Lappeenranta-Lahti University of Technology LUT

School of Engineering Science

Software Engineering

Master's Programme in Software Engineering and Digital Transformation

Vlada Haikara

Searching behavior of experienced users and novices on open data portals

Examiners: Assistant professor Annika Wolff

Assistant professor Antti Knutas

Supervisor: Assistant professor Annika Wolff

TIIVISTELMÄ

Lappeenrannan-Lahden teknillinen yliopisto LUT
School of Engineering Science
Tietotekniikan koulutusohjelma
Master's Programme in Software Engineering and Digital Transformation

Vlada Haikara

Kokeneiden käyttäjien ja noviisien hakukäyttäytyminen avoimen datan portaaleissa

Diplomityö

2022

51 sivua, 24 kuva, 1 liite

Työn tarkastajat: Apulaisprofessori Annika Wolff
Apulaisprofessori Antti Knutas

Hakusanat: avoin data, haku, käytettävyys
Keywords: open data, searching behaviour, usability

Mitä useammin käytämme ohjelmaa, sitä helpompaa meidän on navigoida ja suorittaa monimutkaisempia tehtäviä, mutta päteekö tämä avoimiin data portaaleihin? Tämän opinnäytetyön tarkoituksena oli selvittää, eroaako avoimien data portaalien kokeneempien käyttäjien ja aloittelijoiden käyttäytyminen ja yrittää ymmärtää, miten avoimen datan portaaleja voidaan parantaa ja saman aikaisesti parantaa kokemattomien käyttäjien käyttökokemusta. Tässä tapauksessa ongelman tutkimiseksi ja mahdollisen ratkaisun löytämiseksi suoritettiin kysely, johon osallistui 64 henkilöä ja haastateltiin sen jälkeen 8 käyttäjää. Tulokset osoittivat, että kokeneemmat käyttäjät ovat halukkaampia käyttämään kehittyneempiä hakutyökaluja, kun taas aloittelijat valitsevat yksinkertaisempia polkuja, vaikka ne eivät aina auta. Tutkimuksen ja haastattelujen tulosten perusteella opinnäytetyössä esitetään suosituksia avoimen datan portaalien parantamiseksi.

ABSTRACT

Lappeenranta-Lahti University of Technology LUT
School of Engineering Science
Software Engineering
Master's Programme in Software Engineering and Digital Transformation

Vlada Haikara

Searching behavior of experienced users and novices on open data portals

Master’s Thesis

51 pages, 24 figures, 1 appendix

Examiners: Assistant professor Annika Wolff
Assistant professor Antti Knutas

Keywords: open data, searching behaviour, usability

The more often we use the program, the easier it is for us to navigate it and perform more complex tasks, but does this apply to open data portals? In this thesis, the task was to check whether the behavior of more experienced users and novices of open data portals differs and try to understand how open data portals can be improved so that they will provide better usability for users with little experience. In this case, to research the problem and find possible solution a survey involving 64 people and then interview 8 users were conducted. The results showed that more experienced users are more willing to use more complex search tools, while novices tend to choose simpler paths, although they are not always successful. The thesis provides recommendations for improving open data portals based on the results of the survey and interviews.

ACKNOWLEDGEMENTS

I would like to thank my supervisor Annika Wolff for her patience and support throughout the work on the thesis. Also, I would like to thank all those who participated in the survey and interviews, without you this thesis would not exist. As well as to thank my family and friends for their support and faith in me.

TABLE OF CONTENTS

LIST OF SYMBOLS AND ABBREVIATIONS

API	Application Programming Interface
CSS	Cascading Style Sheets
HRI	Helsinki region infoshare
HTML	HyperText Markup Language
LUT	Lappeenranta University of Technology
PDF	Portable Document Format
XLS	Microsoft Excel file format

1 INTRODUCTION

1.1 Background

Open data has become a frequent topic of conversation or new articles in recent years, but often in the context of the advantages or disadvantages of open data and rarely about how to make open data usable by everyone, regardless of technical background. The fact that not everyone has the technical skills to work with data greatly affects the inequality in the use of open data, especially in developing countries. Since only advanced users can truly use open data for any purpose, it puts less tech-savvy users in a difficult situation. Even if they want to use open data, they will have to learn new technical skills. But at the same time open data portals have loads of data that can be presented in a more readable format for non tech-savvy people.

It is also suggested by literature and research articles that open data portals have drawbacks, for example, creating open data portals is based on the idea that open data users have the resources, expertise, and capabilities to use the data. But in reality, things are a little different. Some data requires the use of statistical techniques, a deep understanding of the underlying data, and an understanding of the types of data relationships. And it turns out that only people with an understanding of statistical methods and other knowledge necessary to process open data can make sense of the data and understand how they can use it [1]. However, there are also many tools to reduce the harm from disadvantages. For example, visualization and recommendations for potential uses of data can help inexperienced users to work with it [2]. This, however, requires ongoing efforts to consider the user's perspective and track needs, ultimately helping users. The literature also describes that in general, in the development of open data portals, there is a limited understanding of the capabilities and needs of users. Also articles conclude that the user's view is largely ignored when developing an open data portal. [3]

Most of the open data portals work as a kind of archive in which the user can search for the datasets, they needs using search, filters or categories. Since the search for datasets is the

main task of the user on the site, the search should be fast and convenient. But often in open data portals, datasets are stored in such a way that it would be convenient to store them, but not so easily support search. Typical problems encountered is that the search only works on the title, and the title itself often doesn't fully match with what is inside dataset. And this confuses the user, especially the inexperienced, and makes them go to look for information in other sources. This inspired the research question "Is the search on open data portals different for users who are tech-savvy and users who are less tech-savvy?". To answer this question, a survey was drawn up and conducted among 64 people. Also, some time after the survey, interviews were conducted with users in a more controlled environment. It also helped to understand user behaviour better and answer another research question "What ways it is possible to find to improve the experience of an inexperienced user with open data portals?"

1.2 Goals and delimitations

In this study, I will focus on the problem of using open data portals by inexperienced users and how exactly it is possible to help them (make using of open data portals easier for them). Open data portals can be difficult to use for people who have little expertise in working with data. Common barriers include that data can be hard to find, also hard to understand if data is suitable and people do not know which tools to use. Previous research has focused on trying to solve this problem by either improving data literacy and teaching people how to use data better or redesigning interfaces for data exploration. However not much research has yet focused on data portal themselves. This research will investigate usability of data portals using survey and interviews.

1.3 Structure of the thesis

Section 2 contains a description of open data and open data platforms. In this section importance of open data as well as opportunities and critiques of it will be discussed. In addition to that the purpose, usability and problems of open data portals will be discussed.

In section 3, methods and materials will be discussed, explanation of methods chosen survey and user interview will be opened. In section 4, information about participants will be given as well as survey and interview results with figures. In section 5, opportunities based on research findings will be discussed.

2 LITERATURE REVIEW

2.1 Open data

2.1.1 Definition

Nowadays many people talk about open data, but how many people know what does it really mean? According to an EU directive, open data as a concept is generally understood to denote data in an open format that can be freely used, re-used and shared by anyone for any purpose. [4]

Speaking about quality of the data, open data should be public information, but at the same time, confidential or otherwise compromising information cannot be published due to the protection of people's privacy rights and public safety. Also, open data should be accessed free of charge. Reason for that is free access is especially important for people exploring the data source for the first time and it also allows people to experiment with data without budget bureaucracy. [5]

In terms of technical accessibility, the data should be published in a format that can be understood by computer programs and that makes it easier to use when developing new web services. Information in PDF documents or most HTML web pages is often difficult to use in derivative works. Publishing data in CSV or XLS formats or providing access to the original data source through an application programming interface (API) are the best solutions. [5]

We can conclude that we can call data open when anyone can freely access, use, re-use and distribute it, for any purpose and without restrictions.

2.1.2 Importance of Open Data

According to the European data portal, open data can improve productivity and improve the efficiency of public services. Improving the efficiency of processes and delivery of public services can be achieved through the exchange of data between sectors, which, for example,

can provide an overview of unnecessary spending and improve using of budget money. [6]

The economy can benefit from easier access to information, content, and knowledge, which in turn will foster the development of innovative services and the creation of new business models.[7]

Social welfare can be improved using open data, because society benefits from more transparent and accessible information. Also, open data improves collaboration, participation and social innovation.[7]

Social services

Open data can improve the quality of social services. We can find interesting information about improving the quality of social services, for example in developing countries with a high level of corruption, some services (including medical and educational) often do not reach the population, and their quality leaves much to be desired. Thus, in Bangladesh, Ecuador, India, Peru and Uganda, the percentage of absent teachers and doctors who are employed and perform their duties according to the papers is 20% and 35%, respectively. [8]

Releasing this data can help reduce corruption and improve public services. In Uganda, researchers provided information to parents by publishing school funding figures in the local press. As a result, the level of corruption has dropped significantly, while the quality of educational services has improved. Thus, providing open information can help the public to discipline public service providers. [9]

Democracy

Thus, open data is useful for democracy. For example, if citizens know the details about the work of the authorities in their state, about expenditures from the state budget, they can bring their own officials to justice. Open data can also help government officials improve their plans, goals for improving people's lives, strengthening economies and improving environmental health.

An illustrative example is Sunlight Labs' Clear spending project, which matched data from US government budgets with data from US contracts. As a result, multiple discrepancies and errors were found. The Sunlight Labs report was reviewed in Congress.[10]

Open data is an important tool for improving the quality of citizens lives. It can make every system more transparent and help to make problems for those who want to hide things from society.[11]

2.1.3 What can be made with open data

What can be made with open data People also may use open data to make useful services for people. For example in Russia, an application has been developed that uses open data to help allergy sufferers avoid allergy attacks based on actual and predicted data on the distribution of pollen in the air. App is called Pollen Club - http://pollen.club/.

Another example is that people can also carry out data analysis, as, for example, did the creators of the trash city project. They decided to create a map that showed how "student's moving week" affects the quality of life and the amount of garbage on the streets of Boston. https://jhaddadin.github.io/trashcity/index.html

Also, for example it is possible to create a website that allows users to examine and compare pollution emissions throughout Europe, using data from the European Pollutant Release and Transfer Register. http://Paastot.fi

Or, there is a portal that displays public information on the quality of learning in each school, county and state in Brazil. It uses Education data from the Ministry of Education in Brazil. For example, parents may use this service when they are deciding where they want their children to go to school. http://www.qedu.org.br

Also, in terms of transparency, projects like the Finnish "Tax Tree" and Britain's "Where Does My Money Go" show how your tax money is spent by the government. This increases the credibility of the government and improves its transparency.

2.1.4 Critics of open data

Open data potentially provides benefits which can improve every sector of people's lives, but everything in the world has two sides and here I would like to describe the barriers against open data.

Financing and sustainability

The key question in case of financing and sustainability is, how are open data projects funded in the absence of a direct revenue stream? Answer for this question is that providing a stable financial base is best achieved through direct government subsidies. This subsidy is driven by factors such as the fact that open data will generate diversified consumer surplus value, creating significant public goods that are worth the investment of government spending. [12]

And similarly, open data will lead to new innovative products that will create new markets, which in turn will lead to additional corporate income and tax revenues[13]. So, for government open data might be an opportunity to make investment, but the problem is that returning of investments is not guaranteed.

Without government subsidies getting money for developing applications might result challenging. The European Commission reported that the average app developer earns only $ 3,000 a year from app sales. That said, 80% of paid Android apps have less than 100 total downloads/purchases. In addition, they noted in the report that even successful apps like MyCityWay, which have been downloaded 40 million times, have not yet made a profit. In the face of austerity and government cutbacks, finding the necessary funds to open data is challenging. [14]

Financially open data initiatives are not protected and need a lot of effort to get funding for innovations and development with open data. So, world might see only initiatives needed for people or organisations with money, this makes risks that open data initiatives will be available only for people with special skills or resources.

Transparency and equality

Open data can both improve equity and worsen it. A very interesting and well-documented example of how open data can become a bad tool in the hands of people who exercise their position and power can be found in the work of Sulli Benjamin and his colleagues studying the impact of digitization of land records in Bangalore. Their findings were that the newly emerging access to information on land tenure in Bangalore is mainly used by middle- and high-income people and organisations to obtain land ownership from the marginalized and poor people. Recently digitized and openly available data has allowed wealthy people to use the information provided as a basis for instructions to land surveyors, lawyers and others to challenge property rights, exploit errors in documentation, and identify opportunities and purposes for bribery. They were able to directly transform their expanded access to information, together with their existing access to capital and skills, into unequal competition over land rights, lawsuits and offers to buy for personal gain, and further deterioration of those already disadvantaged. [15]

Here we can see an example how open data might also work for bad and on the contrary, widening the social gap and worsening democracy. When people with skills and knowledge use data for their own purposes and at the same time people who do not have this knowledge are in a bad position. And unequal open data availability can make this even worse.

2.1.5 Opportunities of open data

The potential of open data is seen in facilitating the production of new knowledge and profit by creating new ways of collecting, producing, visualizing, and analyzing information. From a media perspective, the opportunity to expand data resources to provide opportunities for more informative, transparent, and credible journalism. The role of journalism has also been seen as an intermediary between decision makers and citizens. It should also be noted that the issue of illustrating and enhancing the usability and understanding of data/information is not a completely new issue in discussions related to information and communication technologies. While discussions on democracy and citizen participation have viewed

information from a broader perspective the aim was to find ways and tools for communicating and presenting information in order to enable more open and collaborative decision-making. However, discussions about open data have focused on defining the origin or format of data and developing rules governing the use of data. [16]

2.2 Open data portals

2.2.1 Purpose of open data portals

Open data portals are web-based interfaces designed to make it easier to find re-usable information. Open data portals are like library catalogues or an archive, they contain metadata records of datasets published for re-use, i.e. mostly relating to information in the form of raw, numerical data and not to textual documents. In combination with specific search functionalities, for example search, filters or categories, they facilitate finding datasets of interest. Application Programming Interfaces (APIs) are also often available, offering direct and automated access to data for software applications, which helps users to get easier access to open data and also makes use of it easier. [17]

2.2.2 Usability of open data portals

If we consider the needs of the user of open data, then we can highlight these needs:

Relevance

A study by researchers from the UK found that searches require users to be supported in evaluating datasets according to their relevance, usability, and quality. This can be achieved by providing visual or textual indicators of these aspects in the interface, supported by automatically calculated metrics or user-generated reviews and annotations. [18]

Usability

Users may have a problem related to this need, in order to fix the problem it is necessary to add documentation on labelling, licensing, access and machine readability. [18]

Trustworthiness

To fill this need, it is necessary to add documentation on the methods of data collection, their origin and completeness. [18]

Reuse potential

The solution to this need is to add documentation on the potential use of the dataset and actual examples of reuse of the dataset. [19]

Exploration

This need is about the characteristics of datasets. User problems here can be, for example, lack of ability to download datasets, access to them in a machine-readable form, or lack of licenses to allow truly free datasets to be reused [20].

Analysis

What will help to solve this need is the presence of complete, relevant, georeferenced and disaggregated datasets on the open data portal. [20] [18].

2.2.3 Problems of open data portals

The problem of most of open data portals is that most of organizations or citizens who publish open data tend to take the path of least resistance and publish data in its original format, ignoring the opportunity to make data available in more readable formats[21]. This again leads to problems for end users who want to be able to see the data in a readable format. Usually, open data portals set some quality requirements for data they will publish. [22]

One of the motivations of the open data movement has been making these datasets more accessible to various stakeholders to encourage them by participating and contributing for the public policy. Achieving that public policy will be transparent, open and accessible. But to fully exploit the potential of open data, users usually require a certain level of applied skill. There is still no easy-to-use, proven solution that could help ordinary citizens use open data to make decisions about their own lives or contribute to broader public policy debates [21]. Also, often on open data portals, the data is in its original, unprocessed form and as it is, this data is often not very useful to end users. While some publishers prefer the data in

a human-readable format, others prefer a machine-readable format [21]. This also makes things even more difficult for end users. Open data policies pay little attention to the user's point of view, whereas it is users who create value from open data. The barriers currently facing the open data process are: accessibility and access;the ability to find; usability; the ability to understand; quality; linking and combining data; comparability and interoperability; metadata; interaction with the data provider and discovery and download [23].

Scaling up the use of open government data requires strong leadership commitment and changes in the broader organizational landscape, such as the creation of formal and informal rules and technological developments that stimulate debate about how to use open data. [24]

To conclude it is important to say that most of the problems of open data portals are about data quality and data formats. Also, open data as a term is about making data available for using and re-using for everyone, but open data portals are usually made for people with technical skills. Which This automatically means that open data is not for the use of everyone. Most of the people are not able to use open data from the open data portals and even if they can in many cases it is not easy to use portal and find right data there.

3 METHODS AND MATERIALS

This study is about the user experience of open data portals and for this reason during the planning process it was decided that this study should not be only theoretical, it also should be based on both qualitative and quantitative research meaning survey and user interviews. This is the best way to observe and understand the flaws and opportunities of open data portals. Because this study was conducted in Finland, the Finnish open data portal was chosen as an example portal for survey and user interviews. This portal is Helsinki region infoshare https://hri.fi/, the aim of this portal is to make better use of public data in the Helsinki region. This is not a huge open data portal, but it meets the requirements for the conduction of the study. Also, one of the criteria was an English version of the website, because this study was conducted in English.

While planning, it was also decided that the study should be carried out in two rounds: task-based survey and interviews. The main reason for this is while surveys and interviews have their own disadvantages when combined, the best result can be obtained. Survey covers the part for quantitative research - it allows to take a large sample size and a wide range of information can be collected. Interviews are qualitative research - they are particularly useful for getting a wider understanding of a participant's experiences. Also, it was harder to conduct interviews during Covid-19 times and restrictions, so only interviews were not the best option for this study. Considering all those reasons it was decided to conduct a survey first to cover a larger number of users and when it would be possible, to then conduct interviews in a way more suitable for users online via zoom, teams, or in person in the LUT university's premises.

3.1 Survey Design

The aim of the survey was to understand how novices and more experienced users use open data portals and understand if there are differences in data set searching strategy between them. Each of the participants was asked to complete three tasks. Each person needed to find a dataset according to the instruction of the task. Survey tasks were given in a way that imitate the behaviour of a user of an open data platform. Questions were formed so that it

was hard to find the dataset by just copying questions to the search field on the open data portal because usually, people search by giving one or a couple of words to the search field, not a sentence.

While planning the survey some problems were found, it was hard to understand if the user found exactly the required dataset which he was asked to find in the survey. To cover this part in the survey, questions were added to check that the found dataset was correct. For example, asking the user to compare the dataset they found to the real dataset asked in the survey or provide a link to the dataset that the participant found. After a successful or unsuccessful task attempt, the participant was asked more questions to understand how the user was searching for the dataset, for example, it was possible to know if he/she was searching using filters or categories or just used search field. In addition to that, the survey had open questions aimed to understand user flow and understand the thinking of the user, for example, to question why the user made this choice to use search or filters and which option was successful. as Also, a question aimed at understanding how easy or difficult it was to find the information participant needed.

After the participant completed all three tasks from the survey, they had to answer a multiple-choice question that helped to understand their level of experience with the data. For example, how experienced the user is with performing simple/intermediate/complex data analysis or questions about his experience with open data portals. At the end of the survey, there were open-ended questions aimed at gathering information that could help improve open data portals like asking in general what problems users had with open data portals or what they think can help to improve open data portals. From the point of view of user experience, this survey covered the searching on website part and also contained general questions about the user experience on the open data platform.

3.1.1 Survey tasks

Survey tasks were chosen so that there are easier and harder datasets to find. One task also was not so easy to find because words were different in the title of the dataset and in the task question itself. This was made because users of open data portals don't know the titles of datasets before they see them and the wording of the titles of the datasets can be a little bit

different even if datasets are quite similar. And because different tasks were used in the survey, it was possible to compare easy and hard tasks between themselves which showed a fuller picture of user experience in the open data platform.

Also, in order to get good results, the survey had to check if the participant found the needed dataset by asking the participant to send a link to found dataset. Then, if the participant didn't manage to find the needed dataset, he/she was asked what in their opinion was a problem. Also, the right dataset was shown to participants even if they answered that they did not manage to find the dataset after that participant was asked to tell if this dataset was seen while searching, because sometimes participants saw the right dataset while searching, but didn't recognize that it is the right one and this was a problem. But about problems, we will talk more in the results section.

I would like to show here all tasks of the survey and explain why this task was chosen to be part of the survey. Each task was given in exactly the same form as it is here in quotes.

First task was: “You are wondering how many cyclists are in the Helsinki region.” In this task the participant was asked to find dataset with information about number of cyclists in Helsinki region. This was quite easy task because dataset was possible to find using keyword “cyclists”. It was not only one in search results, but just quick look at the list and dataset was possible to find.

Second task was: “You are interested in information about building of how many apartments was started in Helsinki region in 2013.” In this task the participant was asked to find dataset with information about how many apartments was started to be built in Helsinki region in 2013. This task was the most complicated of all and reason for that was title of dataset didn’t contain any word of task given to participant. But it was possible to find it using keywords “housing started”.

Third task was: “You are interested in what is the temperature of the swimming water in Helsinki.” This task is also quite easy, it was possible to find it using keywords “swimming water”.

3.2 Interview Design

At the end of the survey, there was a contact form and all who wanted to participate in the user interview later could do it by filling out the form. Overall, in the user interview, 8 people participated. In the interview participants were asked to make tasks similar to survey tasks, meaning to find a dataset on the open data portal, but during the interview, I was observing how the process was going in real-time and writing notes to myself. Due to Covid19 restrictions some of the participants did the interview online via Zoom and in this case, the participant shared his/her screen, so it was possible to observe everything happening on the screen. Also, screen capture was recorded, but one recording attempt failed due to technical reasons. Tasks were prepared in advance and can be found in chapter 3.2.1, the interview started with welcoming the person and explaining what and why we are doing and in case the interview was made online, the participant was asked to share the screen and after that, I put the recording on. The participant got tasks in text form through chat in Zoom in case of an online interview, in-person I had tasks ready in written form on paper. In addition to this, notes were always made, for example, it was written in a notebook or to word document how the user behaved on the open data portal, if he was lost, or some decision took more time. Participants were also asked to speak while they were doing the task and explain their decisions, but not all the participants were explaining all their steps and it was decided that it is more important not to disturb participants and ask questions later after he/she finished the task. After completing the task participants were asked questions with the aim to understand how easy or difficult it was for the participant to find the needed dataset and their level of experience with the data. Questions were prepared in advance but depending on the situation there were questions added.

3.2.1 Interview tasks

For the interview two tasks were prepared, it was decided to make only two tasks because of time limitation. After checking survey results it was found that for one task participants sometimes spent more than 10 minutes, especially if they were really interested in finding

right task. So, the decision was made to make only two tasks for interview round. Same as in the survey, here one task was a little bit harder than the other.

First task was: “You are interested in how many of unemployed women were in Vantaa in 2014” Here participant was asked to look for a dataset that contain information about number of unemployed women in Vantaa in 2014.

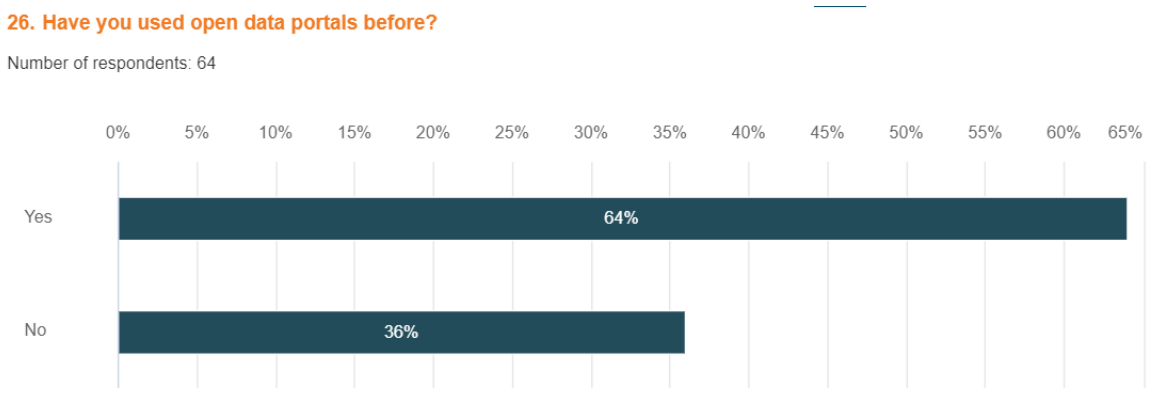
Second task was: “You are wondering how many students were learning A2 Spanish in autumn 2020 in Vantaa” Here participant was asked to find dataset with information about students with A2 Spanish in Vantaa.

4 RESULTS

As soon as the structure of the survey was ready, it was needed to find users to test the open data platform website and complete the survey. For this study, both novices and experienced users of open data portals were needed. The best way to find real experienced users was to distribute a survey among users of open data portals. To make it happen a link to the survey was sent to different Facebook groups related to the open data topic. Also, the link got shared by the official Facebook group of Helsinki region infoshare.

4.1 Participants

To represent novices, it was decided to promote survey links among students in LUT university, although it was also possible that in university, we have some more experienced users of open data portals. The survey was opened by 649 respondents, 85 started answering and 64 of them were surveyed till the end. It is possible to see in figure 1 that overall, out of 64 users 41 used open data portals before they started to fill the survey and for 23 users it was the first time using open data portals. According to this data, we had 41 experienced users and 23 novices.



	n	Percent
Yes	41	64.1%
No	23	35.9%

Figure 1 How many participants used open data portals before.

Also, we know the age of participants, as it is possible to see in figure 2 in the survey most common age of participants was between 21 and 34. But also, we had participants younger than 18 and older than 65.

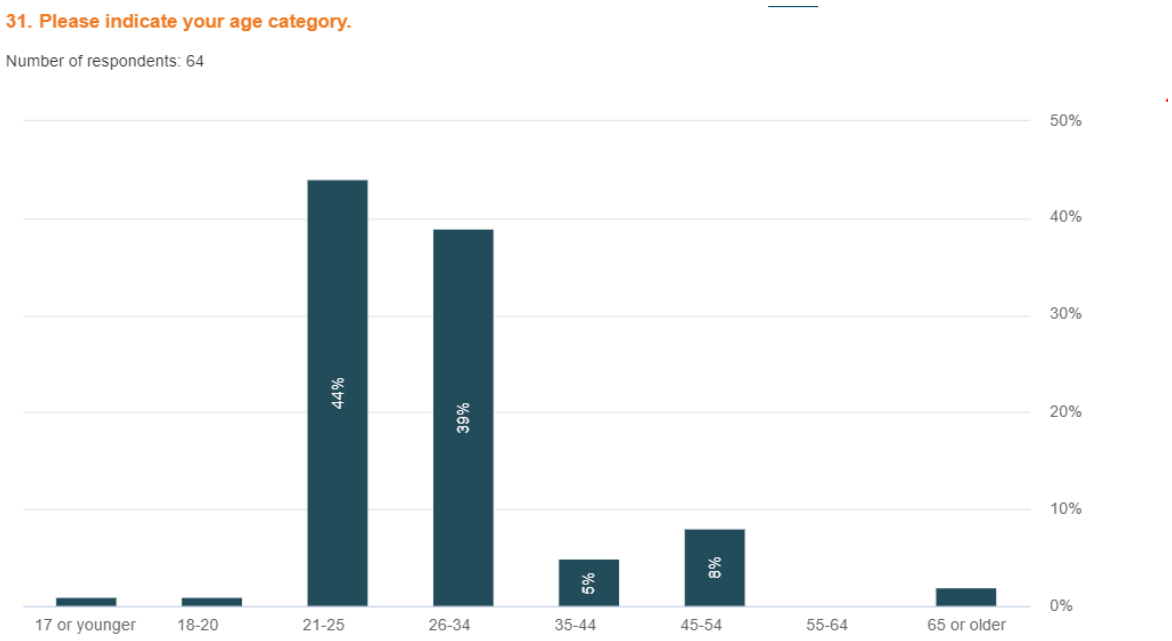


Figure 2 Age categories of participants.

General population statistics, as it can be seen in Figure 3 72% of participants are male and 28% female.

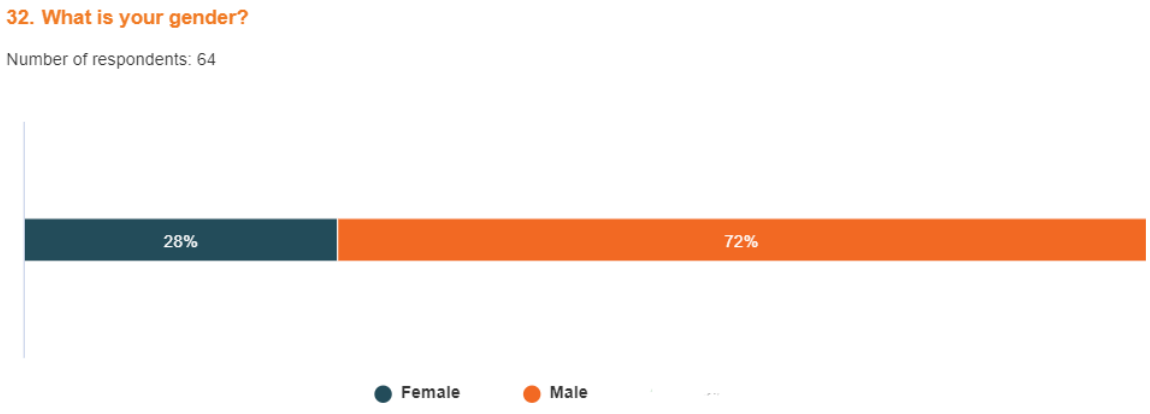


Figure 3 Gender of participants.

The average time respondents took to do the survey and tasks was 20 minutes. The longest session was 56 minutes and the shortest was 8 minutes.

4.2 Survey results

4.2.1 Experienced users and novices

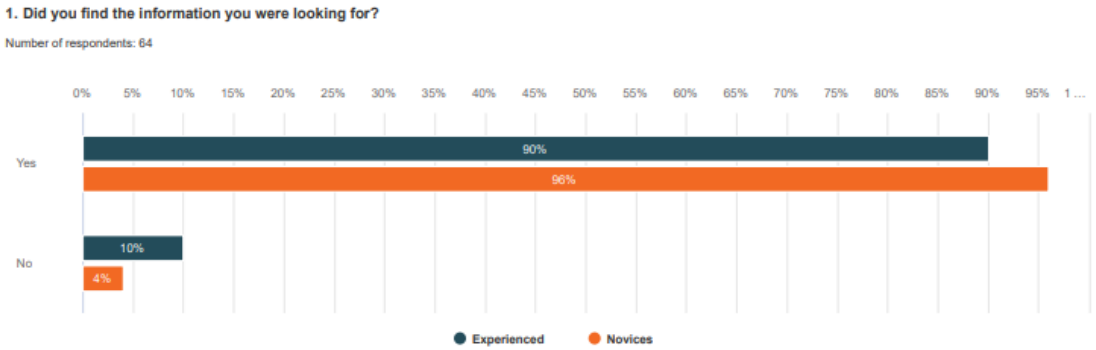
To look more at results from the point of view of novices and experienced users Webropol's data analysing tool was used to compare report groups. The first group was all participants who answered "Yes" to the question "Have you used open data portals before?" and the second group who said "No" to the same question. Data from those two groups was checked about all survey tasks to find some correlations.

The key research question was whether there are differences in strategies between more and less experienced users of open data portals. It was decided to understand the experience in terms of those who had and those who had not used open data portals before. This is an approximation, and in a larger study, there could be a more subtle way to distinguish between these two types of users. Next, there will be information about each task and data from the survey.

4.2.1.1 First task

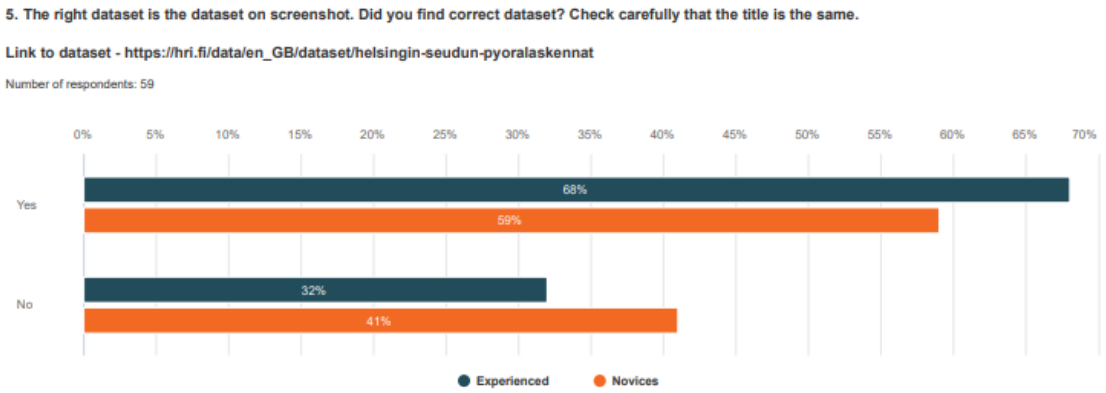
The task was "You are wondering how many cyclists are in the Helsinki region." Here we will see all the data about answers of experienced users and novices.

In figure 4 it is possible to see how many experienced users and novices found the dataset. But more interesting information will be in figure 5 because it will show us how many participants really found the needed dataset.



	Experienced		Novices		
	n	Percent	n	Percent	Total
Yes	37	90.2%	22	95.7%	59
No	4	9.8%	1	4.3%	5
Total	**41**		**23**		**64**

Figure 4 How many experienced users and novices found the dataset for task 1



	Experienced		Novices		
	n	Percent	n	Percent	Total
Yes	25	67.6%	13	59.1%	38
No	12	32.4%	9	40.9%	21
Total	**37**		**22**		**59**

Figure 5 How many experienced users and novices correctly found the dataset for task 1

As was mentioned earlier it was decided to first ask the user if they found the dataset and then ask them again but showing the real correct dataset. If we look at Figure 4, we see that 59 participants said that they found the needed dataset. But if we look at Figure 5 it is

possible to see that only 38 participants out of 59 who said that they found the correct dataset really found the correct dataset. If we look closer to Figure 5 we will see that almost 41% of novices didn't really find the dataset. And if we compare it to Figure 4 we will see that out of 22 participants 9 found the wrong dataset. But same happened with more experienced users in Figure 4, 37 participants said that they found the dataset when only 25 really found it.

The participants who didn't find the dataset were asked to tell what the problem in their opinion was. Here are the results in figure 6. The most interesting comment I found here is "The structure of the location of information on the site is complex" and this comment came from a person who already used open data platforms before. Also, I think the question in the task could be a bit misleading because from comments I understood that some people were looking not just for the dataset, but for a specific number. And of course, they were not able to find it.

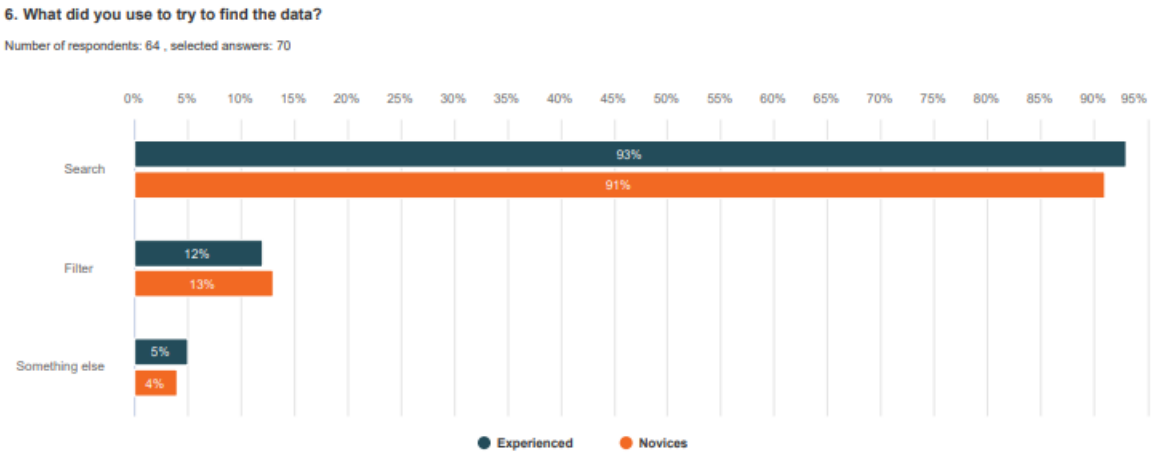
3. What in your opinion was the problem?

Number of respondents: 5

Experienced	Novices
The numbers I found were rather meaningless without more context provided.	I don't know what is helsinki region
the structure of the location of information on the site is complex	
It is because when on the search bar the information provided for number of cyclists in Helsinki was not specified and i could not find it at all.	
I was looking for a aggregated number, which might be just me misunderstanding the task. I did find results for spot counting.	

Figure 6 What was the problem in task 1

Next in our result analyses, we wanted to check what participants used to find the data. In Figure 7 it is possible to find results. But note that Figure 7 represents all user no matter if they found the correct dataset or not. In Figure 8 it is possible to see only participants who found the correct dataset and their answers.



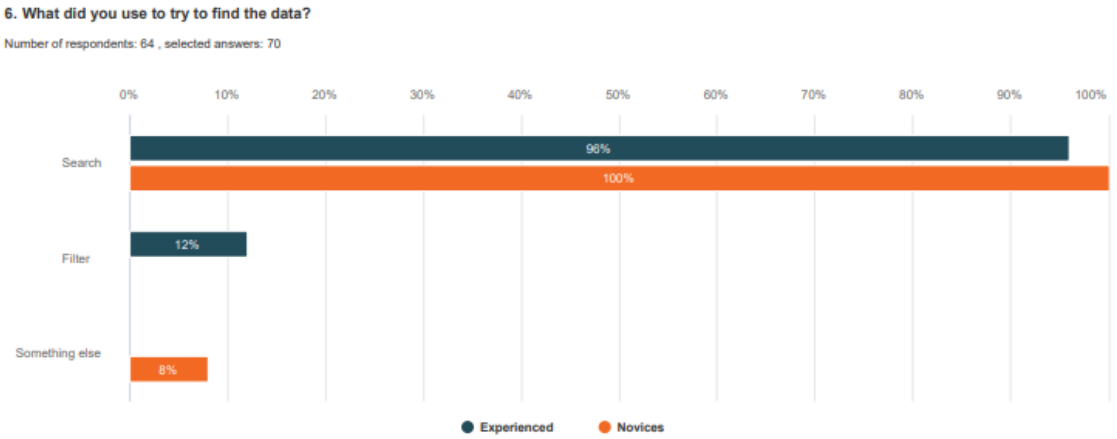
	Experienced		Novices		
	n	Percent	n	Percent	Total
Search	38	92.7%	21	91.3%	59
Filter	5	12.2%	3	13.0%	8
Something else	2	4.9%	1	4.3%	3
Total	**45**		**25**		**70**

Answers given into textfield

Option names	Text
Something else	Filter 'traffic and tourism', then search 'cyclist'
Something else	Searching in dataset pages
Something else	tag

Figure 7 What participants used to find the dataset in task 1

From here we can see that both novices and more experienced users of open data platforms use the same tools to find datasets. But the most popular tool for searching datasets is the search.



	Experienced		Novices		
	n	Percent	n	Percent	Total
Search	24	96.0%	13	100.0%	37
Filter	3	12.0%	0	0.0%	3
Something else	0	0.0%	1	7.7%	1
Total	**27**		**14**		**41**

Answers given into textfield

Option names	Text
Something else	tag

Figure 8 What participants who found correct dataset used to find the dataset in task 1

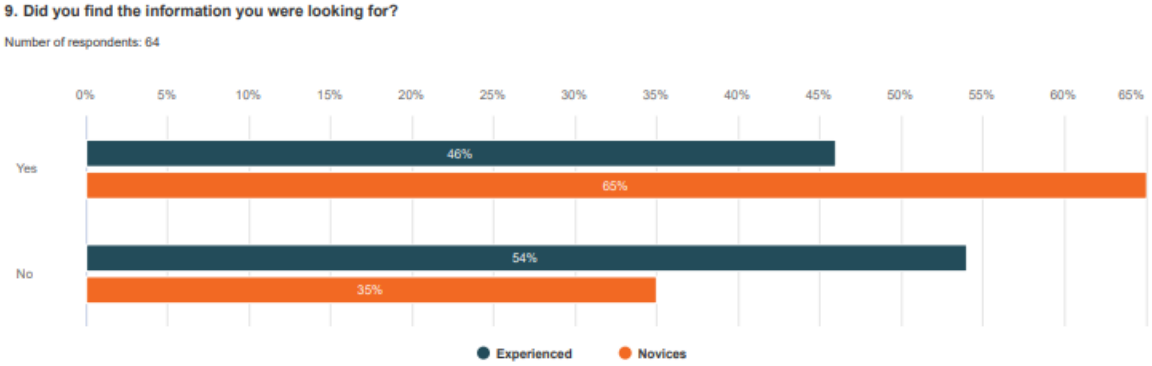
Here in Figure 8, we can see that search is the most popular and the most relevant tool to find datasets on an open data platform. Compared to Figure 7 we can see that filter was not successful for novices, but worked for experienced users, because number of experienced users didn't change in Figure 8, compared to Figure 7.

Also, in the survey there was a question to understand why participants chose one tool or another, I will include this in an appendix, but here I will tell a summary of answers. Experienced users said that they used search because they are lazy, it was the first thing they saw on the webpage, it is the easiest and simplest option, and that search is faster than browsing categories.

4.2.1.2 Second task

The second task was "You are interested in information about the building of how many apartments were started in Helsinki region in 2013." And this task was the hardest one. Here we will see all the data about answers of experienced users and novices.

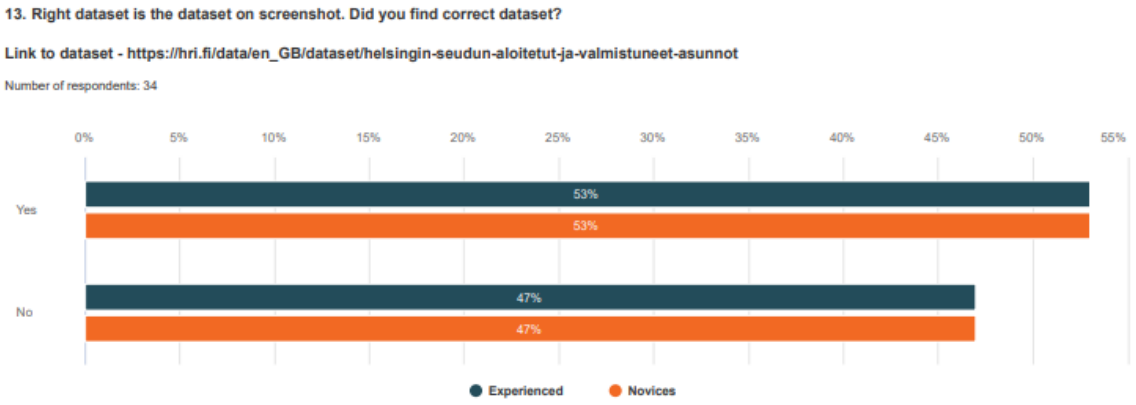
In Figure 9 it is possible to see how many experienced users and novices said that they found the right dataset before they saw the right answer. 34 participants out of 64 found the dataset.



	Experienced		Novices		
	n	Percent	n	Percent	Total
Yes	19	46.3%	15	65.2%	34
No	22	53.7%	8	34.8%	30
Total	**41**		**23**		**64**

Figure 9 How many experienced users and novices found the dataset for task 2

And in Figure 10 how many really got it correctly. As it is possible to see, many users got this dataset incorrectly. Only 18 out of 64 found the dataset correctly. But here I should mention that this question was the hardest one and this is why results are like this.



	Experienced		Novices		
	n	Percent	n	Percent	Total
Yes	10	52.6%	8	53.3%	18
No	9	47.4%	7	46.7%	16
Total	19		15		34

Figure 10 How many experienced users and novices found the correct dataset for task 2

The participants who didn't find the dataset were asked to tell what the problem in their opinion was. The full results are found in the annex, here I will just collect the most interesting answers in it's original form.

Answers from more experienced users:

- The data is only available as a shape file
- Keyword not in correct order
- The problem was in the wrong combination of search words. The phrase "number of apartments in Helsinki region" didn't work, but "started and completed housing in the Helsinki region" worked successfully
- Keywords not adapted in the description

Answers from novices:

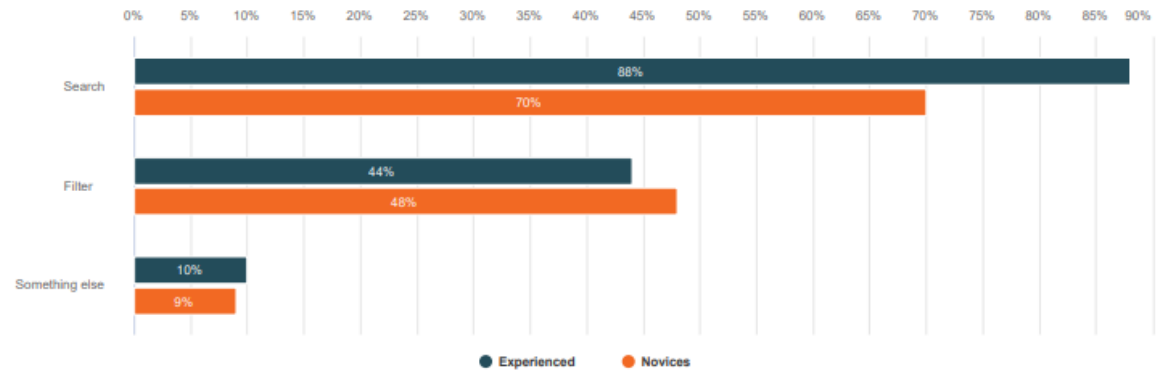
- Fields of tables are not available for reading. It is not possible to identify the information contained in the tables

- I haven't found the information using the search function. Then I tries with the filters, by using the 'housing' and 'helsinki' filter. I have found many things but not the building of apartments in 2013. Maybe the question was not well framed.
- I haven't the right application in my computer, which would open the database HKI rakennuskanta.

Next in our result analyses, we wanted to check what participants used to find the data. And in Figure 11 it is possible to find results. But note that Figure 11 represent all user no matter if they found the correct dataset or not. In Figure 12 it is possible to see only participants who found the correct dataset and their answers. From Figure 11 we can see again that search

14. What did you use to try to find data?

Number of respondents: 64 , selected answers: 87



	Experienced		Novices		
	n	Percent	n	Percent	Total
Search	36	87.8%	16	69.6%	52
Filter	18	43.9%	11	47.8%	29
Something else	4	9.8%	2	8.7%	6
Total	**58**		**29**		**87**

Answers given into textfield

Option names	Text
Something else	Filter 'housing', then search 'building'
Something else	Categories from the top
Something else	Data-Categories-Constructed environment
Something else	Sorting
Something else	tag
Something else	categories

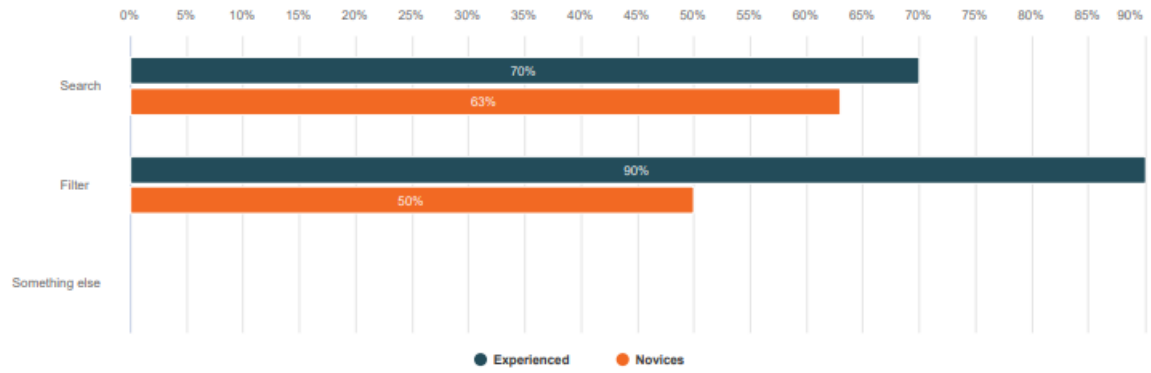
Figure 11 What participants used to find the dataset in task 2

is the most common tool for searching information on open data portals, but from Figure 12 we can notice that the filter option became the most popular for more experienced users. But here it is also important to notice that participants could choose multiple answers to this question and look at results I would say that most probably participants used both search and filters to achieve the goal in this task.

From the open answer "something else" in Figure 11 it is possible to understand that participants used both search and filters during this task. From figure 12 we can see that more experienced users used the filter option more often than novices. As was mentioned before, this question was hardest of them all and filter was more successful for experienced users.

14. What did you use to try to find data?

Number of respondents: 64 , selected answers: 87



	Experienced		Novices		
	n	Percent	n	Percent	Total
Search	7	70.0%	5	62.5%	12
Filter	9	90.0%	4	50.0%	13
Something else	0	0.0%	0	0.0%	0
Total	**16**		**9**		**25**

Answers given into textfield

Option names	Text

Figure 12 What participants who found correct dataset used to find the dataset in task 2

Also, in the survey, there was a question to understand why participants chose one tool or another and because the list is not so long, I will include it here as Figure 13. Here I would

also highlight something of answers from experienced users and novices separately. Many answers of experienced users contain something about filtering and also that participants tried to search first and then decided to use filters when search didn't work. About novices, some of them managed to find datasets just using search and not touching filters, but some of them used only filters.

15. Why did you make this choice? And which option was successful search/filter or something else? 2

Number of respondents: 61

Experienced	Novices
filter building, search helsinki region	Frontpage-Helsinki-Housing-Constructed Environment
Filtering was more useful.	I was not sure how the data set will be named, so I had to use a combination of search and filters
Found by selecting 'Housing' from the front page, then scrolling the results	Search by key words. Apartment/building/start/housing
Because I was not able to find the data by using different search words, I checked what categories similar datasets had and used those. I found the data by using "housing" category and filtering for Helsinki region.	I set category to housing to decrease amount of options. Applied 'Helsinki' search to limit them even more. Then looked through 4 pages to find what I need.
The keyword search provided many options therefore I had to use filter this time. It took couple of tries to get the right page. I wasn't sure if I have reached to the right place therefore I checked the description of data below and saw that it is the right data.	Search
Had to try multiple different methods as the task was to find started housing in 2013, but the headline of the dataset does not tell that. That link was chosen as it seemed to be something that might have the asked dataset in it. Ultimately the search option was the one used to find the answer.	I don't know why, my choice was totally random, filter
Did not know how to search for it so used Helsinki Region & housing as filters.	Wasn't sure what terms to use for search, took some time to narrow down search results, used "helsinki 2013 housing".
I tried search first but I got no results with "apartments". Then I selected the "housing" category and after a quick scroll I found the right one.	Easy to filter
I typed completed housing in the Helsinki and searched from there	
I had to use both options, the key word housing in the search or filter was important to find the relevant dataset	

Figure 13 Why participant used this specific searching tool

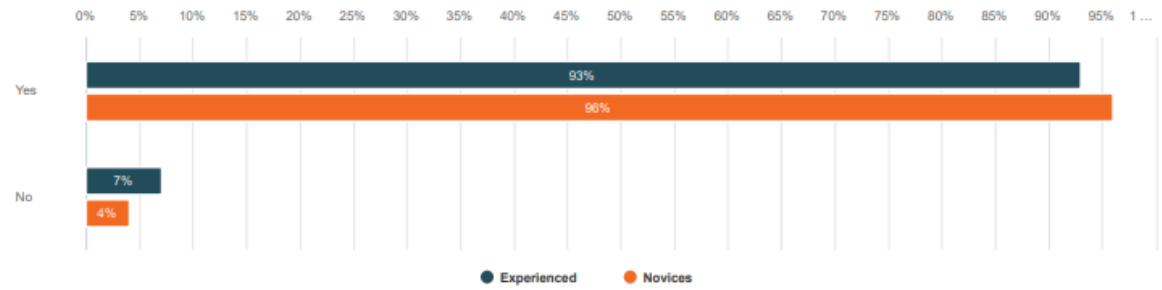
4.2.1.3 Third task

The third task was "You are interested in what is the temperature of the swimming water in Helsinki." Here we will see all the data about answers of experienced users and novices.

In Figure 14 it is possible to see how many experienced users and novices found the dataset. Surprisingly in this task all people who were told in Figure 14 that they found the dataset really found it. This task was the easiest of all three and numbers also confirm this. But still, some people didn't find the needed dataset.

17. Did you find the information you were looking for?

Number of respondents: 64



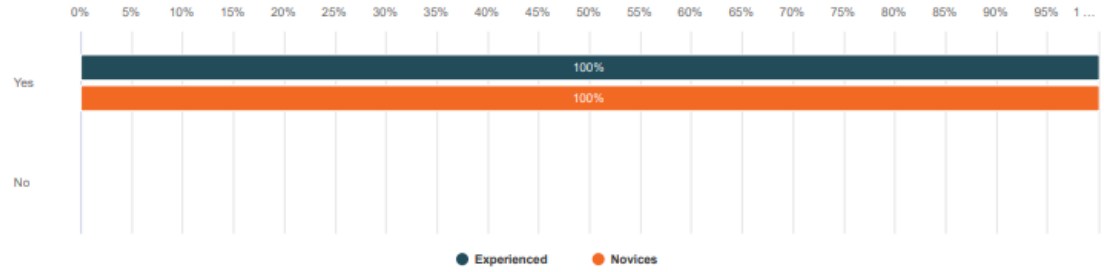
	Experienced		Novices		
	n	Percent	n	Percent	Total
Yes	38	92.7%	22	95.7%	60
No	3	7.3%	1	4.3%	4
Total	**41**		**23**		**64**

Figure 14 How many experienced users and novices found the dataset for task 3

21. Right dataset is the dataset on screenshot. Did you find correct dataset?

Link to dataset - https://hri.fi/data/en_GB/dataset/veden-reaaliaikainen-lampotila-helsingin-uimarannoilla

Number of respondents: 60



	Experienced		Novices		
	n	Percent	n	Percent	Total
Yes	38	100.0%	22	100.0%	60
No	0	0.0%	0	0.0%	0
Total	**38**		**22**		**60**

Figure 15 How many experienced users and novices found correct dataset for task 3

Since this task was easy, I want to show here what was the problem for 4 participants in Figure 14 who didn't find the needed dataset. Their comments can be found in Figure 16. They had problems with keywords and also with an opening dataset.

19. What in your opinion was the problem?

Number of respondents: 4

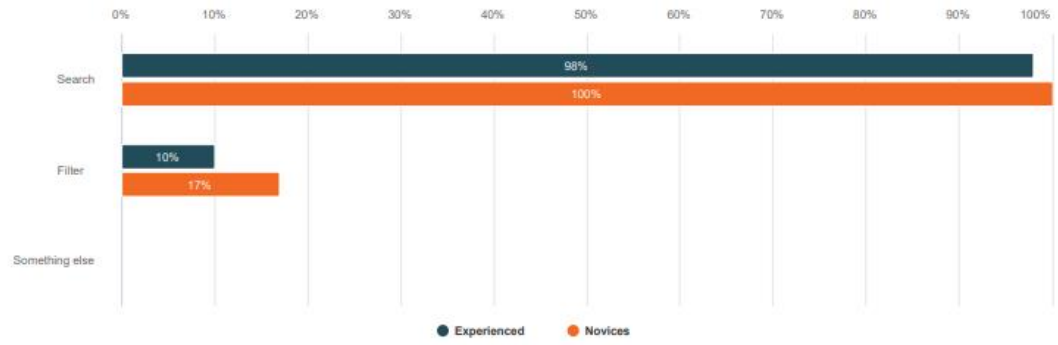
Experienced	Novices
Too confusing	I found the database Real-time swimming water temperature in Helsinki, but when I tried to open it, there was only following list of links "Index of /opendata/uiras/". I tried to open several links, but I couldn't find, what I was looking for.
Keyword search failed	
?	

Figure 16 What was the problem in task 3

Since everyone who searched for dataset number 3 found the dataset in Figure 14. I will show just one diagram of what the participants used to find the correct dataset. Previously, the answers of all survey participants and those who found the correct dataset were shown separately. And on Figure 17 it is possible to see the results. Most of the participants used search and some of them used filters. On Figure 18, we see a slight change in numbers, because it represents only answers of people who found dataset correctly.

22. What did you use to try to find data?

Number of respondents: 64 , selected answers: 71



	Experienced		Novices		
	n	Percent	n	Percent	Total
Search	40	97.6%	23	100.0%	63
Filter	4	9.8%	4	17.4%	8
Something else	0	0.0%	0	0.0%	0
Total	**44**		**27**		**71**

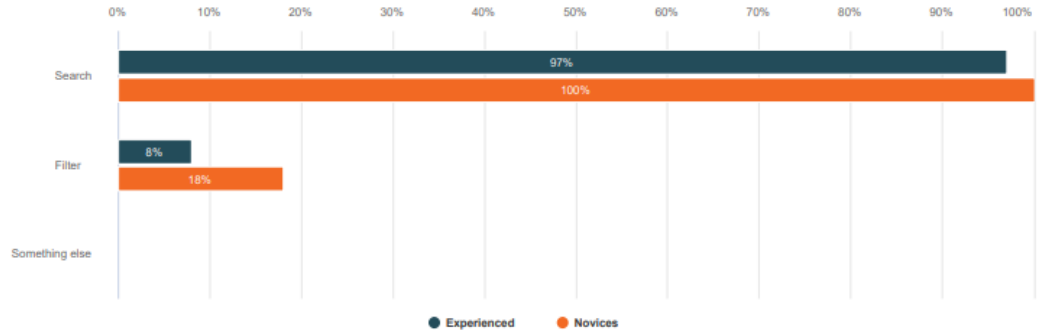
Answers given into textfield

Option names	Text

Figure 17 What participants used to find the dataset in task 3

22. What did you use to try to find data?

Number of respondents: 64 , selected answers: 71



	Experienced		Novices		
	n	Percent	n	Percent	Total
Search	37	97.4%	22	100.0%	59
Filter	3	7.9%	4	18.2%	7
Something else	0	0.0%	0	0.0%	0
Total	**40**		**26**		**66**

Answers given into textfield

Option names	Text

Figure 18 What participants who found correct dataset used to find the dataset in task 3

The interesting thing here is that more novices were using the filter this time. And if we look at the total number, we will see that some of the participants used both filter and search.

Also, in the survey there was a question to understand why participants chose one tool or another, I will include this in an appendix, but here I will tell some most common and interesting answers.

Answers from experienced users:

- I searched "temperature swimming" and found the data as the only result.
- A super specific task so I thought there cannot be many datasets that both have the words "swimming" and "temperature" in them.
- 'Swimming' is good (unique) key-word to search with. I felt it is quicker to start with search on the website and get a quick sense of whether it works.

Answers from novices:

- Because search is easy-to-find and easy-to-use
- Search proved to be a good method and this time it again brought a success. I searched for "temperature"
- Filter to get recreation datasets. Search to get the water

4.3 Interview results

To measure the results of the interviews, notes were taken during the sessions and questions were asked to the participants during the interviews. Here will be a summary of the findings. Overall, 8 people participated in the interview round, 5 of them were experts and 3 were novices from the survey. All of them completed both tasks, but for some, it was hard to find one of the datasets and participants went to google to search for it, didn't find it, and came back to open the data portal. Results of user interviews were the same, both experts and novices started to search for a dataset using the search field, if this didn't help they started using filters. One more experienced person started from filters, but still used search.

4.4 Searching strategies

After careful analysis of the results of the survey and interview, it was possible to see how users navigate and search datasets on an open data portal.

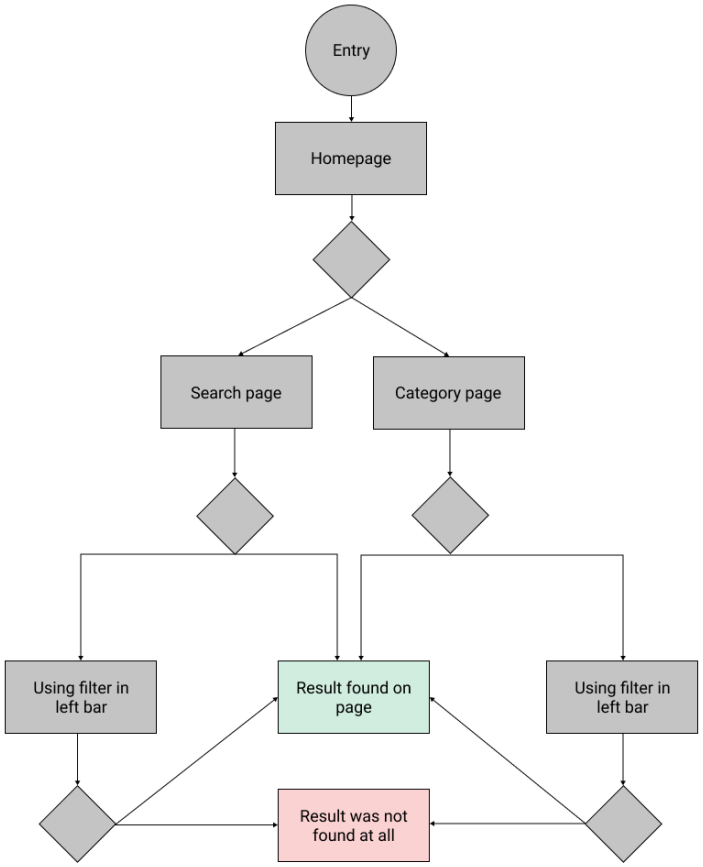


Figure 19 Searching strategy on open data portal

Usually, the strategy was to write a keyword to the search bar and if it didn't work, some people just continued looking for more pages and some started using the filter. Sometimes also participants started to use other keywords. Usually, if the filter didn't help, the person gave up. I think it is worth mentioning that in some cases participants started to try to search the same dataset in Google because they thought that maybe it will be easier to find it there.

4.4.1 Search engine

Based on the results from the survey and interview it is possible to present user paths for search engines.

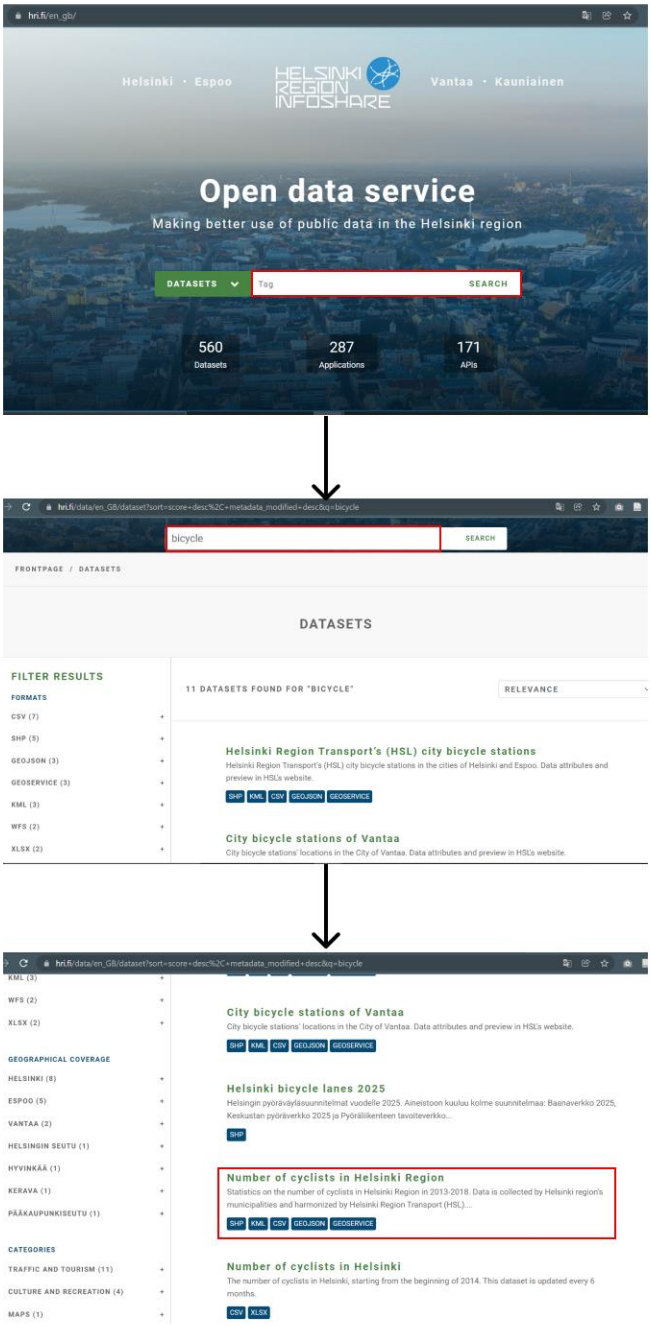


Figure 20 User path for search

In Figure 20 typical user steps of searching using a search engine are presented. The first thing that users see on an open data platform is the search bar, through the search bar it is possible to find all datasets with the keyword. For example, the keyword is a bicycle. On the search page, it is possible to find the dataset that was asked in the survey task, in this case, the dataset is number 4 in the search result. Also, it is noticeable that search engine doesn't help the user in case he made mistake in word. For example, it doesn't suggest the closest options to the keyword. It will only tell that there were no datasets were found for this word, this is shown in Figure 21.

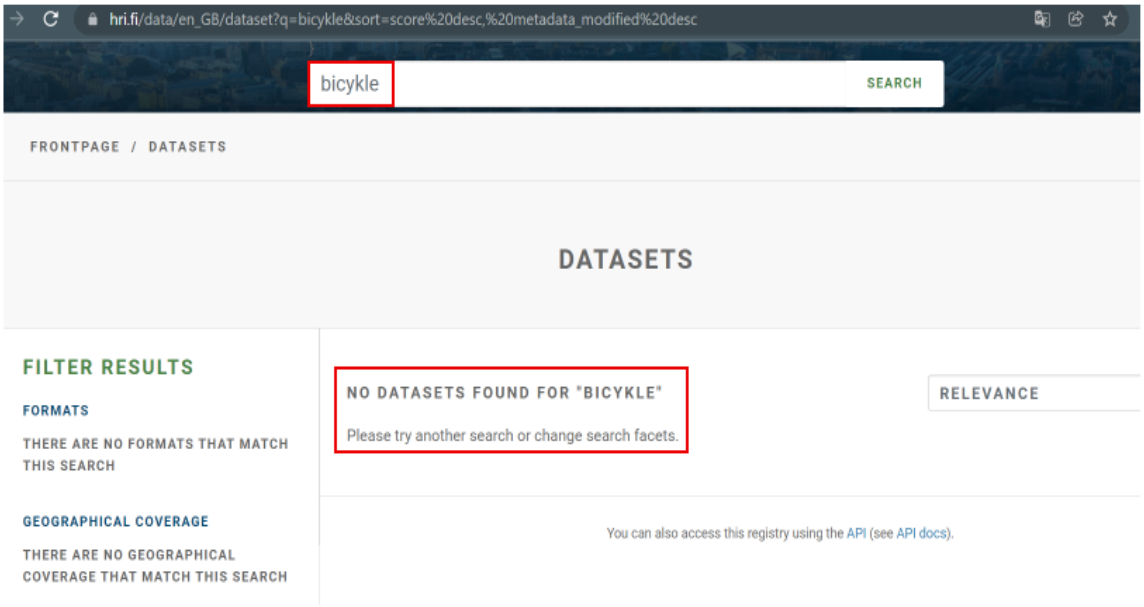


Figure 21 Search doesn't suggest correcting mistake

4.4.2 Filters

Same as with search engines it is possible to show how participants were using filters. In the case with filters usually, if the participants decided not to use search first they took a category from the front page, and after that either used search in a category or used filters in the left bar. This is shown on Figure 22.

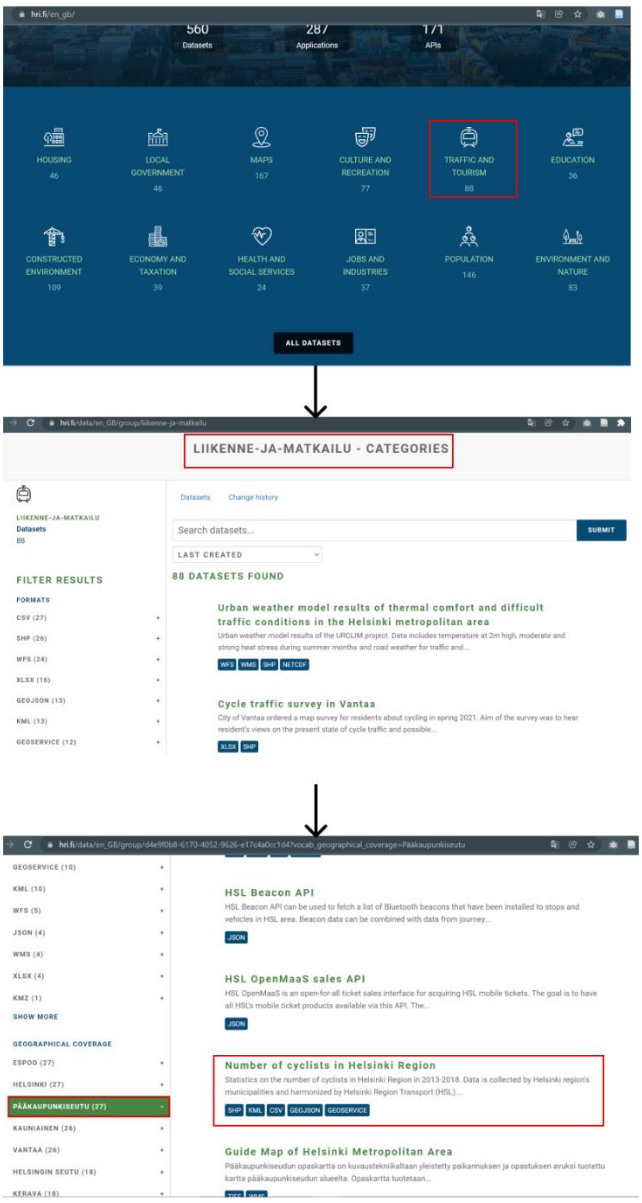


Figure 22 User path for searching using filters

5 DISCUSSION

This thesis was trying to find answers to questions: "Is the search on open data portals different for users who are tech-savvy and users who are less tech-savvy?". And "What ways it is possible to find to improve the experience of an inexperienced user with open data portals?". A survey and interviews were made to understand user behaviour of open data platforms.

From research results, it is possible to see that more experienced users and novices do the tasks in a similar way, but this applies only to easy tasks. When it is time to make more complicated tasks and the dataset is harder to find, experienced users start to use filters actively when novices either give up or continue to use search until they find the correct dataset. This is also confirmed by other studies. Experts tend to experiment with more complex features, while beginners tend to trust very simple features like keyword search, even if they don't work perfectly [25]. This is exactly what was possible to observe on an interview performed during this study, novices try to use search until they find the needed dataset or give up. The same is visible in the survey, but it is harder to notice there as a lot of reading and analyzing is needed.

There is no big difference between experts and novices in performing the task and probably it is because all people now have searching skills which also can be applied to open data portals. The only difference is in what people search.

If we want to improve open data platforms it is needed to concentrate on making search tools easier for users. This study showed that the most important tool for searching for datasets is a search field, but often it doesn't work, especially if the title of the dataset doesn't contain a keyword for searching. This is the most critical thing for novices because they don't start to use other tools so easily. The easiest way is to make search not only by title but for other text such as description or anything related to the dataset. Because this might help users to find even more complicated datasets without using filters, with will make searching for needed datasets easier for novices. For experienced users, it will just make searching faster, which is also a good improvement.

Also, one of the findings from the survey was to make a data preview, so the user can understand if this is the dataset, he is looking for more easily. Because some of the participants didn't understand if they found the needed dataset without downloading and opening it. Data visualization could tell inexperienced people much more at the initial stage of working with data than just a dataset. I will show one example, for this I made two figures Figure 23 which is just a screenshot from hri.fi open data portal and Figure 24 is edited, data visualisation parts is added there. Imagine you are wondering what languages are popular among students. And you look for information about it, find an open data portal and find a dataset with the information you are interested in. What you will see now, at least on HRI, is a page that has a short description about the dataset and a download link (Figure 24). If the open data portal had data visualization, then users would immediately see the result, without even downloading the data (Figure 24). Then they can analyze and process the data himself, but for beginners, the search may stop the moment they find the answer to their question on the open data portal page, and the more experienced will simply make sure that the dataset is what they need.

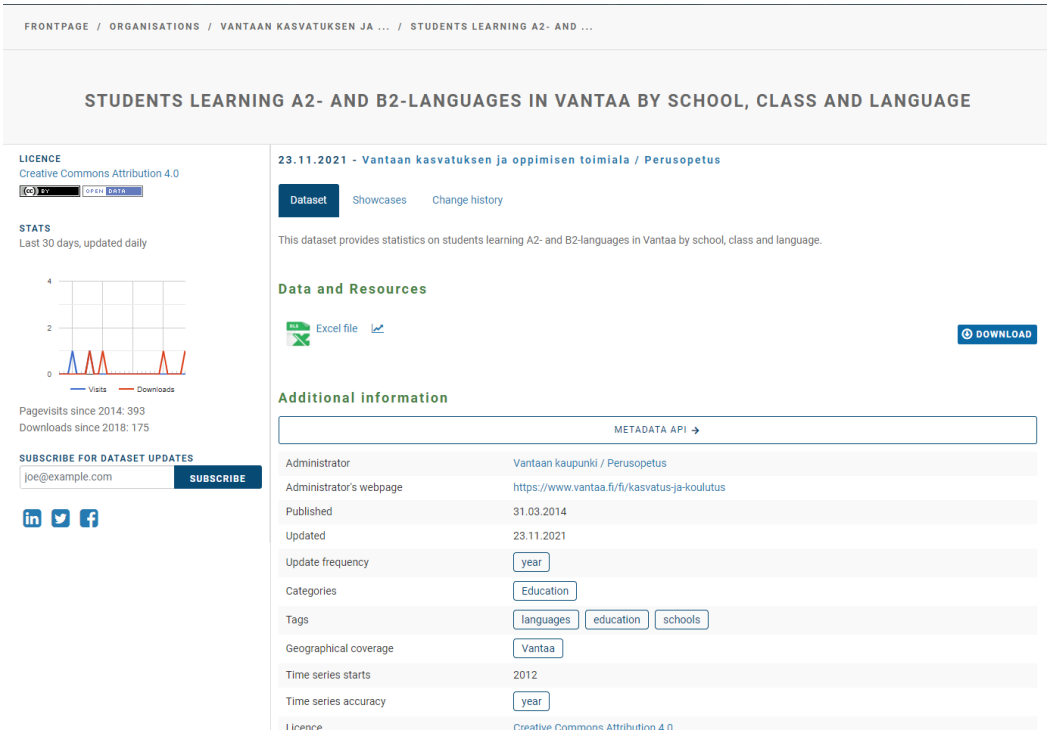


Figure 23 Screen capture from hri.fi open data portal

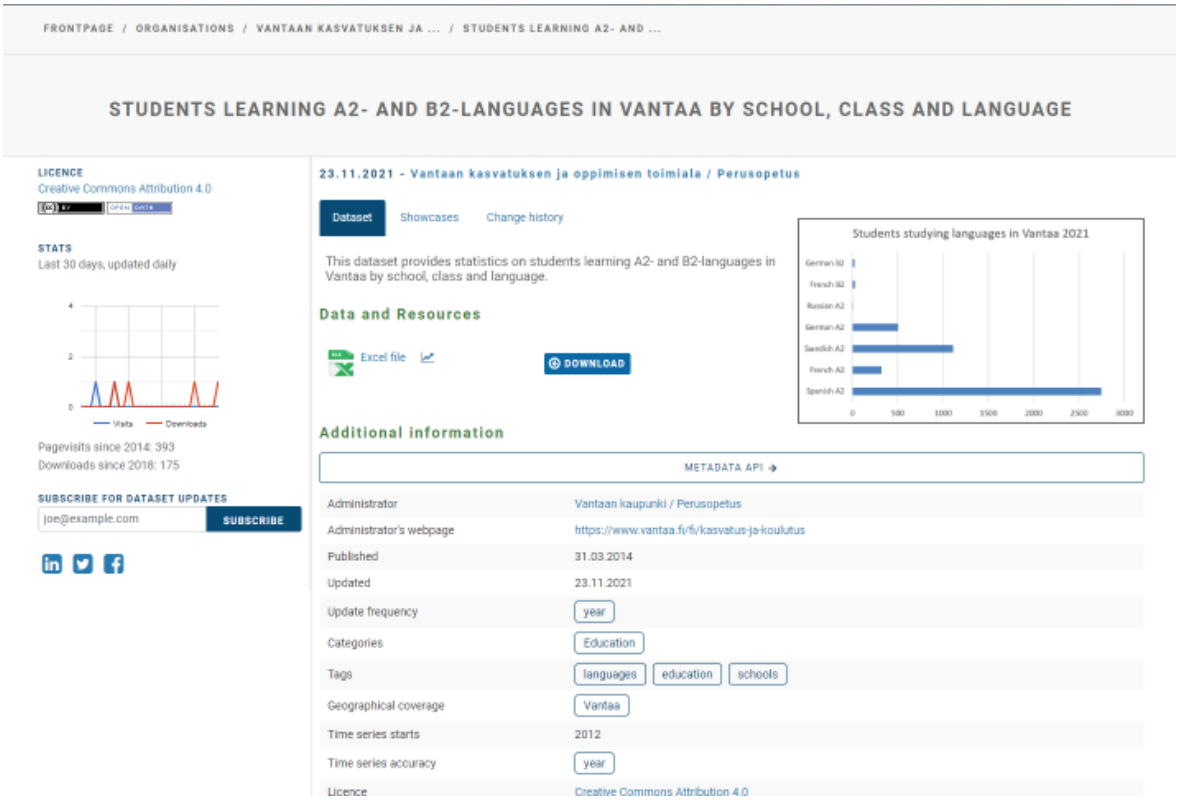


Figure 24 Edited screen capture from hri.fi open data portal. Data visualization added.

Also, one of the studies shows that information visualization helps to make the gap between experts and novices smaller [26]. So, this tool can also be used to improve search on open data platforms which will improve the usability of the open data platform. There can be more research done on how exactly to preview the data and in what form. For example, what is the best way to visualize a dataset with information about school students and what languages they study? Bubble Cloud? Histogram? Something else?

The results of this study relate very well to what the literature suggests. Many different research papers were found that had similar results, but in a different field, not in open data platforms. Overall experts and novices often differ only in how advanced they are in using a particular system, in our case open data platform and searching tools. Can we improve experience of novices and experienced users? Yes, we can, by making data visualizations and by improving search engine.

For the future research I would suggest comparing different open data platforms and understand what the best practices are in making good user experience for both experiences users and novices. Maybe some of big open data portals have some good features with can help users to find what they need and use open data.

6 CONCLUSIONS

In this thesis, a study was made about open data portals and in particular how the behaviour of experienced users and novices differs in finding the right data on the portal. The study was conducted on the Helsinki region infoshare portal. Two studies were carried out, one qualitative and one quantitative. Both showed similar results.

Experienced users and novices in the use of open data portals differ in that more experienced users quickly and easily start using others search tools such as filters, novices most often use search and very rarely start using filters or other tools. This leads to the idea that to make open data portals even more, user-friendly, especially to novices, it is necessary to improve search, because this is the main tool for using an open data portal. There is a need for it to be improved because this study showed that it is now always so obvious how to find needed information in an open data portal. Also, search now doesn't help user in case there is mistake in key word, which often leads to failed search.

Data visualization might help users to understand, if they found needed dataset before downloading and opening it, especially for novices this can save time and make the process easier. One of the problem of open data is that it is not equally available for all people, to understand the data user need to have skills, data visualization can improve equality of novices and more experienced users in using open data portal, because with data visualization tool right on the open data portal, novices do not need skills to work with the data and can find answers to their questions right on the portal.

REFERENCES

[1] E. Ruijer, S. Grimmelikhuijsen, J. van den Berg, and A. Meijer, 'Open data work: understanding open data usage from a practice lens', *International Review of Administrative Sciences*, vol. 86, no. 1, pp. 3–19, 2020, doi: 10.1177/0020852317753068.

[2] R. Ball, G. A. Fink, and C. North, 'Home-centric visualization of network traffic for security administration', in *Proceedings of the 2004 ACM workshop on Visualization and data mining for computer security - VizSEC/DMSEC '04*, Washington DC, USA, 2004, p. 55. doi: 10.1145/1029208.1029217.

[3] M. Janssen, Y. Charalabidis, and A. Zuiderwijk, 'Benefits, Adoption Barriers and Myths of Open Data and Open Government', *Information Systems Management*, vol. 29, no. 4, pp. 258–268, Sep. 2012, doi: 10.1080/10580530.2012.716740.

[4] 'DIRECTIVE (EU) 2019/ 1024 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL - of 20 June 2019 - on open data and the re-use of public sector information', p. 28.

[5] 'Helsinki Region Infoshare - Open data service'. https://hri.fi/en_gb/ (accessed Mar. 08, 2022).

[6] 'What is open data | data.europa.eu'. https://data.europa.eu/en/trening/what-open-data (accessed Feb. 21, 2022).

[7] *Farm data management, sharing and services for agriculture development*. FAO, 2021. doi: 10.4060/cb2840en.

[8] N. Chaudhury, J. Hammer, M. Kremer, K. Muralidharan, and F. H. Rogers, 'Missing in Action: Teacher and Health Worker Absence in Developing Countries', *Journal of Economic Perspectives*, vol. 20, no. 1, pp. 91–116, Mar. 2006, doi: 10.1257/089533006776526058.

[9] 'The Benefits of Open Data – Evidence from Economic Research', *Open Knowledge Foundation blog*, Oct. 05, 2012. https://blog.okfn.org/2012/10/05/the-benefits-of-open-data-evidence-from-economic-research/ (accessed Mar. 08, 2022).

[10] ‘Watchdog finds more than $1T in misreported federal spending’, *FCW*. https://fcw.com/2013/02/watchdog-finds-more-than-1t-in-misreported-federal-spending/249041/ (accessed Mar. 08, 2022).

[11] F. Welle Donker, B. Van Loenen, and A. K. Bregt, ‘Open Data and Beyond’, *ISPRS International Journal of Geo-Information*, vol. 5, no. 4, Art. no. 4, Apr. 2016, doi: 10.3390/ijgi5040048.

[12] E. Ferro and M. Osella, ‘Eight Business Model Archetypes for PSI Re-Use’, p. 13.

[13] R. Pollock, ‘The Economics of Public Sector Information’, Faculty of Economics, Working Paper, May 2009. doi: 10.17863/CAM.5635.

[14] ‘POPSIS, pricing of public sector information study - Publications Office of the EU’. https://op.europa.eu/en/publication-detail/-/publication/d0862188-b657-40b4-b3c0-f3c05d454247/language-en (accessed Apr. 18, 2022).

[15] M. Gurstein, ‘Open data: Empowering the empowered or effective data use for everyone?’, *First Monday*, vol. 16, 2011, doi: 10.5210/fm.v16i2.3316.

[16] P. Lehtonen, ‘Open data in Finland - Public sector perspectives on open data.’, 2011, Accessed: Feb. 21, 2022. [Online]. Available: https://researchportal.tuni.fi/en/publications/open-data-in-finland-public-sector-perspectives-on-open-data

[17] ‘Open data portals | Shaping Europe’s digital future’. https://digital-strategy.ec.europa.eu/en/policies/open-data-portals (accessed Feb. 21, 2022).

[18] L. M. Koesten, E. Kacprzak, J. F. A. Tennison, and E. Simperl, ‘The Trials and Tribulations of Working with Structured Data: -a Study on Information Seeking Behaviour’, in *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, Denver Colorado USA, May 2017, pp. 1277–1289. doi: 10.1145/3025453.3025838.

[19] F. Xiao, D. He, Y. Chi, W. Jeng, and C. Tomer, ‘Challenges and Supports for Accessing Open Government Datasets: Data Guide for Better Open Data Access and Uses’,

in *Proceedings of the 2019 Conference on Human Information Interaction and Retrieval*, Glasgow Scotland UK, Mar. 2019, pp. 313–317. doi: 10.1145/3295750.3298958.

[20] F. Benitez-Paez, A. Degbelo, S. Trilles, and J. Huerta, 'Roadblocks Hindering the Reuse of Open Geodata in Colombia and Spain: A Data User's Perspective', *IJGI*, vol. 7, no. 1, p. 6, Dec. 2017, doi: 10.3390/ijgi7010006.

[21] K. Kapoor, V. Weerakkody, and U. Sivarajah, 'Open Data Platforms and Their Usability: Proposing a Framework for Evaluating Citizen Intentions', in *Open and Big Data Management and Innovation*, Cham, 2015, pp. 261–271. doi: 10.1007/978-3-319-25013-7_21.

[22] 12 Jun 2015 Dr Julia Glidden Feed 366up 1 comment, 'The Citadel reveals open data findings', *Opensource.com*. https://opensource.com/government/15/6/citadel-open-government-data-results (accessed Feb. 21, 2022).

[23] A. Zuiderwijk, M. Janssen, S. Choenni, R. Meijer, and R. Sheikh_Alibaks, 'Socio-Technical Impediments of Open Data', *Electronic Journal of eGovernment*, vol. 10, pp. 156–172, 2012.

[24] E. Ruijer and A. Meijer, 'Open Government Data as an Innovation Process: Lessons from a Living Lab Experiment', *Public Performance & Management Review*, vol. 43, pp. 1–23, 2019, doi: 10.1080/15309576.2019.1568884.

[25] K. Schoeffmann, B. Münzer, M. J. Primus, S. Kletz, and A. Leibetseder, 'How Experts Search Different than Novices – An Evaluation of the Divexplore Video Retrieval System at Video Browser Showdown 2018', in *2018 IEEE International Conference on Multimedia Expo Workshops (ICMEW)*, Jul. 2018, pp. 1–6. doi: 10.1109/ICMEW.2018.8551552.

[26] I.-C. Wu and P. Vakkari, 'Effects of subject-oriented visualization tools on search by novices and intermediates', *Journal of the Association for Information Science and Technology*, vol. 69, no. 12, pp. 1428–1445, 2018, doi: 10.1002/asi.24070.

APPENDIX 1. Results from the survey

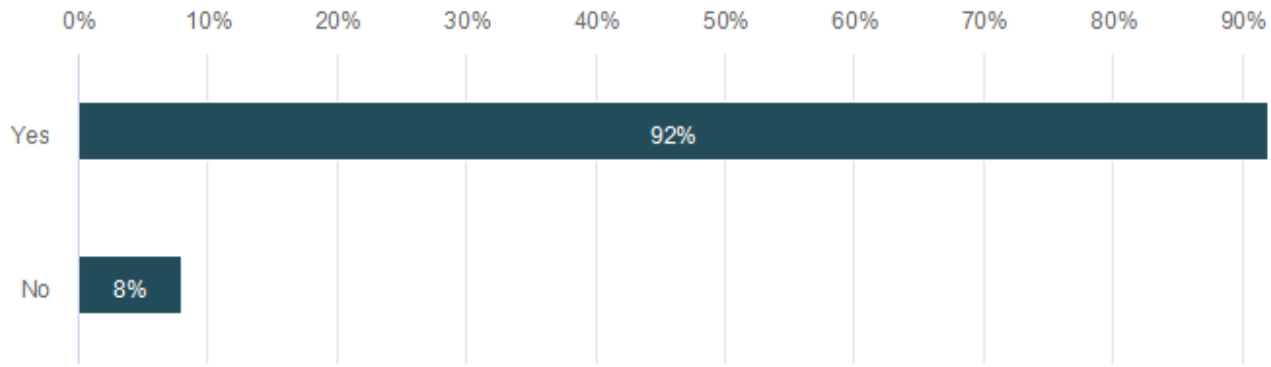
The appendix is a report from Webropol survey.

Open data portals

Total number of respondents: 64

1. Did you find the information you were looking for?

Number of respondents: 64



	n	Percent
Yes	59	92.2%
No	5	7.8%

2. Can you send here a link to the dataset?

Number of respondents: 59

APPENDIX 1. (continues)

Responses
https://hri.fi/data/en_GB/dataset/helsingin-seudun-pyoralaskennat
https://hri.fi/data/en_GB/dataset/helsingin-seudun-pyoralaskennat
https://hri.fi/data/en_GB/dataset/helsingin-seudun-pyoralaskennat
https://hri.fi/data/en_GB/dataset/helsingin-pyorailijamaarat
https://www.hel.fi/hel2/tietokeskus/data/helsinki/ksv/Helsingin_pyorailijamaarat.csv
https://hri.fi/data/en_GB/dataset/helsingin-seudun-pyoralaskennat
https://hri.fi/data/en_GB/dataset/helsingin-pyorailijamaarat/resource/ffc0ba24-08f9-452e-80db-99850ebc7962
https://hri.fi/data/en_GB/dataset/helsingin-pyorailijamaarat/resource/ffc0ba24-08f9-452e-80db-99850ebc7962
https://hri.fi/data/en_GB/dataset/helsingin-seudun-pyoralaskennat
https://hri.fi/data/en_GB/dataset/helsingin-seudun-pyoralaskennat
https://hri.fi/data/en_GB/dataset/helsingin-seudun-pyoralaskennat
https://hri.fi/data/en_GB/dataset/helsingin-seudun-pyoralaskennat
https://hri.fi/data/en_GB/dataset/helsingin-seudun-pyoralaskennat
https://hri.fi/data/en_GB/dataset/helsingin-pyorailijamaarat
https://hri.fi/data/en_GB/dataset/helsingin-pyorailijamaarat
https://hri.fi/data/en_GB/dataset/helsingin-pyorailijamaarat

continues

APPENDIX 1. (continues)

https://hri.fi/data/en_GB/dataset/helsingin-seudun-pyoralaskennat
https://hri.fi/data/en_GB/dataset/helsingin-pyorailijamaarat
https://hri.fi/data/en_GB/dataset/helsingin-pyorailijamaarat
https://hri.fi/data/en_GB/dataset/helsingin-pyorailijamaarat
https://hri.fi/data/en_GB/dataset/helsingin-pyorailijamaarat/resource/ffc0ba24-08f9-452e-80db-99850ebc7962
https://hri.fi/data/en_GB/dataset/helsingin-pyorailijamaarat
https://hri.fi/data/en_GB/dataset/helsingin-pyorailijamaarat/resource/ffc0ba24-08f9-452e-80db-99850ebc7962
https://hri.fi/data/en_GB/dataset/helsingin-seudun-pyoralaskennat
https://hri.fi/data/en_GB/dataset/helsingin-seudun-pyoralaskennat
https://hri.fi/en_gb/
https://hri.fi/data/en_GB/dataset/helsingin-seudun-pyoralaskennat
https://hri.fi/data/en_GB/dataset/helsingin-seudun-pyoralaskennat
https://hri.fi/data/en_GB/dataset/helsingin-seudun-pyoralaskennat
https://hri.fi/data/en_GB/dataset/helsingin-seudun-pyoralaskennat
https://hri.fi/data/en_GB/dataset/helsingin-pyorailijamaarat
https://hri.fi/data/en_GB/dataset/helsingin-pyorailijamaarat
https://hri.fi/data/en_GB/dataset/helsingin-seudun-pyoralaskennat

continues

APPENDIX 1. (continues)

https://hri.fi/data/en_GB/dataset/helsingin-seudun-pyoralaskennat
https://hri.fi/data/en_GB/dataset/espoon-kevyen-liikenteen-opasteviitat
https://hri.fi/data/en_GB/dataset/helsingin-seudun-pyoralaskennat/resource/0e31f4c3-c819-4d14-b481-a7c9661d4e60
https://hri.fi/data/en_GB/dataset/helsingin-seudun-pyoralaskennat
https://hri.fi/data/en_GB/dataset/helsingin-pyorailijamaarat
https://hri.fi/data/en_GB/dataset/helsingin-seudun-pyoralaskennat
https://hri.fi/data/en/dataset/helsingin-seudun-pyoralaskennat
https://hri.fi/data/en_GB/dataset/helsingin-seudun-pyoralaskennat
https://hri.fi/data/en_GB/dataset/helsingin-seudun-pyoralaskennat
https://hri.fi/data/en_GB/dataset/helsingin-seudun-pyoralaskennat
https://hri.fi/data/en_GB/dataset/helsingin-seudun-pyoralaskennat
https://hri.fi/data/en_GB/dataset/helsingin-pyorailijamaarat
https://hri.fi/data/en_GB/dataset/helsingin-pyorailijamaarat
https://hri.fi/data/en_GB/showcase/pyorailijamaarat-helsingissa
https://hri.fi/data/en_GB/dataset/helsingin-seudun-pyoralaskennat
https://hri.fi/data/en_GB/dataset/helsingin-seudun-pyoralaskennat
https://hri.fi/data/en_GB/dataset/helsingin-seudun-pyoralaskennat

continues

APPENDIX 1. (continues)

https://hri.fi/data/en_GB/dataset/helsingin-pyorailijamaarat/resource/ffc0ba24-08f9-452e-80db-99850ebc7962
https://hri.fi/data/en_GB/dataset/helsingin-pyorailijamaarat
https://hri.fi/data/en_GB/dataset/helsingin-seudun-pyoralaskennat
https://hri.fi/data/en_GB/dataset/helsingin-seudun-pyoralaskennat
https://hri.fi/data/en_GB/dataset/helsingin-pyorailijamaarat
https://hri.fi/data/en_GB/dataset/helsingin-seudun-pyoralaskennat
https://hri.fi/data/en_GB/dataset/helsingin-seudun-pyoralaskennat
https://hri.fi/data/en/dataset/helsingin-seudun-pyoralaskennat
https://public-transport-hslhrt.opendata.arcgis.com/datasets/helsingin-seudun-py%C3%B6r%C3%A4laskennat/data

3. What in your opinion was the problem?

Number of respondents: 5

Responses
The numbers I found were rather meaningless without more context provided.
I don't know what is helsinki region
the structure of the location of information on the site is complex
It is because when on the search bar the information provided for number of cyclists in Helsinki was not specified and i could not find it at all.

continues

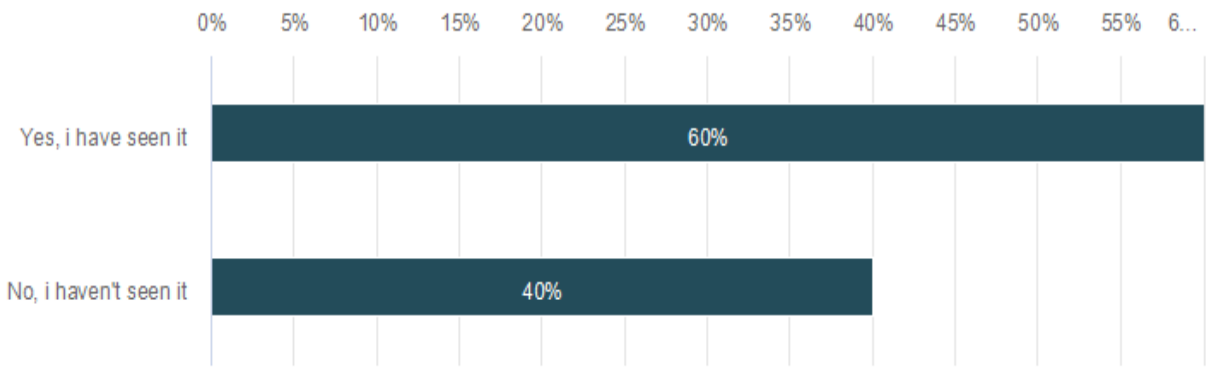
APPENDIX 1. (continues)

I was looking for a aggregated number, which might be just me misunderstanding the task. I did find results for spot counting.

4. Right dataset is the dataset on screenshot. Have you seen it while searching?

Link to dataset - https://hri.fi/data/en_GB/dataset/helsingin-seudun-pyoralaskennat

Number of respondents: 5



	n	Percent
Yes, i have seen it	3	60.0%
No, i haven't seen it	2	40.0%

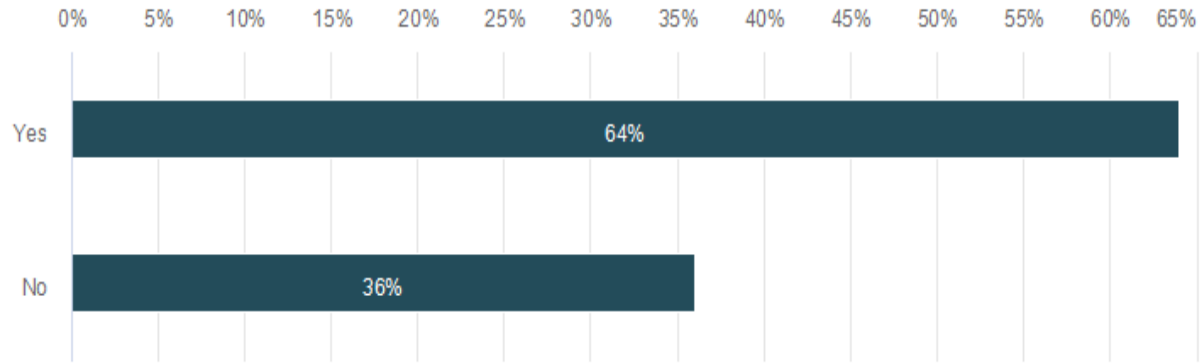
5. The right dataset is the dataset on screenshot. Did you find correct dataset? Check carefully that the title is the same.

continues

APPENDIX 1. (continues)

Link to dataset - https://hri.fi/data/en_GB/dataset/helsingin-seudun-pyoralaskennat

Number of respondents: 59



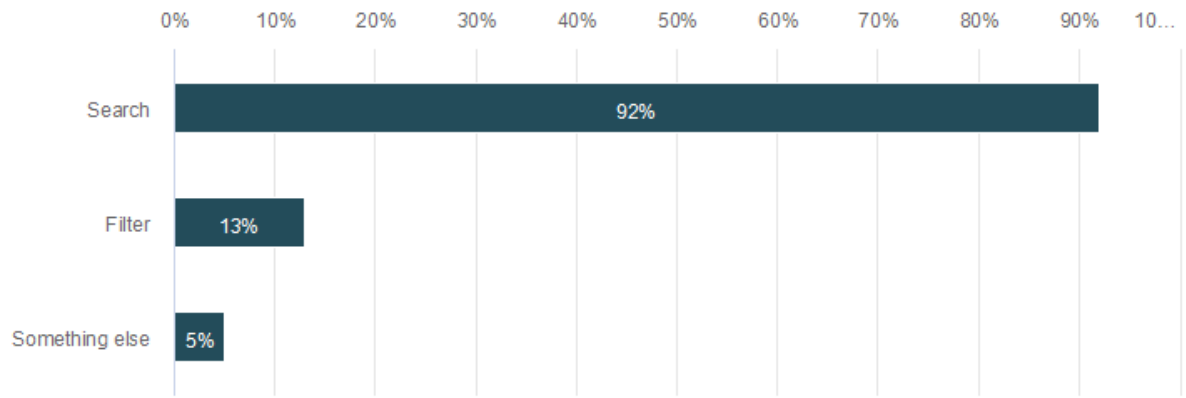
	n	Percent
Yes	38	64.4%
No	21	35.6%

6. What did you use to try to find the data?

Number of respondents: 64, selected answers: 70

continues

APPENDIX 1. (continues)



	n	Percent
Search	59	92.2%
Filter	8	12.5%
Something else	3	4.7%

Answers given into textfield

Option names	Text
Something else	tag
Something else	Filter 'traffic and tourism', then search 'cyclist'
Something else	Searching in dataset pages

7. Why did you make this choice? And which option was successful search/filter or something else? 1

continues

APPENDIX 1. (continues)

Number of respondents: 63

Responses
search transport data - i found tag "cycling" - search by using tag
It was the first thing, which I saw
Because it is easier to use search
It was direct based on my search of number of cyclists ps. Note that the screenshot you gave seems old to me.
Because I am lazy. Keyword cyclist
I misunderstood the question and thought the right answer it just cyclist number in Helsinki, not Helsinki region.
There was only one option
In my opinion it was the fastest way to find the data
I always use search in case of finding information. Put a key word from a question “cyclist”
I chose "Helsinki" as the area first, in the upper part of the page, then typed in "cyclist" in the search box. I thought this would be the quickest way.
It was the first one I saw. I just searched for "cyclist".
I selected 'Traffic and tourism' on the front page, then fortunately the query was in the first result page. Otherwise I would have used the search tool.
I searched "cyclists" and the correct dataset was one of the top results.

continues

APPENDIX 1. (continues)

Combination of both of them, filter first, than search 'cyclists'
just tried the easiest option
The easiest way to find anything - to use the search bar P.S. The question was about Helsinki or Helsinki region?
I usually search with keywords. I did the same and found the link directly instead of browsing for it.
I chose the search option because it was the first thing that came up when I opened the web page. I think the search was successful but I decided to pick the dataset which is updated.
simplest
It seems an obvious decision to me, the search bar is at the center of the front page, it has already been filtered by 'datasets''. Also, usually i don't spend time with filters and always try search at first.
It was easier
Because it is the easiest way to find data needed. Search is easier to use
Search bar seemed the most obvious
It was the first option
Searched cyclist
easy
This page is not used

continues

APPENDIX 1. (continues)

search phrase: cyclists it is regional data
visually understandable and intuitive, "cyclists" was they key word in the search bar
It's easier. I just typed "Cyclists".
Because I think it's better to try to find something using keywords. I have search for the word 'ciclysts'
Cyclists Helsinki
Search, is easier to search by keyworkds
Search was successful.
I used search, because it was in the middle of the screen I searched with "helsinki cyclist" and right dataset was first in results.
I thought, that searching in transport dataset would be more simple.
Gives straight answer
The chapter was correct. But the task was to find the number of cyclists, so I need file that I can open.
Keyword searches are my go-to thing
I put the word "cyclists" in search and in that way find the right database.
There were more than 600 datasets, so it made sense to search for the right dataset.

continues

APPENDIX 1. (continues)

both search and filter were helpful
“Cycling”
I used search option here, it is easy to find with keywords.
cyclists in the Helsinki region
Search was successful because this dataset has clear key words. I just typed “cyclists in Helsinki”
there was a clear search term
Using the search bar was quick and easy choice. I actually found the dataset with the same search when I knew what I was looking for. I used the word "cycling" to search.
By searching and filtering limits the number of information's to be provided by doing so your search results are pin pointed.
Easiest and fastest way to use search
Fastest way to complete the task - if successful
Key words are always efficient
First decided to use a filter for the right category as it was easily visible. Then used a search word in this category to find the right dataset.
Firstly I tried to check traffic, but failed to find. And then I tried search function. Search was successful.
Fastest option.
the easiest way to get the related hits

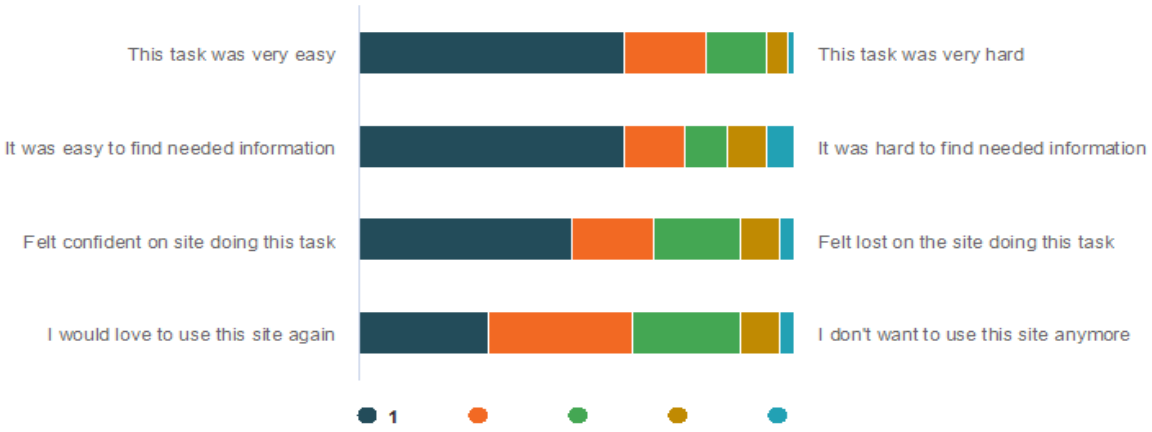
continues

APPENDIX 1. (continues)

Search is faster than browsing categories - I just typed "cyclists" and the third result was the right one
Search for the key word is the best way to find information in my experience. Search for the is a good way, but you must pay attention to the titles, so you won't mix up Number of cyclists in Helsinki and in Helsinki region.
just put the search key-word 'cyclists' in the search bar
I copied your task "Cyclist in Helsinki region" in the search box and found it
I went to data then search data and typed what I am looking for as key words.
I think that use Keywords in data search is faster than the other options.
Search was the most successful

8. Choose a statement, that best describes your experience

Number of respondents: 64



continues

APPENDIX 1. (continues)

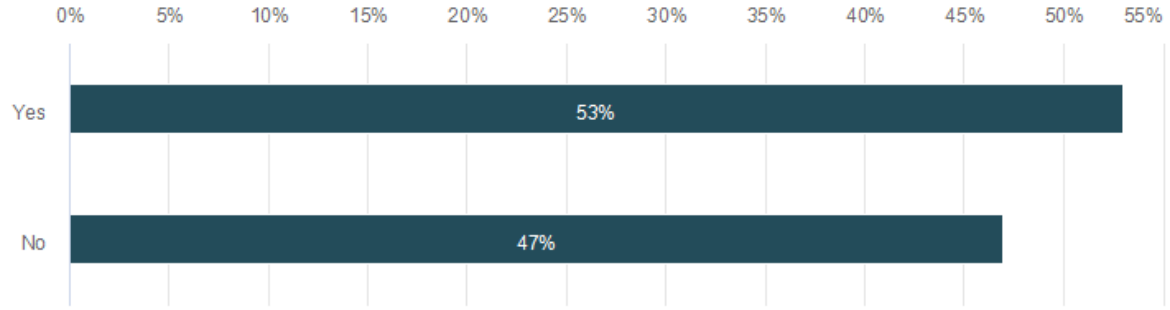
	1						Total	Average	Median
This task was very easy	39	12	9	3	1	This task was very hard	64	1.7	1.0
	60.9 %	18.7 %	14.1 %	4.7 %	1.6 %				
It was easy to find needed information	39	9	6	6	4	It was hard to find needed information	64	1.9	1.0
	60.9 %	14.1 %	9.4%	9.4 %	6.2 %				
Felt confident on site doing this task	31	12	13	6	2	Felt lost on the site doing this task	64	2.0	2.0
	48.4 %	18.8 %	20.3 %	9.4 %	3.1 %				
I would love to use this site again	19	21	16	6	2	I don't want to use this site anymore	64	2.2	2.0
	29.7 %	32.8 %	25.0 %	9.4 %	3.1 %				
Total	**128**	**54**	**44**	**21**	**9**		**256**	**1.9**	**1.5**

9. Did you find the information you were looking for?

Number of respondents: 64

continues

APPENDIX 1. (continues)



	n	Percent
Yes	34	53.1%
No	30	46.9%

10. Can you send here a link to the dataset?

Number of respondents: 34

Responses
https://hri.fi/data/en_GB/dataset/helsingin-seudun-aloitetut-ja-valmistuneet-asunnot
https://hri.fi/data/en_GB/dataset/aluesarjat-hginseutu-ar-rr01-rakennukset-kayttark-rvuosi
https://hri.fi/data/en_GB/dataset/helsinki-asuminen
https://hri.fi/data/en_GB/dataset/helsingin-seudun-aloitetut-ja-valmistuneet-asunnot

continues

APPENDIX 1. (continues)

https://hri.fi/data/en_GB/dataset/helsingin-seudun-aloitetut-ja-valmistuneet-asunnot
https://hri.fi/data/en_GB/dataset/helsingin-seudun-aloitetut-ja-valmistuneet-asunnot
https://hri.fi/data/en_GB/dataset/helsingin-seudun-aloitetut-ja-valmistuneet-asunnot
https://hri.fi/data/en_GB/dataset/aluesarjat-hginseutu-ar-aa01-asuntokanta-rvuosi-talotyyppi
https://hri.fi/data/en_GB/dataset/helsingin-seudun-aloitetut-ja-valmistuneet-asunnot
https://hri.fi/data/en_GB/dataset/helsingin-seudun-aloitetut-ja-valmistuneet-asunnot
https://hri.fi/data/en_GB/dataset/helsingin-seudun-aloitetut-ja-valmistuneet-asunnot
https://hri.fi/data/en_GB/dataset/helsingin-seudun-aloitetut-ja-valmistuneet-asunnot
https://hri.fi/data/fi/dataset/aluesarjat_a01hki_astuot_hper_rahoitus_talotyyppi
https://hri.fi/data/en_GB/dataset/aluesarjat-hginseutu-ar-rr01-rakennukset-kayttark-rvuosi
https://hri.fi/data/en_GB/dataset/helsingin-tilastollinen-vuosikirja/resource/4e178f1a-ceef-4cd0-aaa1-35060e529a4e
https://hri.fi/data/en_GB/dataset/helsingin-seudun-aloitetut-ja-valmistuneet-asunnot

continues

APPENDIX 1. (continues)

https://hri.fi/data/en_GB/dataset/rakennustietoruudukko/resource/41d07943-cd26-43f0-b091-55df34e900b6
https://hri.fi/data/en_GB/dataset/aluesarjat-hginseutu-ar-rr01-rakennukset-kayttark-rvuosi
https://hri.fi/data/en_GB/dataset/helsingin-seudun-pyoralaskennat
https://hri.fi/data/en_GB/dataset/aluesarjat_a01hki_astuot_hper_rahoitus_talotyyppi
https://hri.fi/data/en_GB/dataset/helsingin-seudun-aloitetut-ja-valmistuneet-asunnot
https://hri.fi/data/en_GB/dataset/vantaan-rakennuskanta-ja-asuntojen-huoneistotyypit-rakennuksen-kayttotarkoituksen-mukaan
https://hri.fi/data/en_GB/dataset/helsingin-kaavayksikot
https://hri.fi/data/en_GB/dataset?sort=title+asc&q=Construction+
https://hri.fi/data/en_GB/dataset/asumisen-kerrosala-uusissa-asemakaavoissa-helsingin-seudulla-2012-2015
https://hri.fi/data/en_GB/dataset/rakennustietoruudukko/resource/41d07943-cd26-43f0-b091-55df34e900b6
https://hri.fi/data/en_GB/dataset/rakennustietoruudukko
https://hri.fi/data/en_GB/dataset/aluesarjat-a01s-hki-rakennuskanta
https://hri.fi/data/en_GB/dataset/helsingin-seudun-aloitetut-ja-valmistuneet-asunnot

continues

APPENDIX 1. (continues)

https://hri.fi/data/en_GB/dataset/helsingin-seudun-aloitetut-ja-valmistuneet-asunnot
https://hri.fi/data/en_GB/dataset/helsingin-seudun-aloitetut-ja-valmistuneet-asunnot
https://hri.fi/data/en_GB/dataset/aluesarjat-hginseutu-ar-rr01-rakennukset-kayttark-rvuosi
https://hri.fi/data/en_GB/dataset/helsingin-seudun-aloitetut-ja-valmistuneet-asunnot
https://hri.fi/data/en_GB/dataset/helsingin-seudun-aloitetut-ja-valmistuneet-asunnot

11. What in your opinion was the problem?

Number of respondents: 30

Responses
Fields of tables are not available for reading. It is not possible to identify the information contained in the tables
The task was too hard understand and the search was not return anything by the given keywords
The data is only available as a shape file.
The question was a bit unclear
I tried a lot of synonyms for key words, I suppose it was not enough)

continues

APPENDIX 1. (continues)

... hmm is this screenshot the correct result for that searching query? I might have seen it, but was sure it is not the answer..(
_(o.O)_ _|_/
idk neither of my search request worked properly. navigating via categories couldn't help either.
probably i didn't get the task, tried to search by such requests as "apartments Helsinki", "building Helsinki", "building 2013", and "construction" in different variants
No results if searched via keywords, Filter did not help to find the data too
Searched apartments and did not find
?
there is no information related to apartments in 2013
I haven't found the information using the search function. Then I tries with the filters, by using the 'housing' and 'helsinki' filter. I have found many things but not the building of apartments in 2013. Maybe the question was not well framed.
Keyword not in correct order
Maybe the language used in search was not good.
Searching tags are absent or poorly tuned

continues

APPENDIX 1. (continues)

Hard to narrow down the correct search terms
I haven't the right application in my computer, which would open the database HKI rakennuskanta.
Group of search words resulted in more than 100 datasets. The initial search results didn't gave the suitable results. Also, that search words were in English. It's possible that dataset is in Finnish.
The problem was in the wrong combination of search words. The phrase "number of apartments in Helsinki region" didn't work, but "started and completed housing in the Helsinki region" worked successfully
"building of how many apartments was started" I do not understand what that means, what I am supposed to look for
The data is a group consisting ten years, therefore its impossible to distinguish the difference between 2013 and 2014 for example.
I have to admit i am computer illiterate, besides some information's are in Finnish and i don't understand.
Used wrong keywords
The exact title was not found from the data base titles. I searched for the key words but didn’t find the exact same request
Too complicated search task. Too narrow
search criteria or the key words
I could identify the right key-word/ category to search/ filter. Also, search pages/ results in Finnish language - so may be skipped some results.

continues

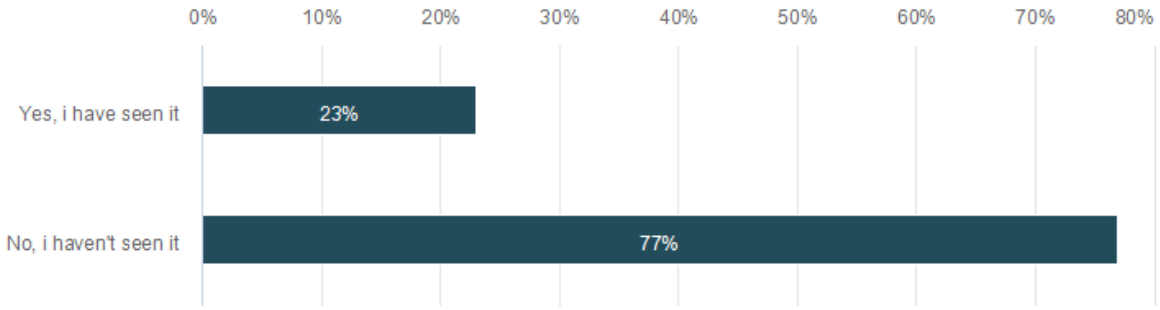
APPENDIX 1. (continues)

there is no a dataset related with this topic
Keywords not adapted in the description

12. Right dataset is the dataset on screenshot. Have you seen it while searching?

Link to dataset - https://hri.fi/data/en_GB/dataset/helsingin-seudun-aloitetut-ja-valmistuneet-asunnot

Number of respondents: 30



	n	Percent
Yes, i have seen it	7	23.3%
No, i haven't seen it	23	76.7%

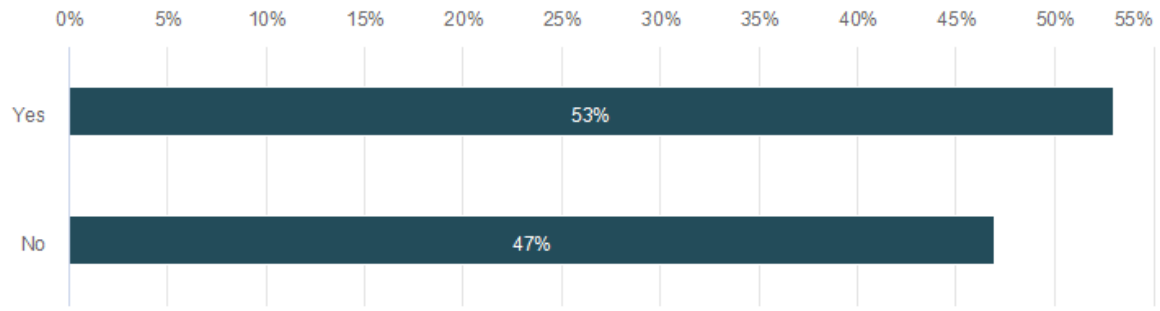
continues

APPENDIX 1. (continues)

13. Right dataset is the dataset on screenshot. Did you find correct dataset?

Link to dataset - https://hri.fi/data/en_GB/dataset/helsingin-seudun-aloitetut-ja-valmistuneet-asunnot

Number of respondents: 34



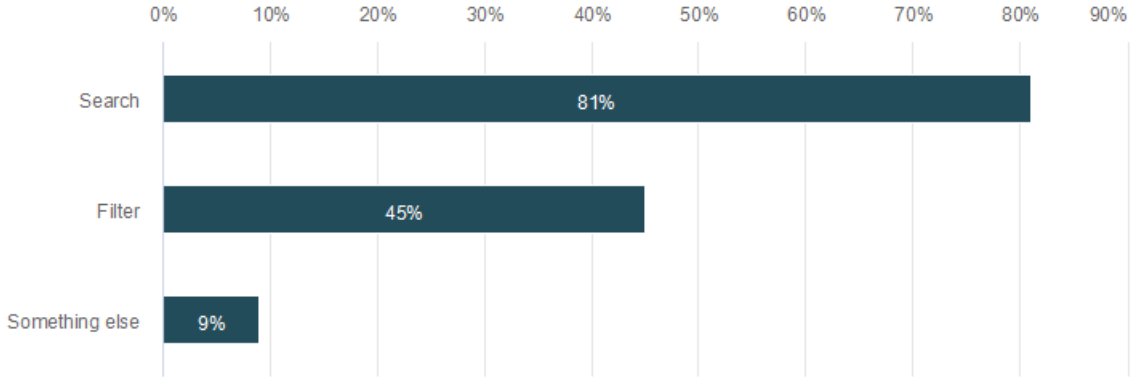
	n	Percent
Yes	18	52.9%
No	16	47.1%

14. What did you use to try to find data?

Number of respondents: 64, selected answers: 87

continues

APPENDIX 1. (continues)



	n	Percent
Search	52	81.3%
Filter	29	45.3%
Something else	6	9.4%

Answers given into textfield

Option names	Text
Something else	tag
Something else	Filter 'housing', then search 'building'
Something else	Categories from the top
Something else	Data-Categories-Constructed environment
Something else	categories
Something else	Sorting

continues

APPENDIX 1. (continues)

15. Why did you make this choice? And which option was successful search/filter or something else? 2

Number of respondents: 61

Responses
Filter for searching tags. Tags for finding a dataset. Filter for taking into account the region and specifics of the dataset. Filter was successful in search.
Frontpage-Helsinki-Housing-Constructed Environment
I dont know where are the filters on this website
I searched few things and could not find.
filter building, search helsinki region
Filtering was more useful.
it was the basic way to begin.
I was not sure how the data set will be named, so I had to use a combination of search and filters
Search by key words. Apartment/building/start/housing
I thought using the search would be the quickest way. I searched first with the term "apartment", no results, then "constuction" and then thought I had found the answer

continues

APPENDIX 1. (continues)

I tried with building 2013 and houses 2013. Didn't find anything
Found by selecting 'Housing' from the front page, then scrolling the results
Because I was not able to find the data by using different search words, I checked what categories similar datasets had and used those. I found the data by using "housing" category and filtering for Helsinki region.
I set category to housing to decrease amount of options. Applied 'Helsinki' search to limit them even more. Then looked through 4 pages to find what I need.
was partially successful
used both, but not logic. nothing.
The keyword search provided many options therefore I had to use filter this time. It took couple of tries to get the right page. I wasn't sure if I have reached to the right place therefore I checked the description of data below and saw that it is the right data.
I decided to use search and filter because there were many more results compared to the first one.
neither
the same
Search
Tried both
Felt more precise
I don't know why, my choice was totally random, filter

continues

APPENDIX 1. (continues)

Front page did have search box but no filters
felt the correct way to do it
treće?
buildings
request for "buildings" was successful, but not related to the desired result
I haven't found the information using the search function. Then I tries with the filters, by using the 'housing' and 'helsinki' filter. I have found many things but not the building of apartments in 2013. Was succesful after that I have searched housing helsinki region"
Housing Helsinki
Its easier to use keywords
Search
Wasn't sure what terms to use for search, took some time to narrow down search results, used "helsinki 2013 housing".
Because previous choice was ineffective
Easy to filter
Just how I search for things most of the time.
I find the first database with search. I tried the words "building statistics", but I couldn't find the right database for some reason.

continues

APPENDIX 1. (continues)

Too many datasets to look through.
neither search nor filter helped to find the database
Construction
I used filter option because it doesn't have any prime key words.
buildings 2013 Helsinki
Search and filter seems efficient but this time there are no clear key words, so only search is not enough and filters are needed.
Clear task - "look for x" - and that's what search engines are for.
I felt it was the most convenient way. I didnt find the right dataset but I found another instead: https://api.aluesarjat.fi/pxweb/fi/Helsingin%20seudun%20tilastot/Helsingin%20seudun%20tilastot__Helsingin%20seutu__Asunto%20ja%20rakennuskanta__Rakennuskanta/Hginseutu_AR_RR01_Rakennukset_kayttark_rvuosi.px/
i thought filtering and search takes you straight to what information you are looking for but i couldn't succeed in any.
Used filter to crop searchs only to Helsinki area.
From previous example I expected the search functionality to be helpful
Because I used it before and it worked
Had to try multiple different methods as the task was to find started housing in 2013, but the headline of the dataset does not tell that. That link was chosen as it

continues

APPENDIX 1. (continues)

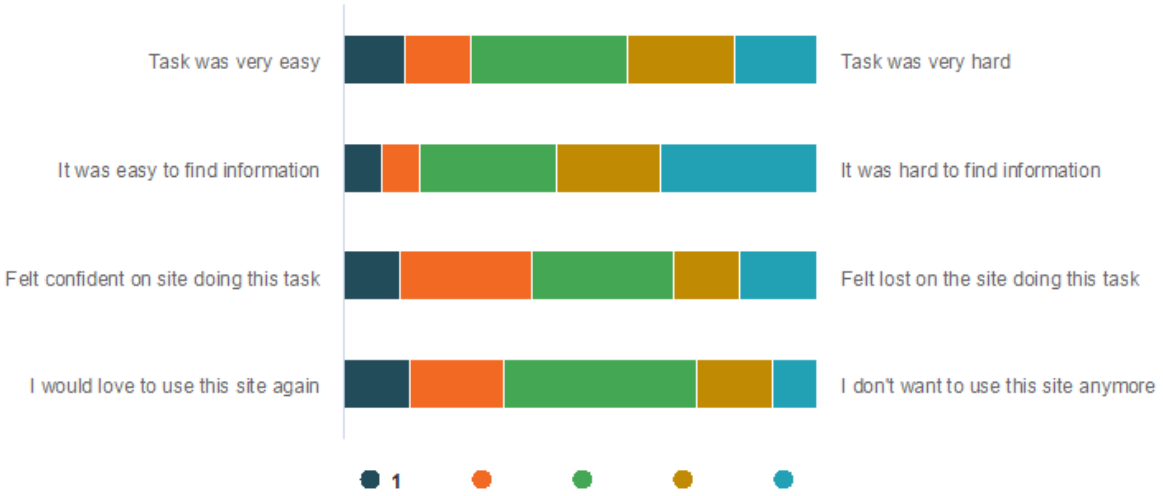
seemed to be something that might have the asked dataset in it. Ultimately the search option was the one used to find the answer.
I used search because it seems to be the only convenient option. However I failed in finding right data set.
Did not know how to search for it so used Helsinki Region & housing as filters.
easiest
I tried search first but I got no results with "apartments". Then I selected the "housing" category and after a quick scroll I found the right one.
I searched for "Building Helsinki region 2013", however I didn't find anything.
I used Filter and then tried Search in the filtered option.
"building of apartments in Helsinki region". I used the search, because it means easier and faster for me
I typed completed housing in the Helsinki and searched from there
To find out with keywords
I had to use both options, the key word housing in the search or filter was important to find the relevant dataset

16. Choose a statement, that best describes your experience

Number of respondents: 64

continues

APPENDIX 1. (continues)



	1						Total	Average	Median
Task was very easy	8	9	21	15	11	Task was very hard	64	3.2	3.0
	12.5 %	14.1 %	32.8 %	23.4 %	17.2 %				
It was easy to find information	5	5	19	14	21	It was hard to find information	64	3.6	4.0
	7.8 %	7.8 %	29.7 %	21.9 %	32.8 %				
Felt confident on site doing this task	8	18	19	9	10	Felt lost on the site doing this task	64	2.9	3.0
	12.5 %	28.1 %	29.7 %	14.1 %	15.6 %				

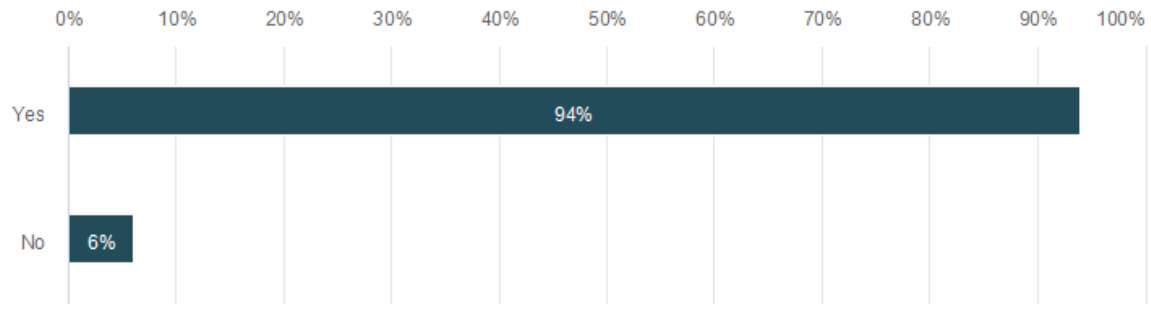
continues

APPENDIX 1. (continues)

I would love to use this site again	9	13	26	10	6	I don't want to use this site anymore	64	2.9	3.0
	14.1 %	20.3 %	40.6 %	15.6 %	9.4 %				
Total	**30**	**45**	**85**	**48**	**48**		**256**	**3.2**	**3.0**

17. Did you find the information you were looking for?

Number of respondents: 64



	n	Percent
Yes	60	93.8%
No	4	6.2%

18. Can you send here a link to the dataset?

Number of respondents: 60

continues

APPENDIX 1. (continues)

Responses
https://hri.fi/data/en_GB/dataset/veden-reaaliaikainen-lampotila-helsingin-uimarannoilla
https://hri.fi/data/en_GB/dataset/veden-reaaliaikainen-lampotila-helsingin-uimarannoilla
https://hri.fi/data/en_GB/dataset/veden-reaaliaikainen-lampotila-helsingin-uimarannoilla
https://hri.fi/data/en_GB/dataset/veden-reaaliaikainen-lampotila-helsingin-uimarannoilla
https://hri.fi/data/en_GB/dataset/veden-reaaliaikainen-lampotila-helsingin-uimarannoilla
https://hri.fi/data/en_GB/dataset/veden-reaaliaikainen-lampotila-helsingin-uimarannoilla
https://hri.fi/data/en_GB/dataset/veden-reaaliaikainen-lampotila-helsingin-uimarannoilla
https://hri.fi/data/en_GB/dataset/veden-reaaliaikainen-lampotila-helsingin-uimarannoilla
https://hri.fi/data/en_GB/dataset/veden-reaaliaikainen-lampotila-helsingin-uimarannoilla
https://hri.fi/data/en_GB/dataset/veden-reaaliaikainen-lampotila-helsingin-uimarannoilla
https://hri.fi/data/en_GB/dataset/veden-reaaliaikainen-lampotila-helsingin-uimarannoilla

continues

https://hri.fi/data/en_GB/dataset/veden-reaaliaikainen-lampotila-helsingin-uimarannoilla
https://hri.fi/data/en_GB/dataset/veden-reaaliaikainen-lampotila-helsingin-uimarannoilla
https://hri.fi/data/en_GB/dataset/veden-reaaliaikainen-lampotila-helsingin-uimarannoilla
https://hri.fi/data/en_GB/dataset/veden-reaaliaikainen-lampotila-helsingin-uimarannoilla
https://hri.fi/data/en_GB/dataset/veden-reaaliaikainen-lampotila-helsingin-uimarannoilla
https://hri.fi/data/en_GB/dataset/veden-reaaliaikainen-lampotila-helsingin-uimarannoilla
https://hri.fi/data/fi/dataset/veden-reaaliaikainen-lampotila-helsingin-uimarannoilla
https://hri.fi/data/en_GB/dataset/veden-reaaliaikainen-lampotila-helsingin-uimarannoilla
https://hri.fi/data/en_GB/dataset/veden-reaaliaikainen-lampotila-helsingin-uimarannoilla
https://hri.fi/data/en_GB/dataset/veden-reaaliaikainen-lampotila-helsingin-uimarannoilla
https://hri.fi/data/en_GB/dataset/veden-reaaliaikainen-lampotila-helsingin-uimarannoilla

continues

APPENDIX 1. (continues)

https://hri.fi/data/en_GB/dataset/veden-reaaliaikainen-lampotila-helsingin-uimarannoilla
https://hri.fi/data/en_GB/dataset/veden-reaaliaikainen-lampotila-helsingin-uimarannoilla
https://hri.fi/data/en_GB/dataset/veden-reaaliaikainen-lampotila-helsingin-uimarannoilla
https://hri.fi/data/en_GB/dataset/veden-reaaliaikainen-lampotila-helsingin-uimarannoilla
https://hri.fi/data/en_GB/dataset/veden-reaaliaikainen-lampotila-helsingin-uimarannoilla
https://hri.fi/data/en/dataset/veden-reaaliaikainen-lampotila-helsingin-uimarannoilla
https://hri.fi/data/en_GB/dataset/veden-reaaliaikainen-lampotila-helsingin-uimarannoilla
https://hri.fi/data/en/dataset/veden-reaaliaikainen-lampotila-helsingin-uimarannoilla
https://hri.fi/data/en_GB/dataset/veden-reaaliaikainen-lampotila-helsingin-uimarannoilla
https://hri.fi/data/en_GB/dataset/veden-reaaliaikainen-lampotila-helsingin-uimarannoilla
https://hri.fi/data/en_GB/dataset/veden-reaaliaikainen-lampotila-helsingin-uimarannoilla

continues

APPENDIX 1. (continues)

https://hri.fi/data/en/dataset/veden-reaaliaikainen-lampotila-helsingin-uimarannoilla
https://hri.fi/data/en/dataset/veden-reaaliaikainen-lampotila-helsingin-uimarannoilla
https://hri.fi/data/en_GB/dataset/veden-reaaliaikainen-lampotila-helsingin-uimarannoilla
https://hri.fi/data/en_GB/dataset/veden-reaaliaikainen-lampotila-helsingin-uimarannoilla
https://hri.fi/data/en_GB/dataset/veden-reaaliaikainen-lampotila-helsingin-uimarannoilla
https://hri.fi/data/en/dataset/veden-reaaliaikainen-lampotila-helsingin-uimarannoilla
https://hri.fi/data/en_GB/dataset/veden-reaaliaikainen-lampotila-helsingin-uimarannoilla
https://hri.fi/data/en_GB/dataset/veden-reaaliaikainen-lampotila-helsingin-uimarannoilla
https://hri.fi/data/en_GB/dataset/veden-reaaliaikainen-lampotila-helsingin-uimarannoilla
https://hri.fi/data/en_GB/dataset/veden-reaaliaikainen-lampotila-helsingin-uimarannoilla
https://hri.fi/data/en_GB/dataset/veden-reaaliaikainen-lampotila-helsingin-uimarannoilla

continues

APPENDIX 1. (continues)

https://hri.fi/data/en_GB/dataset/veden-reaaliaikainen-lampotila-helsingin-uimarannoilla
https://hri.fi/data/en_GB/dataset/veden-reaaliaikainen-lampotila-helsingin-uimarannoilla
https://hri.fi/data/en_GB/dataset/veden-reaaliaikainen-lampotila-helsingin-uimarannoilla
https://hri.fi/data/en_GB/dataset/veden-reaaliaikainen-lampotila-helsingin-uimarannoilla
https://hri.fi/data/en_GB/dataset/veden-reaaliaikainen-lampotila-helsingin-uimarannoilla
https://hri.fi/data/en_GB/dataset/veden-reaaliaikainen-lampotila-helsingin-uimarannoilla
https://hri.fi/data/en_GB/dataset/veden-reaaliaikainen-lampotila-helsingin-uimarannoilla
https://hri.fi/data/en_GB/dataset/veden-reaaliaikainen-lampotila-helsingin-uimarannoilla
https://hri.fi/data/en_GB/dataset/veden-reaaliaikainen-lampotila-helsingin-uimarannoilla
https://hri.fi/data/en_GB/dataset/veden-reaaliaikainen-lampotila-helsingin-uimarannoilla
https://hri.fi/data/en_GB/dataset/veden-reaaliaikainen-lampotila-helsingin-uimarannoilla

continues

APPENDIX 1. (continues)

https://hri.fi/data/en_GB/dataset/veden-reaaliaikainen-lampotila-helsingin-uimarannoilla
https://hri.fi/data/en/dataset/veden-reaaliaikainen-lampotila-helsingin-uimarannoilla
https://hri.fi/data/en_GB/dataset/veden-reaaliaikainen-lampotila-helsingin-uimarannoilla
https://hri.fi/data/en/dataset/veden-reaaliaikainen-lampotila-helsingin-uimarannoilla
https://hri.fi/data/en_GB/dataset/veden-reaaliaikainen-lampotila-helsingin-uimarannoilla

19. What in your opinion was the problem?

Number of respondents: 4

Responses
Too confusing
Keyword search failed
?
I found the database Real-time swimming water temperature in Helsinki, but when I tried to open it, there was only following list of links "Index of /opendata/uiras/". I tried to open several links, but I couldn't find, what I was looking for.

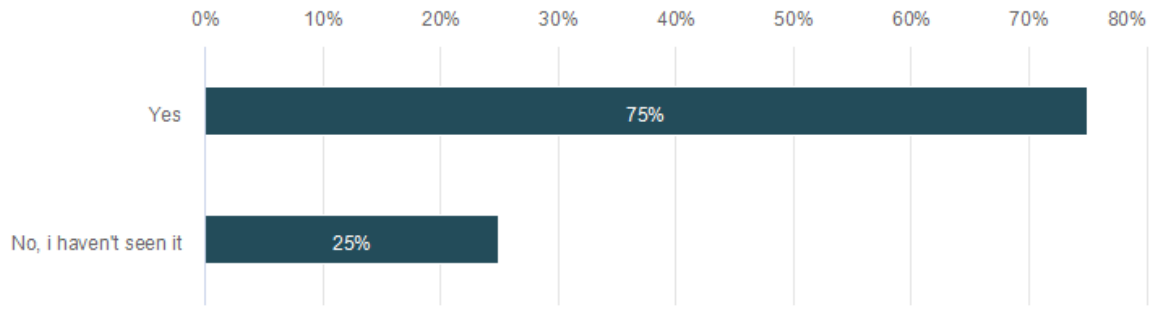
continues

APPENDIX 1. (continues)

20. Right dataset is the dataset on screenshot. Have you seen it while searching?

Link to dataset - https://hri.fi/data/en_GB/dataset/veden-reaaliaikainen-lampotila-helsingin-uimarannoilla

Number of respondents: 4



	n	Percent
Yes	3	75.0%
No, i haven't seen it	1	25.0%

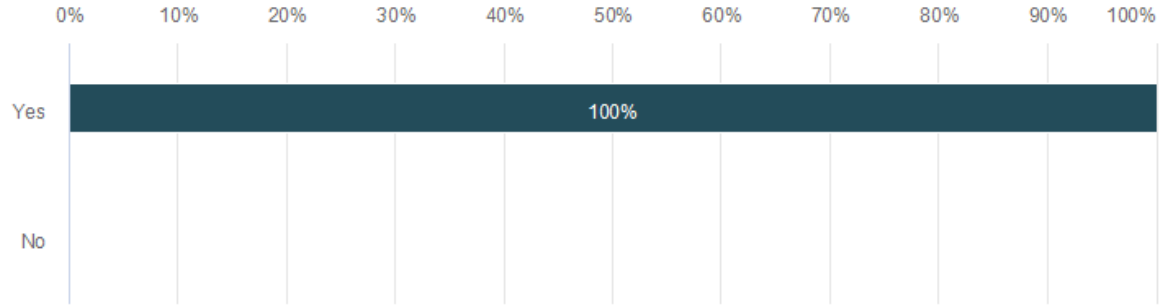
21. Right dataset is the dataset on screenshot. Did you find correct dataset?

Link to dataset - https://hri.fi/data/en_GB/dataset/veden-reaaliaikainen-lampotila-helsingin-uimarannoilla

Number of respondents: 60

continues

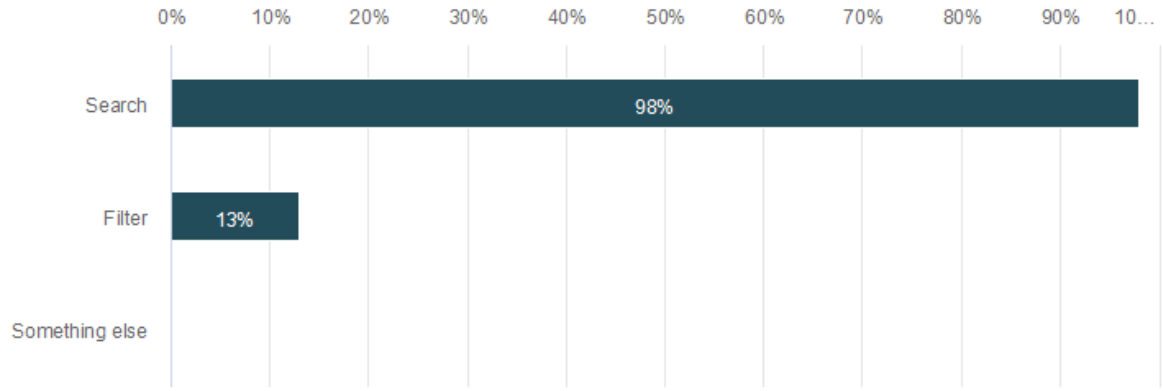
APPENDIX 1. (continues)



	n	Percent
Yes	60	100.0%
No	0	0.0%

22. What did you use to try to find data?

Number of respondents: 64, selected answers: 71



	n	Percent
Search	63	98.4%

continues

APPENDIX 1. (continues)

Filter	8	12.5%
Something else	0	0.0%

Answers given into textfield

Option names	Text

23. Why did you make this choice? And which option was successful search/filter or something else? 3

Number of respondents: 64

Responses
Filter to get recreation datasets.< Search to get the water
I search "temperature water"
because search is easy-to-find and easy-to-use
easy
Prior two searches did not work well using the filter, so searching directly for swimming water temperature was my choice.
water temperature
Searching "water temperature" gave right answer immediately.
There was no other choice

continues

APPENDIX 1. (continues)

I have tried to search for "water" and there were not so many data sets to chose from, so no extra filters were needed
Water temperature in search
First used "Helsinki" filter and then search for "water temperature"
I searched for water temperature
Found from frontpage under 'Culture and recreation'. Using the icon in the frontpage is easy and rewarding.
I searched "temperature swimming" and found the data as the only result.
Set filter, than applied 'water' search.
I learnt I have to use searching bar^) and to be honest, this task was easy
basic key-word "swimming" was enough to find the necessary data.
Again, keyword search "swimming water temperature" gave me one link which was the correct one.
The search was the easiest one and it gave me the right answer so it was not necessary to use any other option.
easiest
the same
search - easy
tried both
Ideal for such a specific query

continues

APPENDIX 1. (continues)

I don't know, my ass advised me search
Searched swim
specific search needed full tag name
https://hri.fi/data/en_GB/dataset/veden-reaaliaikainen-lampotila-helsingin-uimarannoilla
temperature
the easiest way to find the info was to copy the question and then put it into the search bar
It's easier, I just typed "water temperature".
Because I was sure about the information I would like to find. I search for 'temperature water swim'
Swim
Keyword keywords
Search
Searched for "swimming" and filtered for helsinki
Because it's more simple than searching through databases.
Easy
Easy to use

continues

APPENDIX 1. (continues)

Tends to work
I think that I should find what I am looking for with the search.
First option on the page. Also makes sense when you have all the search words available.
It was the easiest and obvious way to search info
“Water “
I used search option creating my own keyword " water temperature" and it is successful
temperature of the swimming water in Helsinki
Clear key words, I typed “Helsinki swimming”
"swimming water"
Easiest way.
Because i wanted to find the temperature for swimming water.
Used keyword "swimming"
~
Search because it’s easier to find based on key words
A super specific task so i tought there cannot be many datasets that both have the words "swimming" and "temperature" in them.
the most easy and quick option.
It was relatively obvious what the dataset would be named like.

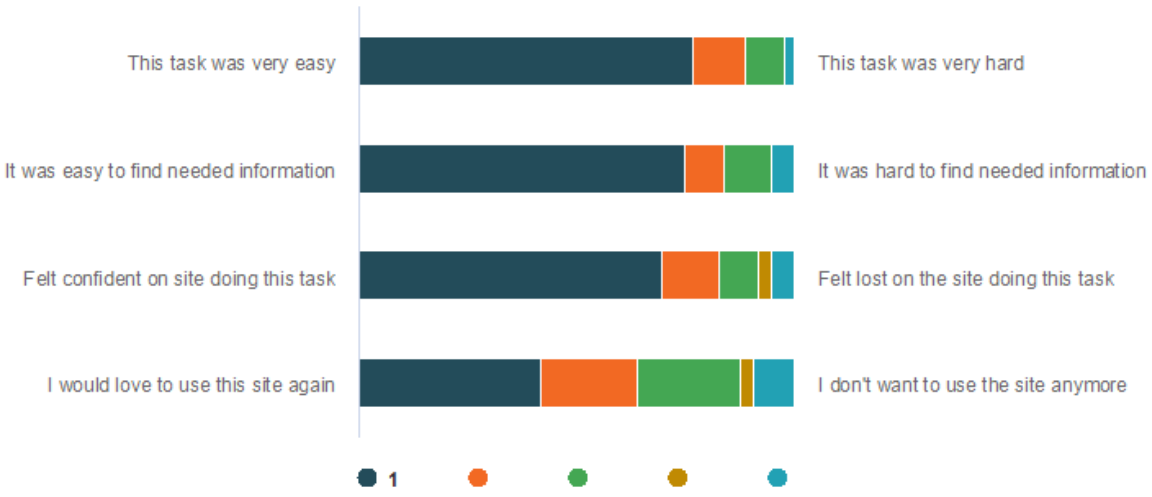
continues

APPENDIX 1. (continues)

easiest
I searched "temperature" and the second result was the right one
Search proved to be a good method and this time it again brought a success. I searched for "temperature"
'Swimming' is good (unique) key-word to search with. I felt it is quicker to start with search on the website and get a quick sense of whether it works.
because it is easier and faster if I use search than filter
Type swimming water temperature is search and it will appear.
Keywords
Search, since it was specific

24. Choose a statement, that best describes your experience

Number of respondents: 64



continues

APPENDIX 1. (continues)

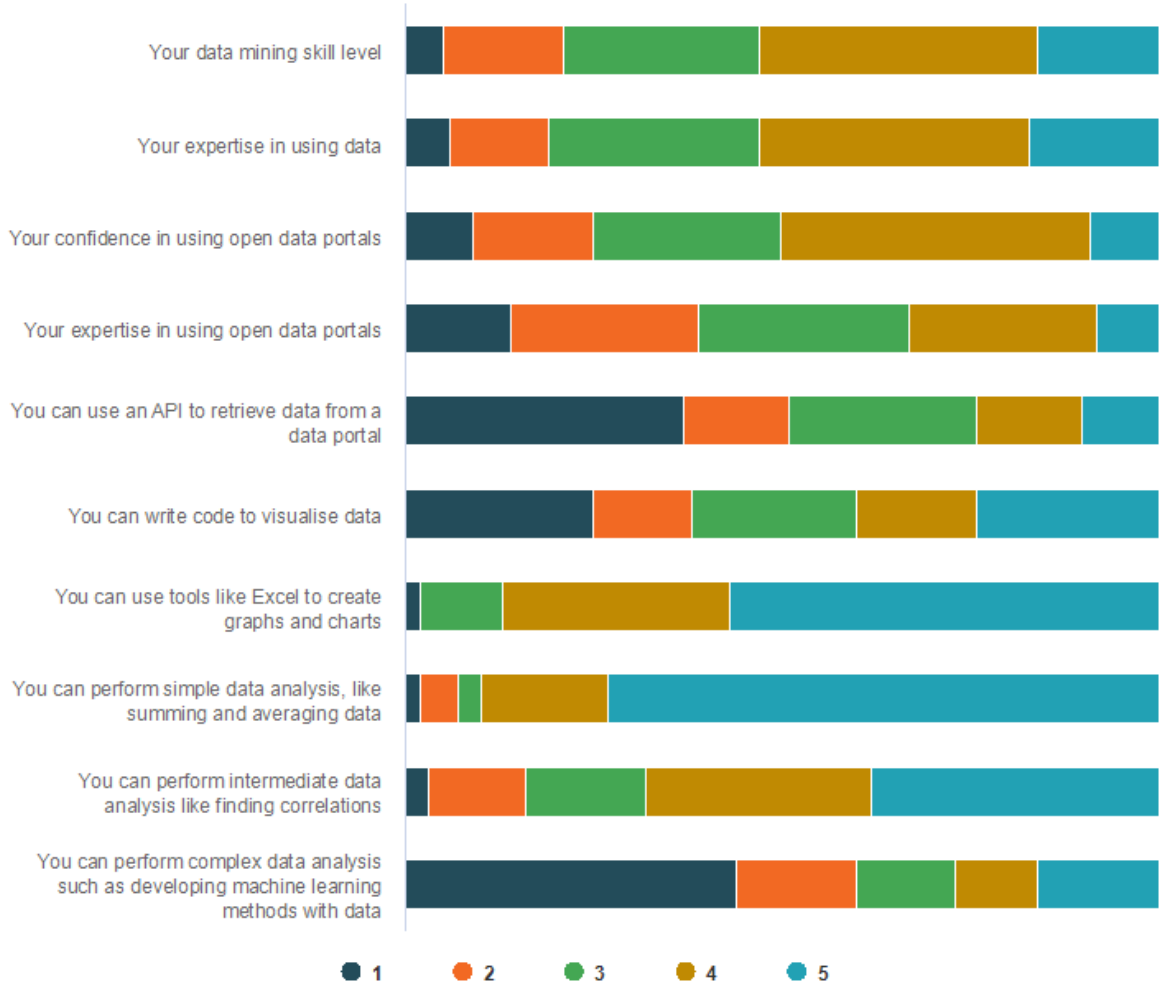
	1						Total	Average	Median
This task was very easy	49	8	6	0	1	This task was very hard	64	1.4	1.0
	76.6 %	12.5 %	9.4%	0.0 %	1.5 %				
It was easy to find needed information	48	6	7	0	3	It was hard to find needed information	64	1.5	1.0
	75.0 %	9.4%	10.9 %	0.0 %	4.7 %				
Felt confident on site doing this task	45	8	6	2	3	Felt lost on the site doing this task	64	1.6	1.0
	70.3 %	12.5 %	9.4%	3.1 %	4.7 %				
I would love to use this site again	27	14	15	2	6	I don't want to use the site anymore	64	2.2	2.0
	42.2 %	21.9 %	23.4 %	3.1 %	9.4 %				
Total	**169**	**36**	**34**	**4**	**13**		**256**	**1.7**	**1.0**

continues

APPENDIX 1. (continues)

25. Using a rating scale of 1-5, please answer the following statements (1 - bad, 5 - excelent)

Number of respondents: 64



	1	2	3	4	5		Total	Average	Median
	3	10	17	24	10		64	3.4	4.0

continues

APPENDIX 1. (continues)

Your data mining skill level	4.7%	15.6 %	26.6 %	37.5 %	15.6 %				
Your expertise in using data	4	8	18	23	11		64	3.5	4.0
	6.3%	12.5 %	28.1 %	35.9 %	17.2 %				
Your confidence in using open data portals	6	10	16	26	6		64	3.3	3.5
	9.4%	15.6 %	25.0 %	40.6 %	9.4%				
Your expertise in using open data portals	9	16	18	16	5		64	2.9	3.0
	14.1 %	25.0 %	28.1 %	25.0 %	7.8%				
You can use an API to retrieve data from a data portal	23	9	16	9	6		63	2.5	2.0
	36.5 %	14.3 %	25.4 %	14.3 %	9.5%				
You can write code to visualise data	16	8	14	10	15		63	3.0	3.0
	25.4 %	12.7 %	22.2 %	15.9 %	23.8 %				

continues

APPENDIX 1. (continues)

You can use tools like Excel to create graphs and charts	1	0	7	19	36		63	4.4	5.0
	1.6%	0.0%	11.1 %	30.2 %	57.1 %				
You can perform simple data analysis, like summing and averaging data	1	3	2	11	46		63	4.6	5.0
	1.6%	4.8%	3.2%	17.4 %	73.0 %				
You can perform intermediate data analysis like finding correlations	2	8	10	19	24		63	3.9	4.0
	3.2%	12.7 %	15.9 %	30.1 %	38.1 %				
You can perform complex data analysis	28	10	8	7	10		63	2.4	2.0
	44.4 %	15.9 %	12.7 %	11.1 %	15.9 %				

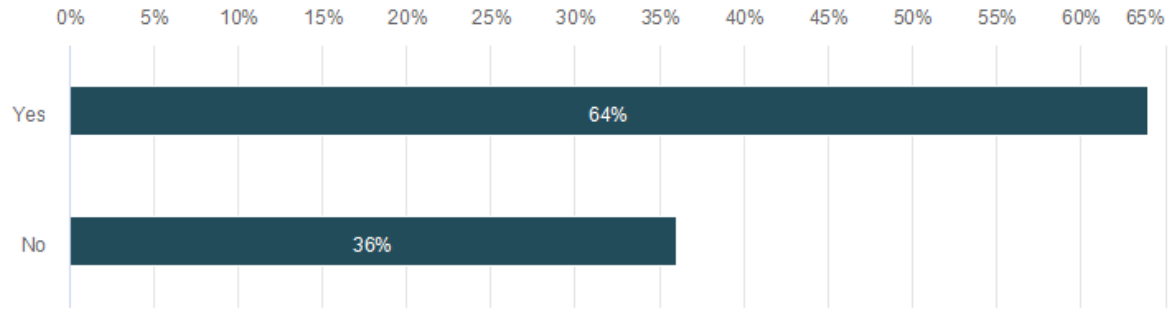
continues

APPENDIX 1. (continues)

such as developing machine learning methods with data									
Total	**93**	**82**	**126**	**164**	**169**		**634**	**3.4**	**4.0**

26. Have you used open data portals before?

Number of respondents: 64

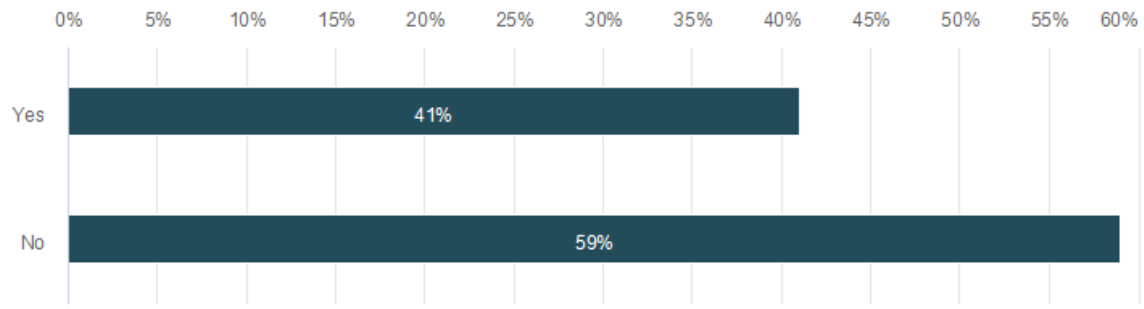


	n	Percent
Yes	41	64.1%
No	23	35.9%

27. Did you have problems with them?

Number of respondents: 41

continues



	n	Percent
Yes	17	41.5%
No	24	58.5%

28. What kind of problems did you have?

Number of respondents: 18

Responses
bad data, missing data, broken links
It is often time-consuming to find what I am looking for. Sometimes I get too frustrated and quit before finding the result.
Your first survey
Sometimes open data portals are not having very clear information about how to use APIs and documentation can be very hard to find.
the picture in the middle of the web site is not a link to the main page!1111111

continues

APPENDIX 1. (continues)

Irrelevant search and invalid download the results
Unstructured data, badly gathered and represented
Could not find data
...
Sometimes it's hard to find information and adopt to new interface.
Hard to find information. But, may be, problem is in my skills
Getting the data in right format. Incomplete data, missing values.
Not reliable data, outdated data
Sometimes there is data, sometimes not. Sometimes data is uptodate, sometimes not. Sometimes data is missing, holes in the data.
Massive amounts of information and problems to find the right data.
Retrieving and evaluating the data.
Had hard time finding what I needed. It was most likely due to not knowing exactly what I was looking for.
Finding the right information

29. Are you more likely to use open data platforms after this brief introduction?

Number of respondents: 23

continues

APPENDIX 1. (continues)

Responses
maybe in future data scraping
No
What is open data platforms?
Yes
If my research work requires it
Perhaps
Maybe
No
Yes
No, I would better kill myself
yes
no, since I don´t need it
Yes sure
More likely to use them form now on
Maybe, yes.
Maybe, if I ever need this kind of data.
When needed yes
No, don't need

continues

APPENDIX 1. (continues)

No, no I am not.
Yes, may be.
Yes
no.
Maybe

30. Do you have something in mind that can improve open data portals?

Number of respondents: 24

Responses
It's good when open data portals have a clear structure, if they contain more than one type of data set.
-
They would be more intuitive to use and maybe have some "tips" and pop-up windows to guide the user when he/she opens the data portal. But not too many pop-ups because they might be irritating, too.
-
No
Clarity of interface and subdivision of data in categories that can be easily related to the query point by the user.

continues

APPENDIX 1. (continues)

At least this one didn't seem to support use of AND, OR, NOT etc. type queries/searches. I think adding support for them would be an improvement.
Few open data portals do not use good number of tags which makes things difficult for researchers to find right data. It would be nicer to use tags of more specific keywords on these portals.
NE
more consistent metadata in all Finnish portals
Better use of keywords
Not right now.
Make more words combination for searching the same item.
More data
-
-
Not really
Not really
Not particularly
I have used data for electricity prices in Nordpool market (https://www.nordpoolgroup.com/Market-data1/#/nordic/table). One improvement could be to provide meta-data such as:- How the data is generated? What is the confidence (accuracy) of the generated data?

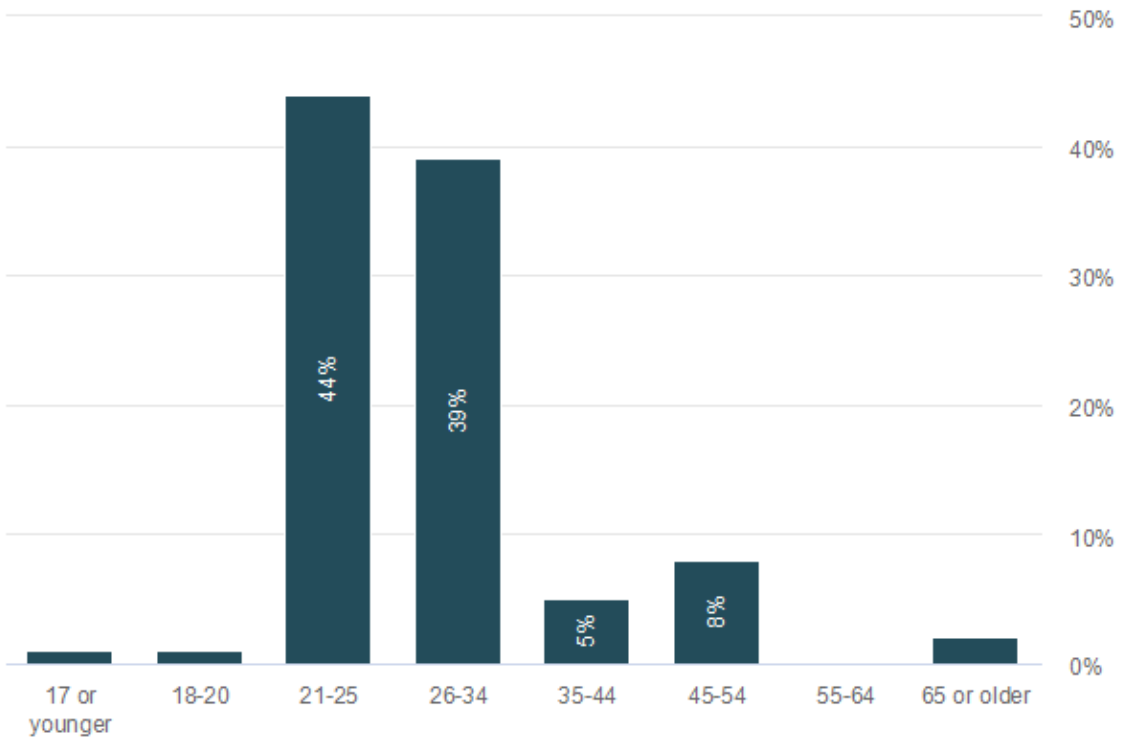
continues

APPENDIX 1. (continues)

etc.
might be more intelligent algorithms of search
Improve the search engines but this seemed perfect
Open data, free access
The key words have to be specific and identified. Adding some visualization to certain topics like housing or temperature to ease the research or finding of relevant data is beneficial. The data portal used was good and information were easily founded

31. Please indicate your age category.

Number of respondents: 64



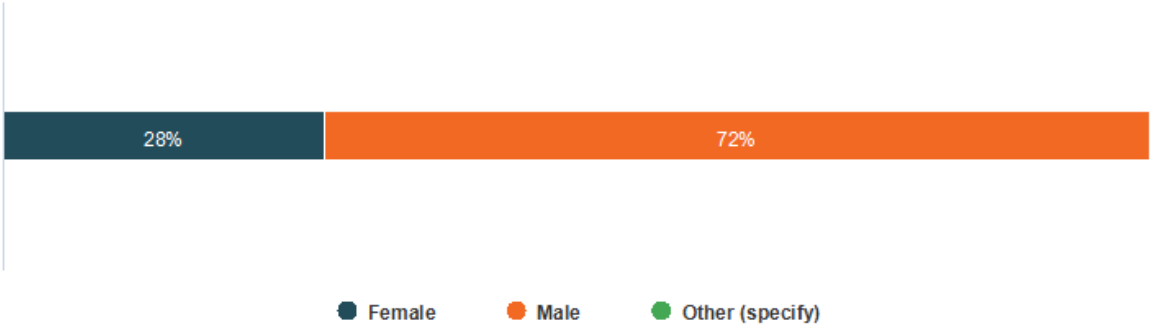
continues

APPENDIX 1. (continues)

	n	Percent
17 or younger	1	1.6%
18-20	1	1.6%
21-25	28	43.7%
26-34	25	39.0%
35-44	3	4.7%
45-54	5	7.8%
55-64	0	0.0%
65 or older	1	1.6%

32. What is your gender?

Number of respondents: 64



continues

APPENDIX 1. (continues)

	n	Percent
Female	18	28.1%
Male	46	71.9%
Other (specify)	0	0.0%

Answers given into textfield

Option names	Text

33. What is your first language?

Number of respondents: 64

Responses
Russian
Russian
Russian
portuguese
German
Russian
Finnish

continues

APPENDIX 1. (continues)

English
Russian
Russian
Finnish
Finnish
Italian
Finnish
Russian
Russian
Russian
English
Finnish
rus
Russian
Russian
Russian
Italian
Russian
Finnish

continues

APPENDIX 1. (continues)

spanish
HR
Finnish
Russian
Russian
Italian
Spanish
Spanish
Arabic
Finnish
Ukrainian
English
Russian
English
Finnish
Urdu
Russian
English
Tamil

continues

APPENDIX 1. (continues)

Turkish
Russian
Finnish
Finnish
Rukwangali
finnish
Russian
Spanish
Finnish
Russian
Finnish
Urdu
Italian
Russian
Hindi
Russian
Arabic
French
Arabic

continues

APPENDIX 1. (continues)

34. What other languages do you speak?

Number of respondents: 64

Responses
English
English, French, Finnish
English
english
English
English, Finnish, French
English and little Swedish
French, Finnish
English
English, Finnish
Swedish, English, Italian, Russian
German, English, Swedish
English, Finnish
English, Swedish, German
English

continues

APPENDIX 1. (continues)

Russian/English
English
Finnish
English, Swedish
rus eng fra esp esperanto ger ita
English
English
English
English
English, Finnish
English
english, russian
EN
English, Swedish, German
English
English, Afrikaans, Spanish
English
English
English

continues

APPENDIX 1. (continues)

Japanese Chinese
English
English, Russian, Ukrainian
Russian, English, Polish
Swedish Finnish
English
Finnish Norwegian French Creole Spanish
English
English Punjabi
English
None
English, Hindi.
English
English, German, Finnish

continues

APPENDIX 1. (continues)

English, Swedish, German
English
English, Oshiwambo, Afrikand and a little bit of Finnish
english
English, Spanish, German
English - Hebrew
English and Swedish
English, Finnish, German
English
English Finnish Punjabi
English
English and a bit of Finnish
English, Marathi
English
Japanese, English and minor French
English Spanish

continues

APPENDIX 1. (continues)

English
French

35. Where do you live (country)?

Number of respondents: 64

Responses
Russia
Finland
Finland
finland
Finland
Finland
Finland
Finland
Finland
Finland
Finland
Finland
Finland

continues

APPENDIX 1. (continues)

Finland
Russian Federation
Russia
Finland
Finland
Finland
russia
Russia
Finland
Finland
Italy
Finland
Finland
finland
HR
Finland
Finland
Finland
Italy

continues

APPENDIX 1. (continues)

Finland
Honduras
Finland
Finland
Ukraine
Finland
Finland
Finland
Finland
Finland
Russia
Finland
Finland
Finland
Finland
Finland
Finland
Finland
Finland

continues

APPENDIX 1. (continues)

Finland
Finland
Finland

continues

Finland
Finland
Finland
Italy
Finland
Finland
Finland
Finland
France
Lebanon