactive visualizations of the dataset
4. An export tab that allows the user to download the data in the available formats
5. An API tab tab that documents an API through which the data can be accessed from a program

So, with features such as displaying data as a user-friendly table and visualizations, OGD portals do not limit themselves to giving access to data, but attempt to present it in a more understandable way for the end-user. Some OGD portals, including the ones proposed by OpenDataSoft, also propose additional visualization features to manipulate the data directly on the portal instead of having to download it. These features will be discussed more extensively in Section 2.4.

2.3 Challenges to OGD Use

However, despite all the efforts made by governments at all levels to publish reusable data and those made by solution providers to offer always more user-friendly and interactive portals, the use of OGD remains low in practice, and the benefits of OGD are not realized to the hoped extent (Zuiderwijk & de Reuver, 2021). Researchers have attempted to discover the reasons explaining this, and many articles have been published to document the challenges to OGD adoption.

This problem is usually addressed under two perspectives. First, the barriers that impede the publication of OGD by governments. Second, the barriers to the use of OGD. The barriers to OGD publication include for example insufficient resources and lack of uniform policy for publishing (Janssen et al., 2012). They are not discussed further in this thesis, as it is more focused on the user perspective.

The barriers to OGD use are multiple. Users can experience impediments such as lack of data or low-quality data (also for metadata) in terms of completeness and timeliness (van Stegeren & Theune, 2018; Warren & Champion, 2014), restrictive legislation, too much time needed to use the data, data behind a paywall, lack of technical skills to use the data, and the list continues (Crusoe et al., 2019; Janssen et al., 2012). Another is the absurdity of the result when you merge or transform the datasets content (Green et al., 2018), but also insensitive or offensive results (G. A. B. Barros et al., 2018). Combining datasets is very challenging and takes a lot of time, cost and effort (Benita et al., 2020). More technically, the endpoints of the provider may not be powerful enough to support large queries (G. A. Barros et al., 2016). Finally, there is the problem of dealing with proprietary technologies that do not provide open interface (Dave et al., 2018).

Even structuring the challenges to OGD use is a challenge in itself in the literature. Usually, authors structure the barriers by grouping them into broad categories. For

example, Zuiderwijk et al. (Zuiderwijk et al., 2012) identified more than 70 barriers to OGD use that they classified into *availability and access*, *find ability*, *usability*, *understand ability*, *quality*, *linking and combining data*, *comparability and compatibility*, and *metadata*. The barriers were identified by a mixed-methods approach, comibning a review of the literature and interviews and workshops with people who had previously worked with open data. For example, Figure 2.4 shows the usability barriers, identified based on literature, interviews and workshop. What is interesting with this mixed-methods approach is that literature can be less comprehensive that the empirical research, and comprehensive overview can lead to improving the open data process and make positive effects. For instance, a single shop for open data was an impediment few years ago, but today most countries have an open data portal. Its shows too that impediments change over time.

Impediment	L	I	W
3) Usability			
Concerns with source trustworthiness.	X		
Data might be incorrect or essential information is missing.	X		
Lack of (domain) knowledge about how to treat the data.	X	X	
Expert advice is needed to use the data.	X	X	
Unclear what new business models should be thought of for the use of open data.		X	
Lack of services given by the data provider to use raw data.		X	
No interoperability of open data infrastructures with other systems.	X		X
The data are not relevant/interesting.	X	X	
No explanation of the applied licences for open data.	X		X
Threat of lawsuits or other violations.	X		
Users are forced to employ various arbitrary data transformations to make data usable and comparable.	X		
Data require substantial human workload to clean them up for machine processing and to make them comprehensible.	X		
Fragmentation of software and applications.	X		
No standard software for processing open data	X		
No time to delve into the details, or no time at all.	X		
No incentives or no added value for users to make use of open data.	X		

Figure 2.4: Part of Table 3 (Zuiderwijk et al., 2012), overview of socio-technical date use impediments that infuence the open data process from the perspective of open data users (L = literature, I = interviews, W = workshops)

Another example is the study by Beno et al. (Beno et al., 2017). The authors identified 54 barriers by distributing a questionnaire among OGD users. The barriers were grouped into 8 categories: *Data quality*, *Legal constraints*, *Data portals*, *Knowledge and experience* for using open data and *Documentation and support*, *Legal constraints*, *Business and strategy*, *Privacy and security* for opening open data. 110 participants from Austria completed the survey, mostly OGD users, but providers participated too. As example, Figure 2.5 shows that the lack of information about the quality and the content of a dataset were rated as biggest barriers. These barriers can discourage users from putting extra effort into evaluating the dataset and determining if it is useful.

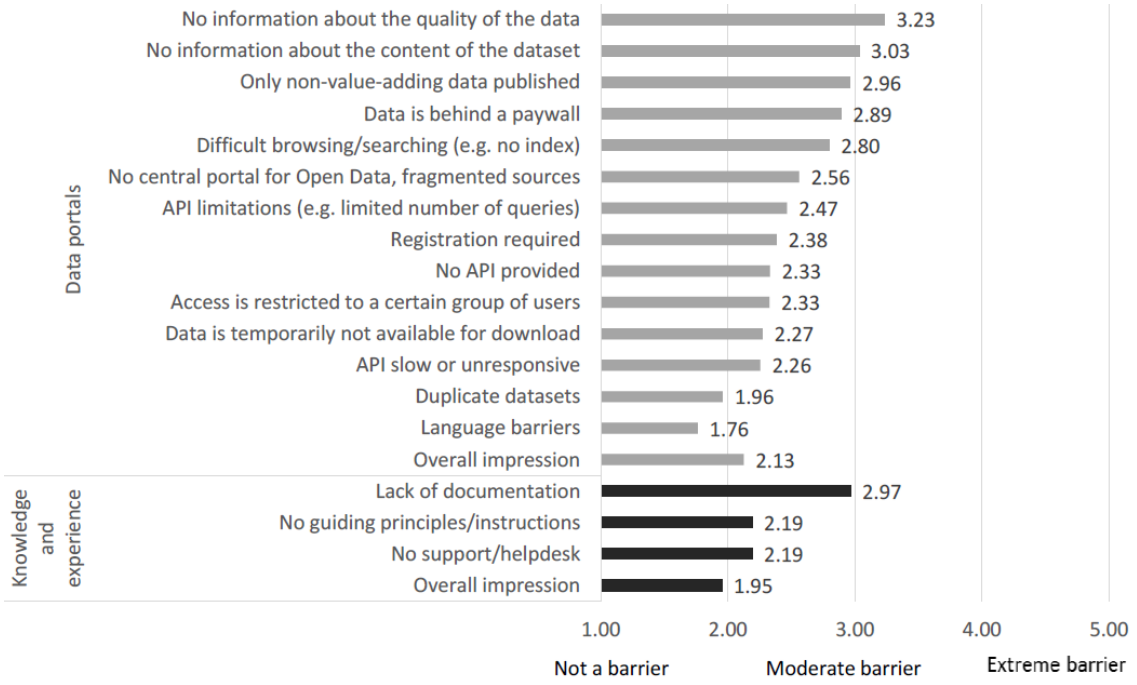


Figure 2.5: Part of Fig. 2 (Beno et al., 2017), barriers to using Open Data

Through investigations with the users on a OGD portal (Beno et al., 2017; Crusoe et al., 2019; Zuiderwijk & de Reuver, 2021), we can go further in understanding users impediments, the barriers they met doing some tasks.

Following Figure 2.6 shows the severity of the barriers using the median (see the red bars for the most severe ones) and, in an other hand, in the second graph (Figure 2.5) we can see an other kind of list of barriers (also the way used in (Zuiderwijk & de Reuver, 2021)).

Crusoe and Ahlin attempt in their articles to define a workflow on how users experiment with OGD through activities and structured a process into phases: start, identify, acquire, enrich, and deploy (Crusoe & Ahlin, 2019). The start varies relating to the intended use of the OGD. In the identify phase, the user is exploring the accessible data to decide if the data are relevant. In the acquire phase, the user is preparing for the delivery of the data from the publisher and receiving it. In the enrich phase, the user is concocting and making something. In the final deploy phase, the user has a product or service that can be provided to end-users. Theses phases can be a way to regroup and organise barriers to OGD use. They also reformulated them into a four-phase process : **start or motivation**, **search and evaluate**, **access and prepare**, **aggregate and transform** (Crusoe et al., 2019). In (Crusoe et al., 2019), the authors have also collected barriers to OGD use via questionnaires and interviews with students who had used OGD to develop an application in the context of a class project. They identified 28 barriers and structured them according to the four phases. This way of structuring barriers is different from the other studies mentioned because it shows at which moment of their use the users will face the barrier. Also, the severity of the barriers was collected from the students and

seven moderate barriers which, according to the article, means that it was “difficult to use the data”. Following Figure 2.6 shows the severity of the barriers using the median (see the red bars for the most severe ones) and, in an other hand.

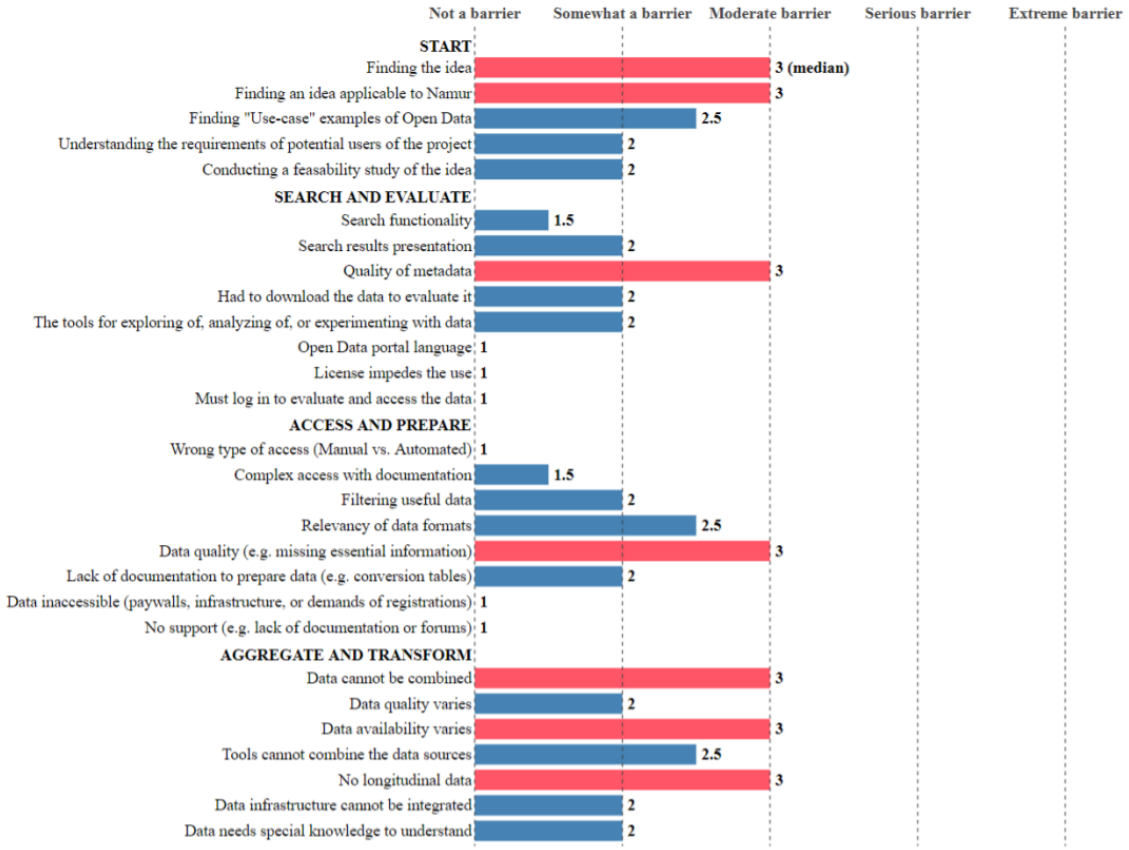


Figure 2.6: Severity of the barriers met by users (Crusoe et al., 2019)

We already listed a lot of impediments reprised in the list above, but we can notice that search and understand the result can be a problem for the user, mostly if you add metadata issues. They did not know how what to do with and how evaluate them.

Another points we can mark is the access of the data, with documentation or not, and the ease to add some filters and get relevant data. It can ask to the user some additionnal knowledges about the data, the context or the field of statistics.

Again, data manipulation can be a challenge, aggregate them can be tricky and ask to the user a lot of efforts. Data quality and availability can varies, which may lead to not seriously considering this dataset, it cans sound awesome but useless.

There are in fact two big categories of barriers. On the one hand, there are barriers regarding the data, that can be solved only on the publisher side. This includes for example the lack of resources and the completeness of the data. On the other hand, once the data is available, there are barriers that can be addressed on the data provision side, such as the difficulty to find the published data and the difficulty

to exploit the published datasets. OGD portals come as a solution to both. On the publisher side, it gives a convenient interface to publish data online. More importantly, on the user side, it provides a user-friendly access point. As said, the barriers related to data quality and availability depend on what publishers choose to (and have to the time to) make available, and cannot be solved by a third-party software development. Therefore, this thesis focuses on the barriers related to the data provision.

The range of features proposed by OGD portals to solve these barriers is wide. For example, it includes searching by keywords and themes, visualization of data as tables, documented APIs. However, adding a portal in the equation can cause other impediments related to its discoverability and usability, such as finding the portal, and understanding how it works (interface, navigation).

As an answer to the lack of technical skills needed to exploit data, some portals offer a well welcome feature: the visualization of the data through charts and maps, directly in the browser. Some portals also allow combining datasets through visualization by creating maps where each layer corresponds to a dataset for example. In the next section, we give an overview of data visualization techniques before going deeper into the visualization features of OGD portals.

2.4 Visualization of OGD

As Tableau says in its article[9] citing The Economist[10], visualization is an increasingly key tool to make sense of the trillions of rows of data generated every day. Data visualization helps to tell stories by curating data into a form easier to understand, highlighting the trends and outliers. A good visualization tells a story, removing the noise from data and highlighting the useful information. The data and the visuals need to work together, and there is an art to combining great analysis with great storytelling. More importantly, visualization reduces the complexity of understanding data, simplifies it and makes it more accessible to users.

[9] https://www.tableau.com/learn/articles/data-visualization

[10] https://www.economist.com/leaders/2017/05/06/the-worlds-most-valuable-resource-is-no-longer-oil-but-data

Figure 2.7: Selection of visualization techniques (Boost Labs, 2021)

Modern browser or desktop and mobile applications allow a lot of new ways to display data to the end-user (Figure 2.7), more rich and detailed than before (Figure 2.8), and by nature portable and shareable. Visualization is not limited to virtual representations, but can exist in the real world, like cards and boardgames (Friberger & Togelius, 2012), and some mix both to add interactivity (Wolff et al., 2017).

Figure 2.8: Lendlease's Open Building Systems Integration (OBSI) platform (Harkins & Heard, 2020)

Figure 2.9: Approaching photoreal interactive virtual environment (Harkins & Heard, 2020)

Through open services and dedicated applications we can use open data for any purpose, like seeing the changes in the environment (Krooks et al., 2014), map navigation like OpenStreetMap or OpenLayers, helper for smart city emergence and

innovations decision-making (Figure 2.9) (Ojo et al., 2015; Smith, 2017), imagination is the limit.

2.4.1 Visualization Techniques

Charts

Some of the most common types of data visualization chart and graph formats include the bar chart, the line chart, the pie chart, the bubble chart (Figure 2.10), and all serve to speed up and improve data readability and interpretation, but not all are appropriate for the same job. For example, line charts are suited to represent trends in time while bar charts are not. Choosing the right visualization technique is the key to prevent user confusion and making sure their analysis is accurate. (Boost Labs, 2021)

Figure 2.10: Chart types

To choose a good one, developers can refer to a visualization catalogue like the datavizcatalogue [11] that guides with a presentation, examples and the functions of each visualization technique.It also provides links to tools that help implementing these visualizations. Another catalogue is From Data to Viz[12] that presents a decision tree based on the type of the data to visualize to propose the best options.

Maps

A lot of data, and especially OGD since they are related to a territory, are geographic in nature and contain either explicit or implicit spatial information (Graham & Shelton, 2013). Furthermore, more and more people rely on freely available user-generated spatial content, known as Volunteered Geographic Information (VGI)

[11] https://datavizcatalogue.com/
[12] https://www.data-to-viz.com/

and Linked Open Data techniques have a role in promoting such online and freely accessible spatial information (Karam & Melchiori, 2013).

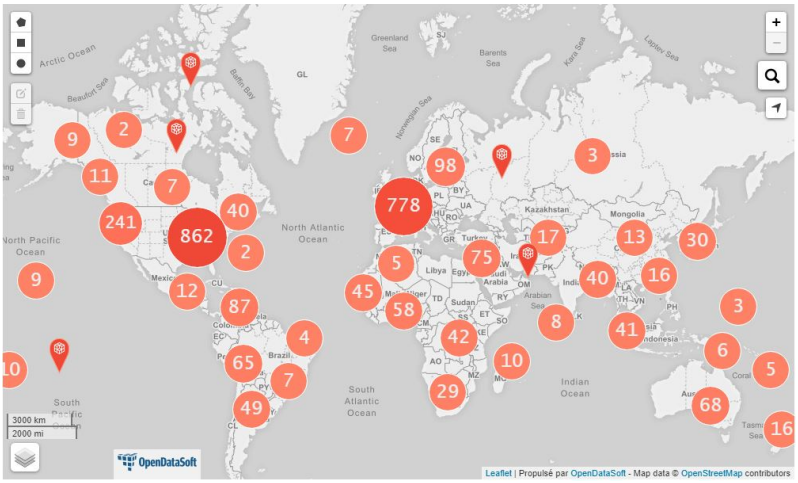


Figure 2.11: OpenDataSoft map sample

Maps can be used to display data about a specific location, a street, or an area with with markers (Figure 2.12). The markers can be represented by a specific icon or by a shape on which information is encoded, for example by the color or the type of shape. There are many different types of map, depending on which entity (point, line, or area) is represented (Unwin, 1981).

Figure 2.12: Map showing information about roads

2D Models

Maps are used to visualize data on a larger area like a district or a street. At a more micro level, 2D models can be useful to display specific data to the user, like room temperature on a floorplan (Figure 2.13) or display on-place alerts on the plan if sensors are connected and provide information.

Figure 2.13: Floorplan showing the temperature in each room (Dave et al., 2018)

3D Models

We have already cited smart buildings (Figure 2.8) and 3D models of them, sometimes on a map (Figure 2.14), but 3D models are also used in simulation rendering (Figure 2.9). 2D rendering context can also use pre-rendered 3D models.

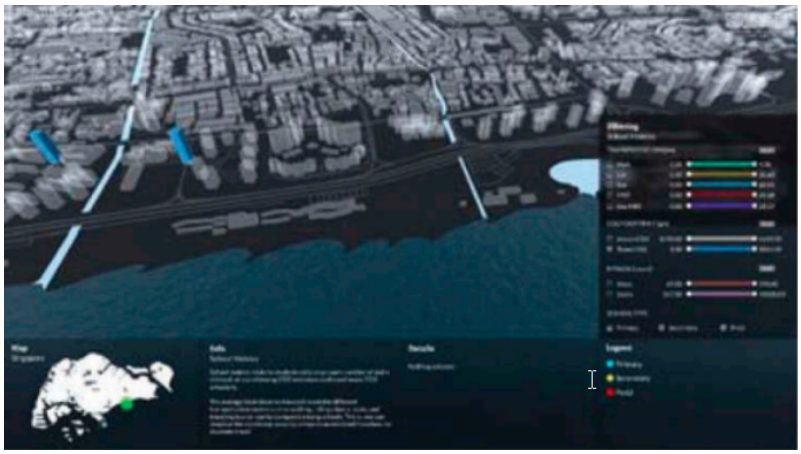


Figure 2.14: 3D model of Singapore (Benita et al., 2020)

Augmented Reality

Using an equipment like Google glasses or through a smartphone photo-like mode with an application layer, users can enjoy reality with additional visual data. In the project (Natephra & Motamedi, 2019) the user can view on his smartphone some environmental data thanks to the coordinates of a detected marker (Figure 2.15).

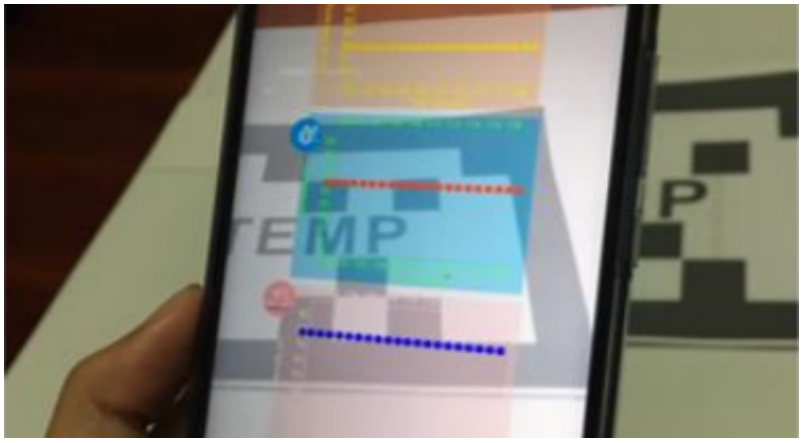
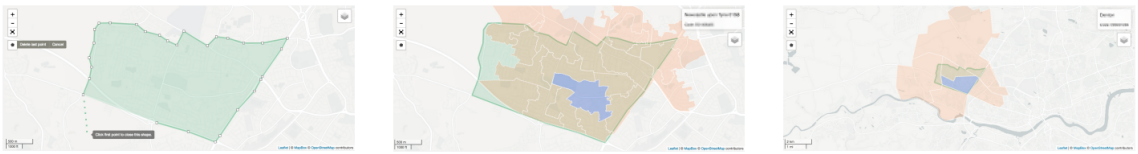


Figure 2.15: Data visualzation on smartphone (Natephra & Motamedi, 2019)

2.4.2 Applications to OGD and OGD Portals

All the techniques mentioned are useful to make data easier to understand, and this applies to OGD too (Eberhardt & Silveira, 2018). Indeed, we have seen in Section 2.3 that some users lack technical skills to use OGD. Graves and Hendler write "for some people, the availability of these datasets is not enough to make use of it. Lack of technical expertise (e.g., programming skills and knowledge on data management) forbids an important proportion of the population to consume such data" (Graves & Hendler, 2013). They proposed to use visualization to help the users lacking this expertise.

Also, Puussaar et al. explain that "while the open data movement now permits citizens to acquire governmental data relating to their communities, little to no effort is made to ensure that these datasets are accessible and interpretable by non-professionals" (Puussaar et al., 2018). So, the authors created a platform that allows citizens to explore OGD by creating their own visualizations (Figure 2.16). They conclude that OGD needs to be nicely presented and contextualized to be used by citizens. Another example is BarcelonaNow, which allows citizens to create their own dashboard (Figure 2.17) using the OGD datasets and the visualization techniques they want (Marras et al., 2018).



(a) Drawing a boundary to define a query

(b) Retrieving intersecting area polygons

(c) zooming out to retrieve new level area polygons

Figure 2.16: Map based queries (Puussaar et al., 2018)

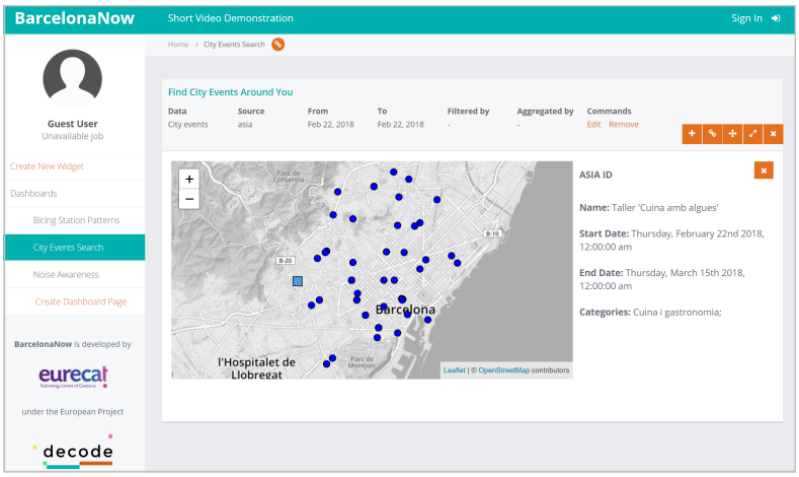


Figure 2.17: Sample dashboard overview (Marras et al., 2018)

So, the literature agrees that visualization is a key to make OGD more usable by the population. This is implemented in practice, visualizations and tools to create them are available on more and more OGD portals. The visualizations are usually shown to present datasets like we will see in the next chapter (Figure 3.10), and visualizations can be created using dedicated analysis tools (Figure 3.11). However, (Marras et al., 2018) note that providing predifined use case without personnalization is not sufficient, and (Graves & Hendler, 2013) found that users find it important to be able to edit visualizations.

The problem is that even if we know that tools to create and edit visualizations are important to have on OGD portals, we do not know what form they should have. The requirements of citizens for tools to create and edit OGD visualizations have not been formally identified. Therefore, this is the goal of this thesis, and the research question is: **What are citizens' requirements toward visualization features on Open Government Data portals?**

3 Requirements Analysis

3.1 Methodology

The goal of this thesis is to understand citizens' requirements about visualization features on OGD portals, and to implement them into an improved prototype. In particular, the focus is put on the usability of these features. Therefore, we opted for a user-centered approach as described in the UX Book written by Hartson and Pyla (Hartson & Pyla, 2012). These authors represent the development of a user-centered system as a process in four steps called the UX wheel (Figure 3.1). First, in the analysis step, the requirements of the users toward the system to be developed are captured. Second, the design step consists in proposing a solution design that fulfill the users' requirements. Then comes the prototyping step where the design is implemented into a prototype that can be evaluated with users in the fourth step, evaluation. It is important to note that this is an iterative process. After the evaluation stage, it is likely that new requirements will emerge or that some identified requirements are not addressed satisfactorily by the proposed design. This should be taken into account to improve the design and validate the improvements with another evaluation.

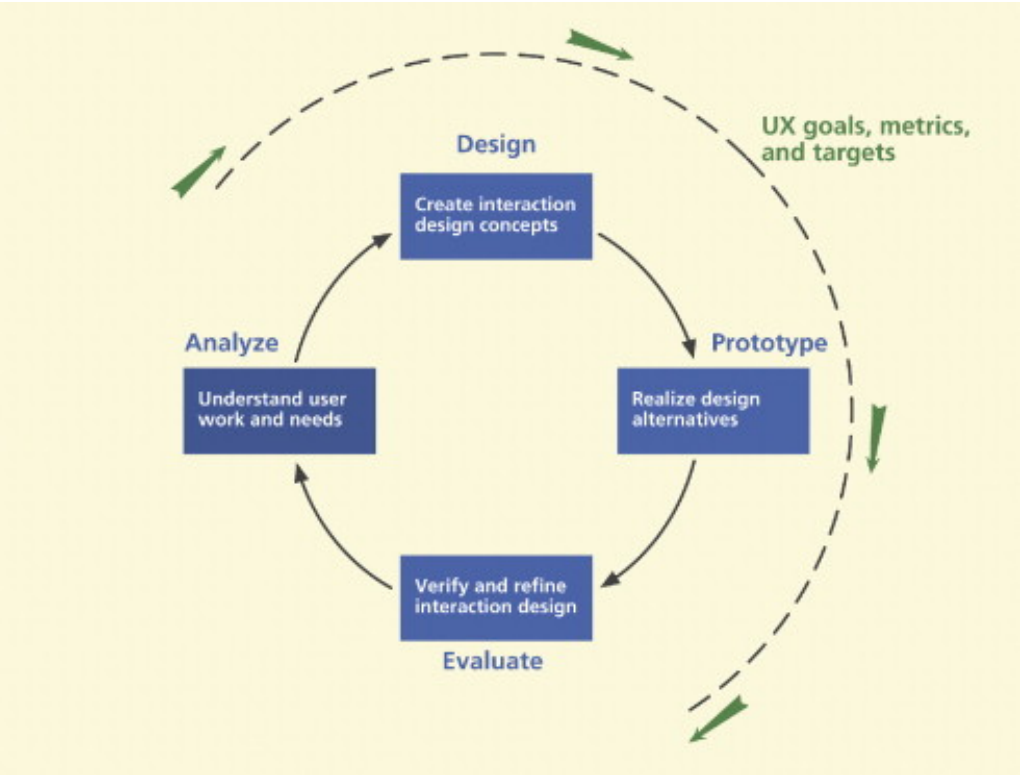


Figure 3.1: UX Wheel process described in (Hartson & Pyla, 2012)

Since there is already a well-established OGD portal that proposes visualization features, we shifted the starting point of the UX wheel. We started by conducting an evaluation of an existing OGD portal that we chose as application case. From there, we could identify requirements and design suggestions that were implemented in a new solution design.

3.1.1 Application Case: The OpenDataSoft Portal of Namur

The first step was to select an Application Case, or in other words an existing OGD portal proposing visualization features on which we could conduct a usability evaluation. Many different OGD portals exist and could have been used as application case, but we opted for an OpenDataSoft (a private provider of OGD portals) portal due to its position on the market, widely used across Europe but also in North America, Mexico, Australia, and the Middle East[1] [2] [3] [4]. For reasons of local relevance, we chose the OGD portal of the city of Namur, which relies on the OpenDataSoft solution.

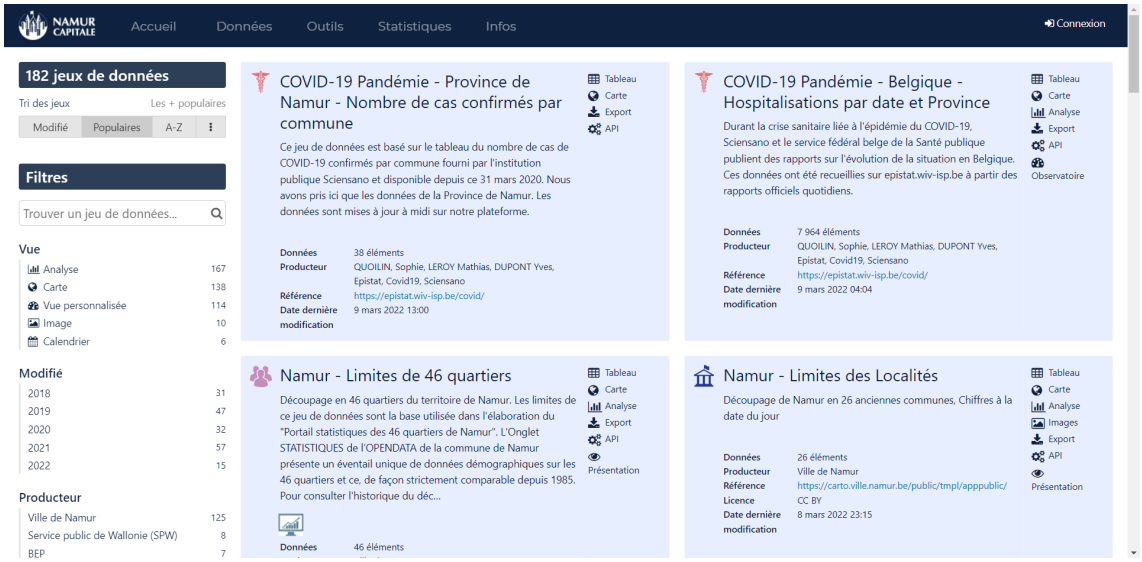


Figure 3.2: Opendatasoft, Namur version, datasets page

Once on a dataset, you can find various tabs like said previously in 2.2 and you start on the table that shows you the dataset and the cross-sectional filters (Figure 3.3). This page is our start point for the sequel of our investigation.

[1] https://enlyft.com/tech/products/opendatasoft

[2] https://tech.eu/2022/01/11/with-a-fresh-25-million-opendatasoft-keeps-the-data-open-plans-to-open-new-markets

[3] https://www.lemagit.fr/actualites/252512042/Opendatasoft-reaffirme-ses-ambitions-a-linternational

[4] https://en.wikipedia.org/wiki/OpenDataSoft

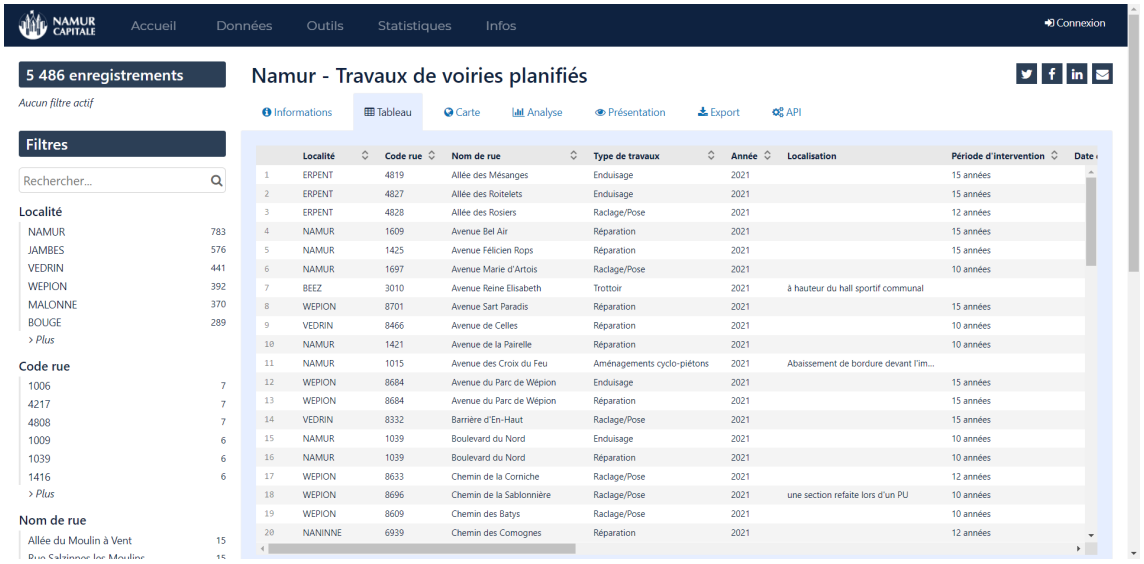


Figure 3.3: Opendatasoft, Namur version, dataset table tab

3.1.2 Data Collection

User Test

Data on the usability of the OGD portal of Namur and on its visualization features were collected with user testing and expert assessment. The user test involved 5 participants, and two experts commented on the OGD portal

A user test was constructed in the form of 2 scenarios driving the participant on the portal to consult and understand datasets, and to be able to answer some open and closed questions afterward.

The idea during the user test is to put the participant in a concrete context with a situation, where he can feel the needs, the objectives and can manipulate the interface himself, thinking aloud and pointing the mouse cursor where his eyes looks. Then, we can prompt him with questions on what he does and why, and what he does not do and why, observing all along his struggle with the portal interface.

To avoid scattered data and issues out of context, we drive the participant on a specific dataset page as a starting point for each scenario. In the questions we target metadata and data throught the available tabs on the dataset page, like information, table view or visualization in different ways and with the help of the available filters aside.

With the consent of the participants, each session of the user test was fully recorded to capture the screen (mouse pointer) and the voice (think-aloud).

Scenario 1 - Opening a new store The first scenario puts the participant in the role of a shopkeeper who wants to choose the best location in Namur to open a

new store. To do that we propose the Namur pedestrian counts dataset[5] that lists 5 streets and we ask to choose between 2 of them: "Rue de l'Ange" or "Avenue de la Gare", but we added a very important event that happens on December 20 at 1:00 p.m.. Therefore, the affluence at this time had to be carefully analyzed by the participants

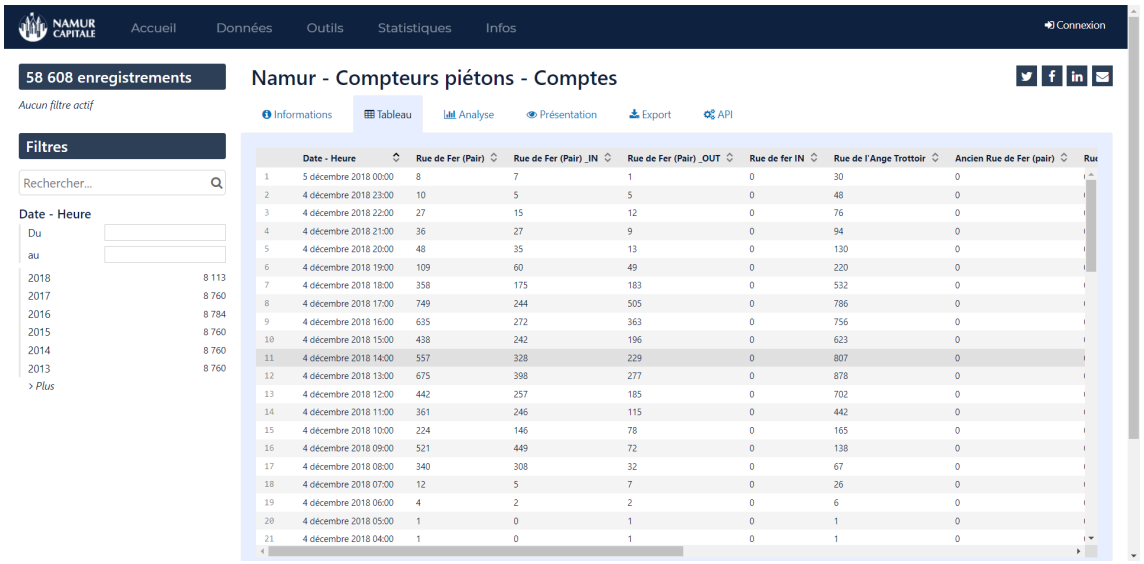


Figure 3.4: Pedestrian counts dataset (Scenario 1) - Data table

This dataset is composed of an information (Figure 3.5) and presentation tabs that regroup metadata, descriptions and some extracts as pictures of the dataset. Data can be read via the table view (Figure 3.4) and the analysis tab (Figure 3.6) that proposes an editable graph mode. Aside you can filter by date (between mode), grouped by year and month.

[5]https://data.namur.be/explore/dataset/namur-compteurs-pietons-comptes/table/?sort=date

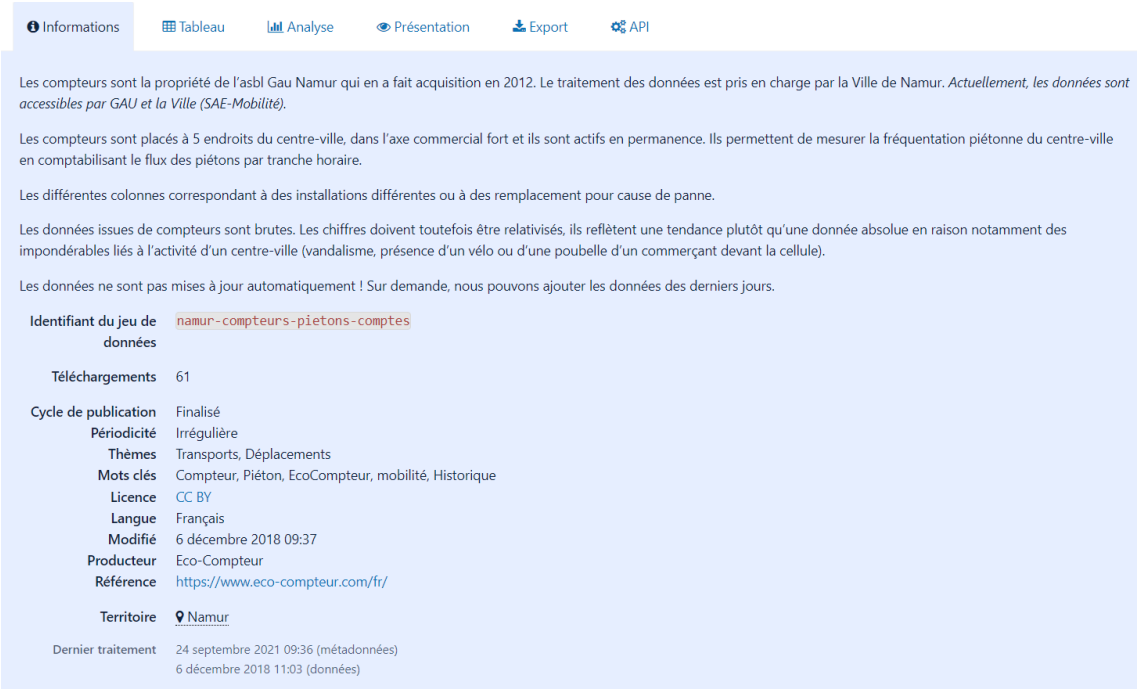


Figure 3.5: Pedestrian counts dataset (Scenario 1) - Information tab

This dataset has as attributes a timestamp field and all the streets that were used to count.

A complex element is the fact that the street's data is not well collected and displayed. Indeed, you can found multiple data column with the same name or with a small name variation, some being clearly understadable such as those mentioning the direction of the pedestrians, others being more ambiguous. From example, "Rue de l'Ange" is described by columns having names like "rue de l'Ange (trottoir)", "rue de l'Ange", "rue de l'Ange rue", "rue de l'Ange twice as nice" or "Rue de l Ange Simple sens (trottoir)". That adds a direct difficulty to the participants.

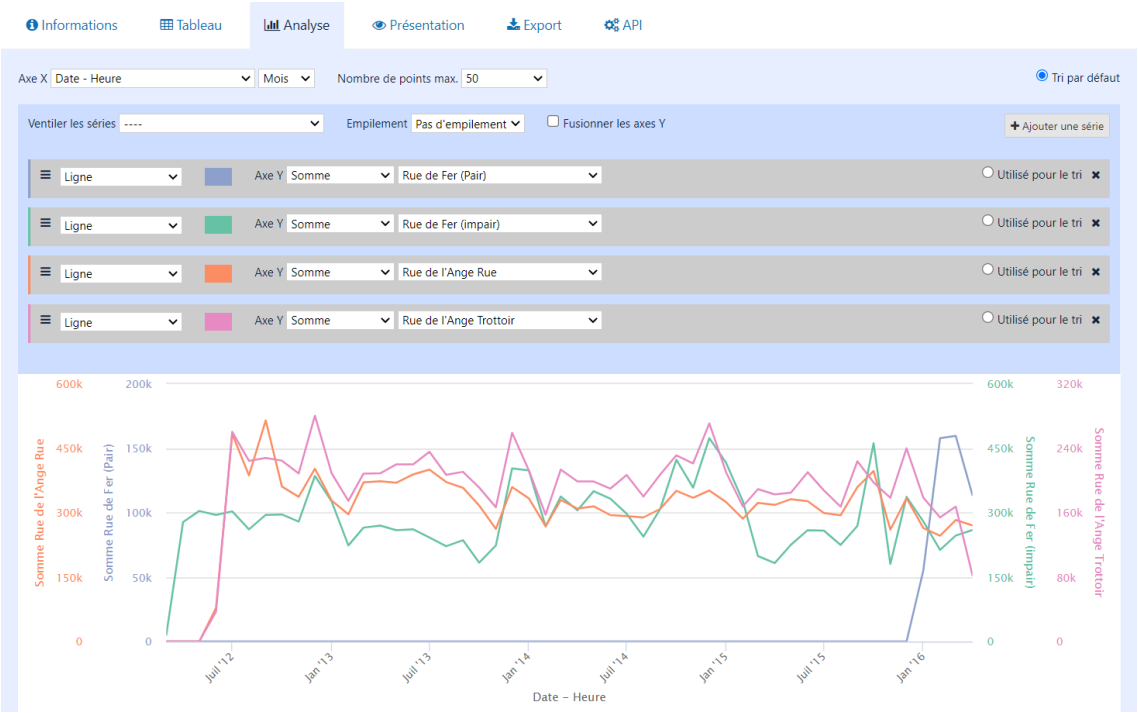


Figure 3.6: Pedestrian counts dataset (Scenario 1) - Analysis tab

Scenario 2 - Selling solar panels The second scenario transports the participant in the Namur districts as a solar panel seller who has to choose 3 districts for his tour tomorrow. Again we give to the participant a dataset, the photovoltaic potential by district of Namur[6] that lists all Namur districts with relatives data.

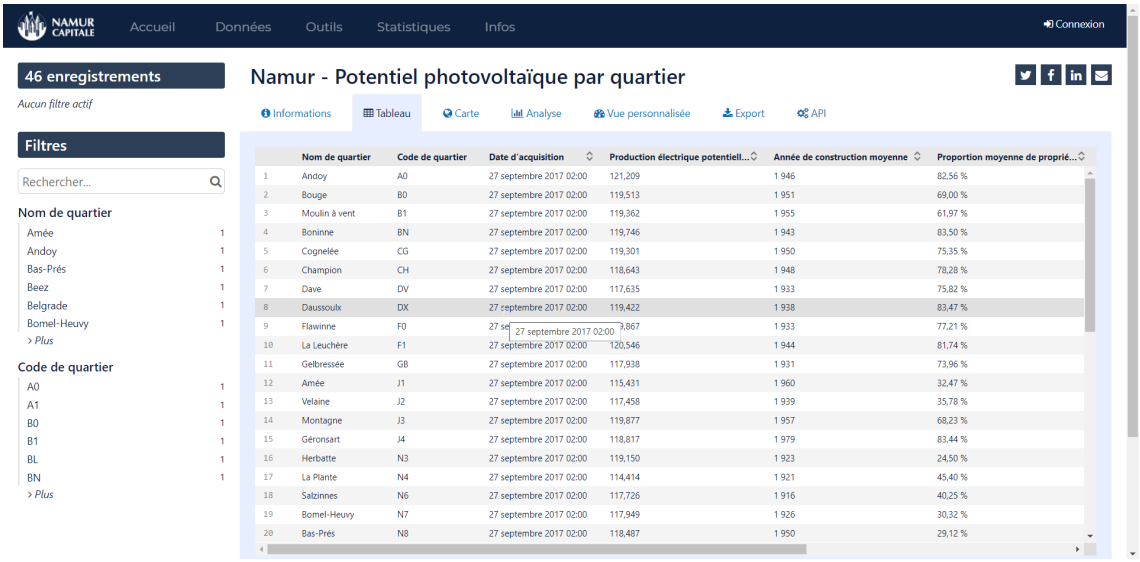


Figure 3.7: Photovoltaic potential by district dataset (Scenario 2) - Data table

This dataset has as attributes a district name and code, a timestamp, the average potential electricity production per square meter of roof, the average year of con-

[6]https://data.namur.be/explore/dataset/namur-potentiel-photovoltaique-par-quartier/table/

struction, the average proportion of owners, the average income and 1 coordinate to locate the district and 1 list of points to delimitate the district.

This time the dataset is provided with one information tab page (Figure 3.8) with the metadata and a custom view tab with graphs and other summarized data (Figure 3.10); the table view is still present (Figure 3.7) like the analysis tab (Figure 3.11) and a map tab is added. Aside you can filter by district code and name.

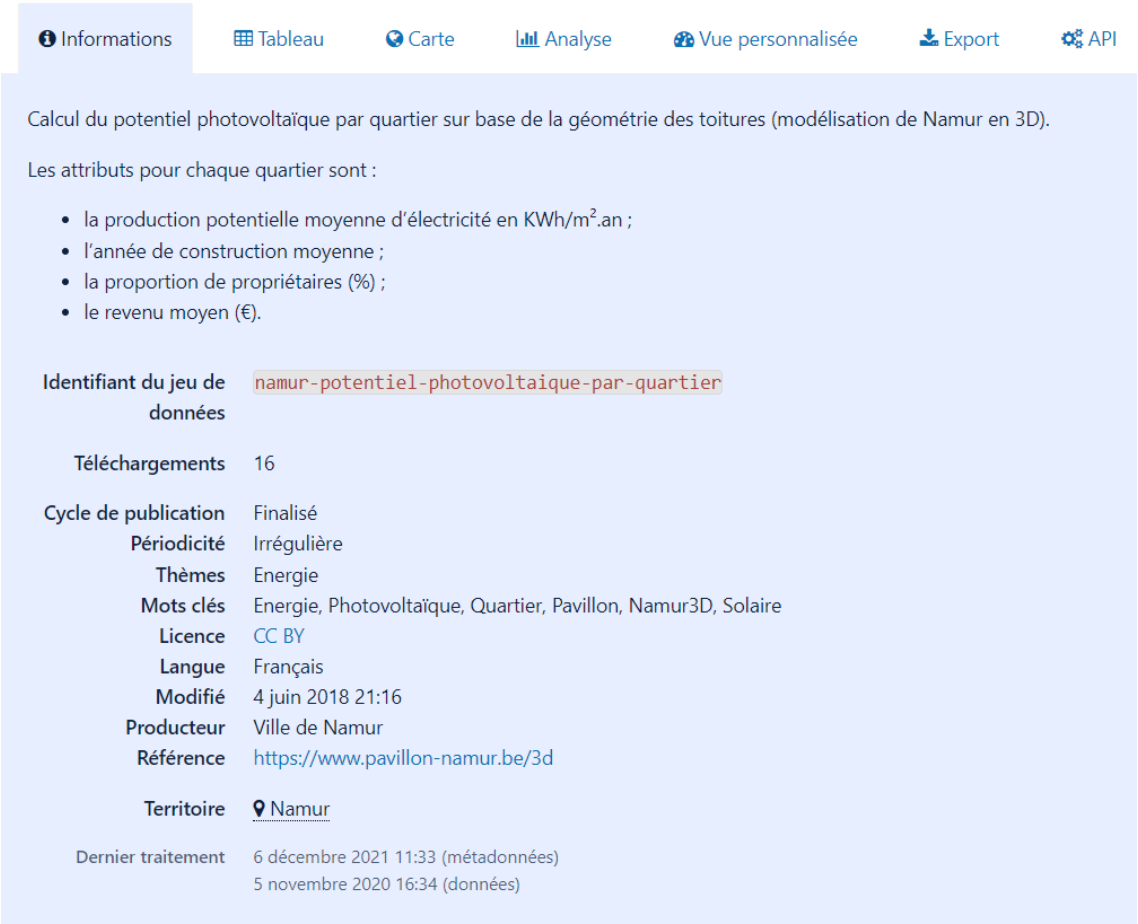


Figure 3.8: Photovoltaic potential by district dataset (Scenario 2) - Information tab

On the map tab (Figure 3.9) we can note that each district is delimited and displayed as a red shape, not representing the information on photovoltaic potential. But on the custom view tab, the map displays this information like a choropleth map and there is a legend that explains how to read it.

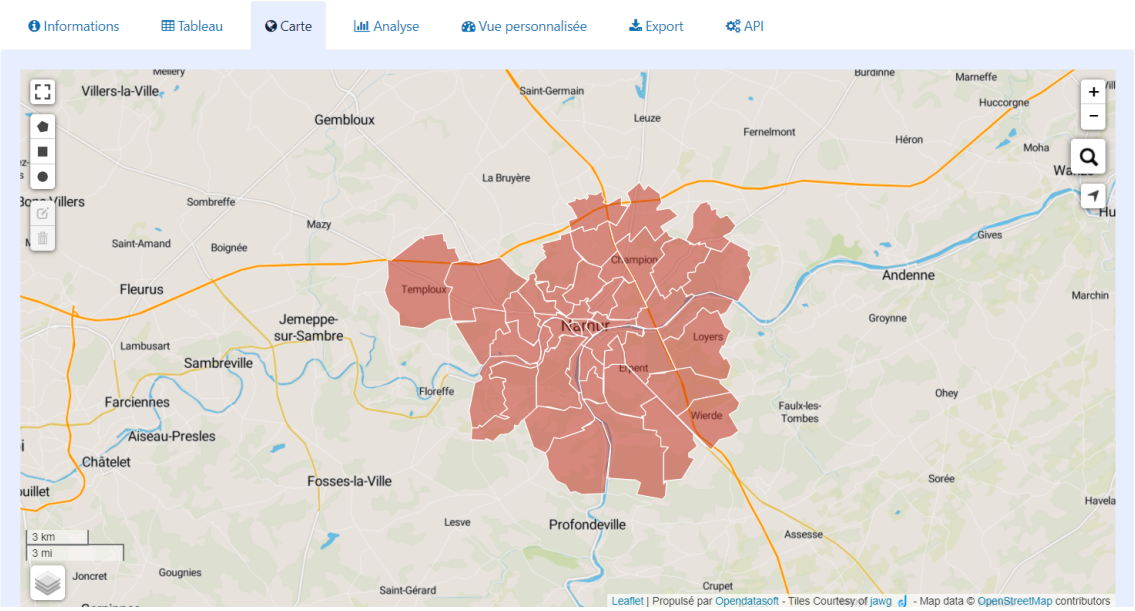


Figure 3.9: Photovoltaic potential by district dataset (Scenario 2) - Map tab

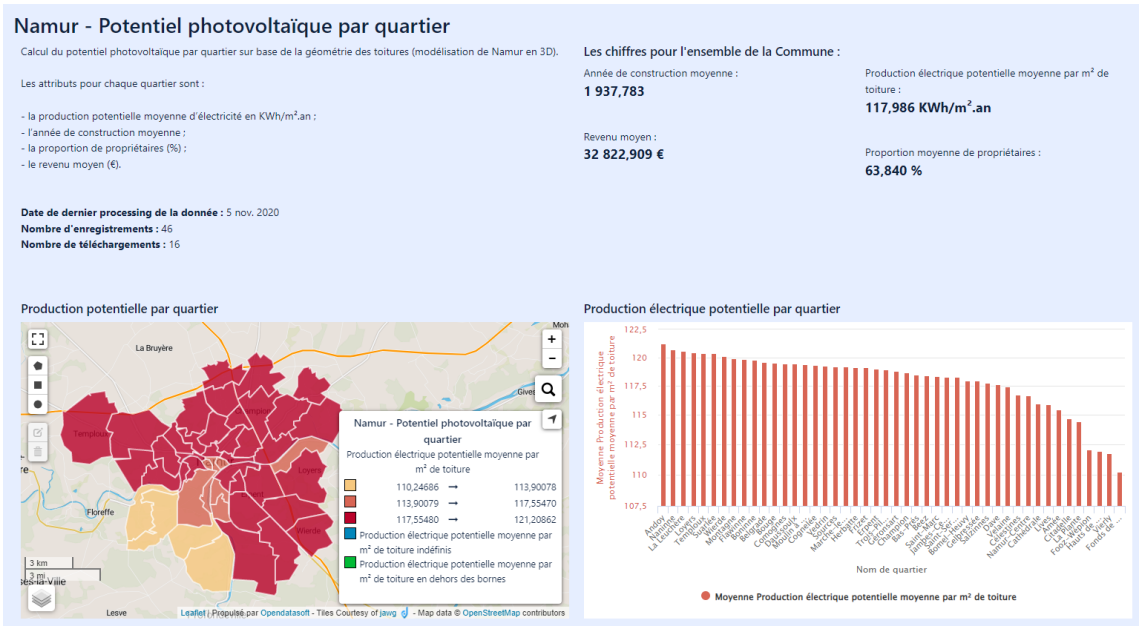


Figure 3.10: Photovoltaic potential by district dataset (Scenario 2) - Custom view tab

After the participant has explored the dataset, we add a step by asking the to cross another dataset with this one in order to have a more valuable information to make the seller's decision more accurate. The dataset in question contains information about the population per district, allowing the seller to mix the photovoltaic potential with the potential number of clients to make his decision. This step asks the particpant to use the interface, search a particular tool available the portal and manipulate both datasets to achieve the goal. For these reasons we ask them how they want to achieve this part, throught an external way or via the existing solution on the portal itself.

Figure 3.11: Analysis tab of the scenario 2

Questionnaire After letting the participant manipulate the datasets and propose a solution, we ask participants to fill a questionnaire. The questionnaire starts by introducing the purpose of the study. Then, it asks contains socio-demographic questions on their profile, questions related to the two scenarios, and retrospective questions on the perceived ease of use and usefulness of the visualization features. At the end of the questionnaire, the participant is thanked and receives karma points to compensate for future evil deeds. The full questionnaire is in Appendix A. Answers were processed anonymously.

The socio-demographic part consists of general questions for statistics review, like gender (or no answer), age range, highest degree obtained and professional situation, plus a self-appreciation of technological affinities measured on a 5-point Likert scale (Likert, 1932) from very low to very high, like common use of technologies or data analysis.

Then, for both scenarios, questions about metadata and data are asked, depending on the available tabs (for example, the presentation tab is only available for Scenario 1, and the custom view tab only for Scenario 2). For metadata we ask a summary of what the data is about, the last update, temporal coverage and how the information tab matches what the participant expect. For scenario 1 we ask how many streets are covered and same for scenario 2 with districts. About information tab, due to the open-ended nature of the question, we prepared some stimulus question, like "do you see superfluous elements?", "which element is the most helpful?" or "what is missing?". For Scenario 1, due to the presentation tab, we ask what is the added-value of this tab. About data, following the available tabs, we ask if and why/why not the participant has consulted them. Then, we ask whether the tab is helpful, enough to get the scenario answer, easy to understand, and depending the case, easy to use.

Finally, with the retrospective questions, we want to assess, on a 5-point Likert scale, from totally disagree to totally agree, the ease of use and usefulness of the visualization features of the portal. We reused and adapted to our specific case the validated 8-item scale of the TAM (Technology Acceptance Model (Davis, 1993)):

1. Using the visualization features on the portal will improve my work
2. Using the visualization features on the portal will improve my efficiency
3. Using the visualization features on the portal will increase my productivity
4. I find the visualization features on the portal to be a useful tool in my work
5. My interaction with the visualization features on the portal is clear and understandable
6. I find it easy to use the visualization features on the portal to do what I want
7. Interacting with the visualization features on the portal does not require a lot of mental effort
8. I find the visualization features on the portal easy to use

Expert Evaluation

In order to have a complementary view of the interface, we asked 2 experts to participate in an additional assessment, in addition to the user test. There are two main ways of conducting an expert evaluation of an interface (Lallemand & Gronier, 2015). The first one is the heuristic evaluation, which consists in checking an interface against a predefined list of usability criteria. The second one is the cognitive inspection, which consists in running a user scenario on an interface to predict where actual users might experience usability difficulties. The experts were asked to evaluate the OGD portal of the city of Namur, just like the participants of the user test, and more particularly the parts involved in Scenario 2 about the "photovoltaic potential by district" dataset. Each expert evaluated the interface in a different way: the first with an heuristic evaluation and the second conducted a cognitive inspection.

3.2 User Test Results

Dear reader, before reading this section, you are welcome to go to the OGD portal, try the scenarios, and fill the questionnaire in parallel and compare your results with those reported here (do not forget to send your questionnaire ☺).

3.2.1 Profile of the Participants

A total of 5 people participated in the user test (3 men and 2 women) with their consent to record audio, screen capture and use the data collected during the user test. This number of people is often enough to discover 80% of the most important problems (Nielsen & Landuer, 1993)[7].

The age of the participants is between 36 and 55, with an average of 42. All are graduated except one who has a high school degree. And all are employed except one who is self-employed.

As the figure 3.12 shows, participants have mostly a common usage of technologies, average or low competencies in data analysis and a large variability of the development skills.

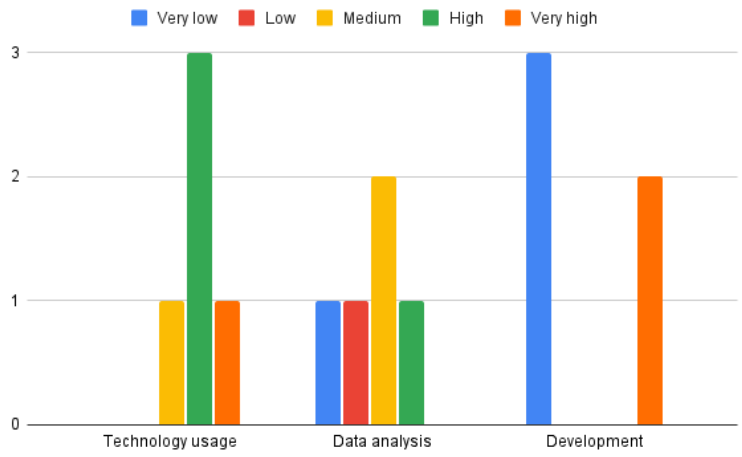


Figure 3.12: Technological affinities

3.2.2 Scenario 1

Resolution of the Scenario

Both the streets was selected as the good option (Figure 3.13). One participant hesitated with “rue de l’Ange (trottoir)” then changed to “avenue de la Gare” and another said both, depending on the shop you want to set up.

[7]https://www.nngroup.com/articles/why-you-only-need-to-test-with-5-users/

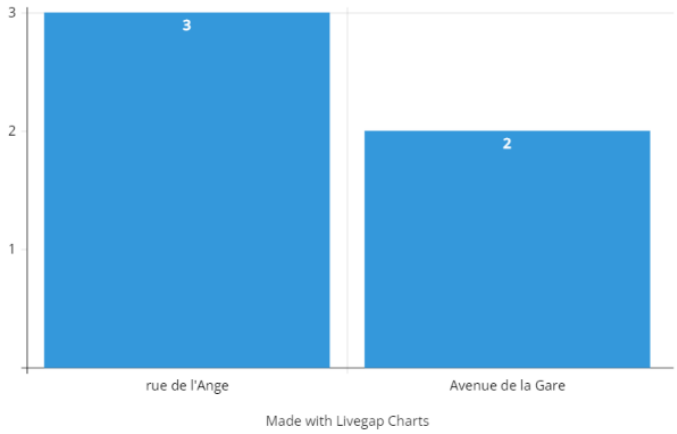


Figure 3.13: Scenario 1 - Preferred street to open the new store

Metadata

Almost all the participants have understood and answered correctly to the questions about the metadata: what the dataset is about, the last update, the streets covered and temporal coverage.

We can just note that 3 of them encounted problems with update and coverage question but, finally, 2 founds and answered correctly and 1 still in confusion on the time coverage and was guided to solution.

We can note too that about the number of streets covered, the known problem of multiple columns for a same street has an impact here. The solutions proposed by the participants are varied and include between 3 and 15 streets.

About the question on the information tab, almost all participants agreed that the page corresponds to what they expect, only one had no expectation but is ok with the page content, and another considered that it is not explicit enough.

Stimulus questions gave a lot of reactions, 3.2.2 shows a list of each item with the number of participants who mentioned it.

Stimulus	**Item**	#
Superfluous elements	Downloads count	2
	Information text (§#1)	1
	Information text (§#5)	1
	Release cycle	1
	Producer	1
	Keywords	1
Most useful elements	Update	3
	Data source	1
	Information text (§#3)	1
	Information text (§#4)	1
	Language	1
	Territory	1
Missing elements	Street names more direct	1
	Street map	1
	Street data (columns)	1
	Pedestrian profile	1

Table 3.1: Reactions to stimulus questions on the information tab, for the dataset of Scenario 1

To refer a specific paragraph of the Information tab text we use the syntax §x, where x is the paragraph number. About "Street names more direct" in *missing elements*, the participant wants to have a direct list of street names without duplication due to count failure or meter change.

One participant considered the "data model" and "reuses" parts at the bottom as useless because they are "empty", but he did not realize that the content is folded and that he had to click to display it.

We can note that some information are considered irrelevant by the participants, like the download count. We can thus divide the metadata elements into useful and superfluous.

For the missing elements, it is heavily linked to the dataset quality and to some improvements that can give more sense and be helpful to the end-user.

About the presentation page, with the same questions, opinions diverged with a tendency to indicate that this page is not useful, does not bring additional useful content, or is redundant (Table 3.2.2). However, one participant considered it is interesting and it gives an idea of how the dataset was created.

Stimulus	Item	#
Superfluous elements	none	0
Most useful elements	2 graphs	2
	Data reliability	1
	How data are collected	1
Missing elements	Animation of how it is working	1
	How to redo the sample graphs	1
	Street map	1
	Sensor map	1

Table 3.2: Reactions to stimulus questions on the presentation tab, for the dataset of Scenario 1

This page is often seen as a copy/paste of the information tab and only brings graphs as new content, but for one participant, once he read it more in-depth, he found it complementary, for the others it brings nothing new. For one partcipant, it can be merged with the information tab.

Data

On the question of knowing which street is the most frequented, the opinion is divided (Figure 3.14). Each answer has a different reason: "rue de l'Ange" has more peaks but "avenue de la Gare" is more widely frequented, or both are the same and the participant picked one arbitrarily.

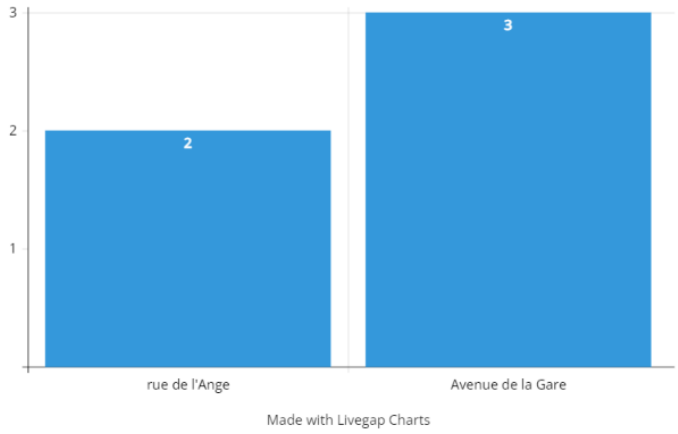


Figure 3.14: Most frequented street according to the participants

Next we asked the event question, which consisted in finding the frequentation for the 20th of December on 13:00, and again answers vary: 1 participant said that "Avenue de la Gare" is twice as more frequented than "Rue de l'Ange", 2 expressed

that “Rue de l’Ange” is more frequented than “Avenue de la Gare” by more than 200 pedestrians, 1 gave another number but found that “Rue de l’Ange” is more frequented, and the last one did not follow the instructions and gave no answer.

To obtain the answers the participants used mostly the table tab, except one who used the presentation tab graphics. Three of them think that this tab is sufficient to find the solution, and one did not understand he could change tab in this exercice. All found the table view is helpful, even if it is not convenient for them, or not understandable enough for 3 of them.

Three participants spontaneously tried to visit the analysis tab, the remaining 2 did so after guidance. Most think this tab can be sufficient to find the answer, One participant did not understand the view and 1 did not feel alright with this. Most participants (4/5) think that this view is helpful, but only 2 of them think its easy to understand and need efforts to manipulate it.

Regarding the exercise, we asked them if they think they have enough data to choose where to set up their store. All said no because they lack data: geolocation, street map, profile of pedestrians and where they come from and by which route (transport), type of business to set up, competition in the street, ease of access to streets, and of course updated meter data. In addition, they believe they have to cross-reference other datasets such as parking, accessibility, urban works, location, frequentation of this type of store and the frequentation of competitors. This confirms the relevance of including dataset combination in the scenarios of the user test. Almost all (4/5) think that a map can be helpful for that.

3.2.3 Scenario 2

Resolution of the Scenario

Different sets of districts was selected (Table 3.3), and for 3 participants, the choice changed while browsing the different tabs.

District	**A**	**B**	**C**	**D**	**E**	**#**
Temploux	x			x		2
Naninne	x					1
Andoy	x		x		x	3
Loyers		x	x		x	3
Suarlée		x		x	x	3
Frizet		x				1
La Leuchère			x			1
Wierde				x		1

Table 3.3: Districts selected by the participants

Metadata

Almost all the participants have understood and answered correctly the questions about the metadata: what the dataset is about, the last update, the number of districts and temporal coverage.

One participant confused last process date and last update, 1 did not find the temporal coverage and was guided; and 4 used the table rows to get the number of districts covered because they did not find it in the information tab, 1 did not found and was guided.

About the question on the information tab, all participants agreed that the page corresponds to what they expect. We can note that Scenario 1 changed their way of reading and understanding this page, and thus the answers. Indeed, it is not the first time that participants encounter an information tab on the portal, unlike during Scenario 1.

Again, stimulus questions gave a lot of reactions, some identified the same elements as in Scenario 1. Table 3.4 shows the list of each item with the number of participants who mentioned them.

Stimulus	Item	#
Superfluous elements	Downloads count	1
	Data model	1
	Release cycle	1
	Producer	1
	Keywords	1
Most useful elements	Update	1
	Attributes	2
	Information text (§#1)	1
	District code	1
	Number of districts	1
Missing elements	Peaks and evolution	1
	More temporal coverage	2
	More attributes	1
	Data source	1
	Data collection method	1
	More information	1

Table 3.4: Reactions to stimulus questions on the information tab, for the dataset of Scenario 2

One participant had the same remark as in Scenario 1 about the "data model" and folding system, which was found unclear and disrupting.

Data

This time, almost all participants consulted all the tabs to answer the question the best way they can. Each had affinities with one or more tabs and used them to produce their answer with the filter on the left for some participants.

About the table tab, 3 participants think it is sufficient and helpful, others use it as a complement. This kind of presentation is interesting for at least 2 participants and 1 used it to get one specific piece of information. On the question about ease of understanding, opinions are divided. Three participants encountered problems with navigation and understanding of the interface. This comes in addition to the concerns about the available data and the tools offered.

The analysis tab had more success here, but not unanimously. The visual effect is attractive but requires some effort to use and understand the data displayed. In

addition, the interface looks complex and one participant even hesitated to touch the "series" part or to change the visualization chosen. Three participants found it easy to use but 2 are not comfortable with it.

The map tab disappointed everyone, as it brings more questions than answers, seems buggy. The participants expressed their opinion in a very direct way: "it is useless". This suggests a bad choice of attribute of the dataset or no choice at all.

The "custom view" tab gave a first impression of redundancy of the information tab, by the attributes among other things, but the graphs it contains attracted attention and forced a deeper reading. However, the opinion remains mixed and, in the end, even if the participants considered it useful, it is not sufficient to answer the question. Two of the participants think that this tab lacks explanations, the other half considers that it is easy to understand.

Datasets Combination

The additional exercise of merging two dataset proved to be a big interface challenge, regardless of the participant's profile. Some started by searching for the dataset they were looking for, others by the merge solution, and 3 of the participants had to be guided to the advanced editing mode in the analysis tab, but once this step was done, all found the button to add another dataset. Note that it was not globally obvious that it should be in this tab. At this point, everyone had the same problem: finding the right dataset with keywords and filters.

It can be noted that they often think of clicking on the "integrate" button (which allows to integrate the code of the visualization in another page) at the bottom of the analysis tab page instead of that of the advanced editing mode. The size and style of the buttons are different and "editing in advanced mode" is much more discreet, like a sublevel link.

When asked to understand how they plan to merge the datasets, 2 participants thought of doing it manually or with an external tool. But generally everyone thought of trying to do it through the portal quite directly.

The opinion on this exercice is negative due to the interface complexity, not intuitive, hard to understand. The participants do not know how to cross-reference information.

3.2.4 Perceived Usefulness and Ease of Use

To conclude the questionnaire with a retrospective, we asked the participants their opinion on the visualization functionalities and their interaction with them.

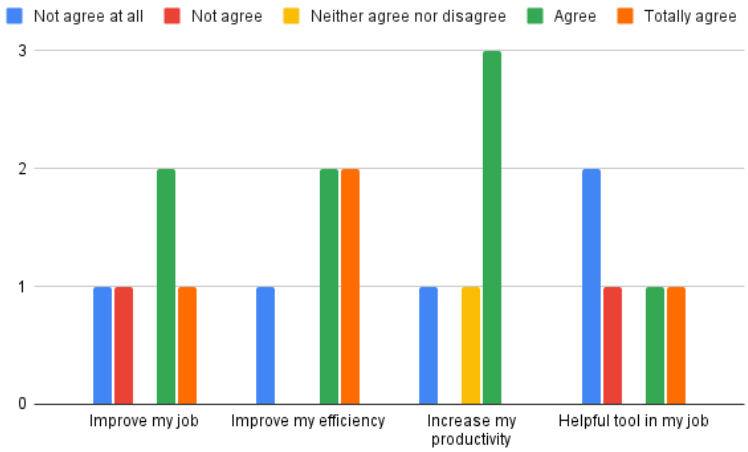


Figure 3.15: Perceived usefulness of the visualization features

Figure 3.15 shows that most participants think that visualization features can help somehow their job and some think it can improve their efficiency and productivity but disagree about how helpful the tool can be.

The differences in results between questions 1 and 4 compared to questions 2 and 3 can probably be explained by the understanding of the "job" word, where participants maybe see it as their profession instead of the job of using the interface to solve the scenario.

The perceived ease of use of visualization features (Figure 3.16) is balanced with a more positive trend.

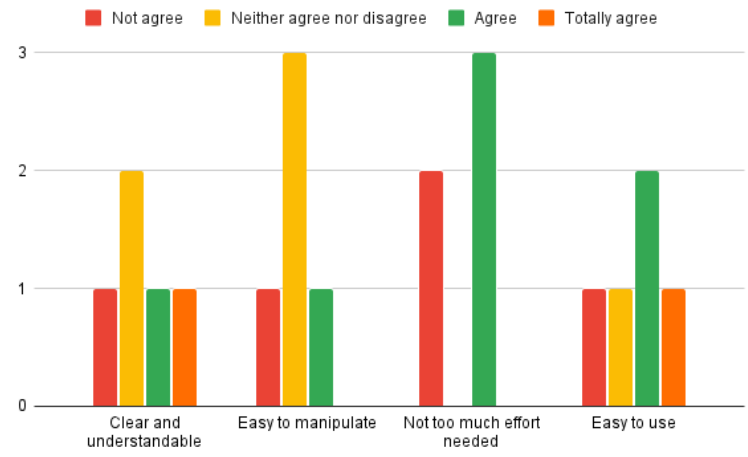


Figure 3.16: Perceived ease of use of the visualization features

Additionally, we asked participants what they wanted to improve on the portal to make scenarios easier to achieve. Three of them want more help in the interfaces and a way to have dynamic samples that they can edit, like images in the "presentation" or "custom view" tabs, and be redirected in the tab "analysis" with the predefined

configuration of the sample. One participant wants to have more “ready to use” scenario, to fastly answer popular questions for each datasets by example. The table tab, the filters, the map tab and the general ergonomics (UX and UI) are also important points and must be reviewed.

3.2.5 Problems List

All problems with the interface reported by at least 1 participant were listed. They are grouped below, by category or by tab.

Table Tab

The table tab that displays the data is too simplistic. Users cannot set a row fixed to help them reading all width, nor reduce the number of columns by selecting them following their interests, nor sort data in a multi-columns way (Figure 3.17).

	Date - Heure	Rue de Fer (Pair)	Rue de Fer (Pair) _IN	Rue de Fer (Pair) _OUT	Rue de fer IN	Rue de l'Ange Trottoir	Ancien Rue de Fer (pair)	Rue de ferc_OUT	Rue de Fer (impair)	Rue de l'Ange Rue
1	5 décembre 2018 00:00	8	7	1	0	30	0	0	15	57
2	4 décembre 2018 23:00	10	5	5	0	48	0	0	41	125
3	4 décembre 2018 22:00	27	15	12	0	76	0	0	93	195
4	4 décembre 2018 21:00	36	27	9	0	94	0	0	105	322
5	4 décembre 2018 20:00	48	35	13	0	130	0	0	127	348
6	4 décembre 2018 19:00	109	60	49	0	220	0	0	373	492
7	4 décembre 2018 18:00	358	175	183	0	532	0	0	807	664
8	4 décembre 2018 17:00	749	244	505	0	786	0	0	1 192	752

Figure 3.17: The table only provides single-column filtering

Map

Like all the tabs, the map is sensitive to current filters, but users cannot choose which attribute of the dataset is displayed on it (Figure 3.18).

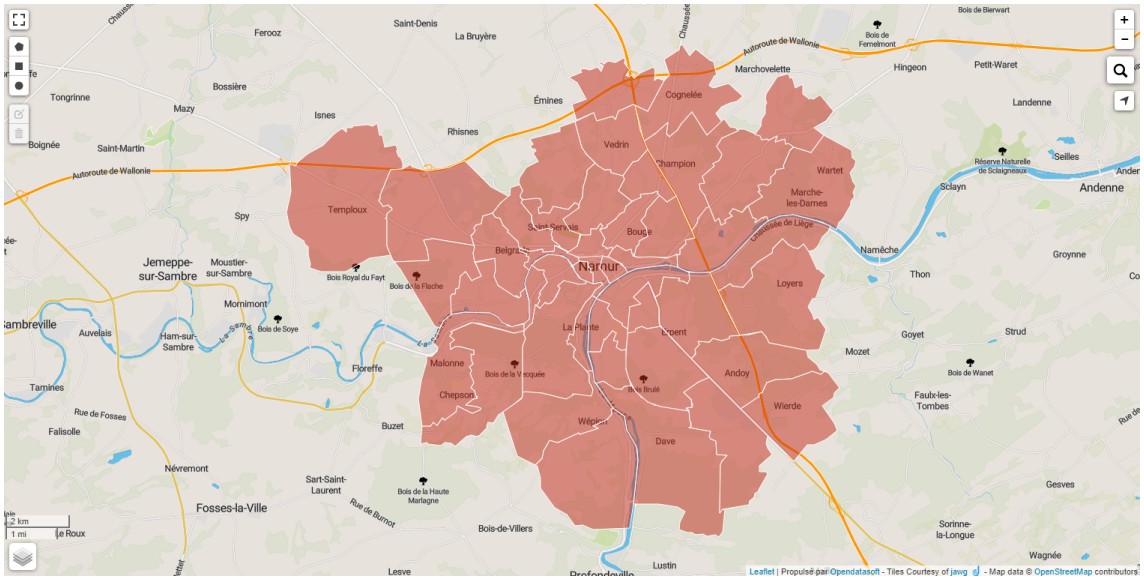


Figure 3.18: Map without legend and no way to select the attribute to highlight

Multi-layers

When users add a series, they can select a visualization, but some cannot be combined correctly and become a source of misunderstanding for the user. For example, a bar graph and a pie or a map and a radar. This happened when the participants tried to find the solution to Scenario 2, due to a lack of knowledge about the datasets, they tried different combinations to see how the data can react and maybe give a clear solution (Figure 3.19).

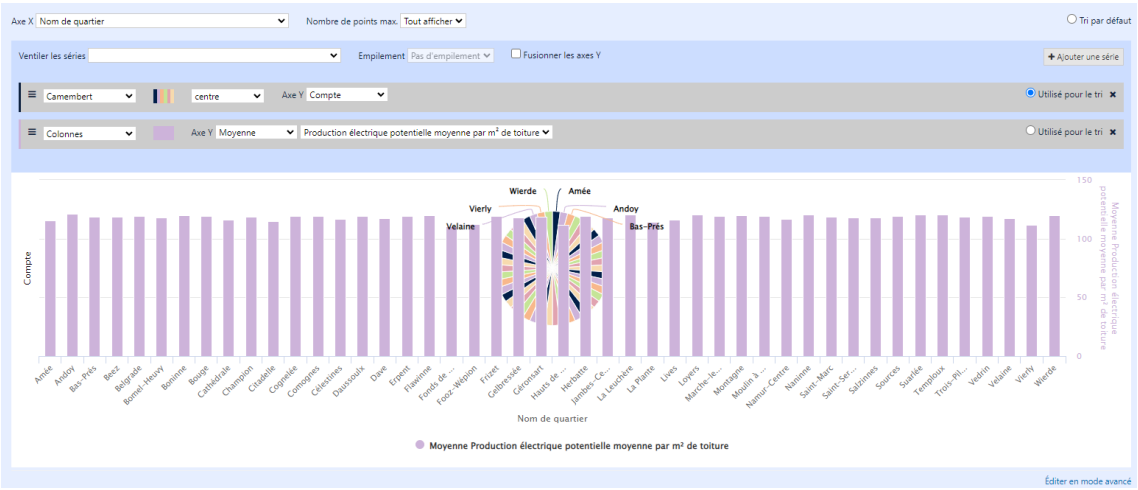


Figure 3.19: Multi-layers overlapping with some combinations

Filters

The filtering system is reduced to only a few fields and it is hard to understand how it works (Figure 3.20). For example, a participant tried to reset the filters and had no clue how to do it. Also, a participant tried to type “ange” to select only the relevant columns.

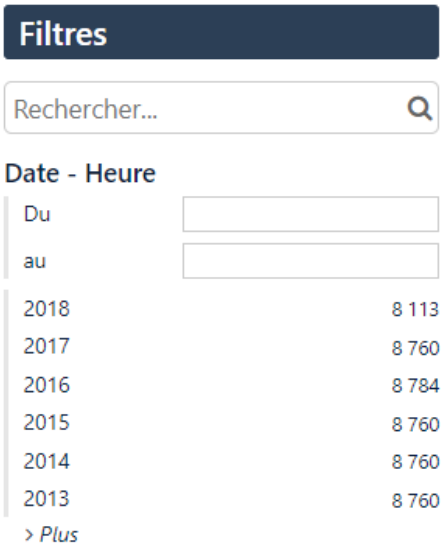


Figure 3.20: Filters area with almost no options

The global search field is not live reactive, it is waiting for the user to press ENTER to reload list with result, if there are any.

Datetime Filter

In the case of the pedestrian counter we got a datetime filter, but users can only choose a date. Despite this misleading name, the date filter is just a range filter, and users cannot select transversely a month or a specific day. In our case, it would be useful to manipulate in a more advanced way the date and the time.

Information

For the participants, metadata are a mix of general data and more useful data and it should be devided to be more clear. On an other hand, a custom tab like "presentation" (Figure 3.21) can be merged with the information tab (Figure 3.22) to regroup same kind of metadata or free content like data collection explanations.

Figure 3.21: Information tab of Scenario 2

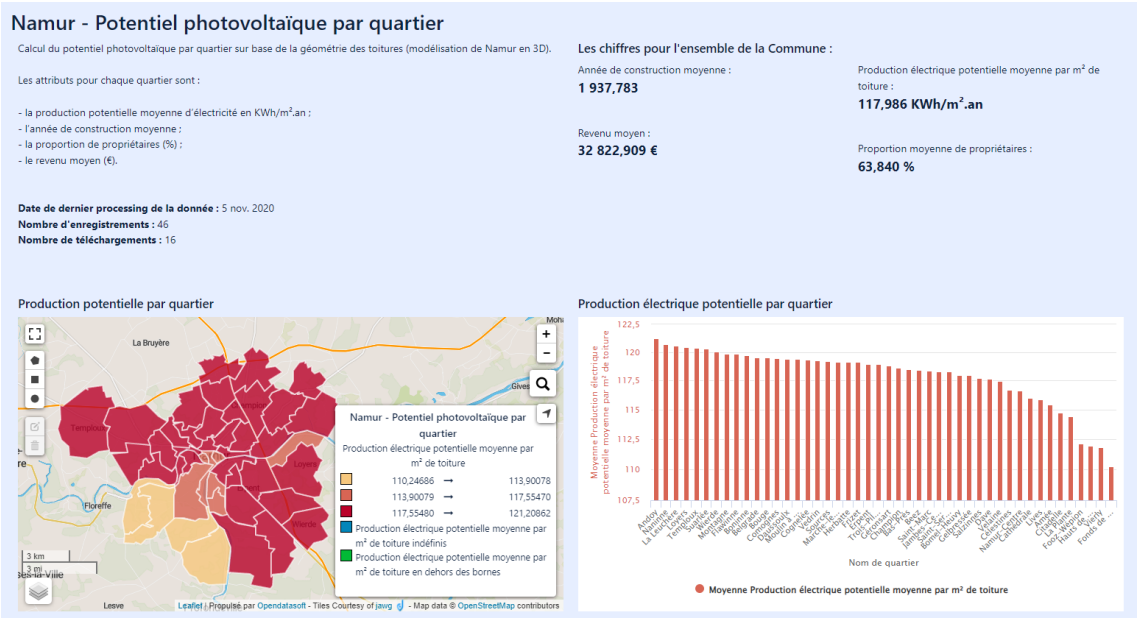


Figure 3.22: Custom view tab of Scenario 2

Buttons and Action Zones

Like the advanced edit mode button, in the dataset combination part of Scenario 2, the look is not adapted to the purpose as a call for action and tricks the user into ignoring it. Moreover, the "date-time" label used in filters or tabs in the external usage area of the analysis tab (at the bottom) are not clear enough for the users and can lead to losing or misdirecting them (Figure 3.23).

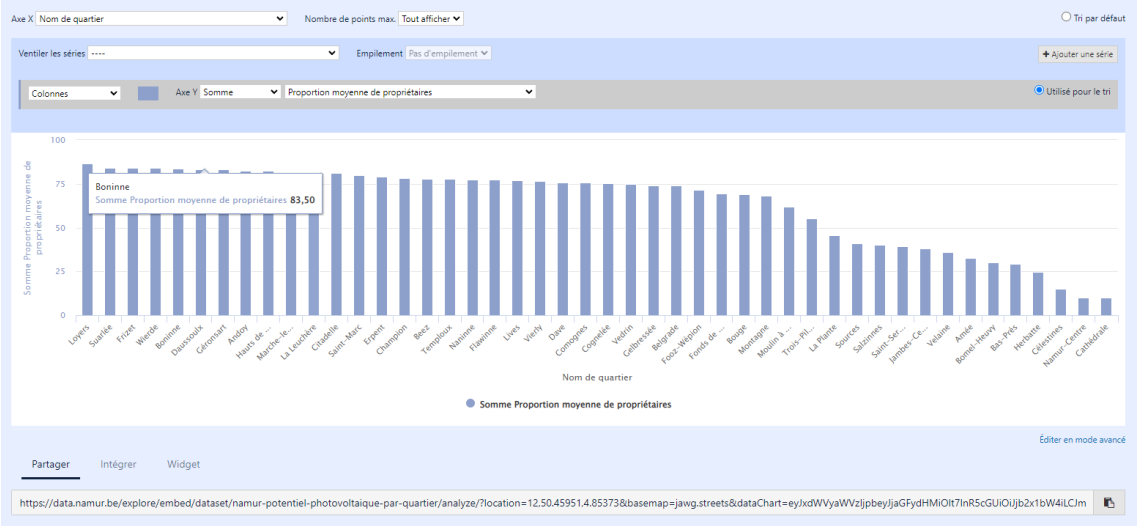


Figure 3.23: UI navigation problem on the analysis tab

Help and Tooltip

Due to lack of understanding, participants searched for contextual help or tooltips but could not find them. Like for map, a legend can be useful.

Navigation

Some participants point the problem to come back to a previous page when they navigate the portal (Figure 3.24).

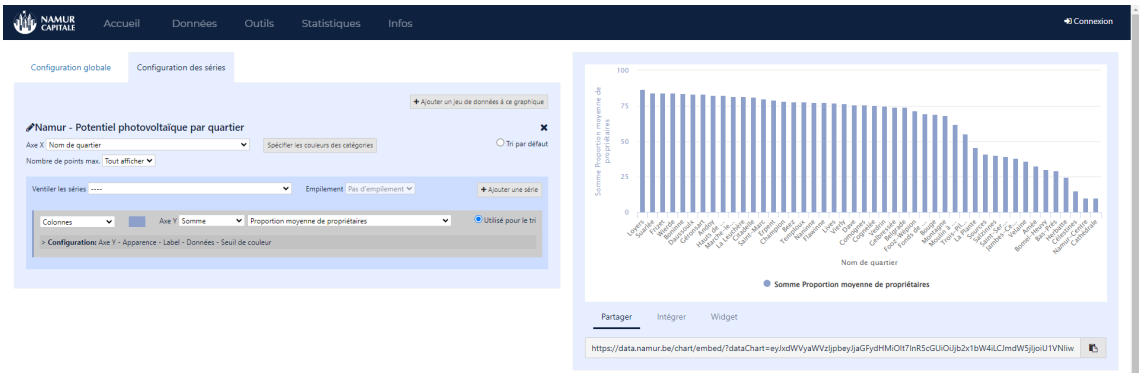


Figure 3.24: There is no way to come back to the analysis tab (except using the built-in browser return button)

Severity of the Problems

Some of the identified problems where a stronger barrier than others. Globally participants were annoyed by these problems, making the experience difficult but

they were not blocked, except few of them on some specific points. One was blocked completely with the table reading and using and prefered skipping to the analysis tab, 2 were blocked on the series usage in the analysis tab and needed some help, and 1 was blocked on the navigation trying to come back to the dataset after going to the dataset merge tool. It took time and effort but they tried, often without the expected success. Table 3.5 shows the severity of the usability problems. Some were a slight annoyance but not really a barrier, others cause discomfort, and others blocked the user in their usage of the portal, and different participants experienced the problems differently, some for example being only a discomfort for a part of the participants and blocking for other. Table 3.5 also shows how the problems were perceived by the participants. For example, the multi-layer was a discomfort for 3 participants and blocking for 2.

Barrier	**Not a barrier**	**Discomfort**	**Blocking**
Table	0	4	1
Map	0	5	0
Multi-layers	0	3	2
Filters	1	4	0
Information	4	1	0
Button & action zones	1	4	0
Help and tooltip	0	5	0
Navigation	2	2	1

Table 3.5: Severity of barriers during the user test

3.3 Expert Assessment

3.3.1 Heuristic Evaluation

First, a professional UI/UX designer did an assessment of the interface (Appendix B) and he has chosen to use the ergonomic criteria of Bastien and Scapin. Ergonomic criteria are characteristics of the interface that will determine its usability. They make it possible to identify the ergonomic problems of an interface[8]. It is a summary of approximately 900 recommendations in the field of ergonomics. This made it possible to produce 18 criteria divided into 8 dimensions: guidance, workload, explicit control, adaptability, error management, homogeneity / consistency, significance of codes and denominations, compatibility (Bach & Scapin, 2005).

Overall the points raised by the expert deal with the difficulties for a user to understand and use the interface in general: meaning of the labels, structure, indications and explanations, insufficient contrasts, but also raise points addressed by the par-

[8]https://www.usabilis.com/criteres-ergonomiques-bastien-et-scapin/

ticipants such as filters, navigation or map usage. The "log in" button, API, and export options will not be adressed in this thesis.

About the suggestions for improvement, we can note some points like:

- the cancel option in filters that is unreadable due to insufficient contrast,
- the information about the current data map representation where the title does not indicate what is observed,
- some use cases for users who have no knowledge of data processing, the used vocabulary or the absence of help or guidance,
- regrouping tabs that use filters in a same tab, bringing together what makes sense (e.g. filters in presentation tabs do not make any sense), and API/Export tabs must be moved with data part.

Parts of these points were also flagged by participants as ideas for improvement.

3.3.2 Cognitive Inspection

Another expert took part in the assesment. He works as a Professor of Human-Computer Interaction. He has a long experience in interface evaluation methodologies and in taking part in interface evaluations as an expert. It must be noted that this expert had some knowledge of the OGD portal beforehand, which may influence the results. However, this is mitigated by:

- the evaluation methodology, which follows the cognitive inspection guidelines, where the expert makes abstraction of their own knowledge of the interface to predict the reaction of new non-expert users, and
- the fact that the expert has not much explored the visualization tools of the portal, which are the main focus of the evaluation.

The expert built his comments step by step exploring the screens and interfaces following the solar panels scenario. He started at the landing page of the second scenario, which is the first tab shown when a user clicks on the "Photovoltaic potential by district" dataset, namely the table tab. The first comment of the expert is striking: "It is quite violent". He would have preferred to come across the information tab first but he assumed that having "Table" as a label instead of something like "Data" or "Dataset" confused users previously and made them unable to find the data from the dataset if they landed in another tab. Therefore, he thinks that the portal designers set the table tab as landing tab to solve this issue. The expert's view is that the issue is with the name of the "Table" label. It is a very important point for the expert and he pointed it out several times throughout the evaluation.

Following this first glance the expert directed his attention to the filters area starting with the code of the municipality, which does not add anything, it should not be highlighted, perhaps only for users who have the domain knowledge to use this information. Also, in the table, the sorting system is displayed differently and raises questions. A step back, the filtering system could have been simplified to only the search input, at a different place to open the display area. Still on the table tab, in the upper left corner, the information about the number of records, according to the defined filter, is sufficiently clear for the expert.

The expert then moved to the map tab, and was once again very direct: it is "the most useless view", we do not know what attribute it is defined on, there is no legend, and a lot of incomprehensible options, such as the selection tools. This is "the most depressing tab", in the expert's words.

Then the expert went to the analysis tab and tried each visualization mode proposed for a series and some gave no result (linear visualization). Then, he discovered that there is no link between the visualization proposed and the type of available data from the dataset. Also the interface does not preserve the viewing configuration when the user tries another one and comes back, the user has to do it all over again. After all these manipulations the expert felt he had "broken" the interface and had to reload the original link by editing the URL on the browser to reset the configuration and start from the default one anew. The sanction is pronounced by the expert: "this thing is unusable". Visualization options are not clear regarding their expected effects and are not at all linked to the data, so this interface seems more dedicated to professional users.

Eventually, the expert arrived at the configuration we expected for the analysis tab in our scenario: district name for the X axis, "average electric potential production by roof square meter" as series. The expert commented on his path toward the correct configuration: "it is a headache". The process should be reversed: first choose the selection, then the axis. For example there are not 2 axes for the pie chart view. Visualisation have to guide the dimensions that give meaning. But unexpectedly (and illogically), the Y axis does have an effect on the pie view. Only the meaningful dimensions, that have interest, should be involved. The expert thinks that people can play around with this tab but will never achieve meaningful results.

Finally, the expert comes to the "Custom view" tab and the name begs the question, what is it and why? The expert thinks it is related to the user's logged in status, since it has to be customized in some way. If true then it is of no interest to the non-logged in users. The expert found the visualizations on this tab meaningful and very interesting. He noted that it is the most valuable tab but unfortunately the one which speaks the least in the sense that its label gives no information on what to find in the tab. Again the expert said that labels are very important and can change everything: the custom view and table tabs are invisible because their label does not mean anything to the user. By observing the charts contained in this tab, the expert found that the Y axis of the bar chart does not start from zero, and that the display obtained by this manipulation modifies the message to the user. He noted

that the choice of starting a Y axis from zero or not should depend on the message the chart tries to communicate.

Out of the context of the scenario, the expert quickly visited the export tab and said that is the more clear and direct of them.

At this state, the expert had completed the first part of the scenario and now knows the districts with the highest production protential. The expert was directed to the second part, which consists in merging two datasets to cross the potential per district with the population per district. His first reflex is to take the two datasets data and put them into Microsoft Excel, but we asked the expert to use of the tools available on the portal, since they are the focus of the evaluation.

Before combining, the expert had to find the dataset that has the information about the population per district, so he went to the data portal menu and searched for one such dataset. The search filter was somewhat unsatisfactory, as the link between keywords and dataset is poor or irrelevant and creates a difficulty to find what is needed. After 3 attempts the expert found the "Borders of the 46 districts" dataset, which contains the needed information. The expert had to look into the content to find out that the population information was available, mentioning once again a label issue with the name of the dataset which does not say that the needed information was there.

Now the goal was to combine it with the dataset on photovoltaic potential by district, using the tool available on the portal. The expert was not able to find the tool and was direct to a discreet link on the bottom right of the analysis tab. Surprised, the expert reacted: "for real?!", followed by a big sigh. After this step he quickly found how to add the population dataset to the photovoltaic potential one. After some manipulations he found another label problem. In order to display the population and photovoltaic potential per district on one combined chart, the expert had to set the X axis of the chart to the district name, for both dataset. However, the name of the district is contained in the "District name" attribute for the first dataset and the "Statistical district" for the second dataset. This lack of consistency in the attribute labels made it difficult for the expert to combine the datasets. Another problem is the understanding of the sort system proposed on series that visually sorts the districts according to their information on the chart, helping to solve the scenario's question by highlighting the best districts. Without success the expert chose to stop at this point. He concluded by adding that there is "no logic at all in the series configuration".

According to the expert, nobody can solve with success this exercise, there is "no reflexion on the [dataset merging] tool conception", and when discussing the series sub inline configuration the expert commented with "of, how awful!" that sends a clear enough message. Configuration has to be contextualized and labels, again, are the most important problem that can be fixed and changed easily the opinion on the portal.

3.3.3 Summary of the Expert Evaluation

Both experts have mentioned the meaning problem several times with labels or contextualization, the map without legend, the filters behavior, the district code meaning in the filters, the table sorting system, or the question and the idea of putting the information tab as entry point of the datasets.

About the number of records their opinion diverge, one says it is clear enough and the other needed more context and explanations. We got the same remark for the export tab. This may be due to the previous knowledge of these elements by the second expert, who had a more positive opinion.

Except the importance of labels and the idea of putting the information tab first, participants met all these issues and are listed in the severity table (Table 3.5). The number of records was never found at this place by the participants, so it can be considered in information point or in a help window or tooltip as local and contextual information for filters as a result indication.

4 Solution Design

Taking care of all feedback from participants, we produced mock-ups illustrating possible solutions. The whole body is a reproduction of the OpenDataSoft canvas, then we defined a small design system about spaces and sizes, shapes and behavior.

Due to the late arrival of expert opinions, their points were not taken into account, apart from a few points in common with the user tests.

In this chapter, we first present a medium-fidelity prototype to illustrate the design ideas. Then, we show how the proposed design could be used to perform Scenario 2.

4.1 Medium-Fidelity Prototype

To make these illustrations, we followed the Atomic Design[1] principle of Brad Frost, that tell us that we build blocks then assemble them to make more complex elements, until we get a full page.

I rewrite the zone order and give back priority to the block reading. Then we start by the dataset itself, then user access the content displayed into tabs and filterable, and finally, user can have options. The main goal is to access and view the dataset, so the body is divided into collapsible columns (Figure 4.1), like we can have in well-known software such as Microsoft Office or Adobe solutions.

[1] https://bradfrost.com/blog/post/atomic-web-design/

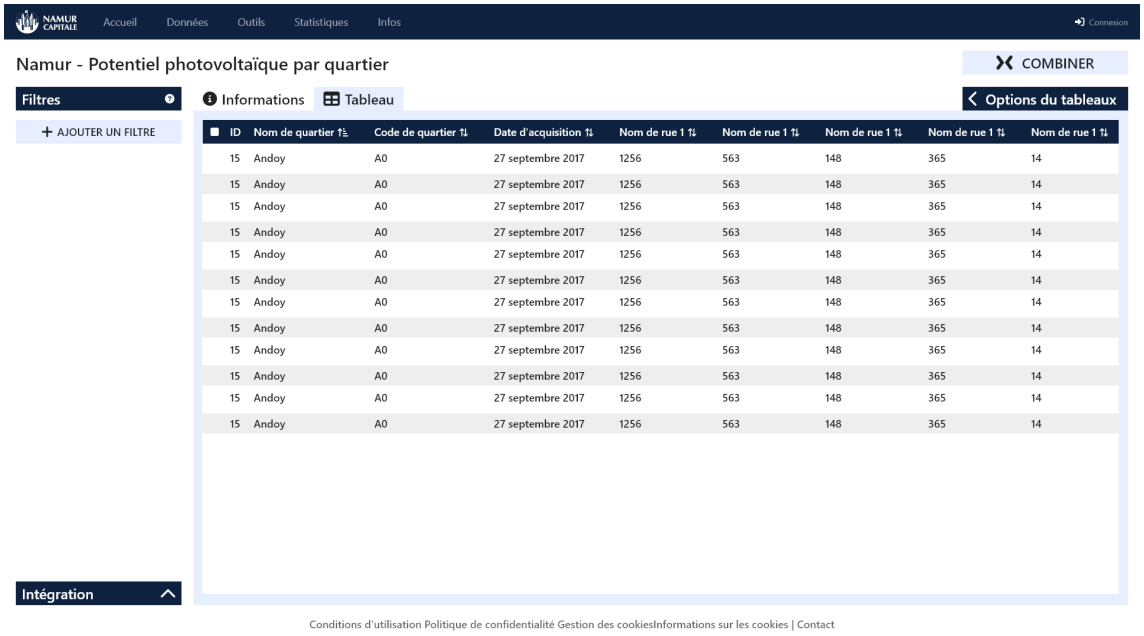


Figure 4.1: Revised table tab with collapsible integration menu at the bottom left and collapsible options menu at the top right, plus a direct access to the datasets merge tool at the top right

Actions are positioned into strategic places, based on common user exeperience. For exemple, the dataset available actions are located at the top-right, like combining datasets or share and integrate, but due to context usage, we made the choice of putting the integrate/share options in the same subsection as filters.

Collapsible columns raise questions about understandability. Can we reduce the column to an icon size like Microsoft VS Code or game launchers such as Blizzard one? Does the user, especially a new comer, need text at first sight? Can we flip columns headers like JetBrains IDE solutions?

In these software, users know the meaning of the icons because they are specific to the activity domain of the users. But on an OGD portal, it is not the case, and the audience is larger. If we keep the icon and the text and flip the whole title, we can successfully reduce the space taken and maintain the understandability (Figure 4.2).

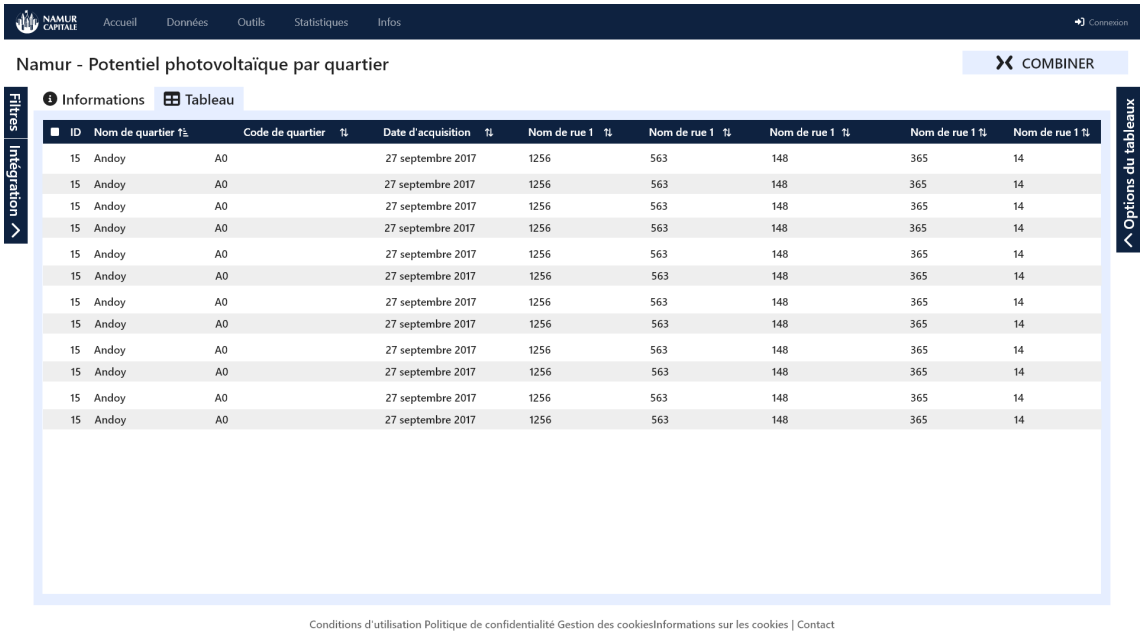


Figure 4.2: Maximum space for contents with all collapsible columns folded

The filtering system had to be rethunk. Users can choose any property of the dataset and, based on the type, can set a filter (Figure 4.3). For example, for a string type they only have to type a content to look at, like a *contains* method, and we can go further by allowing the match mode selection, or for a numeric value, we can propose a direct match or a range selector, and in our case about the datetime filter, we have to propose a new system allowing to choose juste a date or a month through years and months.

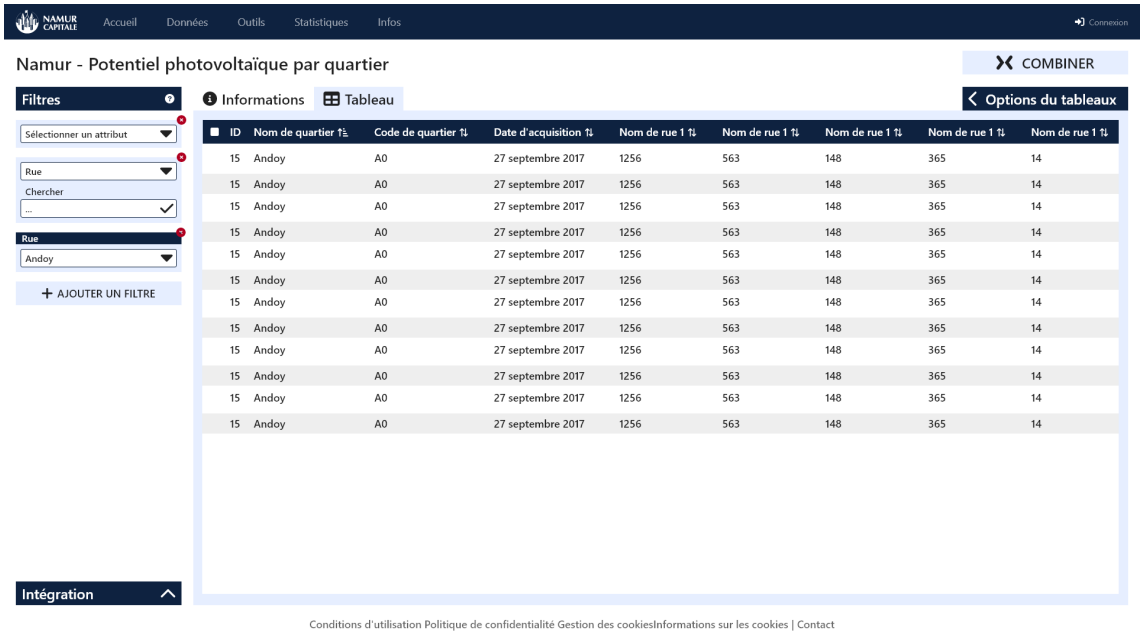


Figure 4.3: Revised filtering system. Users first select an attribute on which to filter and then choose the target value using widgets adapted to the data type of the attribute

The map tab needs only few changes to comply to the feedback, just adding an options column to allow the user to choose which attribute to display. Figure 4.4 shows the map that is displayed in the current OGD portal of Namur, for the photovoltaic potential dataset. The difference is that in the revised mock-up there is a legend that shows that the attribute represented on the map is the district name, which explains why each district has the same color, and there is the possibility to change the attribute to show on the map. In Figure 4.5, the attribute is set to the potential electric production, which allows having a more meaningful map.

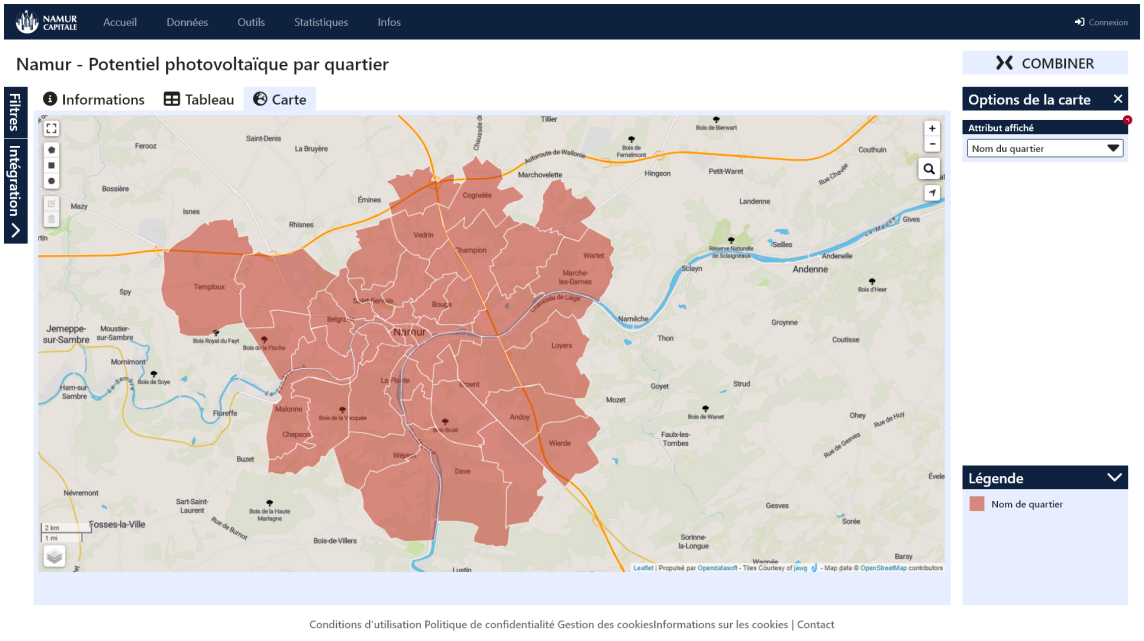


Figure 4.4: Map with options at top right to select displayed attribute and legend at bottom right with color indication

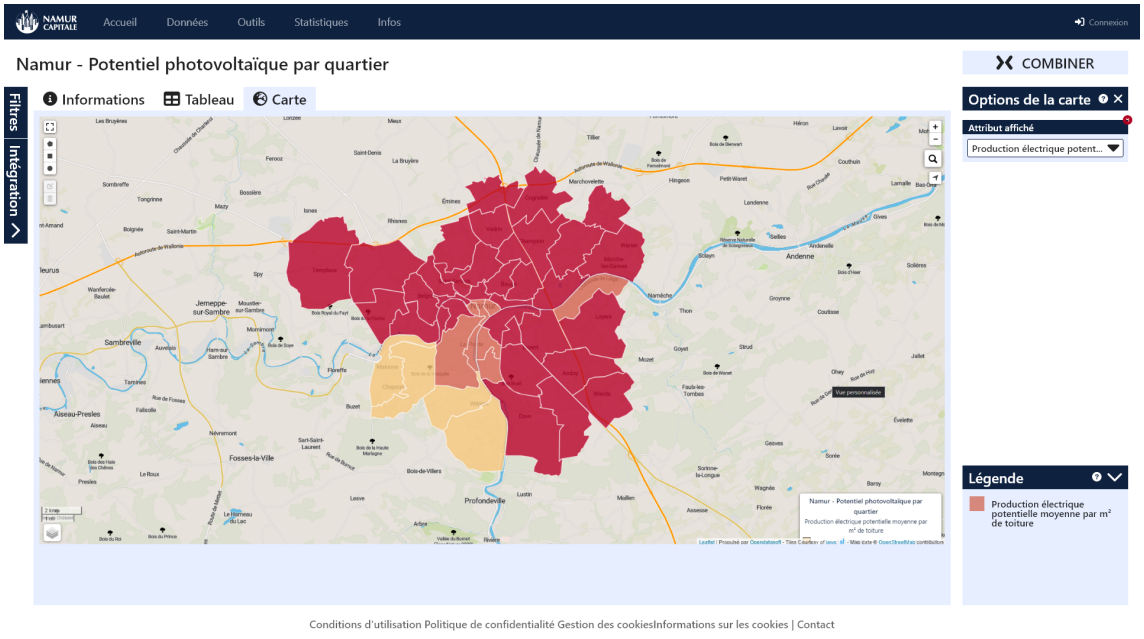


Figure 4.5: Map with another attribute selected

The analysis tab had to be rethunk too and had to propose a better layout that needs to be compatible with the graph making tool when users want to merge datasets. It is not only components, it is about overall consistency, we will see this point below. First we can avoid the multi-layers problem by splitting graphs when they are not compatible, so we can visualize them without loosing readability (Figure 4.6).

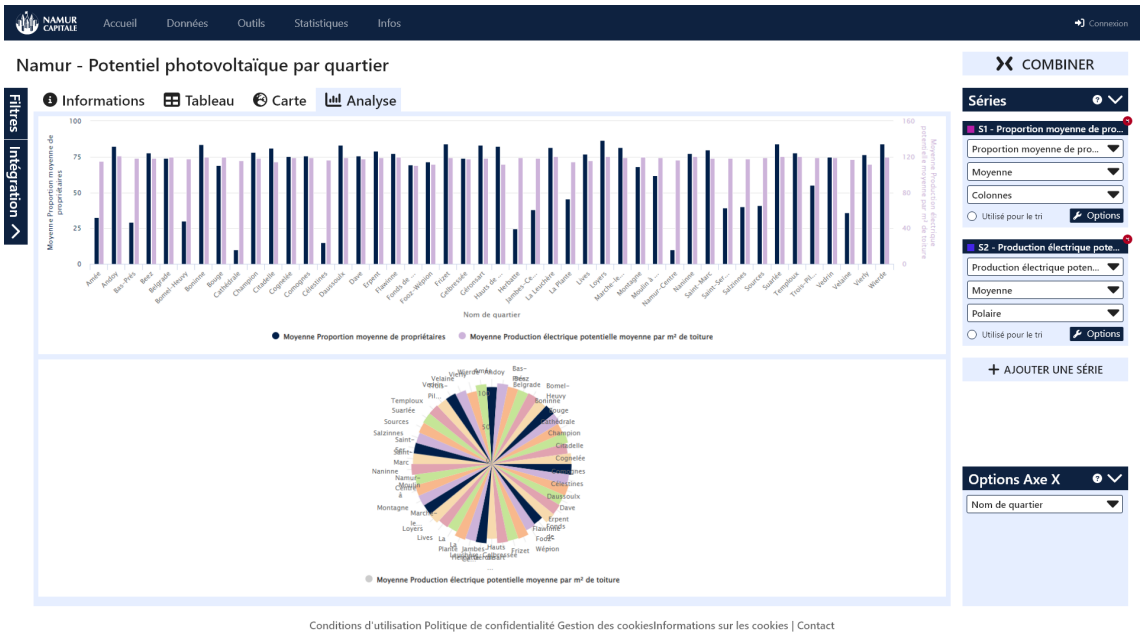


Figure 4.6: New analysis tab design with multiple layers divided by type

Another point is about the series configuration, from inline and formless to a structured and foldable configuration panel (Figure 4.7). In the analysis tab options are not available, but in the datasets merge tool, for the same configuration, you access to a set of various options such as appearence, labels, data modifier, axis modification or threshold color.

This new column is inspired by Autodesk solutions, allowing users to keep an eye on everything and directly view the changes they make. This column appears when you click on the series 'options' button, then it appears at right. Also note the difference between Figure 4.6 and Figure 4.7, the latter being an improvement to group all options instead of having split options at different positions.

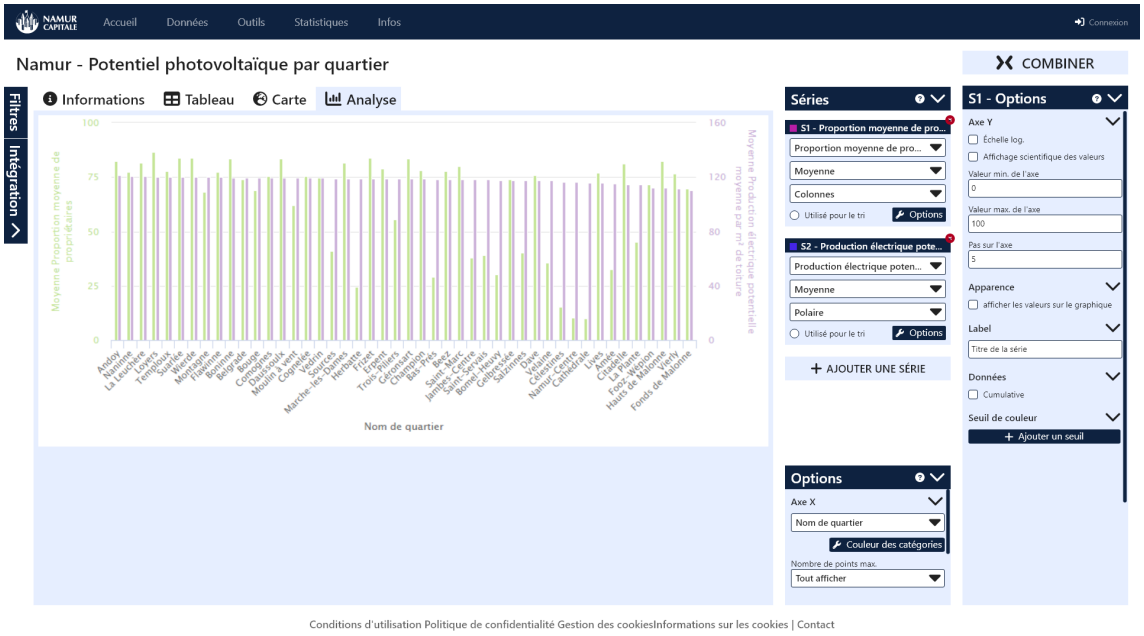


Figure 4.7: Analysis tab - Series configuration

Finally, the graph creation tool with a revisited design to be consistent with the whole concept. Compared to the analysis tab for a dataset, here we have a list of datasets and then the options panel for the selected one, reusing the same system of cumulative columns (Figure 4.8). Again, a configuration panel for the global option regrouping all the scattered options.

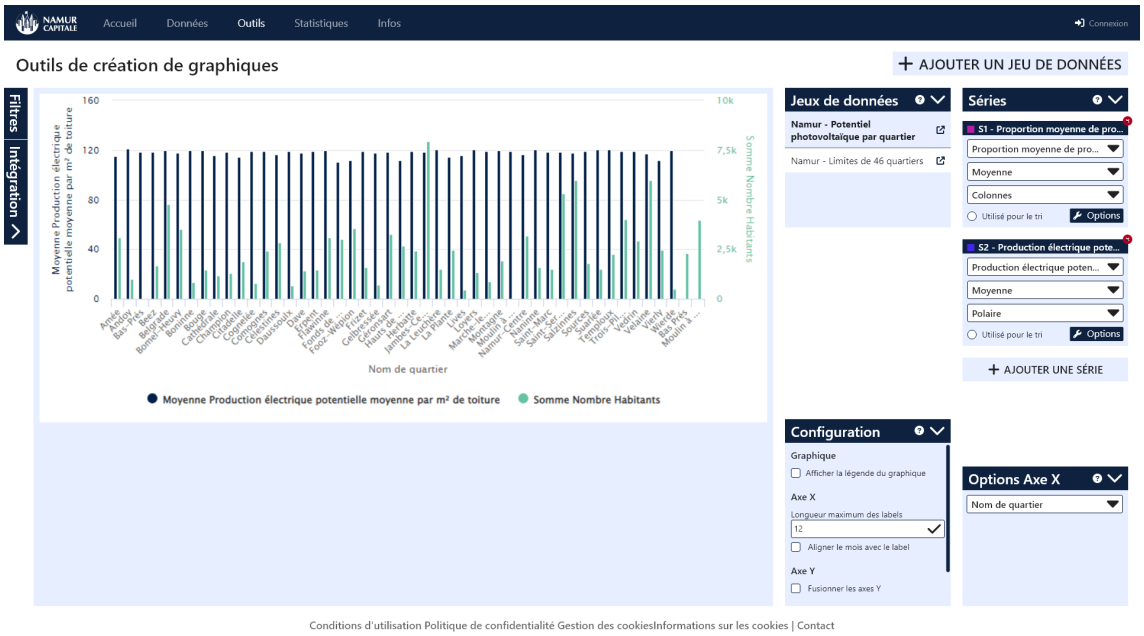


Figure 4.8: Graph making tool revised like the analysis tab, but in multiple dataset mode and related configuration

About help and tooltips feedback, there are in every mock-up rounded question marks icons that are present in almost every header, by this way the user can get

contextual help, as well as by hovering over certain fields or labels to get a tooltip with direct brief explanation.

Last feedback concerns navigation, and for this point, a simple link is added next to the dataset in the list, but it is not a way to go back, like the browser can do, but a link to open the dataset in a new tab in the idea that the user want to explore the dataset at the same time he want to make a graph with the tool.

4.2 Running Scenario 2 with the New Design

To illustrate how the proposed design can solve the barriers identified in the user test, here is an illustration of how Scenario 2 would be performed. Scenario 1 and 2 have the same structure, users are asked to get some information and understanding of the dataset and are asked to solve the problem freely using the tabs they can access, with an extra dataset combination part for Scenario 2. In this scenario the goal is to determine 3 districts of Namur to prospect.

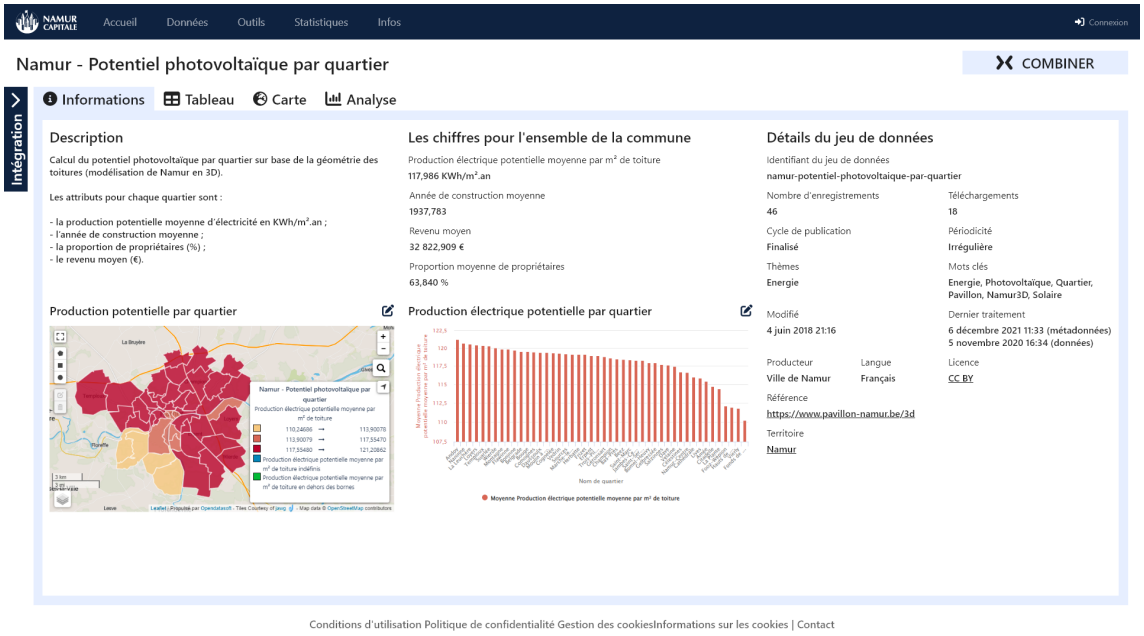


Figure 4.9: The information tab was merged with the custom view tab to show metadata and example graphs on the same tab

First, the users must answer a few questions about the dataset, and with this new version, it has the information and custom view tabs merged with a reorganized structure (Figure 4.9). They can easily answer questions 1 and 2 with description and details “Modifié” and “Dernier traitement”, but question 3 need a better understanding of the label “Nombre d’enregistrements” which can still be improved. The last question is about temporal coverage and this is not answerable by the proposed details on the information tab.

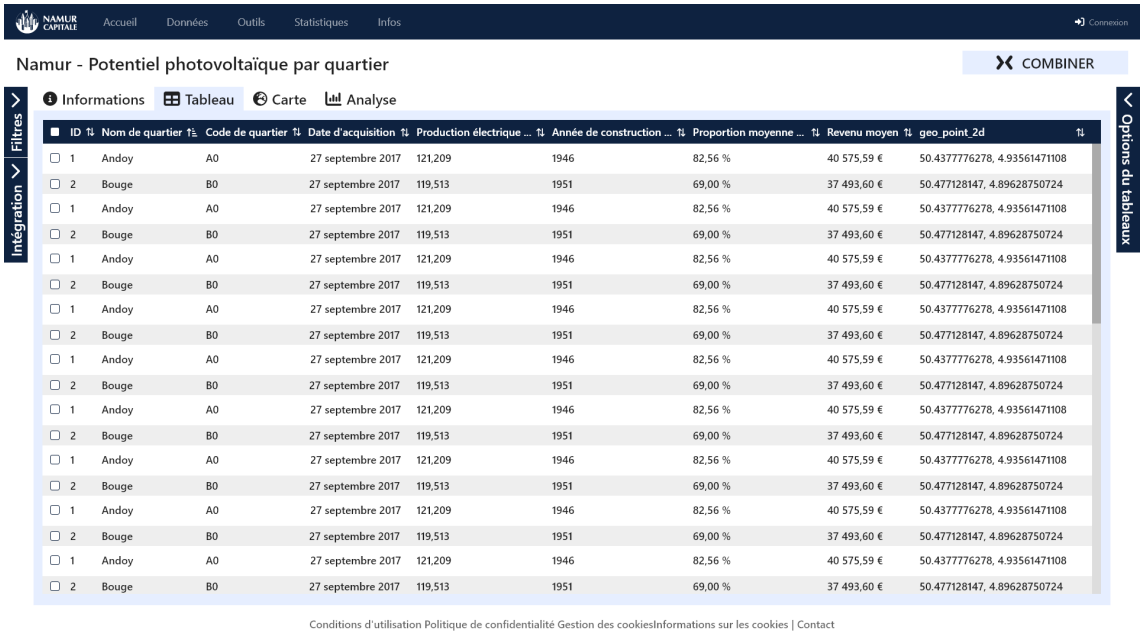


Figure 4.10: Table tab of the photovoltaic potential dataset

On the table tab (Figure 4.10), users can see the column “Date d’acquisition” that answers easily the fourth question since there is only 1 unique date. Then, from there, the participants should look at the data and determine 3 districts with a high photovoltaic potential using all the tabs they want. So to visualize more easily the data, users choose to go on the analysis tab and have a default view with a series about the average proportion of owners (“proportion moyenne de propriétaires”) (Figure 4.11).

Figure 4.11: Default configuration of the analysis tab of the photovoltaic potential dataset

The series allow the users to select 1 data attribute among 4, and one of them is the average potential electricity production per square meter of roof ("production électrique potentielle moyenne par m^2 de toiture") and this is the only one speaking about potential electricity production, so users select it (Figure 4.12). The difference between districts is very small, so we can estimate that the first 3 are the expected solution.

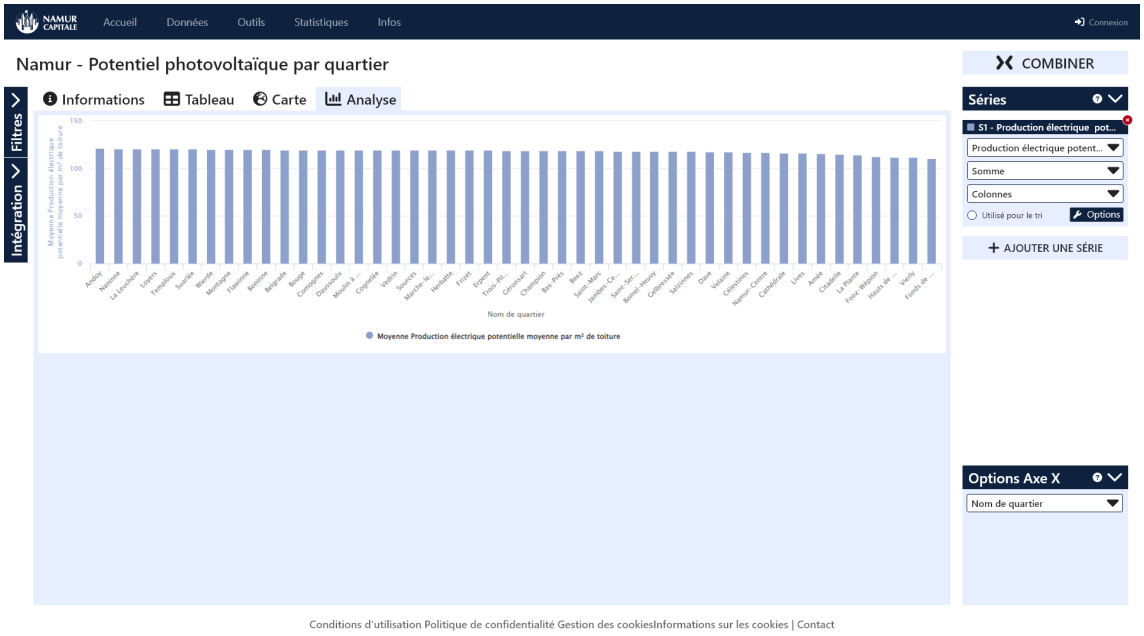


Figure 4.12: Analysis tab configured to show the average potential electricity production per square meter of roof

At this point, the questionnaire asks the participants to make a new step, find another dataset, the district population, and use it to improve the result they already have. From the current dataset, they have an action button at the upper right corner that allows them to combine it with another dataset.

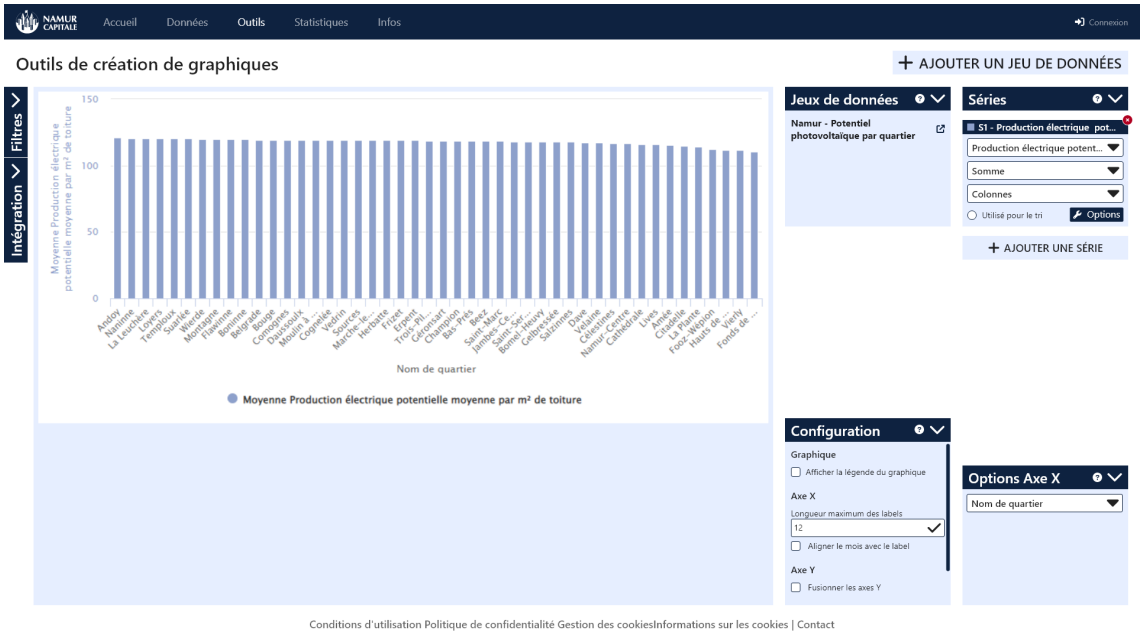


Figure 4.13: Dataset combination tool accessed from the photovoltaic potential dataset

After clicking on this action button, users are redirected to a dataset combination tool which has retained the dataset and configurations set in the analysis tab (Figure 4.13). Now our goal is to find and add another dataset as requested. Again, in the action zone, at the upper right corner, users can find a button to add another dataset. This brings up a dialog box that allows users to find and select a dataset for the combination (Figure 4.14).

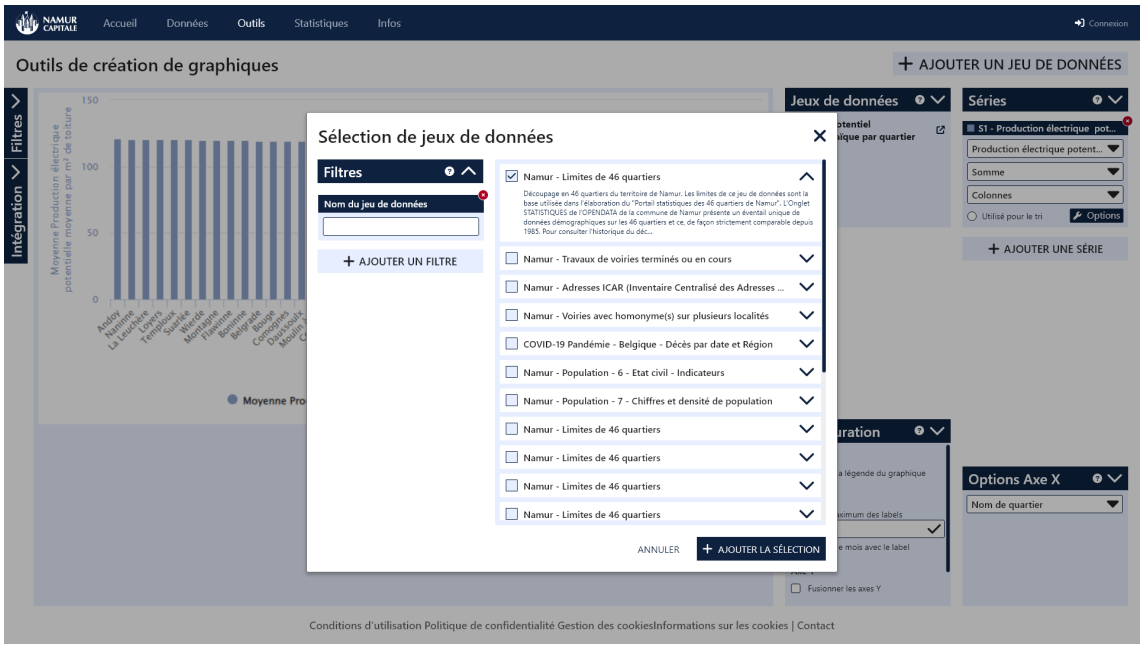


Figure 4.14: Selecting another dataset to merge with the photovoltaic potential dataset

This dialog allows users to filter datasets by certain attributes, such as their name, and displays the entire list by default. By clicking on the chevron of an element, users can read its description. Users choose to select the dataset “Namur - Limits of the 46 districts” because it constains demographic data by district, and in particular the needed information on the number of inhabitants per district.

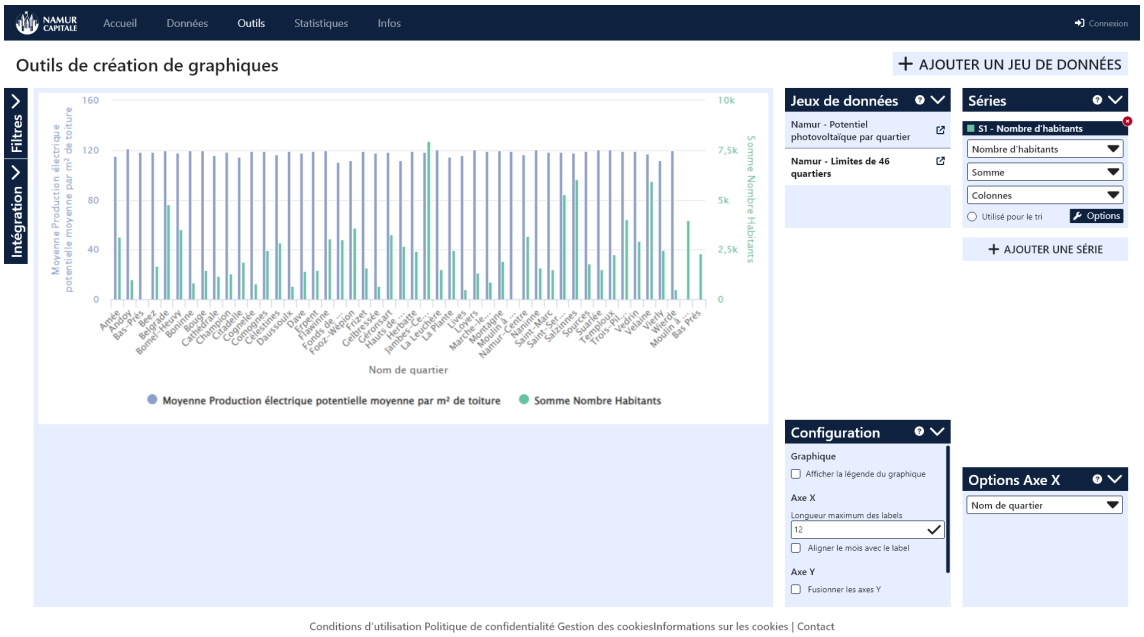


Figure 4.15: Double bar chart created with the dataset combination tool that shows in a single chart the average electric production potential per district (from the photovoltaic potential dataset) and the number of inhabitants per district (from the limits of the districs dataset)

After adding this dataset, the default configuration is interesting for the users, because it shows the number of inhabitants in the same visualization mode (columns). They can see the first set of data with a new perspective and revisit their initial selection of the 3 districts (Figure 4.15). For example, if a user selected Andoy, Naninne and La Leuchère at first, then with the second dataset they could chose Jambes-Centre, Salzinnes and Velaine, which have a slightly lower production potential but a much larger population and therefore much more potential customers.

5 High-Fidelity Prototype

This chapter presents the implementation of the proposed design in a high-fidelity dynamic web application. The choice of developing a web application is to be consistent with the choice of providing open data via web portals in practice.

Due to the time invested in collecting data from both users and usability experts and designing the solution, it was not possible to implement every feature from the medium-fidelity prototype. Thus, we have prioritized the features to implement to focus on those solving the most severe barriers based on the severity table (Table 3.5) and on the visualization focus of this thesis. If a barrier was blocking for at least one participant, if it is about visualization, and if it is a discomfort for enough participants, the barrier was considered for the implementation. The resulting priorization is:

1 Multi-layers

2 Navigation, table and Map

3 Help and tooltip

4 Button and action zones

5 Information

Note that some lower priority points will be addressed by partially solving higher priorities such as buttons or overall ergonomics (action areas).

5.1 Technological Choices

5.1.1 Stack and Philosophy

To build this dynamic web application, we chose to start with Angular which is a popular framework, it is also used by OpenDataSoft for their application, and as a bonus we already know it in the personal and professional context. In addition we selected the PrimeNg[1] library which gives us a list of ready-to-use components such as table, form fields or some data visualization tools using the additionnal charts.js library, but also PrimeFlex for structuring elements.

[1] https://www.primefaces.org/primeng

PrimeNg comes with PrimeIcons but we prefered to use the popular and recognized FontAwesome icon system[2]. And to be sure the web application is cross-browser compatible, we added a reset.css[3] from Eric A. Meyer, so we can define the whole base combining the multiple frameworks and library styles with the aspect and rules we want.

5.1.2 Assets: Data and Metadata Files

To build the prototype we need the data used in both scenarios, we chose to download it using the export tab on each dataset which gives the data, here in a JSON format. However, the metadata is not included in the export, so we have to build a metadata structure by hand, also in JSON format, and copy the information shown in the information tab. These JSON files are assets of the prototype.

5.1.3 Routes and Components

The application is composed of a set of main components, namely pages, that have each an access route (Table 5.1).

Route	Component	Description
/home	HomeComponent	Welcome page
/tools	ToolsComponent	Tools page where we can select the merge tool from the beginning
/infos	InfosComponent	An about page of the application
/datasets	DatasetsComponent	The datasets page where we can access a specific dataset page, instead of knowing by heart the dataset URL
/datasets/:id	DatasetComponent	The page for a specific dataset, the purpose of this prototype. It needs a dataset id which will be resolved.

Table 5.1: Access route of each page of the application

Note that the default route / redirects to */home* and that all route errors are redirected to *PageNotFoundComponent*.

The DatasetComponent Page

Since the application focuses on reworking the tabs of the page of a specific dataset, which was the part of the portal invovled in the scenarios, this section goes into

[2]https://fontawesome.com/
[3]https://meyerweb.com/eric/tools/css/reset/

more details with the DatasetComponent page.

When a user selects a dataset or accesses it by direct URL, the router resolves (Figure 5.1) the id before loading the component, in other cases the user is redirected to the datasets main page. The resolver calls the DatasetService to get the dataset metadata, this is done asynchronously, in this case a JSON file in the assets, then places the result when ready in the data section of the router for the component, which it can access after the 'resolver' part is finished and component running.

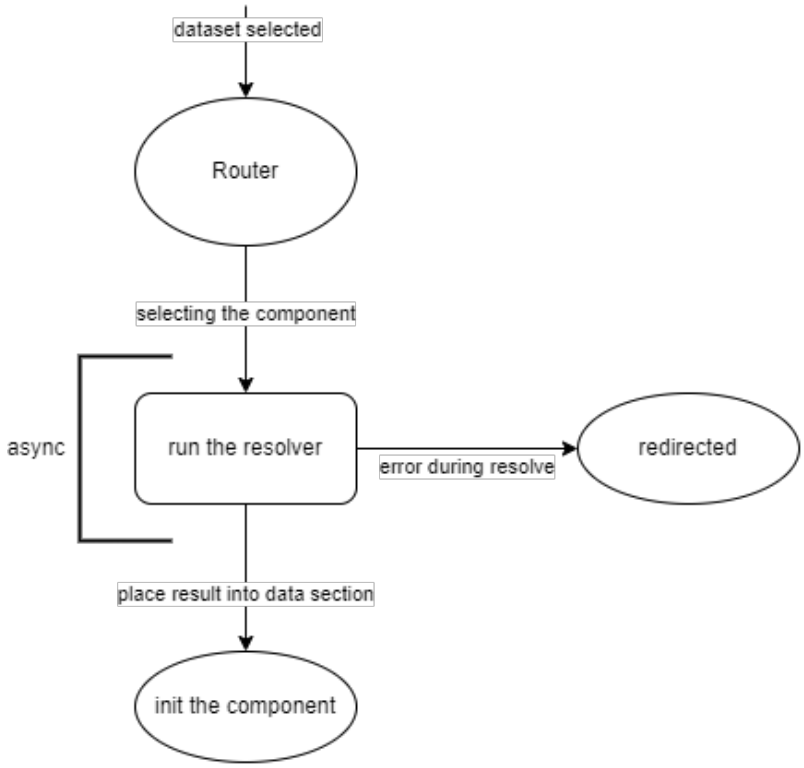


Figure 5.1: Dataset id resolution process in the Angular routing part

The component can start and the user can get the dataset metadata in the router's data section then they can complete it by loading the dataset data through the same service as the resolver. At this point system has the metadata and the data for a dataset, and thus can start to exploit them in the visualization feature.

5.2 Implemented Features

5.2.1 Multi-layers

We start with a blank page and with the multi-layers visualization barrier, which was a big challenge. We have to gather the dataset data and build a structure intended for prepare the visualisation, but also think the link to the interface options and how reactivity can be handled. For the ideal end result, we want a generic component that can take multiple datasets and multiple series as input and manage it, but in this context, we have to manage one dataset.

The multi-layers barrier problem is the graph compatibility, a bar graph cannot take a pie graph (cf. Figure 3.19), so we have to think of a way to group them by number of axes: 1 for a pie chart, 2 for a bar chart or a line chart. To link this idea to the reality we have checked with the visualization component how it handles it. We are lucky, a combo mode allows us to put a bar chart and a line chart together easily and a pie chart will be alone. The map is not a chart so it is a different component to prepare.

To summarize the algorithm, for each series, we look at their type and get the category to group, then we can apply data computing and sorting, then at last look to add this series in an already existing and compatible graph, or create a new one for its type, and it will be translated into the front by the space division for each graph and each one properly set. That is all, the rest is just preparatory calculation of the series data.

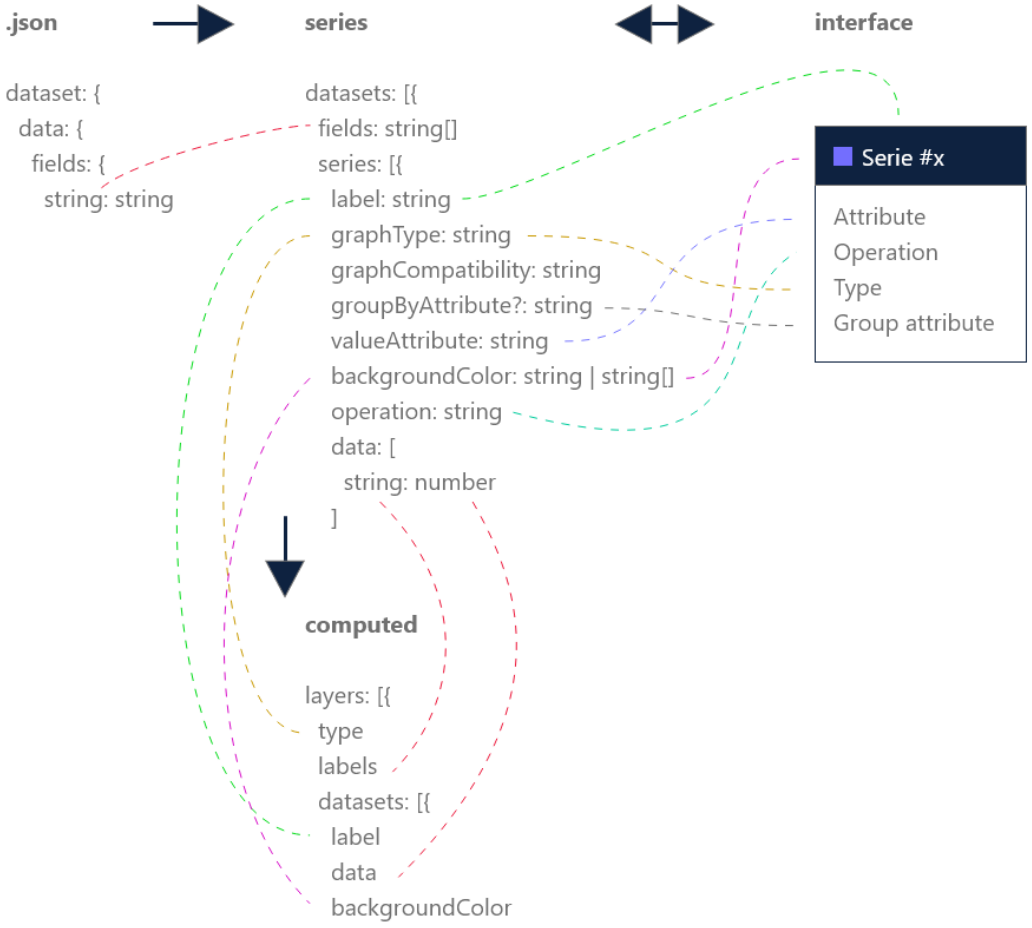


Figure 5.2: Data flow and process till rendering

If we take the data flow point of view (Figure 5.2), we gather the JSON and parse it first to get the fields list that we store in the dataset object, then we create an

arbitrary first series with a bar chart type and selecting first text attribute for X axis and first numeric attribute for value. We complete this configuration of default values to make the link between interface and the visualization component. Series object, series options bloc interface and graph rendering component are linked, the series object is the source of data, options are the modifiers and graph display the result. Data links is illustrated in Figure 5.2 with colored dashed lines.

Aside of this point of view, we have the interface to connect, each series has options that can change the rendering. Each change must run again the data computation and update the rendering, without forgetting to setup specific properties of the graph type or an option attribute mode.

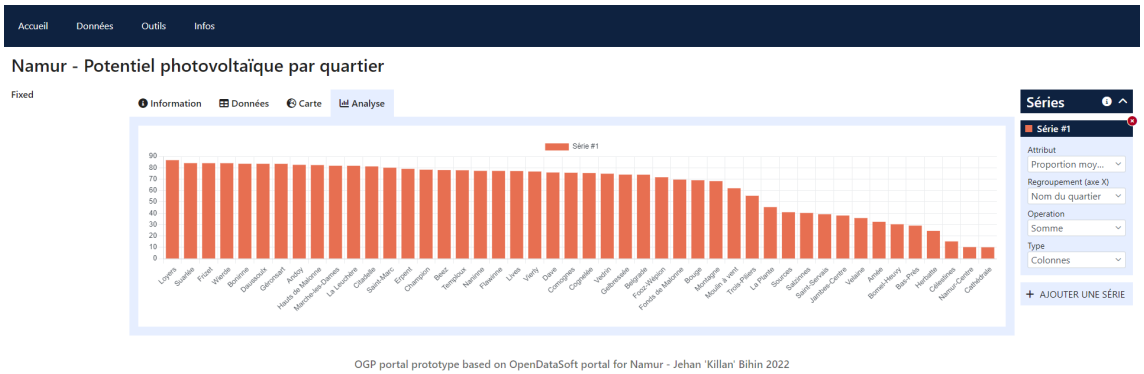


Figure 5.3: One bar graph rendering

First result, a bar graph with its series options box on the right, that users can manipulate to change properties and rendered result (Figure 5.3). Users can also remove series with the upper right icon, or add one with some random and some default options set.

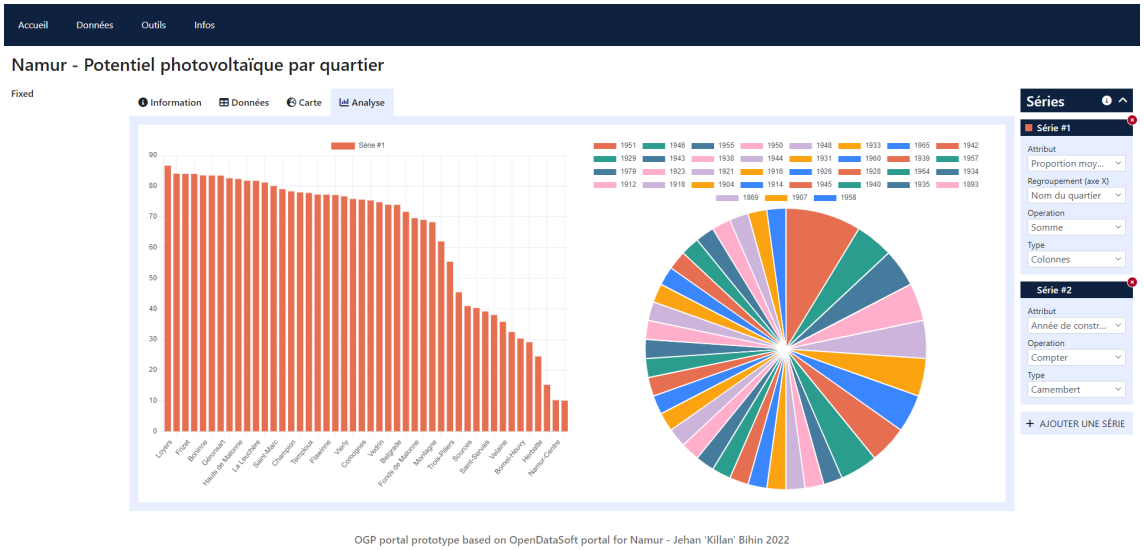


Figure 5.4: A bar chart and a pie chart, happily coexisting

For example, we chose two different types of charts and the algorithm divides them

properly and creates space for rendering (Figure 5.4).

Figure 5.5: 2 series using bar chart merged into one

Last example, if users choose two bar chart, they are merged into one (Figure 5.5). Of course, due to multi-layers barrier and solution, users need to define the same grouping attribute (X axis) to have a coherent merging.

The X axis has been moved into the serie options box, unlike what can be ssen in the medium-fidelity prototype (Figure 4.6), it is due to the fact we can have multiple charts now and so each can have it is own and different X axis.

5.2.2 Table

In another context I wrote an article[4] about PrimeNg and the tables about injecting some dynamic behaviors to the PrimeNg *p-table* component, inspired by AG Grid[5]. I made my article in the form of a tutorial and I made a demo that I put on GitHub[6]. So for this barrier, I have checked out and adapted my ready-to-use custom component based on *p-table* as a starter, plus its service, interfaces and directive.

To use it we just need data, that we already have in the dataset in a subsection, and columns definition, that can be defined by default when we receive data and process the fields list. The columns definition is also a point to adress, allowing users to choose what they want to view.

[4]http://blog.daaboo.net/2022/03/primeng-vs-ag-grid-rendu-et-composant-de-cellule/
[5]https://www.ag-grid.com/
[6]https://github.com/killan/primeng-simple-table

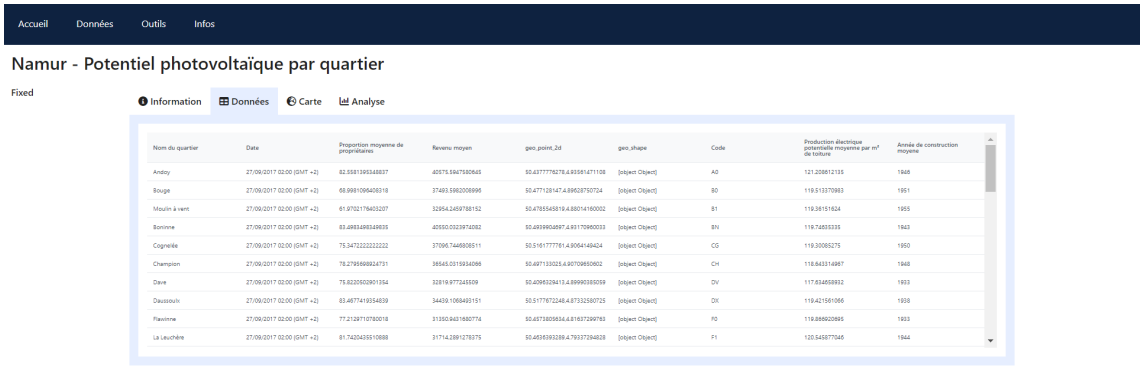


Figure 5.6: Table rendering with default columns and no filters

Cells are rendered as raw values, only dates are processed to illustrate a cell rendering based on type and format detection (Figure 5.6). The base is working and displays well, we can move to columns choice, filtering and sorting.

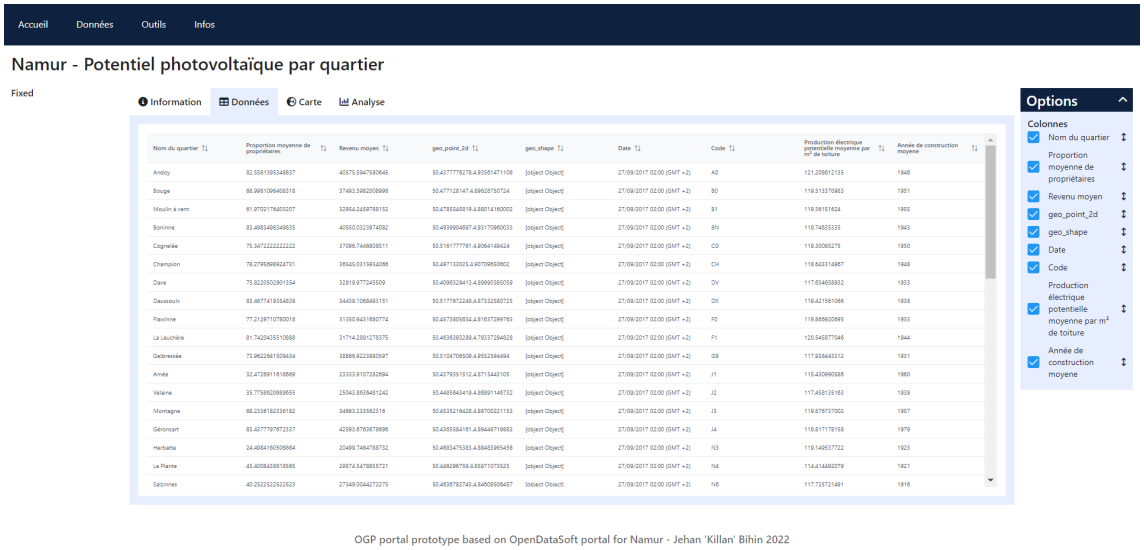


Figure 5.7: Table options to reorder and hide columns

The options for the table contain a list of fields, and users can use the checkboxes to toggle which one they want to display or hide (Figure 5.7). They can also use the up-down arrow icon to move the field to another position and rearrange them as they wish.

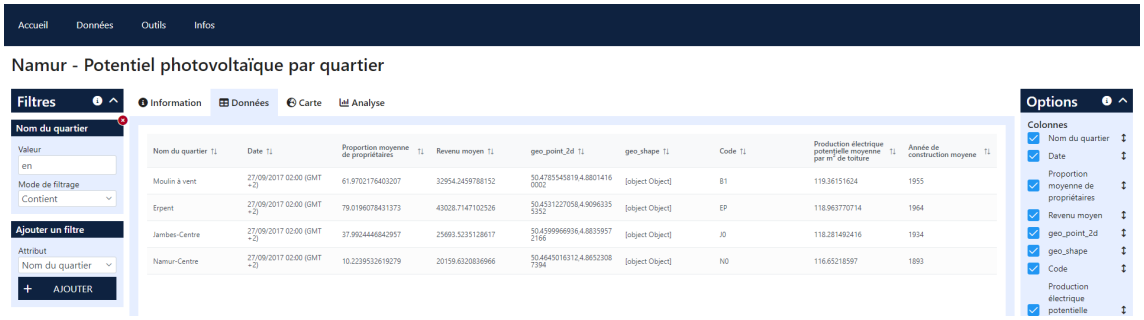


Figure 5.8: Filtering data in the table

The data are filtered according to each filter element, containing here the attribute to compare, a value and the filter mode (*contains*, *equals*, etc.) (Figure 5.9). If a row matches all the filters, it remains selected, otherwise it is removed from the table.

All visualization tabs (table, map and analysis) are linked to the same dataset and can be brushed through the filtering system to create a subset of data, that will automatically trigger an update of the visualizations to display only this subset. This technique is called "Linking and brushing" (Butz, n.d.; Ward, 2009).

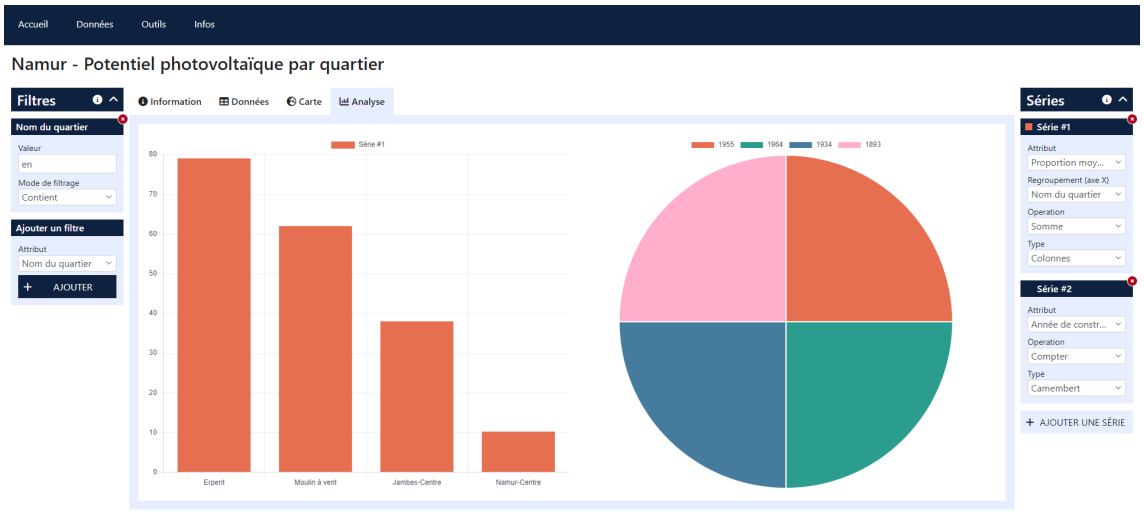


Figure 5.9: Filtering data in the table is reflected in the visualizations, in a link-and-brush fashion

5.2.3 Map

For this tab we selected the open-source library Leaflet[7] that can work with OpenStreetMap (OSM)[8] layers. The goal is to determine among the fields, which are compatible with a map, in other words which fields contains the information of a point (i.e. a specific location) or a polygon (i.e. a specific area) in the datatset.

[7]https://leafletjs.com/
[8]https://www.openstreetmap.org/search?query=namur#map=10/50.4668/4.8662

For example, in the photovoltaic potential dataset, the “Geo Shape” field contains a polygon and the “geo_point_2d” field is a point.

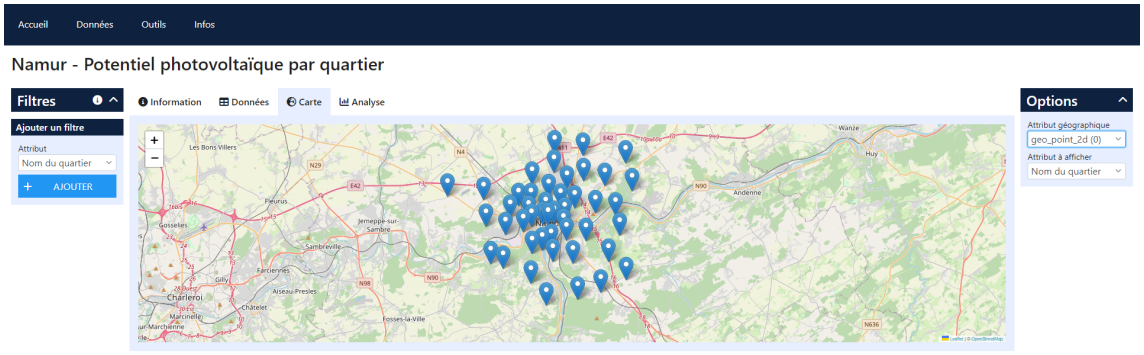


Figure 5.10: Points displayed on a map

Points are simple, the attribute contains an array of 2 numeric values, so it can easily be mapped to a point object (Figure 5.10). The polygon type is an object containing different attributes, especially *coordinates* and *type* that help to determine the kind of data; But there is a trap, the coordinates for the point and the coordinates for the polygon in the dataset are not in the same order: point is *[**latitude**, longitude]* and the polygon structure is *[[[longitude, **latitude**], ...]]*, note the sub-array structure, since a polygon is in fact an array of coordinates, thus an array of arrays.

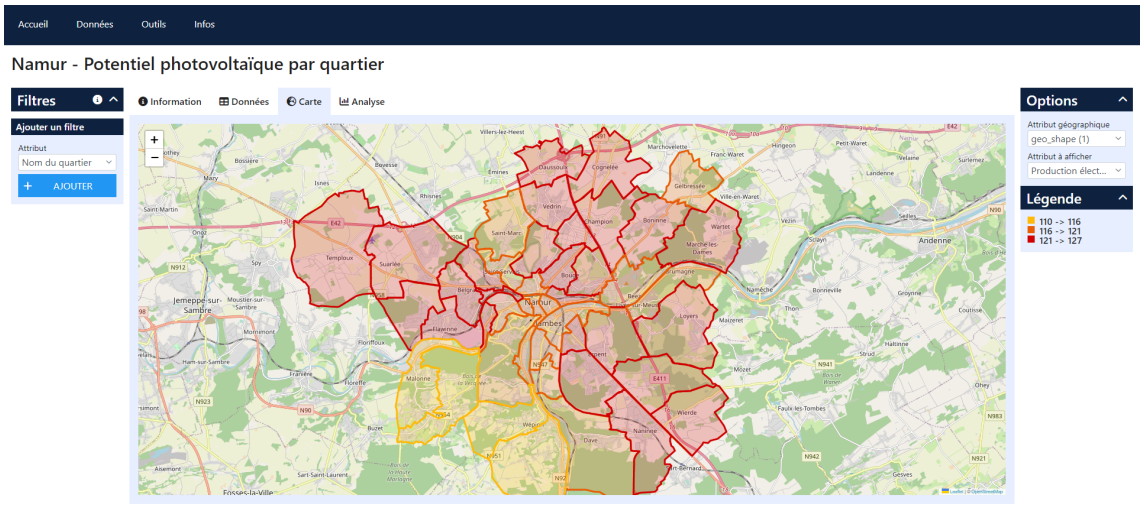


Figure 5.11: Choropleth map showing the electric production potential

Once we can draw the polygons, the question is how to change their appearance to encode information. We have to select an attribute we want to visualize among the list. We need to process the data again and get the min-max values to define the gap, then divide the range into intervals and assign a color to each intervals. To have a color gradient, we chose two colors, yellow and red, then assigned the yellow color to the first interval, the red to the last interval, and a gradient of the two colors for the interval in between. And voilà! (Figure 5.11)

Of course, as with the analysis tab, the filters of the table tab are reflected on the map too (Figure 5.12).

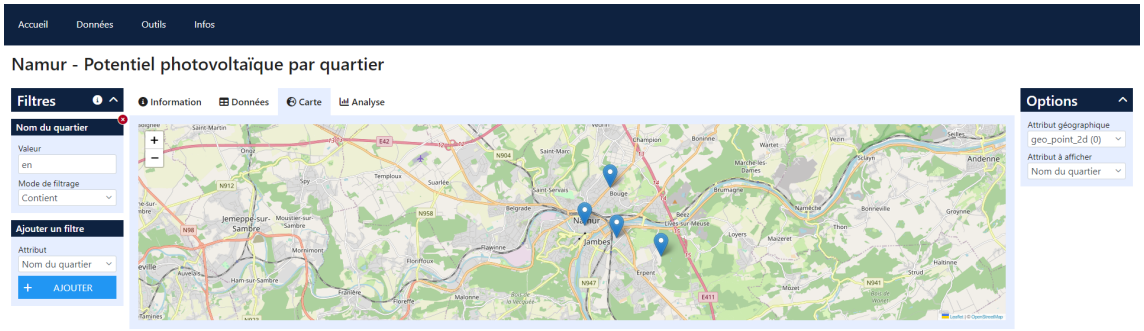


Figure 5.12: Map : filters

5.2.4 Tooltips & Buttons

Help and tooltip, and Buttons and action areas, are handled in previously made interfaces, through contextual help icons (Figure 5.13 and 5.14) and contrast search/color of buttons or checkboxes (Figure 5.13).

Figure 5.13: Contextual help via a tooltip in the options of the table

Figure 5.14: Contextual help via a tooltip in the filters of the table

I first tried using a defined primary and/or secondary color, but the point about contrast and, by extension, making enough differences between things stands out again, so we suggested using a different color (Figure 5.15).

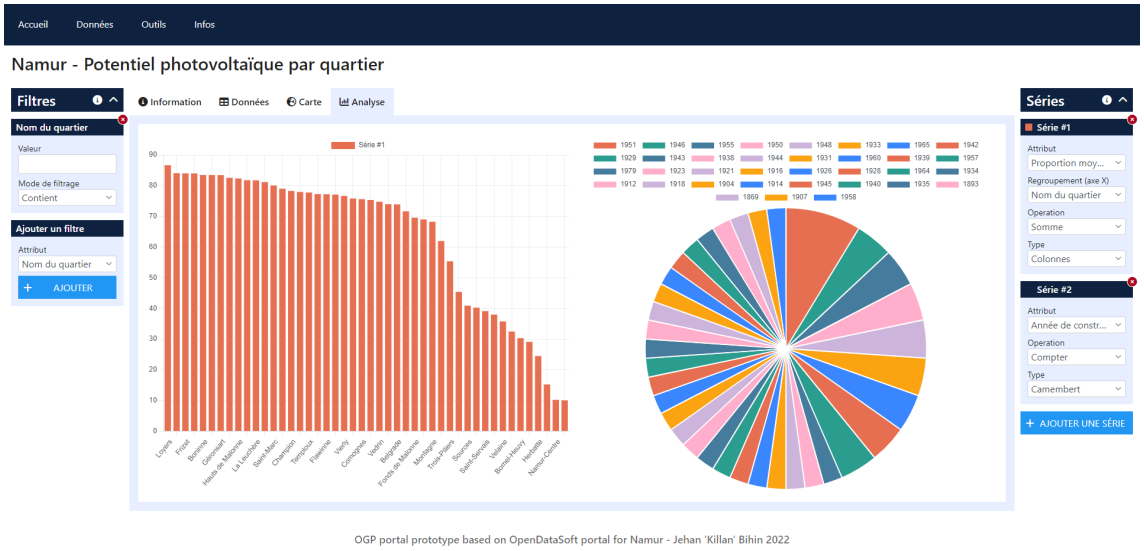


Figure 5.15: Buttons and checkboxes color

6 Future Work

6.1 Further Evaluations

The mock-ups, the second illustrated scenario and the prototype consist of a first proposal based on the feedback from the participants. However, the design of an interface is an iterative process and the implemented solution should be evaluated again with users and experts. In particular, an important barrier for which work remains to be done concerns the labels where you can for example use the card sorting technique (Spencer, 2004) or the WebSort kind of tool (Wood & Wood, 2008). Card sorting approaches allow understanding how users categorize information, so they could indeed be useful to understand what should be grouped together in a tab or an options menu and under what label.

6.2 Further Implementation

6.2.1 General Improvements

Some part of the design can be improved like the elements list, e.g. datasets list (see Figure 4.15), and the drag & drop handling, the dropdown that needs a label and not only a placeholder (see Figure 4.13).

Another point of improvment is the rhythm, because we have reading content it is important to give confort to the reader, plus in the 2x dimensions (columns) and where content has to be aligned. Previously, in my work, I've been inspired with the Gutenberg article of Matej Latin[1] for a rhythm explanation and solution, combined with the 2X grid system of the IBM Carbon design system[2].

I had not enough time to split the *DatasetComponent* code into smaller components like the visualization one or to create reusable interface components like boxes, and to propose a split between data visualization and option panel that can be linked but also used by more than one like filters.

The flipping system for section boxes in left and right columns can be done as a full column toggler, keeping only titles, flipping them and change the position to an absolute anchor, left or right. Without forgetting to hide columns and take all width for the content.

[1] https://matejlatin.github.io/Gutenberg/
[2] https://carbondesignsystem.com/

6.2.2 Tab-Specific Improvements

Multi-layers

The color square can be enhanced into a component that allows a user to click to choose another color from a predefined palette or set the color code. Additionally, in case of a pie chart or a map, we need more than one color, so we can imagine the component being able to draw a color scale in the square to display a summary of the chosen colors.

Sorting options are not proposed and it must take care of the new data structure (multiple datasets, mutliple series), maybe using the same way the *p-table* component is capable of but adapted for each graph.

The consolidation system ignores the different lists of labels on the X axis, in case of multiple series in the same bar/line chart, keeping only the labels of the first series. It needs to detect labels differences then create a new chart to split the series.

Table

Column reordering can be done directly by dragging the column header by setting an option of PrimeNg *p-table*, a callback is provided, so we can imagine merging the displayed and hidden columns, keeping the position of the hidden ones and updating the others.

About the filters, we can keep improving the options of the filters depending on the data type, for example if it is a numeric value, we can set a range value, or we can add a *not* which inverts the filter (does not contain, is not equal, etc.), another example with geographical points, we can define a shape to determine if the point is inside (or not), we can go as far as we want, sky is the limit.

Map

I think we can offer the user a predefined auto range for the value intervals, but allow modifying (add/remove) the intervals and the choice of colors (min-max, interpolation is calculated).

7 Conclusion

The objective of this thesis is to understand what are the citizens' requirements toward visualization features on Open Government Data portals, and to implement them into a prototype focused on the usability of these features opting for a user-centered approach, using user tests and expert reviews. We started by conducting an evaluation of an existing OGD portal as application case: the OpenDataSoft portal of Namur.

A user test was constructed in the form of 2 scenarios driving the participants on the portal and putting them in a concrete context with a situation and an objective. First scenario is about a shopkeeper searching the right place for his new shop among the streets of Namur using a pedestrian counter dataset. Second scenario puts the participant in the role of a solar panel seller looking for the next 3 districts to prospect tomorrow. In Scenario 2, participants have to find and merge another dataset about population by district to improve their answer, while Scenario 1 only involves one dataset.

The evaluation results revealed barriers such as (1) the lack of reading aids for the data table, (2) the difficulty of understanding the map display without legend or attribute indicator, (3) the difficulty of using the graph when mixing types due to overlap, which we refer as the multi-layers problem, (4) filters that only present certain attributes in unclear usage, (5) information tab and second content tab sharing parts of the same information, (6) buttons and action zones confusing users with their differences or meanings, (7) the lack of contextual help or tooltip and (8) unfinished navigation that makes users lose the ability to go back. Then we ranked them by severity based on whether they only generate annoyance or are blocking for users. And first we have the multi-layers barrier, then at the same level the navigation, table and map barriers, then the help and the tooltip, followed by the buttons and action areas and finally the information barrier.

A medium-fidelity prototype was developed to illustrate the new design we propose to address these 8 barriers. Among these, we have selected to implement in a high-fidelity prototype some of the main ones such as multi-layers with a linking and brushing approach, a table with a more flexible system and filters that offer all the attributes, a new tab of map with a legend and the choice of the attribute users want to display. All this in a new container system allowing users to have more space with a collapsible system of option columns.

Some of the reported barriers have not been resolved and require further work, following the iterative UX wheel process, and implementation with user and expert feedback. Considering where we went with the high-fidelity prototype, we can still

go further in implementing the sorting option in the analysis tab with type splitting, or improve the consolidation system on the detection of X axis differences, or give users more options on the map where users can have control over colors and range values, also on table tab with column ordering and filtering.

References

Attard, J., Orlandi, F., Scerri, S., & Auer, S. (2015). A systematic review of open government data initiatives. *Government information quarterly*, *32*(4), 399–418.

Bach, C., & Scapin, D. (2005). *Critères ergonomiques pour les interactions homme-environnements virtuels: Définitions, justifications et exemples* (Doctoral dissertation). INRIA.

Barros, G. A., Liapis, A., & Togelius, J. (2016). Playing with data: Procedural generation of adventures from open data.

Barros, G. A. B., Green, M. C., Liapis, A., & Togelius, J. (2018). Who killed albert einstein? from open data to murder mystery games. *IEEE Transactions on Games*, *11*(1), 79–89.

Bauer, F., & Kaltenböck, M. (2011). Linked open data: The essentials. *Edition mono/monochrom, Vienna*, *710*.

Benita, F., Perhac, J., Tunçer, B., Burkhard, R., & Schubiger, S. (2020). 3d-4d visualisation of iot data from singapore's national science experiment. *Journal of Spatial Science*, 1–19.

Beno, M., Figl, K., Umbrich, J., & Polleres, A. (2017). Open data hopes and fears: Determining the barriers of open data. *2017 Conference for E-Democracy and Open Government (CeDEM)*, 69–81.

Bizer, C., Heath, T., & Berners-Lee, T. (2011). Linked data: The story so far. *Semantic services, interoperability and web applications: Emerging concepts* (pp. 205–227). IGI Global.

Boost Labs. (2021). https://boostlabs.com/blog/10-types-of-data-visualization-tools/

Butz, A. (n.d.). Interaction with visualizations dynamic linking, brushing and filtering in information visualization displays. https://www.medien.ifi.lmu.de/lehre/ws1112/iv/folien/IV-W11-05-Interaction.pdf

Crusoe, J., & Ahlin, K. (2019). Users' activities and impediments from motivation to deployment in open government data–a process framework. *Scandinavian Workshop of e-Government SWEG 2019, the University of South-Eastern Norway (USN), Campus Vestfold, 30-31 January, 2019*.

Crusoe, J., Simonofski, A., Clarinval, A., & Gebka, E. (2019). The impact of impediments on open government data use: Insights from users. *2019 13th International Conference on Research Challenges in Information Science (RCIS)*, 1–12.

Dave, B., Buda, A., Nurminen, A., & Främling, K. (2018). A framework for integrating bim and iot through open standards. *Automation in Construction*, *95*, 35–45.

Davis, F. D. (1993). User acceptance of information technology: System characteristics, user perceptions and behavioral impacts. *International journal of man-machine studies*, *38*(3), 475–487.

Eberhardt, A., & Silveira, M. S. (2018). Show me the data! a systematic mapping on open government data visualization. *Proceedings of the 19th Annual International Conference on Digital Government Research: Governance in the Data Age*, 1–10.

Friberger, M. G., & Togelius, J. (2012). Generating game content from open data. *Proceedings of the International Conference on the Foundations of Digital Games*, 290–291. https://doi.org/10.1145/2282338.2282404

Graham, M., & Shelton, T. (2013). Geography and the future of big data, big data and the future of geography. *Dialogues in Human geography*, *3*(3), 255–261.

Graves, A., & Hendler, J. (2013). Visualization tools for open government data. *Proceedings of the 14th annual international conference on digital government research*, 136–145.

Green, M. C., Barros, G. A., Liapis, A., & Togelius, J. (2018). Data agent. *Proceedings of the 13th International Conference on the Foundations of Digital Games*, 1–10.

Harkins, J., & Heard, C. (2020). Interfacing the city: Mixed reality as a form of open data. *Open cities— open data* (pp. 241–263). Springer.

Hartson, R., & Pyla, P. S. (2012). *The ux book: Process and guidelines for ensuring a quality user experience*. Elsevier.

Janssen, M., Charalabidis, Y., & Zuiderwijk, A. (2012). Benefits, adoption barriers and myths of open data and open government. *Information systems management*, *29*(4), 258–268.

Karam, R., & Melchiori, M. (2013). A crowdsourcing-based framework for improving geo-spatial open data. *2013 IEEE International Conference on Systems, Man, and Cybernetics*, 468–473. https://doi.org/10.1109/SMC.2013.85

Kostovski, M., Jovanovik, M., & Trajanov, D. (2012). Open data portal based on semantic web technologies. *Proceedings of the 7th South East European Doctoral Student Conference*, 1–13.

Krooks, A., Kahkonen, J., Lehto, L., Latvala, P., Karjalainen, M., & Honkavaara, E. (2014). Webgl visualisation of 3d environmental models based on finnish open geospatial data sets. *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, *XL-3*, 163–169. https://doi.org/10.5194/isprsarchives-XL-3-163-2014

Kubler, S., Robert, J., Le Traon, Y., Umbrich, J., & Neumaier, S. (2016). Open data portal quality comparison using ahp. *Proceedings of the 17th International Digital Government Research Conference on Digital Government Research*, 397–407. https://doi.org/10.1145/2912160.2912167

Lallemand, C., & Gronier, G. (2015). *Méthodes de design ux: 30 méthodes fondamentales pour concevoir et évaluer les systèmes interactifs*. Editions Eyrolles.

Likert, R. (1932). A technique for the measurement of attitudes. *Archives of psychology*.

Marras, M., Manca, M., Boratto, L., Fenu, G., & Laniado, D. (2018). Barcelonanow: Empowering citizens with interactive dashboards for urban data exploration. *Companion Proceedings of the The Web Conference 2018*, 219–222.

Natephra, W., & Motamedi, A. (2019). Live data visualization of iot sensors using augmented reality (ar) and bim. https://doi.org/10.22260/ISARC2019/0084

Nielsen, J., & Landuer, T. (1993). A mathematical model of the finding of usability problems. interchi'93. *Conference on Human Factors in Computing Systems*, 206–213.

Obama, B. (2009). Transparency and open government. https://obamawhitehouse.archives.gov/the-press-office/transparency-and-open-government

Ojo, A., Curry, E., & Zeleti, F. A. (2015). A tale of open data innovations in five smart cities. *2015 48th Hawaii International Conference on System Sciences*, 2326–2335. https://doi.org/10.1109/HICSS.2015.280

Open Data Handbook. (2015). What is open data. https://opendatahandbook.org/guide/en/what-is-open-data/

Puussaar, A., Johnson, I. G., Montague, K., James, P., & Wright, P. (2018). Making open data work for civic advocacy. *Proceedings of the ACM on Human-Computer Interaction*, *2*(CSCW), 1–20.

Smith, L. (2017). Benefits of open data for smart cities. https://hub.beesmart.city/solutions/benefits-of-open-data-for-smart-cities

Spencer, D. (2004). Card sorting. *Boxes and arrows*, *7*.

Unwin, D. J. (1981). *Introductory spatial analysis*. Taylor & Francis.

van Stegeren, J., & Theune, M. (2018). Towards generating textual game assets from real-world data. *Proceedings of the 13th International Conference on the Foundations of Digital Games*, 1–4.

Ward, M. (2009). Linking and brushing. https://doi.org/10.1007/978-0-387-39940-9_1129

Warren, R., & Champion, E. (2014). Linked open data driven game generation. *International Semantic Web Conference*, 358–373.

Wolff, A., Barker, M., & Petre, M. (2017). Creating a datascape: A game to support communities in using open data. *Proceedings of the 8th International Conference on Communities and Technologies*, 135–138. https://doi.org/10.1145/3083671.3083686

Wood, J. R., & Wood, L. E. (2008). Card sorting: Current practices and beyond. *Journal of Usability Studies*, *4*(1), 1–6.

Zuiderwijk, A., & de Reuver, M. (2021). Why open government data initiatives fail to achieve their objectives: Categorizing and prioritizing barriers through a global survey. *Transforming Government: People, Process and Policy*.

Zuiderwijk, A., Janssen, M., Choenni, S., Meijer, R., & Alibaks, R. S. (2012). Socio-technical impediments of open data. *Electronic Journal of e-Government*, *10*(2), pp156–172.

Appendices

List of Figures

List of Tables

A Questionnaire

A.1 Évaluation du portail open data de Namur

Ce document décrit la méthodologie d'évaluation du portail open data de Namur. Elle a pour objectif de mesure l'utilité et la facilité d'utilisation des fonctionnalités de visualisation de données disponibles sur le portail. Ce sont en effet les deux facteurs qui déterminent l'intention d'utilisation selon le modèle TAM.

L'évaluation se déroule sous la forme d'un test utilisateur impliquant deux scénarios. Des questions ouvertes et un bref questionnaires sont utilisés pour debriefer après la réalisation des scénarios.

A.1.1 Questions sur le profil du participant

Quel est votre genre ?

- ○ Femme
- ○ Homme
- ○ Je ne souhaite pas le préciser

Quel âge avez-vous ?

- ○ Moins de 18 ans
- ○ Entre 18 et 25 ans
- ○ Entre 26 et 35 ans
- ○ Entre 36 et 45 ans
- ○ Entre 46 et 55 ans
- ○ Entre 56 et 65 ans
- ○ Plus de 65 ans

Quel est le diplôme le plus élevé que vous ayez obtenu ?

- ○ Aucun diplôme
- ○ Enseignement primaire
- ○ Enseignement secondaire
- ○ Enseignement supérieur (bachelier, master, ou équivalent)
- ○ Doctorat

Quelle est votre situation professionnelle ?

- ○ Étudiant.e
- ○ Employé.e ou ouvrier.ère
- ○ Indépendant.e
- ○ Sans emploi
- ○ Retraité.e

Où situeriez-vous votre expertise par rapport aux activités suivantes ?

	Très faible	**Faible**	**Moyenne**	**Élevée**	**Très élevée**
Usage courant des technologies (téléphone, Internet)	○	○	○	○	○
Analyse de données (utilisation de tableurs, statistiques)	○	○	○	○	○
Développement informatique (codage ou programmation)	○	○	○	○	○

A.1.2 Scénario 1

En tant que futur commerçant vous désirez vous installer dans une rue fréquentée. La ville de Namur a installé des capteurs de piétons dans certaines rues, cela pourrait vous aider dans notre choix. Vous décidez de consulter ces données sur le portail open data de la Ville.

https://data.namur.be/explore/dataset/namur-compteurs-pietons-comptes/table/?sort=date

Plusieurs emplacements intéressants sont disponibles à la location, à la rue de l'Ange et à l'avenue de la Gare, il vous faudra donc comparer la fréquentation piétonne des deux rues et son évolution sur les dernières années, et voir s'il y a des différences significatives. Un autre aspect important est votre événement phare, représentant 10% de votre chiffre d'affaires annuel, que vous organisez le 20 décembre à 13h00. Les fréquentations des rues pour ce jour et cette heure vous intéresse donc tout particulièrement.

Questions sur les méta-informations

[Dans la mesure du possible, le répondant doit se souvenir de ses informations ou en avoir pris note. Sinon, il est invité à retourner sur le portail.]

En quelques mots, que représentent ces données?

De quand date leur mise à jour?

Combien de rues sont couvertes?

Quelle est la couverture temporelle des données?

Dans quelle mesure la page "Informations" correspond-elle à ce que vous vous attentiez à y trouver?

- [Relance] Y voyez-vous des éléments superflus?
- [Relance] Quels éléments avez-vous trouvé les plus utiles?
- [Relance] Quels éléments manquent selon vous?

Dans quelle mesure la page "Présentation" correspond-elle à ce que vous vous attentiez à y trouver?

- [Relance] Y voyez-vous des éléments superflus?
- [Relance] Quels éléments avez-vous trouvé les plus utiles?
- [Relance] Quels éléments manquent selon vous?
- [Relance] Apporte-elle quelque chose par rapport à la page "Informations"?

Questions sur les données

Y a-t-il des différences en terme de fréquentation piétonne entre la rue de l'Ange et l'avenue de la Gare?

Quelle est la fréquentation de la rue de l'Ange le 20 à 13h00 pour la dernière année recensée? Est-elle plus ou moins élevée que celle de l'avenue de la Gare?

Comment avez-vous fait pour obtenir ces informations?

- [Relance] Si vous n'avez pas consulté la page "Tableau", pourquoi?
- [Relance] Est-ce que le tableau disponible sur la page "Tableau" vous a suffit à trouver l'information dont vous aviez besoin?
- [Relance] Avez-vous trouvé la présentation en tableau utile?
- [Relance] Avez-vous trouvé la présentation en tableau facile à comprendre?

- [Relance] Si vous n'avez pas consulté la page "Analyse", pourquoi?
- [Relance] Est-ce que les graphiques disponibles sur la page "Analyse" vous ont suffit à trouver l'information dont vous aviez besoin?
- [Relance] Avez-vous trouvé la présentation en graphiques utile?
- [Relance] Avez-vous trouvé la présentation en graphiques facile à comprendre?
- [Relance] Avez-vous trouvé la présentation en graphiques facile à utiliser? [Dans le sens, interagir avec]

Repensez au scénario proposé, qui consiste à décider où implanter un commerce. Les informations que vous avez obtenues vous semblent-elles suffisantes?

- [Relance] Est-ce que des données sur le comptage des piétons mais plus complète ou plus à jour vous auraient été utiles?
- [Relance] Est-ce que d'autres outils d'analyse ou de visualisation des données sur le comptage des piétons vous auraient été utiles?
- [Relance] Est-ce que des données sur autre chose que le comptage des piétons vous auraient été utiles?

A.1.3 Scénario 2

En tant que vendeur de panneaux solaires vous souhaitez déterminer dans quel quartier de Namur il est plus opportun de prospecter. Vous voulez en priorité démarcher les habitants vivant à un endroit à potentiel photovoltaïque élevé. La chance a frappé à votre porte, puisque que la Ville de Namur a mis à disposition sur son portail open data des données sur le potentiel photovoltaïque par quartier, que vous vous empressez de consulter.

https://data.namur.be/explore/dataset/namur-potentiel-photovoltaique-par-quartier/table/

Vous souhaitez déterminer les trois quartiers prioritaires pour votre prochaine tournée.

Questions sur les méta-informations

[Dans la mesure du possible, le répondant doit se souvenir de ses informations ou en avoir pris note. Sinon, il est invité à retourner sur le portail.]

En quelques mots, que représentent ces données?

De quand date leur mise à jour?

Combien de quartiers sont couverts?

Quelle est la couverture temporelle des données?

Dans quelle mesure la page “Informations” correspond-elle à ce que vous vous attentiez à y trouver?

- [Relance] Y voyez-vous des éléments superflus?
- [Relance] Quels éléments avez-vous trouvé les plus utiles?
- [Relance] Quels éléments manquent selon vous?

Questions sur les données

Quels sont les trois quartiers que vous devrez visiter lors de votre prochaine tournée commerciale?

Comment avez-vous fait pour obtenir ces informations?

- [Relance] Si vous n’avez pas consulté la page “Tableau”, pourquoi?
- [Relance] Est-ce que le tableau disponible sur la page “Tableau” vous a suffit à trouver l’information dont vous aviez besoin?

- [Relance] Avez-vous trouvé la présentation en tableau utile?
- [Relance] Avez-vous trouvé la présentation en tableau facile à comprendre?

- [Relance] Si vous n'avez pas consulté la page "Analyse", pourquoi?
- [Relance] Est-ce que les graphiques disponibles sur la page "Analyse" vous ont suffit à trouver l'information dont vous aviez besoin?
- [Relance] Avez-vous trouvé la présentation en graphiques utile?
- [Relance] Avez-vous trouvé la présentation en graphiques facile à comprendre?
- [Relance] Avez-vous trouvé la présentation en graphiques facile à utiliser? [Dans le sens, interagir avec]

- [Relance] Si vous n'avez pas consulté la page "Carte", pourquoi?
- [Relance] Est-ce que les graphiques disponibles sur la page "Analyse" vous ont suffit à trouver l'information dont vous aviez besoin?
- [Relance] Avez-vous trouvé la présentation en graphiques utile?
- [Relance] Avez-vous trouvé la présentation en graphiques facile à comprendre?
- [Relance] Avez-vous trouvé la présentation en graphiques facile à utiliser? [Dans le sens, interagir avec]

- [Relance] Si vous n'avez pas consulté la page "Vue personnalisée", pourquoi?
- [Relance] Est-ce que le tableau disponible sur la page "Vue personnalisée" vous a suffit à trouver l'information dont vous aviez besoin?
- [Relance] Avez-vous trouvé le contenu présenté utile?
- [Relance] Avez-vous trouvé le contenu présenté facile à comprendre?

Vous réalisez soudainement que la population par quartier pourrait être une information utile également. En effet, vous souhaitez aussi prospecter dans les quartiers où il y a le plus d'habitants. Vous décidez donc de créer un graphique mettant en perspective le potentiel photovoltaïque et la population, par quartier.

- [Guide] Cherchez un jeu de données contenant le nombre d'habitants par quartier.
- [Guide] Allez dans le menu "Éditer en mode avancé", disponible sur la page "Analyse".

- [Guide] Ajoutez le jeu de données précédemment trouvé au graphique

Comment vous y prendriez-vous?

- [Relance] En utilisant les fonctionnalités du portail?
- [Relance] En extrayant les données du portail et en créant le graphique par ailleurs?

Avez-vous trouvé la démarche intuitive, complexe?

A.1.4 Rétrospective

Veuillez indiquer dans quelle mesure vous êtes d'accord avec les affirmations suivantes.

	Pas du tout d'accord	**Pas d'accord**	**Ni en accord, ni en désaccord**	**D'accord**	**Tout à fait d'accord**
L'utilisation des fonctionnalités de visualisation sur le portail améliorera mon travail	○	○	○	○	○
L'utilisation des fonctionnalités de visualisation sur le portail améliorera mon efficacité	○	○	○	○	○
L'utilisation des fonctionnalités de visualisation sur le portail augmentera ma productivité	○	○	○	○	○
Je trouve que les fonctionnalités de visualisation sur le portail sont un outil utile dans mon travail	○	○	○	○	○

	Pas du tout d'accord	**Pas d'accord**	**Ni en accord, ni en désaccord**	**D'accord**	**Tout à fait d'accord**
Mon interaction avec les fonctionnalités de visualisation sur le portail est claire et compréhensible	○	○	○	○	○
Je trouve qu'il est facile d'utiliser les fonctionnalités de visualisation sur le portail faire ce que je veux	○	○	○	○	○
L'interaction avec les fonctionnalités de visualisation sur le portail ne demande pas beaucoup d'effort mental	○	○	○	○	○
Je trouve que les fonctionnalités de visualisation sur le portail sont faciles à utiliser	○	○	○	○	○

De manière générale qu'est-ce que vous auriez aménagé sur le portail pour rendre la réalisation des scénarios plus simple?

Avez-vous des dernières remarques ou observations à partager?

Merci pour votre participation, vous avez gagné 17 points de karma!

B Expert assessment - heuristic evaluation

https://data.namur.be/explore/dataset/namur-compteurs-pietons-comptes/table/?sort=date

Menu du haut :

- Trop peu de contraste dans les entrées du menu (https://webaim.org/resources/contrastchecker/)

Affichage du Nombre d'enregistrements :

- Manque un contexte pour comprendre le nombre affichés (de quoi parle t'on, sur quelle période ? ...) Onglets :
- Il manque des titres sous les onglets pour décrire le contenu.

Problème d'affordance :

- Le nombre d'enregistrement, le titre " Filtre ", les boutons de réseaux sociaux utilisent le même type de représentation mais on des fonctionnalités différentes (même forme, couleur de fond, couleur d'indication ...) risque de confusion sur la fonction. Hiérarchie de l'information :
- Pas de claire hiérarchie de l'information, tout les onglets sont au même niveau, l'information est " à plat " sans réelle notion de hiérarchie.
- Pourquoi ne pas regrouper Tableau et Analyse sous " Données " et proposer le filtre dans ce contexte uniquement et pas sur toute la page (intérêt du filtre dans le contexte information, présentation ... ?).
- Export devrait se trouver avec les données
- Api devrait se trouver avec les données

Problème dans les filtres :

- Pas d'explication sur la finalité du filtre (" rechercher " Quel est l'objet de la recherche ?, " date et heure " Quel est l'action, que va-t-on afficher comme information une fois les dates sélectionnées ?

- Difficulté à comprendre le fonctionnement global (les champs et date le picker sont trop proches et sans explications, fonctionnent t'ils ensemble ?) Qu'est ce qui déclenche l'action ?
- Si l'onglet Tableau et Analyse ne sont pas sélectionnés, les actions dans les filtres ne sont pas visibles sur la partie d'affichage (cf proposition de regrouper Analyse et Tableau sous un même onglet
- Le survol de l'option "Tout effacer" n'est pas lisible (aucun contraste, couelur quasi identique). (https://webaim.org/resources/contrastchecker/)
- Dans les filtres en cours, pas d'indication sur la possibilité de ne supprimer qu'un seul filtre (l'option existe)

Onglet analyse :

- Manque un titre clair sur le contenu affiché
- Pas d'aide sur l'utilisation de l'outil d'analyse, ni d'exemples, vue trop, Pourquoi ne pas proposer une vue simplifiée avec des cas d'usages pour les utilisateurs qui n'ont pas de connaissances dans le traitement des données (option de filtres sur les croisements de données les plus courants, à alimenter avec les interviews terrain + analyses des requêtes les plus utilisées).
- Une fois que " Editer en mode avancé " est sélectionné pas de retour en arrière (à part avec le back du navigateur)
- Utilisation d'un vocabulaire technique pas usuel pour la cible (Ventiler les séries ..., Empilement, Treemap, Ecart type, percentil ...)
- Titre éléments du formulaire difficile à comprendre pour non expert (Nombre de points max.)

Tableau :

- Manque un titre clair sur le contenu affiché
- Sens de tri (chevrons à côté du titre) dans les colonnes pas clair (problème de contraste cf norme WCAG)
- La zone de scroll horizontal est peu lisible, si zone de scroll en dessous de la zone de visible, l'utilisateur ne peut pas la voir
- La zone de " Partage, Intégrer, Widget ", l'utilisateur ne sait pas ce qu'il va intégrer (description ? preview ?)

Onglet présentation (conformité au model mental de la tâche) :

- Pourquoi ne pas le mettre en premier, savoir de quoi on parle avant d'afficher les données.

Onglet informations

- Difficile de savoir avant de cliquer la différence de contenu entre Information et Présentation

Onglet Export :

- Mentionner à qui s'adresse le contenu (pour les experts)
- Il manque la description du contenu des exports (Quoi ? Quelle période ? Structure ?)

Onglet API :

- Mentionner à qui s'adresse le contenu (pour les experts)
- Utilisation d'un vocabulaire familier (Jetez un oeil à la documentation de l'API — Consultez la documentation ...)

Bouton "Connexion" :

- Pas d'information sur la destination, but de se bouton (ex : " connexion à " + préciser)

https://data.namur.be/explore/dataset/namur-compteurs-pietons-comptes/table/?sort=date

Idem Lien précédent Spécificité : Onglets :

- Contenu " Vue personnalisée " à un titre, les autres devraient aussi en avoir un.

Onglet Carte :

- Titre onglet pas clair, de quelle carte parle-t'on ?

Filtres :

- Idem remarque lien précédent

- Dans le cas du code quartier, à quoi correspondent les codes ?
- Retour en arrière dans les filtres, pas clair (Nom de quartier et Code de quartier, il faut cliquer sur la sélection pour revenir en arrière, pourquoi ne pas afficher une option de retour ou d'annulation du filtre à proximité de chaque filtre).