

# The Impact of Impediments on Open Government Data Use: Insights from Users

Jonathan Crusoe  
Linköping University  
Linköping, Sweden  
jonathan.crusoe@liu.se

Anthony Simonofski  
University of Namur  
Namur, Belgium  
anthony.simonofski@unamur.be

Antoine Clarinval  
University of Namur  
Namur, Belgium  
antoine.clarinval@unamur.be

Elisabeth Gebka  
University of Namur  
Namur, Belgium  
elisabeth.gebka@unamur.be

**Abstract**—Open Government Data (OGD) is interoperable data that is published on the Internet by public organizations and can be freely used and redistributed by users. OGD is expected to result in several benefits, such as innovative products and services, collective problem solving, and equal access to information. However, several impediments complicate the realization of the benefits. These impediments can impact the process of using OGD that consists of the four phases: motivation, search and evaluate, access and prepare, and aggregate and transform.

In this paper, we examine the case of a data science project conducted by master's students to understand how impediments impact their use. In order to study this case, we use a mixed-method approach combining a questionnaire sent to the 30 students and nine in-depth interviews.

The analysis of the data allowed us to identify and discuss the impact of several impediments on use, such as difficulties in finding an idea and the lack of longitudinal data. Subsequently, we were able to give recommendations to publishers to improve their OGD portals, infrastructure, and data as well as opening up research directions in the OGD field.

**Index Terms**—Open government data, User process, Impediment, Mixed-method, Impact

## I. INTRODUCTION

Open government data (OGD) is interoperable data published on the internet by public organizations to be freely used and redistributed by anyone [1, 14, 17]. The public organizations, who collect and share data, are referred to as publishers, while those who reuse the data are called users [43]. OGD is a change from the traditional system where those who collect and analyze data have been the same [35]. Users can be city managers, businesses, citizens, students, developers, researchers, Non-Governmental Organizations (NGOs), civil society organizations, or journalists [24, 30]. OGD is a collaborative effort to realize several benefits, such as the creation of a new sector, innovative products and services, collective problem solving, and equal access to data [20, 22, 33]. However, both publishers and users can experience impediments in the process, such as low data quality, restrictive legislation, paywalls or loss of income, and no interoperability between systems [10]. The lack of use has been noted as a problem for OGD [16, 19, 23, 30, 37, 40, 41]. This lack means that the opportunities of OGD are not realized, but also a risk for publishers to invest in something that never reaches fruition.

Central to use is the user process, which can be depicted in several ways. Normally, the user process is depicted together

with the publisher process as interdependent [e.g., 1, 42]. In such instances, the processes form a circle where the publisher shares data with the user who later responds with feedback. The process is supported by infrastructure (e.g., application programming interface (API) and OGD portals), which needs to be designed to support users' and publishers' activities. The publishers and the users can experience impediments, such as low-quality data and machine-unreadable formats. Impediments are difficult to solve as they can be connected to each other, be caused by activities earlier in the process, and vary in severity [4, 18, 42].

In this study, the focus is on the user process, as the lack of use and the impediments indicate inadequate infrastructure and a need for further insights into the user process. The inadequate infrastructure can cause bottlenecks in the process, such as unfindable data and inability to combine data [44]. To overcome current impediments to usage, [6] have called for research on the users' motivations, resources, and capabilities. In our research, we take a step towards understanding how impediments impact users, and, therefore, how they can influence users' motivations and need for resources and capabilities.

The purpose of this study is to understand how impediments impact OGD users as they progress through the OGD process. The study seeks to identify crucial impediments and understand how they may change the behaviour of the user. For this purpose, we use a user process framework [9] as a guide for a mixed-method research approach. Our research participants were 30 master's students divided into groups of three, performing a data science project on OGD portals. The study results can give an insight into the current severity of user impediments and inadequacies of the OGD infrastructure. Based on the identification of these impediments, we give recommendations to publishers to improve the OGD portals, data, and infrastructure and open some research directions in the field. The research question is formulated as follows: "*How can impediments impact users as they progress through the OGD process?*"

This article includes the following sections. We first present the background with the user process framework, followed by our methodology. Then, we present the results. Accompanied is a discussion of the results, limitations, and future research. The paper ends with a conclusion that summarizes this study's contributions.

## II. BACKGROUND

There are several types of OGD, that can be used for different purposes and in different ways. Some common categories of OGD are finance and economy, environment, health, energy, education, transportation, infrastructure, employment, and population [36]. In more practical terms, OGD can be national budgets, environmental measurements, weather reports, public transit schedules, crime incident reports, postal codes, consumer complaints, and information about population [1, 2, 28]. Using OGD is believed to lead to benefits, such as increased governmental transparency, democratic accountability, external problem-solving, and evidence-based policy-making [7, 15, 20, 22, 38]. However, publishing OGD for reuse is associated with various risks: violations of privacy, misinterpretations, and misuse [3, 43]. Moreover, use can come in multiple forms, such as studying voting history for specific facts, making graphs or texts to illustrate data, developing searchable interactive digital maps, and combine and clean data to republish through an API [12]. There is a variation in the data as input, use as process, and information, facts, products, and services as output.

The variations in the process of use have been studied by several authors. An early attempt was to connect the input with the output (with a focus on the latter); data to fact, data to information, data to interface, data to data, and data to service [12]. On the other hand, the process can be depicted as a linear value chain, starting with data aggregation, data analysis, and ending with data services and products [29]. Another approach is to view OGD as a cycle where the publisher process and the user process are interdependent. The user process can then be divided into finding the data, analyzing the data, processing data, and discussing and providing feedback on data [42], or, alternatively, data discovery, data exploration, and data exploitation [1]. Thirdly, it is possible to divide the process into several activities and group them into roles, such as data extractor and transformer, data analyzer, and user experience provider [26]. In conclusion, the outcome-based approach misses the variations in the input and the processes, the process approaches tend to miss the variations in the outcomes and the activities, and the role approach solves the aforementioned issues but is less detailed about the activities. In this paper, the user process is viewed as a linear process with variations in the activities and output based on the purpose and data. This approach follows the user process framework of Crusoe and Ahlin [9]. The authors' framework is a recent attempt to synthesise previous research and empirical data on the user process, as such it contains the findings of [e.g., 1, 12, 25, 26, 29, 42, 43, 44]. Their user process consists of phases with activities and relating impediments: *motivation*, *search and evaluate*, *access and prepare*, and *aggregation and transformation* (see Table I).

## III. METHODOLOGY

In this section, we first discuss the context of the study. Then, we detail the data collection techniques we applied to measure our constructs and explore practice. Finally, we

explain how we analyzed the collected quantitative and qualitative data.

Here, we note that this research is part of verifying and further developing the framework proposed by [9]. From this framework, we selected impediments to be studied based on their perceived possible impact on use.

### A. Context

In order to collect information about the impediments that users can experience when using OGD, we examined a specific data science project conducted by 30 master's students between October and December 2018. We argue that these students have high digital literacy and are thus representative of the typical OGD user population. They were asked to use OGD from the cities of Namur<sup>1</sup>, London<sup>2</sup>, Paris<sup>3</sup> or New York<sup>4</sup> to develop an application valuable for the citizens or the public servants working in the administration. The guidelines for the project were reduced to a minimum to stimulate creativity and the real-life use of the OGD portals. The following constraint was nonetheless imposed on the students. The output of their project should be transferable to the city of Namur, provided that the city acquires and makes available the necessary data. In this way, the projects' outputs consisted not only of solutions directly applicable to Namur but also of prospects valuable for the city officials.

### B. Data Collection

To collect data from our case, we followed a mixed-method approach that combines quantitative (one questionnaire) and qualitative (semi-structured interviews) methods. Johnson and colleagues [21] argue that a combination of methods allows having informative, complete, balanced, and useful research results. We chose this combination of methods in order to reach a complete view of the impediments experienced by users. The quantitative insights helped us to understand the importance of some impediments whereas the qualitative insights gave specific details about the impediments. Furthermore, the quantitative insights helped us to frame the interview guide in order to ask questions about the most important reported impediments.

1) *Questionnaire*: We structured the questionnaire based on the user process previously detailed by [9]. This process is divided into four phases that are described in Table I. In the questionnaire and this paper, the *motivation* phase is referred to as the *start* phase. Figure 1 summarizes the four phases, gives the overall structure of the questionnaire, and details the three constructs we evaluated:

- **Resource allocation**: this construct evaluates the amount of time allocated and people involved for each phase. Respondents were asked to rank the four phases according to the time it took to complete them, and also according to how much they contributed.

<sup>1</sup><https://data.namur.be/explore/>

<sup>2</sup><https://data.london.gov.uk/>

<sup>3</sup><https://opendata.paris.fr/explore/>

<sup>4</sup><https://opendata.cityofnewyork.us/>

TABLE I  
THE FOUR PHASES OF THE USER PROCESS FRAMEWORK (BASED ON [9])

Phase	Activities	Sources of impediments
<i>Motivation</i> : The user needs to be aware and motivated to use OGD. In this phase, the user is trying to identify some way to use OGD.	Discover OGD and identify where OGD can be used.	Public advertisement, understand how OGD can be used, and examples of use.
<i>Search and evaluate</i> : The user is searching on the Internet, OGD portals, and publishers' websites for promising data to use. Once promising data has been identified it can be evaluated to determine if it fits with the objectives of the <i>motivation</i> phase.	Engine searching, portal searching, browsing the publisher's website, pre-evaluate data on an OGD portal, and evaluate data on the publisher's website.	Search features, presentation of search results, metadata, must download the data to evaluate, tools to explore and analyze data, language, and licences.
<i>Access and prepare</i> : The user has the goal to acquire promising data and transform it into usable data.	Manual access (e.g., download PDF), automated access (e.g., API), and prepare data.	Access method, documentation, filtering, data format, data quality, accessibility, registration, and support.
<i>Aggregate and Transform</i> : The user has the objective to combine data and transform the new dataset into information, a product, or service.	Aggregate data, analyze data for information, and develop a product or service.	Combine data, quality variations, availability variations, supporting tools, longitudinal data, data infrastructure, and domain knowledge.

- **Perceived usefulness**: this construct evaluates to what extend the phases of the process are perceived as useful for the project's output by the users. A 5-point Likert measure instrument ranging from *not useful at all* to *very useful* was used.
- **Perceived difficulty**: relying on a 5-point Likert measurement instrument, the questionnaire measures the severity of the impediments reported in [9]. We used the term "barrier" in the questions so that the students can answer more easily.

In order to measure the *perceived difficulty*, the severity of the impediments was evaluated by relying on the instrument previously described by [4]. Indeed, the authors also measured some challenges faced by users with the following measure instrument:

- 1) **Not a barrier**: it was easy to use the data.
- 2) **Somewhat of a barrier**: it was still possible to use the data.
- 3) **Moderate barrier**: it was difficult to use the data.
- 4) **Serious barrier**: it was extremely difficult to use the data.
- 5) **Extreme barrier**: it was impossible to use the data.

Besides the constructs, the questionnaire includes contextual questions on the respondents' background, the skills they acquired for their project, and their confidence with programming, data analysis, and OGD portals. The questionnaire ends with a broad open question, allowing respondents to share insights on the class project and OGD in general. The complete questionnaire is available in the appendix to this paper.

2) *Interviews*: In order to complete and better analyze the results from the questionnaire, we conducted in-depth interviews with the nine students. We applied quota sampling to select the interviewees. The quota was based on students' study orientation (computer science, management, or mathematics). We hypothesized that the impediments faced by students throughout the project might vary according to their background. Working with OGD is an interdisciplinary process that can involve idea development, data analysis, and coding. Due to the differences in their interests and curriculum, students from a given background (e.g., mathematics) could be

[RESOURCES]	[RESOURCES]	[RESOURCES]	[RESOURCES]
Start	Search & Evaluate	Access & Prepare	Aggregate & Transform
[USEFULNESS]	[USEFULNESS]	[USEFULNESS]	[USEFULNESS]
[DIFFICULTY]	[DIFFICULTY]	[DIFFICULTY]	[DIFFICULTY]

Fig. 1. Questionnaire constructs. For each phase of the user process framework, the questionnaire measures the *perceived difficulty*, the *resource allocation*, and the *perceived usefulness*

more experienced in some activities (e.g., data analysis). The questionnaire responses were in line with this hypothesis. We limited our study to nine interviews to respect the sampling method (only three mathematics students were involved in the whole project) and because of their limited availability due to exam constraints.

The structure of the interview guide is similar to that of the questionnaire. The interview guide starts with introductory questions on the overall user process, as we were interested to compare the user process framework with the activities conducted by students in their project. Moreover, the severest impediments from each phase that were reported in the questionnaire results were discussed with the interviewees in order to collect deeper insights. The interview guide ends with retrospective questions on the project. This latter part includes, among other things, questions regarding the students' motivation to use OGD again and what support they want from OGD publishers. The detailed interview guide is available in the appendix of this paper.

### C. Data Analysis

1) *Questionnaire*: The questionnaire results consist of the items evaluated on a 5-point Likert measurement instrument for the *perceived difficulty* and the *perceived usefulness*. In order to have a central tendency measure for each of these items, the median was computed since the results consist of Likert-type data [5]. As for the *resource allocation* construct, it was measured by a ranking exercise on the four phases.

In order to obtain a representative ranking, the mode was computed for each phase.

2) *Interviews*: The interviews were analyzed with process and initial coding [31]. The analysis started with summarizing the interviews and then recording them in a data memo. Afterward, the researchers divided the data between each other based on the user process. Each researcher coded a specific phase for each interview. The coding started with skimming the interview to get a sense of the whole, then important sentences were highlighted based on the research question, after the highlights were coded using short sentences to retain context and conceptual relations. The codes were then inserted into a table divided by the interviewees and process phases. As the analysis progressed, researchers could write analytic notes to record insights and thoughts. All coding was conducted in the same cloud-based document, as such the researchers could follow each other's coding process and verify codes if needed.

#### IV. RESULTS

In total, 30 users completed the questionnaire. 22 of them have a management background, four study computer science, and four mathematics. In the following section, the quantitative results from the questionnaire regarding the three measured constructs are successively presented. Then, the discussion on the *perceived difficulty* is refined with the qualitative insights from the interviews. The findings from the questionnaire regarding the constructs are summarized in Table II.

##### A. Perceived Difficulty

Figure 2 presents the results from the questionnaire regarding the severity of the impediments. The severity was computed as the median answer of the 30 respondents. One can observe that none of the process phases is exempt from moderate barriers (median = 3). Hence, the users experienced difficulties in some activity of every phase, since a moderate barrier indicates that the use of data was difficult. In the *start* phase, users struggled to find an idea for OGD use. In the subsequent phase, they faced issues with metadata. Metadata is information about a dataset, such as collection methods and freshness. In the *access and prepare* phase, the problems lied in the quality of the data. Lastly, in the *aggregate and transform* phase, users had trouble with combining data, the variation in data availability, and complained about the lack of longitudinal data.

The questionnaire results also show that some expected impediments were no problem for the users (median = 1). This observation is rather encouraging, as it points out that potential impediments can have been solved by the OGD publishers considered in this study. It is, however, worth noting that these are concentrated in the *search and evaluate* and *access and prepare* phases. Hence, every impediment in the questionnaire for the *start* and the *aggregate and transform* phases impeded the use of OGD to some extent.

##### B. Resource Allocation

In the questionnaire, users were asked to rank the four phases by how much time they had invested in each one.

TABLE II  
QUESTIONNAIRE CONSTRUCTS FOR THE PHASES OF THE USER PROCESS  
FRAMEWORK (BASED ON [9])

Phase	Perceived difficulty	Resource allocation	Perceived usefulness
Start	Find an idea	4 (least)	Useful
Search and evaluate	Quality of metadata	3	Useful
Access and prepare	Data quality	2	Useful
Aggregate and transform	Data cannot be combined Data availability varies No longitudinal data	1 (most)	Very useful

The least time-consuming phase was the *start*, followed by the *search and evaluate*, the *access and prepare*, and, lastly, by the *aggregate and transform* phase, which is, therefore, the most time-consuming. The perceived individual contribution reported by the respondents follows this trend as well. The resource allocation and yield per phase thus appear to increase throughout the process.

Furthermore, 26 of the 30 users reported having to learn new skills required for the success of their project. This self-development was expected since most of the users have a management background. The learned skills were mainly about web-oriented languages (Javascript, CSS, HTML, web libraries), Python, and how to connect a Python script and a web page (the Flask framework was recurrent for this matter). This acquisition is one factor contributing to the high time allocation of the late phases.

##### C. Perceived Usefulness

All phases were perceived as useful (median  $\geq 2$ ). The *aggregate and transform* phase was considered as very useful (median = 1). This outcome was expected, since the final output of the project is delivered at this phase.

##### D. Qualitative Insights on Perceived Difficulty

In the following section, qualitative insights on impediments are presented following the phases of the user process framework.

1) *Start*: The in-depth interviews revealed that the process of ideation was mainly based on the personal needs, uses, or intuitions of the users. Then, users visited the OGD portals, looking for datasets, and, as a result, changed or gave up their first ideas as data was missing. Sometimes this was repeated several times. This process of divergence and convergence, using creative techniques like mind-maps and brainstorming, was constrained by the availability of datasets in the chosen portals. As one user expressed: “We’ve got a lot of ideas by looking at the datasets’ name in 20 to 30 minutes of brainstorming. But the difficulty was to realize them. When we opened the datasets, important information was missing. Thus, we had to give up some ideas.”

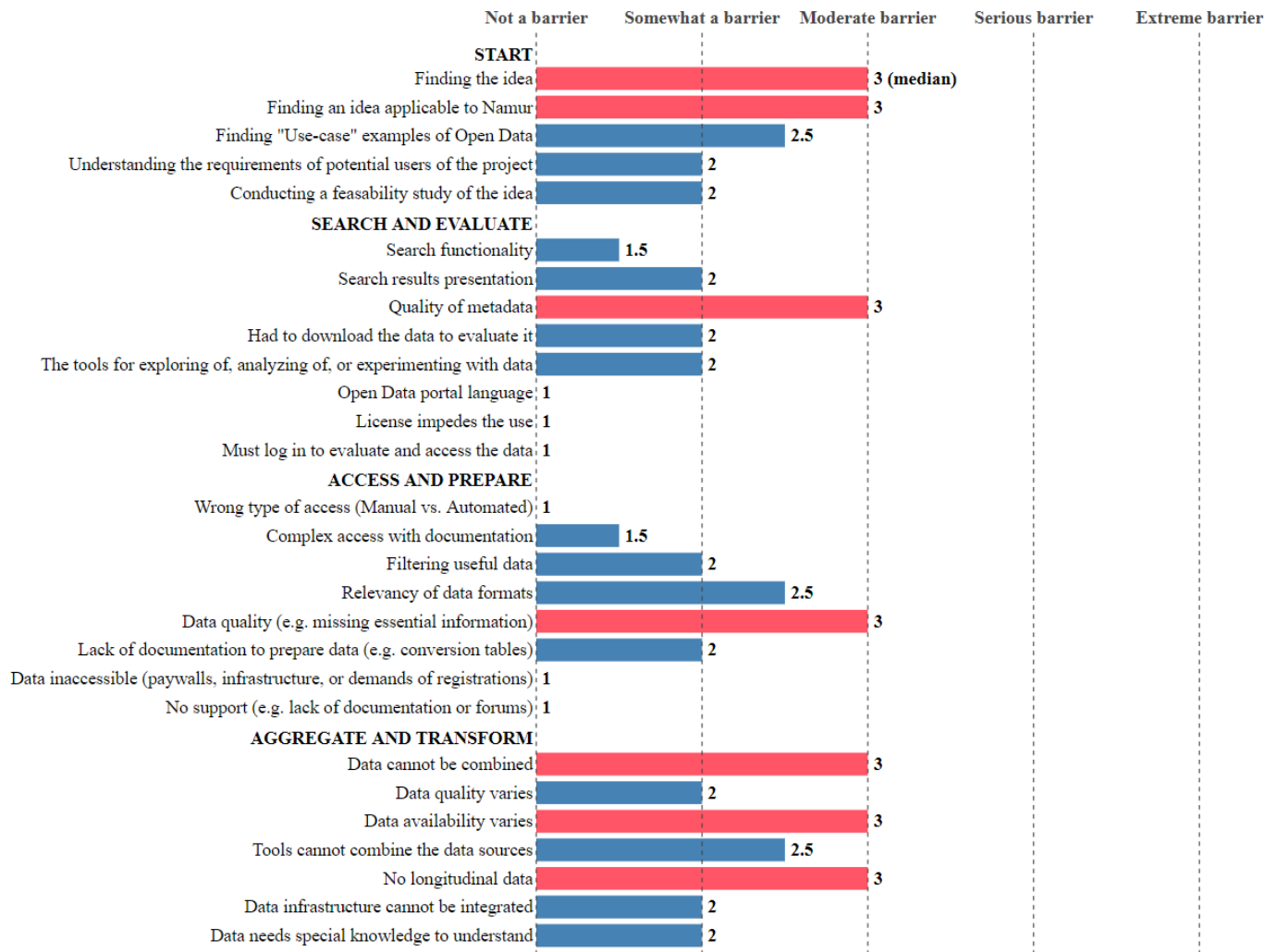


Fig. 2. Median for each impediment computed from the questionnaire results. The impediments colored in red are the most severe for their respective phase.

All the interviewees faced this issue, and it shaped their output. Two teams of users started directly from the OGD portal and with the combination of datasets to save time. Seven of them mentioned the limited number of datasets as a major constraint for their project development. They said they had more innovative ideas than data to develop them. As one user expressed: *"I have the impression that the OGD portals were pretty empty in valuable datasets, or the datasets could be usefully combined with other sets that we didn't have."*

Users were frustrated they could not exploit more than the city's OGD portal for their project since the data was too limited and they had little insights into Namur's challenges or priorities, citizens' needs, and market opportunities. The interviewees reported being discouraged by the abundance of existing apps, the lack of domain knowledge and given examples, the absence of precise demand, and the lack of useful datasets. As one user explained: *"It would have been helpful to have ideas from the citizens"*.

2) *Search and evaluate*: The interviews confirmed that the main impediments for the *search and evaluate* phase are in

the *evaluate* activity. No interviewee reported issues with the interface of the OGD portals they used. They found that the portals were well-designed in this regard and provided adequate features for searching data, such as filtering and suggestion of related datasets. The *evaluate* activity, however, was more challenging. The interviews confirmed that the main impediment in the phase is the lack or inadequacy of metadata, as observed when analyzing the questionnaire answers. Six users stated that they encountered issues with metadata. The interviews allowed us to refine this point, and uncovered that the *evaluation* of data can be examined at different granularity levels.

Firstly, three users faced issues with the **data features** (e.g., columns in a dataset) names, which they found uninformative and badly described. Unexplained columns meant that the user could not use the dataset. For example, one user encountered a population census dataset holding a number of girls and a number of women and he had to make calculations with census data to determine whether the girls are also counted as women. Another user complained: *"In our datasets, half to*

*three-quarters of the features we didn't know what they meant, and it wasn't explained anywhere [...] We completely ignored these columns".*

Secondly, metadata at the **dataset** level were mentioned by three interviewees as an issue. One user was disappointed that *"some dataset titles are awesome, but there is nothing exploitable in them."* In order to compensate for appropriate metadata, users had to resort to other evaluation methods. One user downloaded the data for further examination. Even worse, one user reported feeling frustrated, as the lack of informative metadata prevented exploiting the full potential of the available data. On the other hand, one user used the API request functionality to explore a dataset. Nonetheless, some features offered by portals were helpful in the evaluation of data. One user noted that the reuse examples comforted him with the potential of the data and increased his motivation to use data. Another user mentioned the usefulness of visualizations showing an overview of data as helpful.

3) *Access and prepare*: The in-depth interviews allowed us to explore the data quality impediment reported in the questionnaire. On a general note, the data quality was very fluctuating across datasets. There was a lot of data in the datasets that three interviewees respectively qualified as *"missing"*, *"irrelevant"* or *"corrupted"*. Two interviewees mentioned the lack of longitudinal data as the main impediment for regression analysis. Finally, a lack of consistency in the datasets was also reported by four interviewees. These impediments impacted the output of the work. One user noted the **impact** as *"because of the few relevant datasets our application was not as valuable as we wanted. Our application issues abnormal recommendations, such as: fewer trees will lead to better air quality!"*.

No major issue was reported in terms of data formats as most of the datasets were available in JSON and CSV. Some case-to-case issues were still reported, such as some irrelevant geographic data formats and the presence of the "string" data type instead of "integers" data type.

Also, the interviewees experienced no major problem with the API:s (one API was missing, and some errors were present in the requests). However, three respondents declared they did not use the API but rather downloaded the data directly. This approach may reveal that they did not see the added value of using this channel to access the data. It indeed often seems faster for the users to just download data. The approach may also indicate that the users did not seek to develop a sustainable solution.

4) *Aggregate and transform*: In the in-depth interviews, the interviewees explained some activities in the *aggregate and transform* phase. For the *aggregation* activity, users can select specific datasets based on criteria, select columns in the dataset, merge datasets to fill out gaps, and use scripts to clean the data. Merging can be time-consuming and involve a "one at a time" error solving approach. As one user told us: *"I spent a very large part of my time just on data aggregation since it was a monster mess."* For transformation, users can calculate averages, integrate data into web applications, think

about the end-user experience, and seek support. For example, for a specific technique, such as machine learning. However, there were also a few impediments. To make data combinable, sometimes keys needed renaming, data needed normalization as datasets were inconsistent (e.g., different metrics), and, in other occasions, several datasets had to be merged to form a complete dataset. As one user explained: *"[Datasets] didn't use the same way of localizing things. They all used neighborhood names, but the issue was that they didn't use the same neighborhood names. [...] It was nearly impossible to do the matching; they didn't use the same partitioning at all."* In the worst case, data lacked unique identifiers, datasets could use different unique identifiers, or feedback could be hard to leave, making an increase in data quality improbable. On the other hand, *transformation* could be impeded by slow data delivery, a need to add exceptions, and technical complexity. As a result, data needed more preparation to be combinable, which likely contributed to the high time consumption of this phase. One explanation is that when the users once started to transform data or combine data, they noted the work needed to prepare the data. Moreover, the in-depth interviews indicated that it was easier to use one dataset than several, or a single category of data (e.g., photos) than combining several.

## V. DISCUSSION

In the following section, we first discuss the quantitative and qualitative data from the users' perspective and give managerial recommendations for publishers.

### A. User Process: Theory versus Practice

Following the observations from our study, the user process can be represented as a U shape with the *access* activity at its bottom. The *start*, *search*, and *evaluate* are the left side of the U, while *prepare*, *aggregate*, and *transform* are the right. The U-shape is a representation of the number of and the severity of the studied impediments through the process to gives insight into where users can experience impediments impacting their use. In both the first and final phase (*start*, *aggregation*, and *transform*), none of the measured impediments were reported as *not a barrier*, while there was more variation in the middle. The interviews confirmed that the first and final phases of the user process were the most challenging to the users overall, but different in time consumption and perceived contribution. At the middle of the U-pattern, it is easier to find (*search*) and *access* OGD than to understand (*evaluate*) and make it usable (*prepare*). The impediments reported as *moderate barriers* after the *start* phase seem to come from the data itself or are concerned with the data. Their source is likely the publishers' different practices of data management [10]. Figure 3 illustrates the described U-pattern.

In the descending portion of the U-pattern, users tended to backtrack, while in the ascending part they tended to forward think. Backtracking involves going back from the current activity to a previous activity to resolve an issue in one or both. In our study, we observed users starting with an idea, and then visit the OGD portals for data, only to find none or few

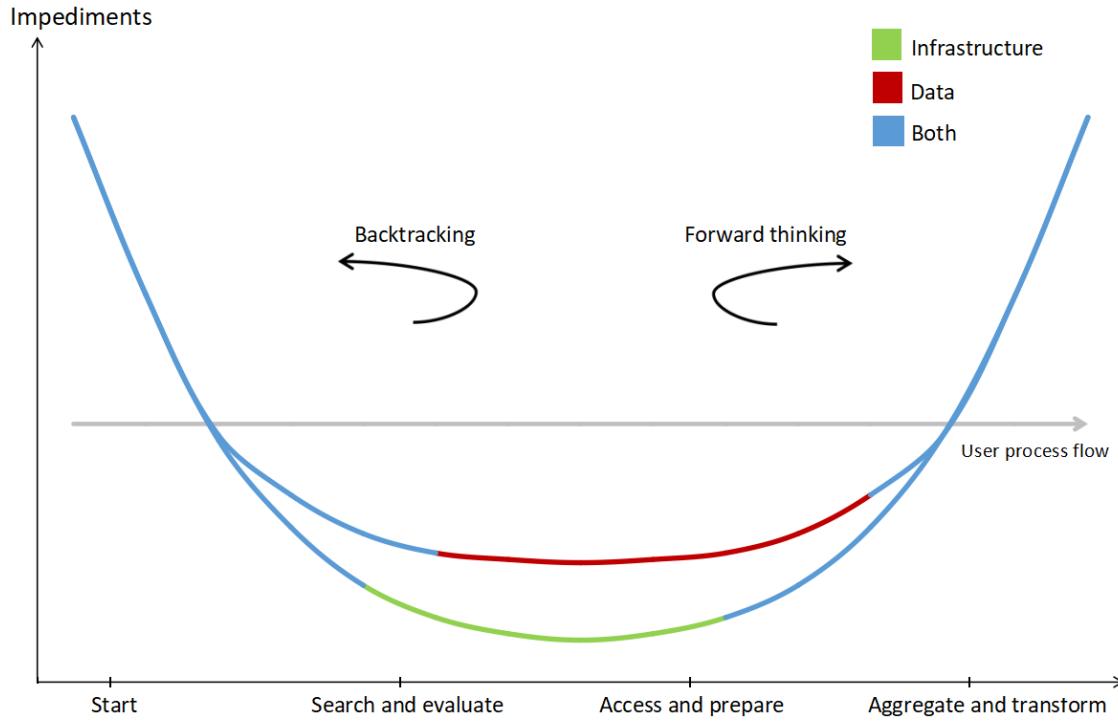


Fig. 3. U-pattern representation of the user process phases. In the first half of the U, users tended to backtrack whereas in the second half they tended to forward-think.

datasets fitting with their needs. Then, they returned to the *start* phase to change their idea to adapt to the new information. The lack of data and low quality of metadata are known impediments [10, 44]. Metadata is important for evaluating data and the lack of it can make an analytic investigation impossible [1, 27, 32, 42]. We observed that the availability of large and comprehensive sets of quality data, documentation, and metadata influences the experience and the motivation of the users, especially, in the earlier activities (*start* and *search*). The specific constraints of the class project revealed the importance to able to combine datasets and OGD sources from outside the city's OGD portal to keep stimulating the creativity of the users and maximize the outcome's value.

Forward-thinking involves thinking about what needs to be done in a certain activity to be able to succeed in the next activity. For example, cleaning, merging, and normalizing datasets in preparation for their aggregation. As we observed in our study, the low quality of data meant a need to dedicate more resources to cleaning and normalization of data [27]. The low quality is a known impediment [10, 18, 42] that can make data so difficult to analyze that it is no longer meaningful [39]. The impediments in the second portion of U-pattern likely contributed to the high time consumption we observed for the late phases, as users needed to fix several small problems sequentially and then transform the data into the outcome through development.

A common theme through the user process was the attrition on value experienced from impediments by users. In this paper, attrition is the reduction or wear down of the output's value.

Few datasets and incompatible data made certain ideas and solutions impossible to implement, meaning that users had to go for something else or ignore important data (we are not arguing that every idea is a valuable idea). The lack of unique identifiers can impede the ability to combine datasets and scale the product [27]. At the same time, incomplete data, the lack of longitudinal data, and varying data availability caused the outcome to have a lower value. The attrition encourages the use of one dataset or of a single category of data (e.g., photos). In worse cases, it can encourage the use of a few columns in a dataset. Impediments do not always attrit an idea; sometimes the user is able to adapt. The backtracking above is one example, another was when a user downloaded a dataset to further evaluate it as the metadata was lacking, and a third was when a user renamed keys and merged datasets to overcome data quality. However, adoption can still mean changes in the value of the outcome.

In previous research, similar to [4] we have identified that not all impediments are perceived to be equally severe. Impediments are often interrelated and can create such a high threshold that "open" data is still private in practice [20]. The threshold can make certain value creation impossible. In our case, impediments were discovered to also attrit the idea, as such they are more than a binary threshold. Moreover, the phases of the user process are interdependent [42] and we identified that activities in previous phases can be used to negate impediments in the subsequent phases, and not only create them. For the OGD user process, quality of data, availability of data, and infrastructure are important [20, 30], which



our research highlights. The large number of diverse data structures can make comparison and aggregation impossible [1].

### B. Managerial recommendations

The results show that the impediments can restrain and discourage even skilled users. A publisher aware of the user impediments can develop strategies to support users or change the infrastructure or data to ease usage - strategies that increase the likelihood of better return of investment. The areas of improvement are not limited to technical aspects. We suggest publishers to perform several actions to enhance the user experience and encourage projects development.

Firstly, we suggest publishers to introduce feedback mechanisms for the users to report incomplete data and lack of longitudinal data or metadata. A more advanced feedback mechanism could empower the users to improve the metadata and the data themselves.

Secondly, publishers can follow common standards in how the data is structured (e.g., formats and content) to allow for easier harmonization between datasets. This approach can lessen the time needed to prepare the data for aggregation and use.

Thirdly, publishers could enrich the portals with datasets from other public actors, institutions, and local organizations. The portals could also be enriched with data collected through sensors by several stakeholders. Enriched portals could enable the creation of more innovative ideas.

Fourthly, publishers can involve the users in the development of the portals. This participation can be implemented through several methods (e.g., workshops, interviews, living labs). The user participation can help publishers better understand the needs of the users and understand what support and features they need (e.g., tutorials, projects examples, reporting systems). This continuous exchange between users and publishers can support the improvement of both the portal and the re-use projects.

Finally, publishers could give users the opportunity to innovate and compete around real problems and needs. This opportunity can increase the match between output and real-world application. OGD can provide an opportunity to build a community around problem-solving in the city. One way to achieve this is the organization of hackathons by the city in collaboration with other stakeholders (universities and businesses).

## VI. LIMITATIONS

Our study presents some limitations mainly based on the specific sample of 30 students. However, we also argue that this limitation does not introduce fundamental bias in the study as students constitute a re-user group of OGD [24, 30]. Still, the limitations introduced by the selected sample are discussed below.

Firstly, our sample consisted of skilled students with computer science, management, or mathematics backgrounds. Therefore, their digital literacy is quite high. We argue that

this sample is representative of the OGD users but not of the full population. Thus, alternative studies with users with low or normal digital literacy could be interesting to perform in order to compare the findings. Alternative studies can also extend the sample to study OGD users at the city level to generalize the findings.

Secondly, the motivation of our sample was biased as it is an imposed class project that the students had to perform. Therefore, we were unable to capture information about what could motivate and drive users to visit the OGD platforms in the first place.

Thirdly, the conditions and rules of the class project itself introduced limitations. The students had limited time, and the project was based only on OGD city portals, which restricted the number of datasets. It does not represent the real practices of users, who can be used to web-scraping or combining datasets from various publishers. This constraint could misrepresent the perceived difficulties through the process, especially, at the *start* phase. As a consequence of time constraint and educational context, priorities were put on the search, access and aggregation phases. The objective of the students was to deliver on time a visualization or an application to pass the class project, whatever its market viability or power of advocacy. The time dedicated to each phase should consider variables in skills, the profile of the users, objectives of the project and avoid generalization.

## VII. FUTURE WORK

In this section, we describe solutions that can be performed to solve the issues raised in this paper. Also, the text will open for future research.

### A. Questionnaire validation

In the proposed questionnaire, several impediments were measured for each project phase. Although the objective was to identify individual impediments, we wondered if the scores for each impediment of a given phase could be used as a reliable measurement of the difficulty experienced for this phase. For this purpose, we computed the Cronbach's alphas [8] for each of the four phases. The results displayed by Table III shows that the alpha is acceptable for the *search and evaluate*, *access and prepare*, and *aggregate and transform* phases. These numbers encourage us to envision the design and validation of a questionnaire aiming to measure the difficulty for each phase of an OGD project as a future contribution. In particular, a reliable measurement scale measuring the difficulty encountered in the *start* phase needs to be developed in order for the current questionnaire to reliably assign a difficulty indicator for each phase.

### B. User Process Framework Refinement

The user process framework [9] is still under development, but has been peer-reviewed by colleagues and in a workshop. However, the framework needs further development. The identified impacts of the impediments need to be integrated into the framework as they are adaptations to impediments, rather



TABLE III  
CRONBACH'S ALPHAS FOR EACH PROJECT PHASE

Phase	Cronbach's alpha	Internal consistency
Start	0.466	Unacceptable
Search and evaluate	0.706	Acceptable
Access and prepare	0.760	Acceptable
Aggregate and transform	0.805	Good

than expected activities. At the same time, the interaction of phases can cause new impediments and behaviours, such as backtracking. This relationship means that the phases are more interdependent than a linear process. More importantly, we also found that the phases have different perspectives and different “problem pictures” (e.g., find an idea that is feasible at the start or increased time-consumption at the end).

#### C. Ecosystem View on Open Government Data

Valuable and sustainable reuse of the OGD not only depends on the users' capabilities or the publishers' infrastructure but on their interactions. The supply-driven approach of the OGD platforms shows its limits. The impediments and feedback reported by the interviewees outline the importance of considering them as stakeholders instead of data consumers. In an ecosystem with OGD as a shared resource, each stakeholder has his own perspectives and expectations: political, economical, technological, or bureaucratic [13]. Research should be conducted on the different support methods needed to stimulate collaboration between actors. What model of partnership or cooperation can enhance the value creation in multiple perspectives? What model of platform can promote interaction in the system? This last question supposes not only new features on a platform but a change in the knowledge and value creation paradigm [11].

#### D. Support for Idea Generation

The OGD reuse is a project with prospective benefits for citizens, communities, or markets. More research is needed to support the users' needs in the *start* phase since finding a valuable idea was reported as a key challenge by respondents. The application of several citizen participation methods, as described by [34], to stimulate the idea generation might be useful. Indeed, using creativity techniques to stimulate discussions among citizens or taking advantage of ideas submitted in participation software are two relevant examples. With these ideas, needs, and requirements properly elicited, it can allow the development of products and services aligned with actual issues faced by citizens.

### VIII. CONCLUSION

The publishing of Open Government Data (OGD) by public organizations has promising possibilities, such as the creation of a new data-driven sector and collaborative efforts toward innovative products and services, with data accessible to all. However, the actual use of OGD remains low [30] and its potential is largely under-exploited.

In this article, we selected a user process framework as a theoretical baseline to study the impediments experienced by 30 students who conducted a class project using OGD. Through a mixed-method research approach involving a questionnaire and semi-structured interviews, we were able to identify the impediments to OGD use students experienced, for each phase of their project. Our results show that finding an innovative use for OGD, the lack of metadata describing OGD properly, and the lack of support to combine OGD datasets are the most detrimental difficulties faced.

In our research, we studied how impediments could impact the user as they progress through the OGD process. We identified that impediments caused the user to adapt by, for example, doing more backtracking and forward-thinking. In cases where adaption was not possible, the end-product would experience attrition (lower or change in value). For backtracking, the user needed to have the capability to be creative with the limited available datasets, while for forward-thinking, the amount of needed resources increased. Together, the impediments increased the need for motivation. Large and comprehensive sets of combinable data, documentation, and metadata are important for the OGD use process.

Subsequently, we discussed the theoretical implications of our findings, and we provided managerial recommendations to publishers in order to encourage them to consider these impediments before publishing OGD, with the goal to foster use by users (e.g., citizens, organizations, and entrepreneurs).

Our findings encourage governments to reverse the local supply-driven approach towards a contextualized OGD strategy as a collective endeavour. A user-driven platform beside a continuous communication of the city challenges and priorities could help to raise interest and engagement (top-down) and sustain spontaneous, efficient, and valuable reuse of the OGD (bottom-up). At the same time, working to find common standards and formats for the data can allow value creation outside a single dataset and region.

#### ACKNOWLEDGMENT

The authors would like to thank the European Regional Development Fund (ERDF) and the Belgian Federal Science Policy Office (BELSPO) for financially supporting this research, respectively for the Wal-e-Cities and the FLEXPUB (subsidy no. [BR/154/A4/FLEXPUB]) project. The authors also thank Pr. Benoît Frénay, who allowed us to collect data in the context of the class project he supervised, and the students who accepted to take part in the interviews despite their busy schedule. Finally, we are thankful to the program committee of the conference for not taking advantage of April 1st to make a joke about the acceptance notification.

#### REFERENCES

- [1] J. Attard, F. Orlandi, S. Scerri, and S. Auer. A systematic review of open government data initiatives. *Government Information Quarterly*, 32(4):399–418, 2015.

- [2] L. B. Ayre and J. Craner. Open data: What it is and why you should care. *Public Library Quarterly*, 36(2):173–184, 2017.
- [3] E. Barry and F. Bannister. Barriers to open data release: A view from the top. *Information Polity*, 19(12):129152, 2014.
- [4] M. Beno, K. Figl, J. Umbrich, and A. Polleres. Open data hopes and fears: determining the barriers of open data. In *E-Democracy and Open Government (CeDEM), 2017 Conference for*, pages 69–81. IEEE, 2017.
- [5] H. N. Boone and D. A. Boone. Analyzing likert data. *Journal of extension*, 50(2):1–5, 2012.
- [6] J. Brugger, M. Fraefel, R. Riedl, H. Fehr, D. Schneck, and C. S. Weissbrod. Current barriers to open government data use and visualization by political intermediaries. *Proceedings of the 6th International Conference for E-Democracy and Open Government, CeDEM 2016*, page 219229, 2016.
- [7] W. Carrara, M. Nieuwenhuis, and H. Vollers. Open data maturity in europe 2015. Retrieved from the European Data Portal website: [https://www.europeandataportal.eu/sites/default/files/edp\\_landscaping\\_insight\\_report\\_n1\\_-\\_final.pdf](https://www.europeandataportal.eu/sites/default/files/edp_landscaping_insight_report_n1_-_final.pdf), 2015.
- [8] L. J. Cronbach. Coefficient alpha and the internal structure of tests. *psychometrika*, 16(3):297–334, 1951.
- [9] J. Crusoe and K. Ahlin. Users activities and impediments from motivation to deployment in open government data - a process framework. In *Scandinavian Workshop of e-Government SWEG 2019, the University of South-Eastern Norway (USN), Campus Vestfold, 30-31 January, 2019*. [Work in progress; not published].
- [10] J. Crusoe and U. Melin. Investigating open government data barriers. In *International Conference on Electronic Government*, pages 169–183. Springer, 2018.
- [11] L. Danneels, S. Viaene, and J. Van den Bergh. Open data platforms: Discussing alternative knowledge epistemologies. *Government Information Quarterly*, 34(3):365–378, 2017.
- [12] T. Davies. Open data, democracy and public sector reform. *A look at open government data use from data.gov.uk*, 2010.
- [13] F. Gonzalez-Zapata and R. Heeks. The multiple meanings of open government data: Understanding different stakeholders and their perspectives. *Government Information Quarterly*, 32(4):441–452, 2015.
- [14] O. D. Handbook. What is open data. Retrieved 17th February, 2015.
- [15] M. Hartog, B. Mulder, B. Spée, E. Visser, and A. Gribnau. Open data within governmental organisations: effects, benefits and challenges of the implementation process. *JeDEM-eJournal of eDemocracy and Open Government*, 6(1):49–61, 2014.
- [16] A. Heise and F. Naumann. Integrating open government data with stratosphere for more transparency. *Web Semantics: Science, Services and Agents on the World Wide Web*, 14:45–56, 2012.
- [17] M. A. Hossain, Y. K. Dwivedi, and N. P. Rana. State-of-the-art in open data research: Insights from existing literature and a research agenda. *Journal of organizational computing and electronic commerce*, 26(1-2):14–40, 2016.
- [18] R. Huang, T. Lai, and L. Zhou. Proposing a framework of barriers to opening government data in china: A critical literature review. *Library Hi Tech*, 35(3):421–438, 2017.
- [19] A. Jaakola, H. Kekkonen, T. Lahti, and A. Manninen. Open data, open cities: Experiences from the helsinki metropolitan area. case helsinki region infoshare www.hri.fi. *Statistical Journal of the IAOS*, 31(1):117–122, 2015.
- [20] M. Janssen, Y. Charalabidis, and A. Zuiderwijk. Benefits, adoption barriers and myths of open data and open government. *Information systems management*, 29(4):258–268, 2012.
- [21] R. B. Johnson, A. J. Onwuegbuzie, and L. A. Turner. Toward a definition of mixed methods research. *Journal of mixed methods research*, 1(2):112–133, 2007.
- [22] J. Kucera and D. Chlapek. Benefits and risks of open government data. *Journal of Systems Integration*, 5(1):30–41, 2014.
- [23] K. Kuhn. Open government data and public transportation. *Journal of Public Transportation*, 14(1):5, 2011.
- [24] J. Lassinantti, A. Ståhlbröst, and M. Runardotter. Relevant social groups for open data use and engagement. *Government Information Quarterly*, 2018.
- [25] D. Lee. Building an open data ecosystem an irish experience. 2014.
- [26] J. Lindman, T. Kinnari, and M. Rossi. Business roles in the emerging open-data ecosystem. *IEEE Software*, 33(5):54–59, 2016.
- [27] G. Magalhes and C. Roseira. Exploring the barriers in the commercial use of open government data. *Proceedings of the 9th International Conference on Theory and Practice of Electronic Governance - ICEGOV 15-16*, page 211214, 2016.
- [28] K. Okamoto. Introducing open government data. *The Reference Librarian*, 58(2):111–123, 2017.
- [29] E. D. Portal. Re-using open data, 2017.
- [30] I. Safarov, A. Meijer, and S. Grimmelikhuijsen. Utilization of open government data: A systematic literature review of types, conditions, effects and users. *Information Polity*, 22(1):1–24, 2017.
- [31] J. Saldaa. [32] *The coding manual for qualitative researchers*. Sage, 2015.
- [32] S. Saxena and I. Muhammad. Barriers to use open government data in private sector and ngos in pakistan. *Information Discovery and Delivery*, 46(1):6775, 2018.
- [33] B. Schrier. Government open data: Benefits, strategies, and use. the evans school review, 2017.
- [34] A. Simonofski, M. Snoeck, and B. Vanderose. Co-creating e-government services: An empirical analysis of participation methods in belgium. In *Setting Foundations for the Creation of Public Value in Smart Cities*, pages

- 225–245. Springer, 2019.
- [35] G. TANANBAUM. Adventures in open data. *Learned Publishing*, 21(2):154156, Apr 2008.
  - [36] N. Veljković, S. Bogdanović-Dinić, and L. Stoimenov. Benchmarking open government: An open data perspective. *Government Information Quarterly*, 31(2):278–290, 2014.
  - [37] G. Vickery. Review of recent studies on psi re-use and related market developments. *Information Economics, Paris*, 2011.
  - [38] S. Virkar and G. V. Pereira. Exploring open data state-of-the-art: A review of the social, economic and political impacts. In *International Conference on Electronic Government*, pages 196–207. Springer, 2018.
  - [39] V. Weerakkody, Z. Irani, K. Kapoor, U. Sivarajah, and Y. K. Dwivedi. Open data and its usability: an empirical view from the citizens perspective. *Information Systems Frontiers*, 19(2):285300, 2017.
  - [40] A. Whitmore. Using open government data to predict war: A case study of data and systems challenges. *Government Information Quarterly*, 31(4):622–630, 2014.
  - [41] A. Zuiderwijk and M. Janssen. A coordination theory perspective to improve the use of open data in policy-making. In *International Conference on Electronic Government*, pages 38–49. Springer, 2013.
  - [42] A. Zuiderwijk and M. Janssen. Barriers and development directions for the publication and usage of open data: A socio-technical view. In *Open government*, pages 115–135. Springer, 2014.
  - [43] A. Zuiderwijk and M. Janssen. The negative effects of open government data-investigating the dark side of open data. In *Proceedings of the 15th Annual International Conference on Digital Government Research*, pages 147–152. ACM, 2014.
  - [44] A. Zuiderwijk, M. Janssen, S. Choenni, R. Meijer, and R. S. Alibaks. Socio-technical impediments of open data. *Electronic Journal of e-Government*, 10(2), 2012.

#### APPENDIX A

##### FULL QUESTIONNAIRE

Please refer to Table IV.

#### APPENDIX B

##### INTERVIEW GUIDE

The interview starts with a brief reminder of the phases. The questions about the phases are answered. The interview is structured in six parts: introductory questions, phase-specific questions (for each of the four phases) and concluding questions.

#### OVERALL

- 1) To warm up, give us a brief presentation of your project and how you worked with it.
- 2) In the survey we had the phases start of the project, search and evaluate, access and prepare, and aggregate and transform. How do you think this fitted with how you worked? Is any activity missing? If so, which?

- 3) Rank the four phases by difficulty.

#### START

*Main barriers:* finding the idea, finding an idea applicable to Namur, and finding “Use-case” examples of Open Data.

- 1) Was it difficult to find an idea of application for open data? Why? What was particularly tough?
- 2) If not, did you experience any barriers at the start of the project? If so, what barriers and how did they impact your work?

#### SEARCH AND EVALUATE

*Main barriers:* quality of metadata.

- 1) What was your overall barriers when using the Open Data Portal? Give an example.
- 2) Did you experience any barriers with the quality of the metadata? If so, how did they impact your work? Give examples.
- 3) If not, did you encounter any barriers when searching for the right data? If so, what barriers did you encounter? How did they impact your work? Give examples.

#### ACCESS AND PREPARE

*Main barriers:* relevance of data formats and data quality.

- 1) Did you experience any barriers from the data formats or data quality. If so, how did they impact you work? Give examples.
- 2) If not, did you encounter any barriers when accessing and preparing data? If so, what barriers and how did they impact your work? Give examples.

#### AGGREGATION AND TRANSFORMATION

*Main barriers:* data cannot be combined, data availability varies, tools cannot combine the data sources, and no longitudinal data.

- 1) Did you experience any barriers when trying to combine and use the data in the product? If so, tell us about it. How did they impact your work? Give examples.
- 2) If not and if you did experience any other barriers, what barriers did you encounter and how did they impact your work? Give examples.

#### END ON A HIGH NOTE

- 1) If the barriers persisted, would you be motivated to continue developing a product or service on open data? Why?
- 2) What is your dream scenario for working with open data?
- 3) What do you need to get there? How can publishers and open data portals support you?
- 4) Is there anything you think we have forgotten to ask about? Or something relating to the project you want to tell us about?
- 5) Did you experience any barriers not captured by the survey or mentioned in the interview? If so, give some examples.

TABLE IV

Part	Question	Measure instrument
Introduction	Can you provide your name or a brief description of the project you worked on? (this will be used to aggregate the answers by project)	Free text
	What is your background?	Multiple choice (management, computer science, mathematics, science, other)
	How confident are you with <ul style="list-style-type: none"> <li>• Programming</li> <li>• Data analysis</li> <li>• Open data portals</li> </ul>	5-point Likert <ul style="list-style-type: none"> <li>• Very confident</li> <li>• Confident</li> <li>• Neutral</li> <li>• Not confident</li> <li>• Not confident at all</li> </ul>
Start	For each of the tasks below, please indicate how difficult it was with the following scale. <ul style="list-style-type: none"> <li>• Finding the idea</li> <li>• Finding an idea applicable to Namur</li> <li>• Finding “Use-case” examples of Open Data</li> <li>• Understanding the requirements of potential users of the project</li> <li>• Conducting a feasibility study of the idea</li> </ul>	5-point Likert <ul style="list-style-type: none"> <li>• Not a barrier</li> <li>• Light barrier</li> <li>• Moderate barrier</li> <li>• Serious barrier</li> <li>• Extreme barrier</li> </ul>
	For each of the tasks below, please indicate how difficult it was with the following scale. <ul style="list-style-type: none"> <li>• Search functionality</li> <li>• Search results presentation</li> <li>• Quality of metadata</li> <li>• Had to download the data to evaluate it</li> <li>• The tools for exploring of, analyzing of, or experimenting with data</li> <li>• Open Data portal language</li> <li>• License impedes the use</li> <li>• Must log in to evaluate and access the data</li> </ul>	5-point Likert <ul style="list-style-type: none"> <li>• Not a barrier</li> <li>• Light barrier</li> <li>• Moderate barrier</li> <li>• Serious barrier</li> <li>• Extreme barrier</li> </ul>
Search and evaluate	For each of the tasks below, please indicate how difficult it was with the following scale. <ul style="list-style-type: none"> <li>• Wrong type of access (Manual vs. Automated)</li> <li>• Complex access with documentation</li> <li>• Filtering useful data</li> <li>• Relevancy of data formats</li> <li>• Data quality (e.g., missing essential information)</li> <li>• Lack of documentation to prepare data (e.g., conversion tables)</li> <li>• Data is inaccessible because of paywalls, limitations of infrastructure, or demands of registrations</li> <li>• No support (e.g., lack of documentation or forums)</li> </ul>	5-point Likert <ul style="list-style-type: none"> <li>• Not a barrier</li> <li>• Light barrier</li> <li>• Moderate barrier</li> <li>• Serious barrier</li> <li>• Extreme barrier</li> </ul>
	For each of the tasks below, please indicate how difficult it was with the following scale. <ul style="list-style-type: none"> <li>• Data cannot be combined</li> <li>• Data quality varies</li> <li>• Data availability varies</li> <li>• Tools cannot combine the data sources</li> <li>• No longitudinal data</li> <li>• Data infrastructure cannot be integrated</li> <li>• Data needs special knowledge to understand</li> </ul>	5-point Likert <ul style="list-style-type: none"> <li>• Not a barrier</li> <li>• Light barrier</li> <li>• Moderate barrier</li> <li>• Serious barrier</li> <li>• Extreme barrier</li> </ul>
Access and prepare	For each of the tasks below, please indicate how difficult it was with the following scale. <ul style="list-style-type: none"> <li>• Data cannot be combined</li> <li>• Data quality varies</li> <li>• Data availability varies</li> <li>• Tools cannot combine the data sources</li> <li>• No longitudinal data</li> <li>• Data infrastructure cannot be integrated</li> <li>• Data needs special knowledge to understand</li> </ul>	5-point Likert <ul style="list-style-type: none"> <li>• Not a barrier</li> <li>• Light barrier</li> <li>• Moderate barrier</li> <li>• Serious barrier</li> <li>• Extreme barrier</li> </ul>
	For each of the tasks below, please indicate how difficult it was with the following scale. <ul style="list-style-type: none"> <li>• Data cannot be combined</li> <li>• Data quality varies</li> <li>• Data availability varies</li> <li>• Tools cannot combine the data sources</li> <li>• No longitudinal data</li> <li>• Data infrastructure cannot be integrated</li> <li>• Data needs special knowledge to understand</li> </ul>	5-point Likert <ul style="list-style-type: none"> <li>• Not a barrier</li> <li>• Light barrier</li> <li>• Moderate barrier</li> <li>• Serious barrier</li> <li>• Extreme barrier</li> </ul>
Aggregate and transform	Please rank the four phases from most-time consuming (1) to least time-consuming (4).	4-point ranking
	Can you rank the four phases from the one in which you contributed the most (through your skills) (1) to the one in which you contributed the least (4).	4-point ranking
Resource allocation	Do you think the time invested in the following phases was useful to the final output of your project? <ul style="list-style-type: none"> <li>• Start</li> <li>• Search and evaluate</li> <li>• Access and prepare</li> <li>• Aggregate and transform</li> </ul>	5-point Likert <ul style="list-style-type: none"> <li>• Very useful</li> <li>• Useful</li> <li>• Neutral</li> <li>• Not very useful</li> <li>• Not useful at all</li> </ul>
	Have you applied the 4 phases above sequentially or iteratively?	Binary choice
Perceived usefulness	Did you have to learn new technologies, techniques, concepts to carry out the Data Science project?	Free text
	Which open data portals did you use?	Multiple choice (Namur, London, New-York, Paris)
	Are there any reasons why you did not use specific portals?	Free text
	Do you have additional comments on open data portals?	Free text