Users' activities for using Open Government Data: A process framework

Jonathan Crusoe and Ahlin Karin

Tweet

LINKÖPING UNIVERSITY

Users' activities for using Open Government Data - a process framework

Jonathan Crusoe[1] and Karin Ahlin[2]

[1]Division of Information Systems, Linköping University, 581 83 Linköping, Sweden, jonathan.crusoe@liu.se
[2]Department of Computer and System Science, Mid Sweden University, 851 70 Sundsvall, Sweden, karin.ahlin@miun.se

Abstract

Purpose - This research aims to develop a user process framework with activities and their variations for the use of open government data (OGD) based on empirical material and previous research. Open government data (OGD) is interoperable data that is shared by public organisations (publishers) for anyone (users) to reuse without restrictions to create new digital products and services. The user process was roughly identified in previous research but lacks an in-depth description. This lack can hamper the ability to encourage the use and the development of related theories.

Design/methodology/approach - A three-stage research approach was used. Firstly, a tentative framework was created from previous research and empirical material. This stage involved three different literature reviews, data mapping, and seven interviews with OGD experts. The empirical material was analysed with inductive analysis, and previous research was integrated into the framework through concept mapping. Secondly, the tentative framework was reviewed by informed OGD experts. Thirdly, the framework was finalised with additional literature reviews, eight interviews with OGD users, and a member check, including all the respondents. The framework was used to guide the data collection and as a tool in the analysis.

Findings - The user process framework covers activities and related variations, where the included phases are: start, identify, acquire, enrich, and deploy. 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 is 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.

Research limitations/implications - The framework development has some limitations: the framework needs testing and development in different contexts and further verification. The implications are that the framework can help guide researchers towards relevant and essential data of the user process, be used as a point of comparison in analysis, and be used as a skeleton for more precious theories.

Practical implications - The framework has some practical implications for users, publishers, and portals. It can introduce users to the user process and help them plan for the execution of it. The framework can help publishers understand how the users can work with their data and what can be expected of them. The framework can help portal owners to understand the portals role between users and publishers and what functionality and features they can provide to support to the user.

Originality/value - In previous research, no user process with an in-depth description was identified. However, several studies have given a rough recall. Thus, this research provides an in-depth description of the user process with its variations. The framework can support practice and leads to new research avenues.

Keywords Open Government Data, User, Process, Phase, Activity, Framework, Reuse, Use, Concept Mapping, Descriptive Theory

Paper type Research paper

1   Introduction

This paper presents a process framework for using open government data (OGD) from start to deployment. OGD is interoperable data that is shared by publishers over infrastructure for unrestricted reuse by users (Attard et al., 2015; Handbook, 2015; Hossain et al., 2016). The supporting infrastructure is organised by several different actors, where publishers, the Internet, and OGD portals are the base (Davies, 2010; Zuiderwijk and Janssen, 2014a; Lindman et al., 2016). While publishers are public organisations, users can be anyone (e.g., businesses, city managers, citizens, civil society organisations, developers, journalists, NGOs, researchers, or students (Safarov et al., 2017; Lassinantti et al., 2019)). OGD is believed to result in several benefits,

such as increased transparency, democratic accountability, external problem-solving, and economic efficiencies (Janssen et al., 2012; Kucera and Chlapek, 2014; Hartog et al., 2014; Carrara et al., 2015). The economic efficiencies are based on further use and development of OGD, such as commercial applications and time-based efficiency for the user. However, publishing OGD is also associated with risks, such as misinterpretations, misuse, and privacy violations (Barry and Bannister, 2014; Zuiderwijk and Janssen, 2014b).

In recent years, the world-wide OGD movement has slowed and is showing several worrying trends, such as decreasing speed in the publishing of OGD, earlier leaders faltering, and the view that publishing OGD is a side project(Barometer, 2018). One reason for the slowdown can be the noted lack of use (Kuhn, 2011; Vickery, 2011; Heise and Naumann, 2012; Zuiderwijk and Janssen, 2013; Whitmore, 2014; Jaakola et al., 2015; Safarov et al., 2017). This lack may be a result from impediments experienced by the users in the process of transforming OGD into products and services (e.g., Crusoe and Melin, 2018). The inability to respond to impediments can originate in a nescience about the user process.

Several authors have studied the user process and depicted it differently. Use can be described as data to an output (Davies, 2010), a straight value chain (Portal, 2017), one important side of a cycle (Janssen and Zuiderwijk, 2012; Zuiderwijk and Janssen, 2014a; Attard et al., 2015), a life cycle in relation to another life cycle (Charalabidis et al., 2018), and a series of interdependent roles (Lindman et al., 2016). Previous descriptions of the user process are often general or abstract, which can impede richer insight and understanding of what is happening. However, use can take a variety of forms, for example, create a visualisation from datasets, combine and clean datasets to be provide through an API, and a service relying on OGD (Davies, 2010). As a result, there is a possible discrepancy between what is known about the user process and the knowledge generated about it.

In previous research, it is known that several heterogeneous actors can be users and the user process can vary in both execution and output. In a conducted literature search, no previous research that describes the user process with its variations was identified. As a result, it is time synthesise the user process to reveal its complexities. To take a first step towards creating a descriptive theory (see Gregor, 2002) for the user process.

Therefore, this research seeks to synthesise a user process framework from previous research and practice. This framework can promote an understanding of the use and support the re-design of the needed infrastructure and published data. At the same time, researchers and practitioners can reflect on the user process to identify areas of improvement. A process framework is a tentative or incomplete theory describing a set of concepts and ideas and

their proposed relationships (Maxwell, 2012). Here, the aim is guided by the following research question:

1. How can the activities in the user's OGD process be conceptualised based on previous research and practice?

The article includes the following sections. First a presentation of the research approach to develop the user process framework, followed by a section describing the framework, including previous literature and empirical results on activities. Accompanied is a discussion on the frameworks implication on practice and research. The paper ends with a conclusion, followed by limitations and future research. Presented in appendices are an overview of interviewees and examples of identified user processes.
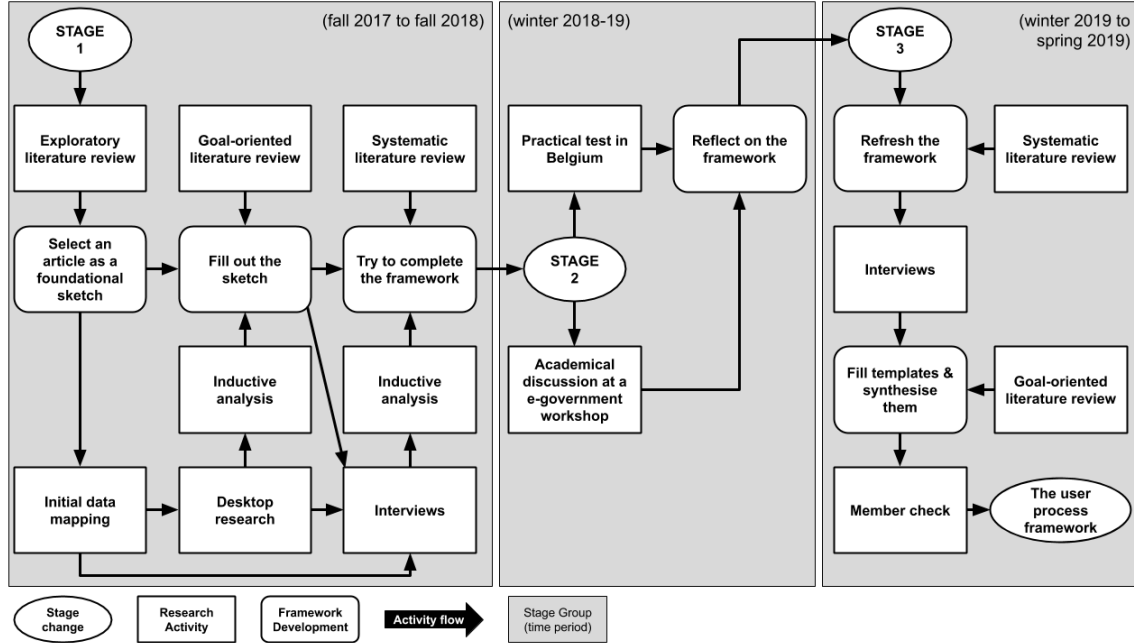
2   Method

A qualitative research approach was used to develop the user process framework (Maxwell, 2012), with the purpose to understand and map users' activities in the OGD process. The choice of qualitative research is required as a wish for a deeper understanding of a specific phenomenon (Myers, 2013), favoring a rich interpretative description from informed OGD users. For here, the design was based on a focus on the experiences of informed users and experts in the area of OGD and the novelty of the framework. Overall, the conducted research design was iterative and initiated in the literature to understand what was previously known, followed by interviews with informed users. The reoccurring move between literature and empirical findings gave this study robustness in providing a framework with adequate knowledge. The research process was conducted in three stages, see Figure 1. Described in the following sections are: first, the used method for the process framework development, then the history of the development, and, finally, the methods for each of the three stages.

2.1   Development of the process framework

The process framework was developed using concept mapping (Maxwell, 2012). A process framework, similar to a conceptual framework, is a tentative or incomplete theory describing a set of concepts and ideas and their proposed relationships. The structure of a process framework intends to capture or model a process in the world with how and why. Specifically, a *process framework* is concerned with events, situations, activities, starts, ends, and their relationships (Maxwell, 2012). A process is a flow of limited activities

Figure 1: Detailed view of the research process.

with a specific purpose (Sörqvist, 2004). Concept mapping as an analytical method includes defining concepts and their relationship. The end product is a visualisation of the tentative or incomplete theory. In this research, the article of Zuiderwijk et al. (2012) was used as a starting point for the development of the process framework. This choice was based on the articles content as it outlines a user process and clear thematic divisions of impediments with several examples. As the process framework developed and gaps were spotted, a need grew for further empirical investigation and exploration of previous research. The development was driven by generative questions, such as why, how, and what.

   While developing the framework, various data sources were used. The respondents all had the roles of active users of OGD or national experts in the area of OGD with varying backgrounds (e.g., journalists, researchers, and developers) which gave this study a broad empirical picture. Added to the interview material was data from other studies, and documents, heading for triangulation and getting a richer picture (Myers, 2013). Other inputs were gathered from the evaluation of the current framework by several of the respondents to fulfil the goal of the study. In stage 2, the framework was also discussed, tested, and commented by external researchers.

2.2   Framework Development History

The first stage of the framework development started in mid-fall of 2017 and resulted in a paper that was submitted to an e-government workshop at the end of 2018 (Crusoe and Ahlin, 2019). At the time, the framework contained the phases motivation, search and evaluate, access and prepare, and aggregate and transform. The phases were a series of activities with impediments identified in the empirical material and previous research. However, the relationships between the activities were mainly identified in previous research. **The second stage** started after the preceding peer-review of the e-government workshop. This stage consisted of testing the framework in Belgium (Crusoe et al., 2019), reflecting on the process activities, and collecting feedback from the workshop. In the testing and the e-government workshop, it was identified that the framework needed further development and focus. Besides, in the testing, it was identified that the users could use backtracking and use forward-thinking and that some activities were more interdependent than others. Finally, **the third stage** started once the testing was finished. In this stage, the framework was tested in practice with seven OGD users, and further enhanced with previous research. The outcome of the third stage is this paper.

2.3   Stage 1 - Creating a Tentative Process Framework

The first stage focused on user activities with impediments but shifting then to user activities and their variations, see Figure 1. Appendix 6 presents an overview of all collected data in their first stage, such as synthesised facts about the respondents. The creation of the tentative framework followed an abductive approach that jumped between the exploration of previous research, development of the framework, empirical data collection, and analysis. Previous research was found on Scopus, Google Scholar, Google, and the two universities' libraries forming a literature review (Machi and McEvoy, 2012). This stage can be summarized as an attempt to sketch, fill, and complete the tentative framework.

Literature Review. The literature review was conducted in several steps, where the initial step aimed for exploration to generate a sketch of the user process, the second for filling the gaps in the framework, and the last to complete the tentative framework before testing it. The initial step was conducted as an exploratory literature review, searching for previous studies on the OGD user process. After analysing the articles, a decision was made to use one article (Zuiderwijk et al., 2012) as the foundation for a tentative user

process and to start gathering empirical material (see Initial Data Mapping below). The focus was to understand the current knowledge on the user process for OGD by searching for literature on Google Scholar and Scopus. The findings, in total 40, were analysed according to their relevance for this work; first individually by the researchers and then in collaboration. The literature was graded into four grades: very useful (10), useful (13), might be useful (8), and not useful (9). The grades were firstly based on the abstract, followed by a brief reading of the entire article. The knowledge searched for were various user processes and related activities. The second step, the filling, was a goal-oriented literature review, focusing on filling the gaps of the framework with the previous literature search. The third step in the literature review was a systematic literature review. This literature review took place after the interviews, completing the framework based on findings in the empirical material, not previously addressed. The search for literature was based on the search query: TITLE-ABS-KEY(open AND government AND data AND user AND process). Identified in this search were 264 articles, which were analysed according to the tentative user process. Nineteen new articles were identified and further analysed, besides seven already identified articles (to a total of 26 articles). Besides this, the goal-oriented literature review was conducted to fill gaps in the user process and to find further knowledge to build upon. The result from the literature searches was the theoretical foundation, where the articles were analysed and used in an iterative process during the development of the framework.

Initial Data Mapping. An initial data mapping was conducted to identify potential empirical material for the tentative user process by conducting desktop research and interviewing national open data experts. The desktop research was conducted by investigating open data portals, open data forums (both on a national level and domain-specific), and other interesting open data websites. The interview respondents were both informed users and publishers, as well as private OGD organisations. In total six interviews were conducted, with a total length on seven hours. This initial data mapping was used to guide future data collection and framework development.

Data Collection. The primary empirical material originated from seven interviews with informed Swedish OGD experts, where four of them were semi-structured (Whiting, 2008; Myers, 2013), conducted via Skype, and three of them were email interviews (Mccoyd and Kerson, 2006; Gibson, 2010; Bowden and Galindo-Gonzalez, 2015; James, 2016). An interview guide was created, which was used for both Skype and email interviews. The guide

included the following themes: (1) the respondents work role relation and use of open data, (2) encountered impediments and (3) rank of impediments. The Skype interviews lasted between 50 to 90 minutes and were all transcribed verbatim. The email interviews were reused from a previous study (Crusoe, 2019). The respondents were selected for their experience with or knowledge about OGD use in Sweden. This way, this selection resulted in 5 experienced developers and two experts on OGD user, which was deemed enough for the initial development of the process framework.

Data Analysis. An inductive analysis was used to search for concepts to develop the process framework in the empirical material (Alvesson and Skldberg, 2008). The concepts were activities and impediments from the empirical material. The analysis followed these steps; underline, condensation, coding, categorisation, and conceptualisation (Hsieh and Shannon, 2005; Forman and Damschroder, 2007; Elo and Kyngäs, 2008; Bengtsson, 2016; Erlingsson and Brysiewicz, 2017). The analyses started with identifying meaning units in the empirical material and synthesised to various concepts. The concepts were then used in the development of the process framework. The analysis was first conducted individually by the researchers then the concepts were discussed and analysed together as they were used in the development of the process framework. This two-step process is one way to ensure the quality of the findings for content analysis (Hsieh and Shannon, 2005; Bengtsson, 2016).

2.4   Stage 2 - Verifying and Reflecting on the Framework

The second stage included three parallel development processes for the user process framework, see Figure 1. (1) The first draft of the framework (Crusoe and Ahlin, 2019) was sent to an e-government workshop and received peer-review comments from two researchers. The comments were discussed, and the framework went through minor adjustments. The framework was then presented at the workshop and received rich feedback by a dedicated opponent. The researchers later discussed the feedback and decided that impediments need to be dropped from the framework to focus more on activities and their variations. (2) This draft of the framework was also tested on a data science project of 30 master students in Belgium (Crusoe et al., 2019). The testing was conducted as a mixed-method study and focused on how impediments impacted user activities. Thirty students answered a questionnaire and nine of those students were in-depth interviewed. Consequently, the result showed that the framework needed to be further developed focusing on richness in details. (3) Therefore, the phases in the process framework

were broken apart into activities, and those activities were then grouped into activities in renewed phases to develop the understanding of the user process further. As a result, it was noted that the phases were to big and general and were broken down into the activities: start, search, evaluate, access, deliver, prepare, aggregate, transform, and deploy.

2.5   Stage 3 - Apply and Enhance the Framework

The third stage aimed to test the framework with a few users and enhance the framework with previous research, see Figure 1. This stage can be summarised as refresh, test and add finishing touches, and practically check the framework. The previous research was identified as described in stage 1. The result is presented in this paper.

Literature Review.   An additional literature review was conducted to look for newly published studies and to see if more relevant articles existed. This literature review was conducted in two steps where the initial one was conducted as a systematic literature review, systematically searching for literature according to (Machi and McEvoy, 2016). The foundation for this literature review was on Google Scholar and Google. The review was based on the search words "use process" and "reuse process" in combination with "open data" or "open government data". The findings were limited to the first 20 hits, in total 157, were analysed according to their relevance for this work by one of the researchers. Walters (2011) has identified that for simple keyword searches, Google Scholar has a good coverage, with a recall and precision that is above average, as such it was used at this stage. Google was used in hopes to identify missed research, but also more empirical material (e.g., white papers). The arguments looked for were the user process and how various user groups performed their process and its variations. The analysed material were then categorised and related to the existing user process framework. The second step in this literature review was a goal-oriented literature review, based on the existing user process framework. Here, the search aimed for specific findings related to the various phases and activities in the framework. The result from this literature review was that nothing of relevance was found.

Data Collection.   The primary empirical material for this stage originated from eight Swedish semi-structured interviews conducted via Skype and phone. An interview guide was created based on the framework and contained the themes: (1) user background, (2) the user process, and (3) dream scenario for

9

the process. At the end of the interview, interviewees were asked if any questions were missing, if they knew any critical documents for the study, and if they recommended someone for further interview. The interviews lasted between 30 to 45 min; however, one lasted for one hour and 20 min. All interviews were recorded. This stage used the same criteria for respondent selection as described in the first stage. Two experienced users were revisited for in-depth interviews. This was done to ensure the framework would still be anchored in their experiences and they were selected based on their perceived in-depth knowledge of the topic.

Data Analysis. In this stage, the framework was used to develop a template of the user process. Appendix 7 presents a few of the final templates filled from the empirical material. The researchers listened to all recorded interviews separately, and one to several framework templates were filled with information about the activities per interview. The researchers then discussed the filled out templates to integrate them into a final set. The synthesised results were then used to develop the user framework.

Member Check. For the member check, a popular science article was written, including the analysis of the user process and e-mailed to all respondents for verification, in total twelve. This approach helped to evaluate the findings. Three of them responded to the e-mail, saying that they agreed with the findings. One of them took a particular interest and recognised himself. Another respondent verified the findings and added details about the transformation of existing foreign services to domestic ones. A supplementary member check was conducted at a domestic hackathon, serving more than 500 users and 90 public organisations. Several users and experts on OGD were asked for their view on the process framework, adding not more than some details to the existing user process.

3   User Process Framework

The following section presents the user process framework developed in this study. The process is described as linear for a pedagogic presentation. In reality, the users work in iterations and move back and forth between the included phases and activities. In the following text, Harvard citation style is used to cite previous research, but for the empirical material IDs are used, for example, (R8) (see Appendix 6).

Figure 2: Overview of the user process.
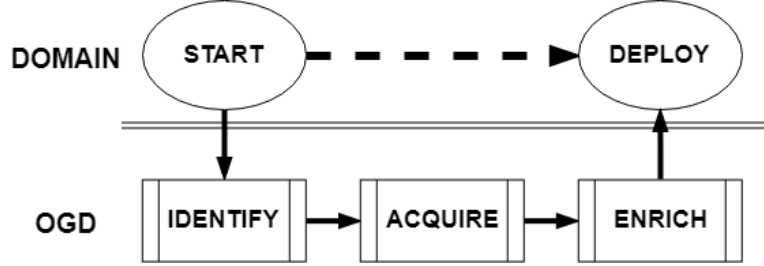
3.1 Process Overview

Figure 2 presents an overview of the user process. The process is initiated in the start phase, and the goal is to end in the deploy phase. For this reason, the user has to traverse the identify, the acquire, and the enrich phases.

In the process there are some **Common activities:**

Learning is an activity where the user has to work to with understanding the data, the used language, the data delivery systems, and metadata by, for example, acquiring domain knowledge through reading documentation, reading metadata, experimenting with the data, or talking with experts (R1; R2; R3; R4; R7; R10; Zuiderwijk and Janssen, 2014*a*; Beno et al., 2017; Martin et al., 2013; Zuiderwijk et al., 2012) As expressed by R7: "Data I do not understand I drop."

Backtracking is an activity where the user goes back from the current activity to a previous activity to resolve an issue in one or both (Crusoe et al., 2019). This activity was observed for R7 and involved jumping between and inside the start, and the identify phase. It could extend into the acquire and the refine phases.

Forward-Thinking is a strategy where the user thinks about what must be done in one activity to be able to succeed in coming activities (Crusoe et al., 2019). This strategy was observed for R4, R7, R8, R9, R11, and R12 in either the acquire or the enrich phase. However, if the strategy was not feasible, it could end the process, such as needing an API for a notification app but is only provided information on a website (R7).

3.2 Start Phase

The start phase is a situation where users are experiencing conditions that initiate their will to execute the user process. Therefore, it can take many

different forms (e.g., sitting in a morning meeting, reading a book, or participating in an OGD hackathon (R7; R10; Attard et al., 2015; Safarov et al., 2017)). Often, users have a hunch or idea of an end situation (the deploy phase), for example, to understand the world through data visualisation in the classroom or automate the hunt for local news in statistics (R5; R7; R8; R9). Once started, the user can jump between the phases and activities. The start phase can broadly be divided into the start points (1) the user has an idea, (2) user runs into data, and (3) the user starts with a need for data (R10; R12):

The **idea-driven** start point begins in the identify phase as the user has an idea with a hunch of what can be done and the data needed to reach that goal (R4; R5; R7; R8; R10). As explained by R10: *"[We] have questions that need answering, and data can answer them."* The prerequisites for this starting point are that the user needs to (1) know about OGD, (2) have an idea of what can be done with OGD, and (3) be able to come in contact with OGD (R2; R5; R12; Ubaldi, 2013; Attard et al., 2015; Safarov et al., 2017; Janssen et al., 2012; Huang et al., 2017; Tan, 2016). It is possible that a user identifies data and then introduces it to someone else (R7; R10).

The **data-driven** start point begins in the acquire phase as someone has introduced data to the user (R5; R7; R10). For example, in a digital OGD forum where someone writes "this dataset can be used" or as part of a workshop teaching the user how to use an API (R7; R8; R10; R12). After the introduction, the user might either attempt to continue the use of OGD or leave. Continuing, the user can twist-and-turn the data to identify something meaningful or possible application (R4; R8).

The **demand-driven** start point begins in the enrich phase as the user is working with something and realises that OGD can contribute (R3; R6; R7; R12). As explained by R6: *"The reason for why I started using OGD was that I wanted to make a price-searching site for train tickets."* This point has the same prerequisites as the idea-driven start point added by the user's knowledge on what to do and how OGD can support the end-user.

For the above *starting points*, the users need to have (1) skills, (2) resources, and (3) motivation to carry through the user process and overcome any impediments (R4; R7; R12). These users may require skills on an academic level or OGD specialist level, domain-specific knowledge for OGD, and the domain or general knowledge about data (Huang et al., 2017; Beno et al., 2017; Brugger et al., 2016). Resources can be monetary, time, tools, and alternative data sources (R5; R7; R8; R9; R10; R11). Motivations can vary and, for example, be activism, anti-corruption, data analytics, data journalism, decision-making, developing new services, development of functionality for smart cities, innovation, and research purposes (R4; R5; Safarov et al.,

2017).

3.3 Identify Phase

The identify phase consists of two main activities; exploring and assessing. This phase starts when the user wants to find data for use and ends if the user cannot find the data or has identified the data. If the user identifies the data, the user can move over to the acquire phase. This phase is based on infrastructures organised by more than one actor, such as OGD portals and publisher's websites.

Exploring is an activity where the user is trying different activities to discover data. Users can (1) *contact* publishers, experts, and researchers (R4; R7; R9; R10), (2) *search* on the Internet, publisher's websites, or OGD portals (R1; R4; R6; R7; R8; R9; Charalabidis et al., 2018), or (3) *engage* in digital forums, hackathons, or other social groups (R2; R4; R7; R8). The infrastructure supporting OGD searching needs to provide a clear interface to show search possibilities and provide clear navigation (Zuiderwijk and Janssen, 2014a). In sum, exploration for OGD is best explained by R7: *"a hunt after whom I can ask to find some data."*

Assessing is an activity where the user is trying to judge and decide if the discovered data can be used. The user may need to weigh the qualities and properties of the data and the delivery method against the activities in the refine phase and the value it can contribute to the deploy phase (R4; R7; R8). Users can assess data by (1) *appraising* it through questioning (What can I do with you?), reflecting on possible end-user value, and be inspired by use cases (R2; R7; R8; R9), (2) *experimenting* with the data (e.g., use visualisation tools) (R1; R4; R12), (3) *study* metadata and data models to verify that the data has the right properties (e.g., fresh and granular) (R2; R4; R7), and (4) *talk with experts* and publishers (R7; R10). R9 gave an example: *"Is this interesting and relevant for end-users and society? Is the data granular and fresh?."* Important for assessing are tools that can be used to explore, analyse, and create mashups to help the user make sense of the data (Attard et al., 2015; Susha et al., 2015). For example, visualisations allow users to evaluate the OGD without requiring technical skills and can even motivate them to continue their work (Graves and Hendler, 2013). At the same time, the user might acquire the sought information by using the tools and software and leave the user process (Davies, 2010). Another approach to assessing the data is for publishers to support interaction and conversation about data and potential applications to encourage the use of OGD (Colpaert et al., 2013; Susha et al., 2015).

3.4   Acquire Phase

The acquire phase consists of two main activities: access and delivery. This phase starts when the user wants to acquire data from a publisher and ends if this is impossible or the data is delivered. If the data is acquired, the user can move to the enrich phase. The acquire phase is based on the infrastructure provided by the publisher. R2 stressed the importance of this phase to be quick and easy, supported by automatic technical support. R5, R7, and R9 stated that it is important that users have access to the delivery of raw data to allow diverse use.

Access is an activity that is concerned with setting up the situation surrounding the data delivery. Access can be as easy as a click on a link to download a text file, to be more time consuming as data has to be requested over mail, or to come with a few extra steps by registering an account to get an API-key by mail (R1; R2; R6; R7; R12). Also, the user needs to read the license (R2; R5). The user can prepare to store data locally or connect to data storage and access it as a flow. Preparing to *store* the data locally can be if the delivery method is, for example, mail with an EXCEL-file or download a PDF-file. The user can prepare the local storage for both manual and automated delivery. Preparing to *connect* to a data storage (e.g., through API) can need the setup of infrastructure by reading documentation, writing code, and register for API-key (R1; R2; R4; R7; R9; R12).

Deliver is the transfer of data from publisher to user. Data transfer can be scrapping the publisher's website, receiving emails with digital files, retrieving paper documents, downloading files as JSON or PDF, or accessing the data through an API (Charalabidis et al., 2018). The delivery can be manual or automated. *Manual* delivery is when the users request the data from the publishers (e.g., like paper (R4; R10)), download the data in a none machine-processable format, such as PDF (R3), or read the data through visualisations as the raw data cannot be accessed (R5). Commonly, manual delivery methods have a need for the users to store the data locally since data is easiest accessed by humans and can need extensive preparations before use. On the other hand, *automated* delivery is when the users request data from a data storage (e.g., through API) (R2; R5; R7; R8; Charalabidis et al., 2018), download the data in a technical format (e.g., CSV, XML, or JSON (R6)), or manually trigger scrappers (R4; R9; Davies, 2010; Charalabidis et al., 2018). These methods seek to automatically retrieve and process the data, as such machine-processable formats are important (R1; R7; R12). Also, the methods can either download the data to store (e.g., EXCEL) or work with it as a flow (e.g., API) (R3; R4; R7; R8; R10; R11). However, the methods require the users to spend time and resources on setting up the infrastructure

by, for example, reading documents and writing code for a retrieval interface (R4; R7; R9). In sum, access, delivery, and how users intend to use data are highly interdependent.

3.5   Enrich Phase

The enrich phase consist of the main activities: prepare, concoct, and make. The phase starts when the user has acquired data and seeks to do something with the data. The phase ends when the user cannot do what is intended or makes something that can be deployed. At a minimum, the user downloads data to identify some fact to be shared with others or to make an informed decision (Davies, 2010).

Prepare is an activity where the user is making the acquired data ready for concocting and making, as such the user may need to work with data extraction, data casting, data mapping, data restructuring, and data pruning. The users may need to *extract* the data from the delivery package, such as visualisations, PDFs, JSON, and EXCEL (R5; R7; Davies, 2010; Lindman et al., 2016). Data might need to be *casted* to other value systems, such as coordinate systems or change decimal comma to decimal dot (R6; R8), *mapped* to other datasets and therefore be prepared for such, for example, by finding common identifiers (R5; R6), or be *restructured* to fit with the intended use, for example, inserting the data into a database or upload to a server (R3; R4; R7; R11]. As R3 explained: *"I took the data and put it in a database so I could make things with it. I started with traffic safety regulations since I needed it for my education. I wanted it in a database to avoid browsing documents."* By doing this, the user has made the data ready for processing. Other options can be to convert the data to XLS, CSV, or RDF (Liu et al., 2011). Together data casting, data mapping, and data restructuring can be viewed as normalisation of the data. The normalisation includes to prepare the data to be combined with other datasets (R4; R9; Lindman et al., 2016). Normalisation can allow for aggregating local data to regional (R9). Finally, data might need some *pruning*, which involves cleaning data, splitting data fields, and repairing corrupt data or missing content (R3; R7; R8; R11; Davies, 2010; Zuiderwijk and Janssen, 2014a; Lindman et al., 2016). It is important to understand the dirt in the data to remove it (R4; R7; R10; Charalabidis et al., 2018). As R7 explained: *"A lot of garbage, different codes, and different content. First, I had to filter each data point. What does this field mean? Oh, it means this. Then I pick it out and remove the rest. To start with something and get something to function."* For the processing in make, the preparation can be made part of the solution (R7). As such, prepared data can be provided to other users, for

example, converting text to RDF that is provided through an API (Davies, 2010; Berends et al., 2017). At the same time, the same prepared data can be used in both the analysing and processing to create a variety of products (R4; Davies, 2010).

Concoct is an activity where the user is attempting to make something by adding several different parts together (e.g., code, tools, and datasets). For this purpose, the user may analyse the data or process it. *Analysing* data can involve calling experts or the publishers to understand the results, ask generative questions, search for a story, visualise the data, start with simple questions and then move upwards, statistical analysis, explore the data by twisting-and-turning it, and analysing data on a local level and continuing the analysis for more significant trends (R4; R8; R10; Zuiderwijk and Janssen, 2014a; Berends et al., 2017). The user may analyse one to several datasets at a time (Lindman et al., 2016). Analysing data is often supported by some tool, for example, Excel, Python, Panda, R, Ruby, Google Fusion Tables, or GapMinder (R4; R8; R10; Brugger et al., 2016; Charalabidis et al., 2018). The tools can be used for different purposes; for example, Python and panda can be used to write articles, while Excel can be used to teach journalists about data analysis (R10). At the same time, scripts can be used to combine and compare data (R8; Zuiderwijk and Janssen, 2014a). The line between analysing data and processing data can be thin, for example, a user can use Python and Jupiter Notebook to step-wise develop an interactive visualisation (R8). As a result, the user can move from analysing to processing data (R4; R8; Zuiderwijk and Janssen, 2014a; Brugger et al., 2016). *Processing* data is an activity where the user is developing or creating something (e.g., smartphone applications, desktop applications, and server software) to achieve a result with the data (R4; R7; R8; R9; R10; R11; Lindman et al., 2016; Berends et al., 2017). This activity is useful when there is too much data for a human to analyse (R10). No matter the output, if other users approach the analysed data, it can be essential to write FAQs, method descriptions (R4), and even validation controls are relevant (R10; Charalabidis et al., 2018). Processing can be anything between observing datasets to analysing datasets for patterns automatically (R7; R9).

Make is an activity where the user is capturing the result from concocting to produce something. This activity focuses on either adding value to or solidifying the results into something deployable. The activity has two variations: capture and develop. *Capture* is an activity where a user tries to solidify the results from the analysis through writing research basis, articles, reports, blog posts, or creating graphical representation (R4; R8; R10; Davies, 2010; Charalabidis et al., 2018). Instead of capturing the data, the users can *develop* products on the result from the processing, such as digi-

tal maps, websites, timelines, visualisations (e.g., maps, plots, and charts), APIs, and interactive diagrams (e.g, clusters) (R4; R8; Davies, 2010; Berends et al., 2017; Charalabidis et al., 2018). The developed product can have several functions, such as simulation, calculation, enable collaborative data tagging, notify about interesting divergent trends or patterns, notify about important events, allow for the exploration of visualised data, or integrate with an existing interactive map (R4; R7; R8; R9; R10; R11; Davies, 2010; Lindman et al., 2016). It is possible for the user to develop a product that can be used by other users in their analyse activity (R11). However, not all development will have the same demand on preparation, processing, and analysing. One example is that for analysing, users may need to harmonise the data, while it might not matter for notifications (R7).

3.6   Deploy Phase

The deploy phase is, if everything has gone well, the endpoint of the user process. The user has information, services, or products that can be distributed for usage by others. The product can also be provided as a service (R9). Often is the domain of deployment specific, such as local news agencies, the life of everyday citizens, and high schools (R7; R8; R9). In all interviews, it was identified that the outcome either acted or intended to **serve** the end-users in someway.

4   Discussion

The focus for the discussion is on the entire user process, developed in this study and shown in Figure 3. Examined are variations on the user process and implications for the previously mentioned roles, such as users, publishers, and portals. With previous research, this studys user process is dynamic and detailed, including the overall phases, activities, and selected variations of activities. This studys approach to the user process differs from, for example, Davies (2010) who has a more static approach and Zuiderwijk and Janssen (2014a) who has a more overarching approach. Thus, this studys approach to the user process is a contribution to the understanding of the user process where the start, the full process with activities and variations, and the deployment are viewed as necessary. The following sections discuss each of the phases in the user process and end with implications.

The starting points

In this research, it was identified that the user process could start as idea-driven, data-driven, or demand-driven. Past research tends to circulate around the first starting point, such as in (Beno et al., 2017). They de-

Figure 3: The user process with activities and variations.
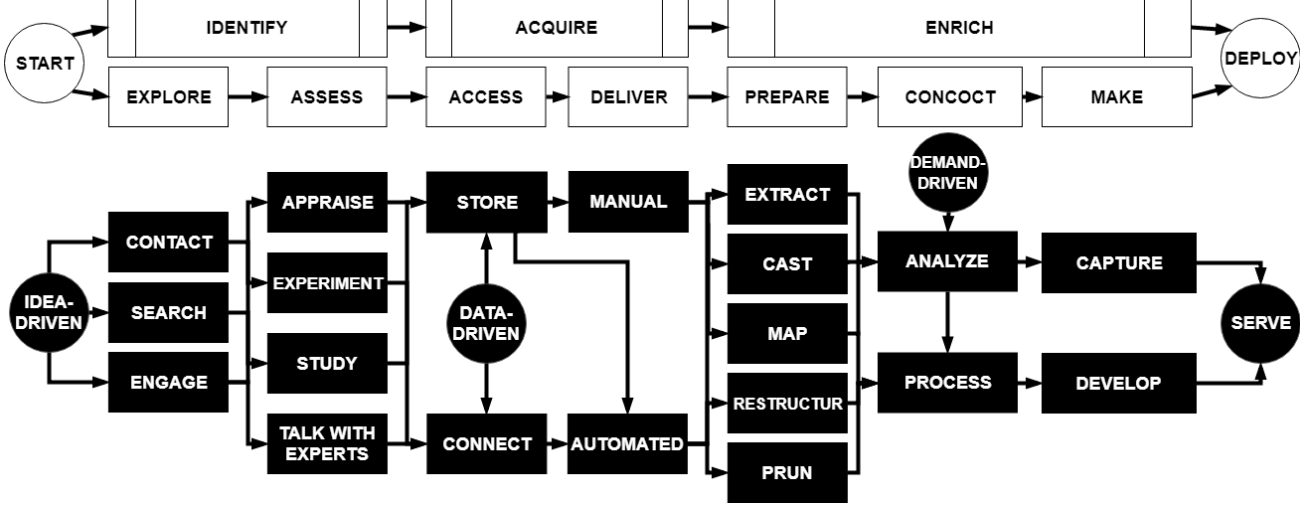
scribe the user process as starting with the user searching for data on an OGD portal or a publishers website, intending to download the data. In the literature review, no studies using varying starting points for the user process were identified. To support a deeper understanding of the different starting points, found similarities and differences will be described. One similarity is that all starting points initiate the user processes and might to some degree need to iterate to the identify phase to find more data. The idea-driven and demand-driven starting points varies in their objectives; the first seems to be adapted by innovators and newcomers, while the second seems to be adapted by experienced data users. The idea-driven and data-driven starting points focus on the exploration of the data to identify what and how it can be used. However, origin of the starting points differs. The idea-driven starts with an idea and seeks to identify data for it (or modify the idea following existing data (Crusoe et al., 2019)), while the data-driven starts with data and has to identify an idea or possible use for it. Moreover, essential for the starting points are that they have different contexts and conditions for the users.

The context for the *idea-driven* starting point can be an innovation event, such as a hackathon. This point puts high demand on the user to be knowledgeable about OGD in terms of knowing about OGD, where to find it, how to identify what can be done with the data, and how it can support end-users through products, services, or information. This initial knowledge demand seems to be the highest of all starting points. Even though previous research has focused on this starting point (e.g., Janssen and Zuiderwijk, 2012; Zuiderwijk and Janssen, 2014a), there is still a need to study it to lower the threshold for newcomers. One approach could be to provide an

easy-accessible open data education, focusing on the newcomers, e.g. on the publisher's website or OGD portal (e.g., Zotou et al., 2018).

The context for the *data-driven* starting point can be data labs, workshops, forums, or online communities where someone is introducing the user to specific data. This starting point requires that the user identifies what can be done with the data, what for, and how the data can be used. In comparison to the idea-driven starting point, the user does not need to have an idea that the data can satiate; instead, the user needs to generate or identify one. While this starting point is initiated in the acquire phase, the OGD portal or the publisher's websites can still play a role as the user might seek to understand the introduced data further. It is possible that this starting point could be used as an introduction to OGD for newcomers.

The last starting point, the *demand-driven*, can contextually be data analysis projects or international services. The user starts with a product or service that needs further data to be improved or work (or a need for information). The user is more entrenched in the data user domain as they already know what and how data can be used. This category of user could faster produce a product or a service that can be used as they have several of the puzzle pieces in place (e.g., knowledge, idea, skills, and possible benefits). For this starting point, the identify data phase is more about identifying data for the intended purpose and as such the OGD portal can play a role.

The identify phase

User initiating the user process from the idea-driven starting point begins in the identify phase where activities, such as exploring for and assessing of OGD are central. These activities include variations; here described as contact, search, and engage for the exploration and appraise, experiment, study, and talk with experts for the assessing. The context for exploration and assessing can involve social and technical elements; social as the user may need to be involved in a community, read and write, or have discussions with publishers and technical as the user may need to use the Internet to find and acquire OGD.

Deepening the social elements of the identify phase, Xiao et al. (2019) claim that the exploration of the OGD is one activity that includes obstacles for the user. Identified in this study is that the user practices various ways to identify required OGD, where one can describe them as mainly social, such as reading on a forum or discussing with publishers. The variations in exploring might be a result of the mentioned obstacles, where the users depend on the publishers and their ways of publishing and marketing OGD. The most common answer from the publishers to the user's requirement of exploration is a centralised point, declaring the need for a portal. Mainly, this portal is based on a specific domain or geographical limitations and including

various structured ways to explore OGD (Attard et al., 2015; Susha et al., 2015). Charalabidis et al. (2018) are on the same path for how exploration is conducted; describing the importance of including contextual descriptions of the OGD as well as the users experience and good examples of previous use of the identified OGD. Another way for the publisher to help the user to explore the OGD is to be available for discussion in suitable channels and that the user is familiar with those channels.

Moving over to the assessing activity, the focus for the user is on the decision: is this OGD the right OGD for the intended use? This decision can be based on the data quality and the delivery. It can be formulated as *"can I do what I want to do with this data."* Therefore, one can believe that the decision should be based on both facts and interpretations based on good examples, whereas Attard et al. (2015) describe data quality as based on interpretations. The decision is, thus, an analysis of the data quality in comparison to the goal of the outcome based on facts, and interpretations. This analysis can be approached in several different ways (e.g., experimenting or appraisal). One possible support for the user could, therefore, be OGD standardisation of data quality as well as descriptions of good examples on the use of datasets.

The acquire phase

Once the user has identified interesting data, it is time to acquire it. To acquire OGD, the user conducts the activities access and deliver. Access is where the context for the delivery is setup and includes the variations store and connect. The delivery is the transfer of data from the publisher to the user, divided into manual or automated delivery.

The user has a wish for efficiency in this phase, and many publishers do focus on the technical aspects while publishing. The technical aspects are decided by the publishers, where their internal processes can decrease efficiency for the users, such as having problems with instant access to the data. Like this, can the external requirements of efficiency from the user be denied by requirements from the publisher's internal processes. However, many publishers do focus on the technical aspects while publishing, which can satisfy the user's requirements for this phase. While ending this phase, the user is leaving the close technical dependency on the publisher and is entering a dependency on the content of the OGD and the intended content of the service, product, or information.

Xiao et al. (2019) stress various forms of impediments for the user while accessing the data, such as inability to find the right documentation. This studys access activity is divided into store or connect with a focus on the technical aspects of accessing. The user can choose to access OGD from one or several suppliers. However, the supplier sets the conditions for access.

The user has less impact on the supplier for this phase in comparison to the identify phase, where the user can request OGD.

Attard et al. (2015) emphasise the complex ways of delivering OGD as a problem and the publishers power to decide the way to publish OGD. The user is focusing on how the delivery takes place and its alignment with the intended service, where, for example, rarely published OGD can be used for specific purposes. The delivery puts a demand on the user, who needs to investigate efforts in accepting and using the delivery method (Xiao et al., 2019). The main view is that there is a contradiction between the manual versus automated delivery method, locking the user into one situation. The freedom for the user is often forgotten based on the publishers decision on how to publish the OGD. Rarely is the OGD delivered both as automatic and manual, which would increase the users freedom.

The enrich phase

The enrich phase includes many more variations than the identify phase. The enrich phase includes activities: prepare, concoct, and make. Prepare includes the variations extract, cast, map, restructure, and prune, the concoct of analyse and process, and the make activity capture and develop.

The preparation phase can be done in many ways, which means the most variations. The user has to handle the content of the delivered OGD in one or several ways, opening up for another relationship with the publisher than in the acquire phase. Found in the analysis is that the user can use visualisation to identify the quality of the data, whereas Attard et al. (2015) and Graves and Hendler (2013) focus on the visualisation to understand the data or identify values. Researchers discuss the impediments of the content of the OGD in various forms, such as low data quality (Attard et al., 2015; Zuiderwijk and Janssen, 2014a; Susha et al., 2015), the data structures non-fitting for comparison (Attard et al., 2015; Susha et al., 2015; Graves and Hendler, 2013), the OGDs understandability for non-domain users (Zuiderwijk and Janssen, 2014a; Xiao et al., 2019; Susha et al., 2015; Graves and Hendler, 2013). All of these impediments do affect the preparation phase resulting in many variations for the user, like pruning the OGD for missed or lacking variables or casting to fit for comparison. These impediments seem to increase the need for forward-thinking and resource allocation (Crusoe et al., 2019). The many variations give the user the freedom to prepare the data as both wished and needed, where freedom of choice could be part of one way to reflect on the variations. At the same time, this freedom puts demands on the user to be knowledgeable about the content of the data and how it is produced. Therefore, and in comparison, with the acquire phase, the relationship with the publisher is more affected by the OGDs content and the user's intention with how to use the OGD than the technical aspects, solely

provided by the publisher.

One crucial part of the concoct activity is the attempt to find the path forward in alignment with the outlined intentions by combining various datasets. However, the combination of various datasets might not fit together, which is not strange. The publisher is publishing their datasets and does not know about the intention of the present and future users. The user is seldomly discussing their needs for combining any published datasets with the publishers, which leaves the publishers with an extensive set of possible combinable datasets. They might be unable to provide data for combination, as they do not know how the user intends to combine the data. Standardisation could be one answer to this problem, used in domains such as public transportation (Attard et al., 2015; Graves and Hendler, 2013). For analysing, the user might be dependent on the need for domain support from the publisher or relying on the user's domain knowledge. Added to this is that there is various software for use, leaving the user with many choices and demands to have access to the software.

(Zuiderwijk and Janssen, 2014a) describe the barriers for the user while capturing and developing the data; omitting the responsibilities of the user in this phase. Not described is that these responsibilities grows, exemplified by the users obligation to document the service for future use. To document the service, the user needs to understand how the service is supposed to add value. This way, the user needs to turn more focus on the intended end-user and the intended use. The broader focus for the user includes both personal use and external end-users. The user, therefore, needs to understand the intended audience as well as their needs, both at present and in the future.

The deployment phase

For here, the deploy is described as the end of the user process, where in many cases it is solely the phase where the product or service can serve external end-users or be used for personal gains. For the OGD user, the deployment can be governed by large organisations in order to reach a broad audience or limited audience in a small-scale deployment, adding another focus. Zuiderwijk and Janssen (2014a) refer to activities at the end of the user process as discussing with publishers, whereas the intention with this user process is that the service should be able to serve someone. If a user succeeds in deploying a product or a service, they often re-iterate and start the user process all over again. This finding is in contradiction to the discussion of (Zuiderwijk and Janssen, 2014a) that found few practitioners re-iterating the user process. This difference indicates an increase in OGD maturity and learning from previous failures. One statement could be that the population of skilled and knowledgeable users are increasing as more reiterating practitioners were identified.

4.1   Implications

The presented user process framework has several implications, both for practice and academia. In practice, the user can refer to and understand the current and coming activities in a more detailed way while developing a new service, products, or information based on OGD. The process can inform users about the included phases, activities, and alternatives and, thereby, allow them to make other decisions than previously. For the publisher or organisers of events related to OGD, the framework gives more profound insight into the activities of the user and how the process can be supported. For academia, the framework gives detailed knowledge that can act as a foundation for the investigation and identification of distinct phases as well as synthesising them with results from research about related impediments. Specifically, there are a few recommendations for each of the phases to practitioners and academia.

For the *start phase*, it is recommended that researchers and publishers consider all starting points. Still, there is a need to develop the idea-driven starting point further to offer several possibilities. The possibilities of today are often given in hackathons and other community activities. These possibilities could be to offer opportunities for describing and discussing ideas on a public website in alignment with various datasets. The data-driven starting point is discussed as not being included in todays OGD portals, where they could take a more active role. Identifying and supporting users in the demand-driven starting point has the potential to realise benefits with OGD faster. This could be done by informing possible users about OGD (e.g., researchers and journalists).

For the *identify phase*, it is recommended that publishers support the users' exploration for data in different ways, such as providing contact information (to them and data experts), search opportunities (e.g., portals and platforms), and engage with the users in different communities, forums, and events. It is essential to recognise the social side of identifying data and that technical infrastructure can play a role. Websites, where OGD can be found, should be standardised to increase discoverability and efficiency. Today, the standardisation exists, even though adopted differently. Moreover, it is recommended that publishers recognise that users can assess data in several ways and should attempt to provide several off them, like tools, expert contacts, useful examples, and documentation. It can be essential to display the qualities and properties of the data to the user in a transparent manner to ease their judgment. The researcher can approach the identification of OGD as an exploration and assessment activity similar to prospecting.

For the *acquire phase*, it is recommended that publishers should ease the

access and delivery of the data to as few steps as possible and describe how to access the data, and provide several delivery methods. Research on technical solutions and standards could help with easing access and delivery.

For the *enrich phase*, it is recommended that the publisher focuses on the content of OGD as the user needs several skills to prepare it. One such is the data quality, rarely described and when, not in a standardised way. The user also needs clear descriptions of contextual content to translate it in an understandable way for the end-user. Another issue is that of correlation between OGD from the various publisher, offering standardised ways to use the same identifier throughout a specified domain. It is also recommended that the user focus on their end-users, for example, by adding documentation to the service or product and adjusting the value delivery. Researchers should take note of how users are trying to enrich data and what activities they are using.

For the *deploy phase*, it is recommended that the user thinks of it as integration into something; not the end or goal of achievement. The user may need to add more value by updating the service or product. Researchers can study how OGD and data products and services are used.

4.1.1 How can the user process framework be used...

Besides, the above implications, there are some possible uses of the framework in practice and research. Both will be described below.

... in practice?   Users, publishers, and owners of portals can use the developed user process framework for different purposes:

Users can use the framework to plan and make strategies of their usage of OGD and introduce themselves to the process and what they may be needed to do to turn data into something.

Publishers can use the framework to understand how the user can work with data and also what can be expected of them (e.g., manual and automated delivery).

The owners of the portals can use the framework to understand its role between users and publishers and what functionality and features they need to offer to support both parties' activities.

... in research?   The developed user process framework can be used to understand the user's activities and support the development of a theory of understanding, but should not be used to predict the future (see Gregor,

2002). Specifically for case studies, the framework can be used as an initial guide, part of an iterative process, and part of a final product (see Walsham, 1995; Eisenhardt, 1989). Each type of use is given an example below.

As an introductory guide the framework can give researchers an insight into the possible user process. This insight can help direct the researcher towards relevant and important data.

As part of an iterative process the framework can be used as a point of comparison to identify similar activities and more variations. The framework offers a base for knowledge contribution.

As a final product the framework could be integrated as a lightweight skeleton upon which further details and descriptions are added.

5   Conclusion

The OGD user process is in focus since the OGD movement has slowed down in many countries and earlier developed user process frameworks are general and abstract. Therefore, this study has focused on developing a detailed user process framework with variations (see section 3) using both previous research and empirical data. Included in the development has been testing in practice.

This study's developed user process is framed by the phases: start, identify, acquire, enrich, and deploy. To previous research does this framework consists of phases with activities and their variations, including additional variations and richer details. The identified activities show that the user process is complex, requiring skilled users with knowledge, such as technical and domain, and available resources. The variations in the activities highlight the different requirements related to the users, the infrastructure, and the publishers. One main finding from this study is that users can approach the process through several different strategies, such as idea-driven, data-driven, or demand-driven, creating new implications for both theory and practice. One other main finding is that the user can approach the activities in varies ways, depending on the goal for the OGD or available resources. For example, to assess data the user can appraise, experiment, study, and talk with experts. Moreover, once the users are able to realize something based on OGD, they seem to be more willing to re-start the user process.

The significant contribution of this research is the detailed framework of the user process with variations in the users activities. This user process can act as a foundation for future research about OGD use (see subsection 4.1).

At the same time, the framework can be a first step towards the creation of a descriptive theory (see Gregor, 2002) for the usage of OGD. The identified variations of the user process allow for a broader approach to what OGD use can be and how it can be conducted. This breath can help us understand the different strategies users employ and how they can be supported.

5.1   Limitations and Future Research

The framework development contains some limitations. The literature reviews in stage 1 and stage 3 of the method were either systematic or goal-oriented, which ensured that several critical articles were identified. However, the major part of the literature review was goal-oriented and important literature can still be out there. The framework was mainly developed on Swedish empirical material, but was peer-reviewed in a Scandinavian workshop and tested in Belgium, as such the framework needs further testing and exploration in other contexts to add more activities and variations (but should still be usable in them). In the final stage, the framework was reported as a popular science article to the participants of the study, but few did respond. The received responses were positive, and when verifying the framework at a domestic hackathon, it received positive feedback. The framework is in line with what was previously known and has synthesised a richer understanding of the user process.

The framework research has opened several avenues of future research, which are listed below:

- The different starting points and how they impact the user process.

- Activities and their variations in different contexts and for different user groups.

- Requirements and prerequisites for the activities.

- Connect different types of data with different sets of user activities and outcomes.

- Connect impediments and barriers with different user activities, outcomes, and user groups.

References

Alvesson, M. and Skldberg, K. (2008), *Tolkning och reflektion - Vetenskaps-filosofi och kvalitativ metod*, Studentlitteratur.

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

Barometer, O. D. (2018), 'Open data barometer'.
   URL: *https://opendatabarometer.org/ (accessed 2018-11-18*

Barry, E. and Bannister, F. (2014), 'Barriers to open data release: A view from the top', *Information Polity* **19**(12), 129152.

Bengtsson, M. (2016), 'How to plan and perform a qualitative study using content analysis', *NursingPlus Open* **2**, 8–14.

Beno, M., Figl, K., Umbrich, J. and Polleres, A. (2017), Open data hopes and fears: determining the barriers of open data, *in* P. Parycek and N. Edelmann, eds, 'International Conference for E-Democracy and Open Government 2017 (CeDEM17)', the IEEE Computer Society, pp. 69–81.

Berends, J., Carrara, W., Engbers, W., and Vollers, H. (2017), *Re-using Open Data: A study on companies transforming Open Data into economic & societal value*, European Data Portal.
   URL: *https://www.europeandataportal.eu/sites/default/files/re-using_open_data.pdf (accessed 2019-04-19)*

Bowden, C. and Galindo-Gonzalez, S. (2015), 'Interviewing when youre not face-to-face: The use of email interviews in a phenomenological study', *International Journal of Doctoral Studies* **23**(2), 493–501.

Brugger, J., Fraefel, M., Riedl, R., Fehr, H., Schneck, D. and Weissbrod, C. S. (2016), Current barriers to open government data use and visualization by political intermediaries, *in* P. Parycek and N. Edelmann, eds, 'Proceedings of the 6th International Conference for E-Democracy and Open Government (CeDEM 2016)', the IEEE Computer Society, p. 219229.

Carrara, W., Nieuwenhuis, M. and Vollers, H. (2015), 'Open data maturity in europe 2015'.
   URL: *https://www.europeandataportal.eu/sites/default/files/edp_landscaping_insight_report_n1_-_final.pdf   (accessed 2019-04-19)*

Charalabidis, Y., Zuiderwijk, A., Alexopoulos, C., Janssen, M., Hchtl, J. and Ferro, E. (2018), *The World of Open Data*, Springer International Publishing.
   URL: *http://www.springer.com/gp/book/9783319908496*

Colpaert, P., Joye, S., Mechant, P., Mannens, E. and Van de Walle, R. (2013), The 5 stars of open data portals, *in* T. Janowski, J. Holm and E. Estevez, eds, 'Proceedings of the 7th International Conference on Methodologies, Technologies and Tools Enabling E-Government (MeTTeG13)', ACM Press, pp. 61–67.

Crusoe, J. (2019), 'A national ecosystem of faults and frustration the case of open data in sweden', *High ranking e-government journal* . Manuscript submitted for publication.

Crusoe, J. and Ahlin, K. (2019), 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'. [In preparation and Peer Reviewed].

Crusoe, J. and Melin, U. (2018), Investigating open government data barriers, *in* P. Parycek, O. Glassey, M. Janssen, H. Jochen Scholl, E. Tambouris, E. Kalampokis and S. Virkar, eds, 'the 17th IFIP WG 8.5 International Conference on Electronic Government 2018 (EGOV 2018)', Springer, pp. 169–183.

Crusoe, J., Simonofski, A., Clarinval, A. and Gebka, E. (2019), The impact of impediments on open government data use: Insights from users, *in* 'Proceedings of the 13th International Conference on Research Challenges in Information Science (RCIS)', IEEE. Accepted.

Davies, T. (2010), 'Open data, democracy and public sector reform', *A look at open government data use from data.gov.uk* pp. 1–47.

Eisenhardt, K. M. (1989), 'Building theories from case study research', *Academy of management review* **14**(4), 532–550.

Elo, S. and Kyngäs, H. (2008), 'The qualitative content analysis process', *Journal of advanced nursing* **62**(1), 107–115.

Erlingsson, C. and Brysiewicz, P. (2017), 'A hands-on guide to doing content analysis', *African Journal of Emergency Medicine* **7**(3), 93–99.

Forman, J. and Damschroder, L. (2007), Qualitative content analysis, *in* L. Jacoby and L. Siminoff, eds, 'Empirical Methods for Bioethics: A Primer', Emerald Group Publishing Limited, pp. 39–62.

Gibson, L. (2010), Using email interviews. Working paper.

Graves, A. and Hendler, J. (2013), Visualization tools for open government data, *in* S. Mellouli, L. Luna-Reyes and J. Zhang, eds, 'Proceedings of the 14th Annual International Conference on Digital Government Research', ACM Press, pp. 136–145.

Gregor, S. (2002), 'A theory of theories in information systems', *Information Systems Foundations: building the theoretical base* pp. 1–20.

Handbook, O. D. (2015), 'What is open data?'.
URL: *http://opendatahandbook.org/guide/en/ what-is-open-data/* (accessed 2018-11-15)

Hartog, M., Mulder, B., Spée, B., Visser, E. and Gribnau, A. (2014), 'Open data within governmental organisations: effects, benefits and challenges of the implementation process', *JeDEM-eJournal of eDemocracy and Open Government* **6**(1), 49–61.

Heise, A. and Naumann, F. (2012), 'Integrating open government data with stratosphere for more transparency', *Web Semantics: Science, Services and Agents on the World Wide Web* **14**, 45–56.

Hossain, M. A., Dwivedi, Y. K. and Rana, N. P. (2016), '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.

Hsieh, H.-F. and Shannon, S. E. (2005), 'Three approaches to qualitative content analysis', *Qualitative health research* **15**(9), 1277–1288.

Huang, R., Lai, T. and Zhou, L. (2017), 'Proposing a framework of barriers to opening government data in china: A critical literature review', *Library Hi Tech* **35**(3), 421–438.

Jaakola, A., Kekkonen, H., Lahti, T. and Manninen, A. (2015), '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.

James, N. (2016), 'Using email interviews in qualitative educational research: creating space to think and time to talk', *International Journal of Qualitative Studies in Education* **29**(2), 150–163.

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

Janssen, M. and Zuiderwijk, A. (2012), Open data and transformational government, *in* 'Transforming Government Workshop'.

Kucera, J. and Chlapek, D. (2014), 'Benefits and risks of open government data', *Journal of Systems Integration* **5**(1), 30–41.

Kuhn, K. (2011), 'Open government data and public transportation', *Journal of Public Transportation* **14**(1), 5.

Lassinantti, J., Ståhlbröst, A. and Runardotter, M. (2019), 'Relevant social groups for open data use and engagement', *Government Information Quarterly* **36**(1), 98–111.

Lindman, J., Kinnari, T. and Rossi, M. (2016), 'Business roles in the emerging open-data ecosystem', *IEEE Software* **33**(5), 54–59.

Liu, Q., Bai, Q., Ding, L., Pho, H., Chen, Y., Kloppers, C., McGuinness, D. L., Lemon, D., de Souza, P., Fitch, P. and Fox, P. (2011), Linking australian government data for sustainability science - a case study, *in* D. Wood, ed., 'Linking Government Data', Springer New York, New York, pp. 181–204.
URL: *https:// doi. org/ 10. 1007/ 978-1-4614-1767-5_ 9*

Machi, L. a. and McEvoy, B. T. (2012), *The Literature Review - Six Steps to Success*, Corwin Press Inc.

Machi, L. A. and McEvoy, B. T. (2016), *The literature review: Six steps to success*, Corwin Press Inc.

Martin, S., Foulonneau, M., Turki, S. and Ihadjadene, M. (2013), Open data: Barriers, risks and opportunities, *in* E. Ferrari and W. Castelnovo, eds, 'Proceedings of the 13th European Conference on eGovernment (ECEG)', Academic Conferences and Publishing International Limited, pp. 301–309.

Maxwell, J. A. (2012), *Qualitative research design*, SAGE Publications Inc.

Mccoyd, J. L. and Kerson, T. S. (2006), 'Conducting intensive interviews using email: A serendipitous comparative opportunity', *Qualitative Social Work* **5**(3), 389–406.

Myers, M. D. (2013), *Qualitative research in business and management*, Sage Publications Inc.

Portal, E. D. (2017), 'Re-using open data'.
   URL: *https: // www. europeandataportal. eu/ sites/ default/ files/ re-using_ open_ data. pdf* (accessed 2019-04-19)

Safarov, I., Meijer, A. and Grimmelikhuijsen, S. (2017), 'Utilization of open government data: A systematic literature review of types, conditions, effects and users', *Information Polity* **22**(1), 1–24.

Sörqvist, L. (2004), *Ständiga förbättringar: en bok om resultatorienterat förbättringsarbete, verksamhetsutveckling och Sex Sigma*, Studentlitteratur.

Susha, I., Grönlund, Å. and Janssen, M. (2015), 'Organizational measures to stimulate user engagement with open data', *Transforming Government: People, Process and Policy* **9**(2), 181–206.

Tan, J. (2016), 'Obstacle factors of open government data: analysis based on the toe theoretical framework', *Journal of Intelligence* **35**(8), 175–178. in Chinese.

Ubaldi, B. (2013), 'Open government data - towards empirical analysis of open government data initiatives', *OECD* **22**(22), 1–60.

Vickery, G. (2011), 'Review of recent studies on psi re-use and related market developments', *Information Economics* pp. 1–44.

Walsham, G. (1995), 'Interpretive case studies in is research: nature and method', *European Journal of information systems* **4**(2), 74–81.

Walters, W. H. (2011), 'Comparative recall and precision of simple and expert searches in google scholar and eight other databases', *portal: Libraries and the Academy* **11**(4), 971–1006.

Whiting, L. S. (2008), 'Semi-structured interviews: Guidance for novice researchers.', *Nursing Standard* **22**(23), 35–40.

Whitmore, A. (2014), 'Using open government data to predict war: A case study of data and systems challenges', *Government Information Quarterly* **31**(4), 622–630.

Xiao, F., He, D., Chi, Y., Jeng, W. and Tomer, C. (2019), Challenges and supports for accessing open government datasets: Data guide for better open data access and uses, *in* L. Azzopardi, M. Halvey, I. Ruthven, H. Joho, V. Murdock and P. Qvarfordt, eds, 'Proceedings of the 2019 Conference on Human Information Interaction and Retrieval', ACM Press, p. 313317.

Zotou, M., Tambouris, E., Tarabanis, K., Tarrant, D., Vacher, E., Hermans, P., Ryberg, T. and Camacho, H. (2018), Open data education: Curriculum, course models and learning model, *in* S. Virkar, P. Parycek, N. Edelmann, O. Glassey, M. Janssen, H. Jochen Scholl and E. Tambouris, eds, 'Proceedings of the International Conference EGOV-CeDEM-ePart 2018', Donau-Universitt Krems, pp. 153–162.

Zuiderwijk, A. and Janssen, M. (2013), A coordination theory perspective to improve the use of open data in policy-making, *in* M. Wimmer, M. Janssen and H. Jochen Scholl, eds, 'the 12th IFIP WG 8.5 International Conference on Electronic Government 2013 (EGOV 2013)', Springer, pp. 38–49.

Zuiderwijk, A. and Janssen, M. (2014*a*), Barriers and development directions for the publication and usage of open data: A socio-technical view, *in* M. Gasc-Hernndez, ed., 'Open government', Springer, pp. 115–135.

Zuiderwijk, A. and Janssen, M. (2014*b*), The negative effects of open government data-investigating the dark side of open data, *in* G. P. Cid, S. P. Robertson, J. Zhang and J. R. Gil-García, eds, 'Proceedings of the 15th Annual International Conference on Digital Government Research', ACM Press, pp. 147–152.

Zuiderwijk, A., Janssen, M., Choenni, S., Meijer, R. and Alibaks, R. S. (2012), 'Socio-technical impediments of open data.', *Electronic Journal of e-Government* **10**(2), 156–172.

6 Overview of Collected Data

Table 1: Interviewees participating in the development of the framework

7   User Processes

Table 2: [9] user process for their news notification service. Machine refers to machine processing.

Table 3: [R10] user process for writing articles.

Table 4: [R4] use process for telling a data story.

Table 5: [R11] use process with developing an analytical tool.