

Using open data for digital innovation: Barriers for use and recommendations for publishers

Marit K. Natvig,^{1*} Shanshan Jiang,² Erlend Stav³

^{1*} ORCID Nr: 0000-0002-5885-029X

SINTEF AS, P.O. Box 4760 Torgarden, N-7465 Trondheim, Norway, marit.k.natvig@sintef.no

² ORCID Nr: 0000-0001-6862-5187

SINTEF AS, P.O. Box 4760 Torgarden, N-7465 Trondheim, Norway, shanshan.jiang@sintef.no

³ ORCID Nr: 0000-0001-9957-6744

SINTEF AS, P.O. Box 4760 Torgarden, N-7465 Trondheim, Norway, erlend.stav@sintef.no

Abstract: Open data from the public sector can fuel the innovation of digital products. This paper investigates barriers and success factors regarding use of open data in such innovations, and how the public sector can increase the value of published data. A multimethod approach was used. An initial study identified aspects of relevance through interviews, a system development experiment, and a focus group. An in-depth study used the insight to perform interviews and a survey targeting innovators. Details on data needs, discovery, assessment, and use were found as well as barriers regarding use of open data in digital product innovations. Associated recommendations to data owners are provided regarding how they can increase the innovation capacity through appropriate licenses and service levels; convenient access mechanisms; publishing channels and infrastructures; transparency and dialogue; data, metadata, documentation, and APIs of high quality; harmonization and standardization.

Keywords: Open government data, open data barriers, open data recommendations, open data publishing, product innovation

Acknowledgement: The work has been funded by The Research Council of Norway, grant number 257153 and 295788. A CC BY or equivalent license is applied to any Author Accepted Manuscript (AAM) version arising from this submission, in accordance with the grant's open access conditions.

1. Introduction

Open data is data that can be freely used, reused, and redistributed by anyone¹. According to the European Open Data Directive (Union, 2019) and Inspire Directive², government data must in general be open data and available to the public, and the ITS Directive (Union, 2010) defines requirements on open transport data. One important motivation for the opening of public data is to fuel the economy - open data should facilitate the innovation of new digital products (digital services included), create new jobs, increase value creation, and contribute to a better and more prosperous society. European studies like (Huyer & Knippenberg, 2020) and (Berends et al., 2020) show the huge potential of open data regarding economic and societal aspects. (Berends et al., 2020) estimates the market size for the period 2016-2020 to be 325 billion Euros. For 2020, the savings in public administrations are estimated to 1.7 billion Euros, and 7000 lives are estimated saved due to quicker emergency responses. (Huyer & Knippenberg, 2020) estimates the market size in 2025 to be up to 334 billion Euros, and that there will be up to 1,97 million direct and indirect open data related jobs in 2025. A long list of economic and non-economic benefits is mentioned for many sectors of the society, e.g., that open data in one year will save 1425 lives due to decrease in road fatalities, 629 million hours in traffic congestions, and 5.8 million tons oil equivalents in households.

(Blank, 2019) points out that European countries have recognized the potential of open data, but there are large variations in the maturity with respect to policy, infrastructure, impact, and data quality. To unlock the potential of open data, the right decisions must be taken, and good solutions must be implemented.

Digital innovation is the process of creating novel products, services, business processes, and business models using digital technologies (Yoo et al., 2010) (Fichman et al., 2014) (Nambisan et al., 2017). In this paper, we focus on digital product innovation that leads to digital products, both tangible and intangible (such as services), and investigate how open data can be best utilized in the digital product innovation process.

Open data can fuel digital innovations that are crucial both with respect to solving many of the societal challenges related to environment and inclusion, and for economic prosperity. The emergence of digital products can happen through evolutionary incremental growth, for example when new smart city services are to be developed, as described by (Abella et al., 2017). (Huyer & Knippenberg, 2020) provides many examples of how open data can contribute to e.g., energy saving measures, more sustainable energy use, and better travel services. Real-time travel services may for example, be developed by combining data on public transport (addressed by the ITS Directive) with data on parking, city bikes, charging for electric vehicles, car sharing, etc. The innovations, e.g., products such as Apps and systems, will be delivered by both public sector and commercial actors. They will combine data from many different sources. Real-time data streams e.g. emerging from

¹ Open Knowledge Foundation, Open Data Handbook, <https://opendatahandbook.org/guide/en/what-is-open-data/>

² <https://inspire.ec.europa.eu/>

Internet of Things (IoT) infrastructures (providing a variety of sensor data), will fuel even more innovations, including innovations based on artificial intelligence and big data processing (Sergey A. Yablonsky, 2019). The innovation may for example predict upcoming events and situation, facilitate pro-active measures to reduce negative impacts, assist contextual decisions, and support the handling of undesired situations.

This paper addresses open data from a user perspective, which according to (Zuiderwijk et al., 2014) is paid little attention to, *with focus on open data users involved in digital product innovation*. The aim is to give detailed insight into the problems they experience and to recommend measures that can mitigate the problems and maximize the usability and value of open data. The research questions answered by this paper are:

- RQ1: What barriers exist with respect to the use of open data in digital product innovations?
- RQ2: What measures should public data owners take to support the use of open data in digital product innovations?

The barriers are identified through a survey with data collection from a questionnaire and semi structured interviews among software developers and digital product innovators. The data analysis is qualitative and quantitative, and the results are analyzed in the context of digital innovation. Based on the findings, recommendations are provided on how the public sector should publish data.

2. Related work

The user perspective regarding open data is to our knowledge usually neglected when open data infrastructures and tools are established. This is also confirmed by the literature review provided by (Zuiderwijk et al., 2014). The study confirms that open data enable innovation in both the public and private sector, but the processes to create, open, find and use the data are however complex, and this complexity is often not taken into account. 118 barriers are identified in current literature regarding 1) availability and access, 2) find-ability, 3) usability, 4) understand-ability, 5) quality, 6) linking and combining data, 7) comparability and compatibility, 8) metadata, 9) interaction with the data provider, and 10) opening and uploading. In general, there is a lack of open data, the open data provision is unpredictable, and there are technical problems with released data sets (Janssen et al., 2012). More than 75% of open data users have concerns about the stability and quality of the data (C. Martin, 2014). There is also a need for feedback mechanisms that allow users to provide input on data quality and the usability (Janssen et al., 2012), and there is a need for a dialogue regarding data updates (S. Martin et al., 2013).

Several papers address the barriers related to the publishing of data. (Conradie & Choenni, 2014) requests better understanding of how internal conditions and processes in the public sector influence the publication of open data. The release of data is among others affected by fear of false conclusions being drawn from the data, the overhead related to the administration of licensing fees, concerns about privacy issues and copyrights restrictions, unknown data locations, and low prioritizing of the opening of data. The current literature does however not address how the data should be published to meet the challenges faced in digital product innovation.

Open data portals for access to open data are required to support the discovery of open datasets (Toots, M. et al., 2017). Such portals are established on a European level (see data.europa.eu), national level (e.g., data.norge.no and geodata.no, data.gov.uk, catalog.data.gov) and local levels (mainly larger cities), and open data is published through these portals.

Metadata is recognized as important for the discovery of open datasets, in for example, open data portals. Such portals, e.g., those based on the Comprehensive Knowledge Archive Network (CKAN)³ open source, in general provide mechanisms for metadata. Standard vocabularies describing datasets in data catalogues also exist, e.g., the Data Catalogue Vocabulary (DCAT)⁴, and the DCAT application profile (DCAT-AP)⁵ is used for data portals in Europe. The importance of metadata is also emphasized in a recent a study on transparency and portals, and a literature review identifies relevant aspects such as, among others, data quality, accessibility, APIs, documentation, and communication between data publishers and users (Lnenicka & Nikiforova, 2021). However, the use of and the quality of metadata and the open data is not sufficient. More than 75% think there is a lack of metadata, and about 75% think it is difficult to interpret the data (C. Martin, 2014). These problems are also widely recognized in other publications (Toots, M. et al., 2017). Usually, the meta data provided is incomplete or of poor quality, and there is for example usually no information on how the data originally is used, no meta data or context information supporting the interpretation of the data, and no information on data provenance. Metadata may however not be the sole solution. The understanding can be supported through visualization of the data (S. Martin et al., 2013), and tools have been developed to support such visualizations (Folmer et al., 2019).

The understanding of the data may also be affected by the interfaces to the data and data formats, especially when the users do not have detailed domain knowledge. Easy access to data (through APIs) is crucial. About 75% think that the interfaces provided are not user friendly (C. Martin, 2014), and the formats are often proprietary and messy (Toots, M. et al., 2017). Data fragmentation, which implies the need to combine data from several sources, also reduces the usability (Toots, M. et al., 2017).

The use of open data in the digital innovation process is as far as we know not described in the literature. The innovation process in general is however, widely addressed. (Taylor, 2007) describes the stages of the innovation life cycle. These stages are *strategy*, *design*, *transition*, *operation*, and *continual improvement*. (Smith & Sandberg, 2018) has studied the innovation barriers in the context of these stages and different open data user types to see how the user types play different roles in the innovation ecosystem based on open data.

A review article on the *digital* innovation process (Helmer et al., 2021) identifies six generic process steps that take part in a highly iterative and nonlinear cycle of divergent and convergent activities, repeated in an unpredictable way over time. The *opportunity identification* step identifies potential customers and their needs through market research, customer interviews, and studies of new trends. The *ideation and idea management* step is about the idea generation and scoping as well as idea

³ <https://ckan.org/>

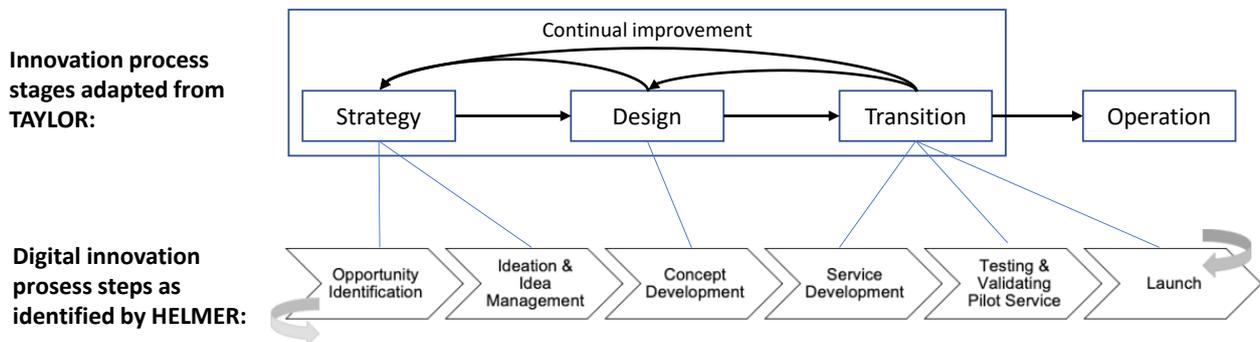
⁴ <https://www.w3.org/TR/vocab-dcat-2/>

⁵ <https://joinup.ec.europa.eu/solution/dcat-application-profile-data-portals-europe/release/11>

assessments, prioritizing and selection. In the *concept development* step, the idea is enhanced to a detailed concept that is described, selected, and verified. The *service development* step includes among others the software development, design activities, and the prototyping of a pilot service. The *testing and validating pilot service* step is about the installation and deployment of the developed software, as well as testing and validation. The *launch* step may include commercialization.

Figure 1 provides a mapping between the generic stages of the innovation process (Taylor, 2007) and the digital innovation process steps (Helmer et al., 2021). Continual improvement includes iterations with bug fixing, refinements, extensions, and initiation of new innovations. The link between the process steps and the use of open data is as far as we know not properly addressed in the literature, except for aspects regarding use of open data in early prototyping (Ferry et al., 2019), linked to the service development step.

Figure 1 Innovation process stages based on (Taylor, 2007) and the digital innovation process steps - subfigure from (Helmer et al., 2021)



3. Methodology

As described above, there have been studies on barriers related to open data, both from the government's and data user's point of view, and the main challenges are the availability, the quality, and the usability of the data. We conducted a survey as summarized in Table 1 to extend earlier findings with aspects of relevance to the digital product innovation process.

Table 1 Survey on barriers and drivers of easily usable open data – data collected and data analysis approach.

Steps and methods	Description	Result
Initial study: Literature study	Search for ("open data" OR "open governmental data") AND barriers in Scopus and ScienceDirect. Snowballing.	Input and knowledge on <ul style="list-style-type: none"> • State of the art and gaps • Open data in system development process • Challenges • Usability of open data • Effects of open data on innovation processes
Initial study: 2 interviews	Interview with representative from municipality organizing hackathon and researcher using open data.	
Initial study: Experiment and focus group	6 students used open data to develop a travel assistance application. Focus group to collect experiences.	

Initial study: Workshop	Initial use cases were refined in workshop with stakeholders.	Detailed use cases on <ul style="list-style-type: none"> • Data publishing • Data use
In depth study: Questionnaire 27 respondents	Questionnaire to ITS Norway members, members of relevant meetup groups, participants at hackathons and members of other relevant networks.	Input and knowledge on <ul style="list-style-type: none"> • Experiences from open data discovery and use • Effects of open data on innovation processes • Data user needs • Challenges to solve
In-depth study: 13 interviews	Interviews with developers and innovators in existing companies and start-ups and hobby developers.	
Analysis: Quantitative	Input collected through questionnaire (Likert scale approach) were used.	New knowledge
Analysis: Qualitative	Input collected through questionnaire and interviews were analyzed.	Answers to research questions

The study approached Norwegian data users. Thus, the focus is on data from Norwegian sources. However, some of the users also refer to experiences from their use of data from other countries.

3.1. Initial study

The initial study established the knowledge needed for the elaboration of the strategy for the in-depth study. It started with a literature study, and two semi structured interviews were carried out to get further insight. The first interview was with a representative from a large Norwegian municipality who in collaboration with an incubator had organized a hackathon with focus on mobility in 2016. The interviewee prepared and published data from the municipality and guided and supported the hackathon participants in the use of the data. The second interview was with a researcher doing concept and prototype development with extensive use of open data.

Input from the interviews guided the planning of the next step of the initial study - a system development experiment with use of open data. This was a setup with six students in the 4th year of the 5th years master's degree program in Computer Science at the Norwegian University of Science and Technology (NTNU). As a part of their mandatory "customer driven project", the researchers (the authors of this paper) played the customer role, and a traveler assistance application with novel functionality was requested. The links to some commonly known open data sources were provided, and the students were also encouraged to search for and use other open data, as the researchers' intent was to study the use of open data. The students' focus was however on system engineering tasks and on concept and application development. The use of open data was just a tool used in the system development of an application supporting people going hiking in mountains and forests⁶. A focus group with the students was organized after the end of the project according to the methodology defined by (Krueger, 2014).

The interview provided detailed insight into the barriers experienced and lessons learned. It was clear that open data can play an important role in early prototyping of digital products (Ferry et al.,

⁶ <https://github.com/OpenTransportDataProject/turplan>

2019). Detailed use cases for data publishing and use of open data in innovations were established, and these were verified and refined at a workshop with relevant stakeholders to cover aspects of relevance for different stages of the system development process.

3.2. In-depth study

The in-depth study focuses on the challenges faced when using open data in digital product innovation and software development. The insight and use cases established in the initial study guided the design of a questionnaire and semi structured interviews with respect to the topics addressed, the terminology used, and the answer alternatives when multiple-choice or Likert scale questions were used in the questionnaire.

The questionnaire was distributed through several digital channels where we knew we could reach users of open data. The main channels were the member network of ITS Norway (organizations and companies in the transport sector – software companies developing products within the transport domain included); a meetup group on open data in the Oslo region in Norway; and the networks of government organizations providing open data, among others the Norwegian Public Road Administration and the Norwegian Mapping Authority. Google forms were used for all these channels. In addition, respondents were also recruited among participants at hackathons which are popular events for users of open data. They received a paper version of the questionnaire.

To get more in-depth details on the use of open data, related challenges and wishes, semi structured interviews with software developers and innovators were carried out. The interviewees were recruited at hackathons aiming for the use of open data from the public sector in the development of digital innovations. Potential interviewees were asked if they had experiences with the use of open data, and only those that had been involved in the digital product innovation with open data were selected. Most interviews were carried out at the hackathons. A few were however just agreed upon and carried out via telephone after the hackathons. The overall questions asked that triggered the responses were: 1) What is your background with respect to education? 2) What is your experience with use of open data? 3) Which data have you used? 4) Which data have you searched for, but not been able to find? 5) What are the most important problems you have experienced regarding use of open data? 6) How has the use of open data influenced product and functionality ideas? The researchers also asked follow-up questions during the interviews to ensure that relevant issues were addressed.

One researcher carried out the interview whilst another observed, recorded the interview (audio recording and notes), and addressed missing issues. The AudioNote tool was used for the recordings, allowing notes to be linked to the timeline of the interview, and thus arranging a tight link between notes and statements in the audio recordings. This simplified the transcription and analysis of the interviews.

The questionnaire provided input to both quantitative and qualitative analysis, whilst the interviews were input to qualitative analysis. The qualitative input was analyzed and structured according to the use cases from the initial study. The input from the questionnaire and the interviews were

analyzed together since the issues addressed were overlapping. In general, the input from the interviews were much more detailed and provided better insight.

4. Results from in-depth study of user barriers

This section presents the input from the questionnaire respondents (from now on called respondents) and the interviewees. As described above, the questions asked were guided by use cases identified in the initial study (addressing topics such as find data, select data, access data, and use data). In addition, the effects of open data on the innovation capacity were also addressed.

4.1. Background of respondents/interviewees

To get an overview of the background of the respondents and the interviewees, they were asked to provide information on experiences and knowledge of relevance to the study. Table 2 summarizes the results. Most respondents/interviewees had relatively long experience with work of relevance to the use of open data, like programming and concept development, and they also had much or very much experience with the use of open data. The majority also claimed to have medium, much, or very much knowledge about the challenges related to the use of open data.

Most respondents/interviewees reported to have used open data in system development. However, use of open data in research, market surveys and information gathering related to system development were also mentioned. Most of the questionnaire respondents had used open data at work. All interviewees had used open data at hackathons (the interviewees were done at hackathons), but many of the interviewees had also used open data at work.

Table 2 Use of open data among questionnaire respondents (R) and interviewees (I)

Experience with and knowledge about open data	For what have open data been used? ⁷		Where have open data have been used?					
	R	I	R	I	R	I		
≥ 3 years with programming	59%	62%	System development	93%	69%	Hackathons or similar	37%	100%
≥ 3 years with concept development	60%	69%	Research	21%	0%	At work	81%	77%
Much/very much experience with open data	67%	59%	Market survey	0%	15%	During education	41%	23%
Medium/much/very much insight into open data problems and challenges	88%	64%	Information gathering	21%	8%	As a hobby	70%	38%

⁷ In total only 14 responders answered this question since it was added at a later stage.

4.2. The data of interest

In general, there is little knowledge on which open data is used in digital innovation and which open data is requested but not found. Thus, both respondents and the interviewees were asked about which data they had used and which data they have searched for but not been able to find. A long list of datasets is used, and an equally long list addresses datasets which are missing. When nothing else is mentioned, the data are Norwegian datasets.

4.2.1. Data used

In addition to meteorological data and data on climate issues, many different datasets on transport and traffic issues have been used. Real-time data from cars, city bikes and buses were used as well as data on public transport like timetables and routes. In addition, time, and space related characteristics of the traffic flow such as volume, speed, occupancy, congestions, driving conditions, incidents, accidents, travel times and traffic messages were mentioned, as well as statistics on road accidents and electric vehicles. The physical transport infrastructure was also addressed through data on the road network in general, and more specifically data on toll roads and bridges. Automatic identification system (AIS) data from the tracking of sea vessels were also mentioned.

Geographical data from among others the official mapping authority were commonly used, especially topographical land maps with different resolutions, data on hiking routes and tourist cabins, data on public toilets, place name data, zip code data (coordinates included), data on properties and buildings, data on warnings (flooding, avalanches, landslides, and rock avalanches), and coastlines. Data on different types of earth observations from satellites and satellite orbits were also used as well as satellite pictures. Other pictures from national registries were also used.

Human geographical data and statistics were mentioned, such as statistics on population density, salaries, divorces, criminality, export, import, education, and results from elections, as well as data on property value taxation, enterprises, and job announcements.

Data from municipalities were scarcely used except for the data collected by national agencies. The data used were data on economy, priorities, services, processing times, etc. In addition, data from the National Archives and Records Administration had also been used as well as data on news, art (paintings, pictures, movies, music, tec.), cultural monuments, and nature values.

The most common data sources in general were the official mapping authority, the public road administration, the statistics authority, the environment agency, and the agency responsible for public registries related to legal entities, enterprises, company accounts, etc.

The interviewees also reflected upon the data availability in general and stated that the availability of data is improving, but they also pointed out large variations between countries. In general, the availability of open government data was considered as good in Norway compared to other countries. Some countries do not provide open address data. The data users reported that they had to pay for important datasets from England, and some of them are very expensive. Providers of international services can get useful data in some countries, but in other countries they must collect their own data.

4.2.2. Data not found

It was reported that data on the climate and historical meteorological data were missing. For transport and traffic issues, the respondents/interviewees wanted data on variable signs, real-time data from buses, data on traffic flows and travel times for different transport modes, annual average daily traffic data, data on the use of bikes, traffic messages for pedestrians and cyclists, and historical data on delays for air transport and railways. They also requested data on train stations and routes, parking spaces, detailed road geometry, and real-time data on service vehicles doing snow clearing and road salting. Better access to data from the national registry of road vehicles was also mentioned. Today these data are available for free only via FTP or on a CD.

For geographical data, data on heights, maps with high resolutions, vector data maps, cadastral maps, relief models, and more specific place names were missing as well as air photos and lidar data. Data on properties and property boundaries, and details on buildings were also requested.

Human geographical data and statistics that were not found, were data on population density, ownership of cars, employment, jobs, age distribution, education, day care facilities and playgrounds for children, residence types, and criminality.

Data from municipalities were also requested, such as data on applications for day care places, shift schemes, economy, procurement, parking spaces for bikes, water supply and sewage, the condition of sewerage pipes, industry, and businesses. It was however pointed out that to ease the use of local data in innovations for a larger market, the data from different municipalities must follow the same format for each year and the data must also be harmonized and provided in a common way across different municipalities.

A few respondents/interviewees reported that they also were missing data on laws, regulations and principles of law, data from the national labor and welfare administration, data on advertisements, data on emergency responses carried out by the police, the emergency medical services and the fire brigade, and geographical information on emergencies.

It was pointed out that real-time data are of particular interest, preferably data streams related to geographical locations, areas, and distances. Some real-time data are available today but are needed for more locations. Real-time data of particular interest are data on traffic flows, travel time, the position of buses, parking space availability, and road maintenance (snow clearing, salting, etc.).

It is a paradox that some of the data reported as missing in fact were available when the surveys were carried out, but for different reasons the respondents/interviewees have not been able to find them. This is for example the case for population density, education, real-time data from buses, traffic flows and travel times.

4.3. Find and select data

During the work on the innovation idea and when the concept is defined, it is crucial to find and assess datasets that may include the data needed. This process is supported by metadata. In the

following, we present the questionnaire results, and input from the interviews is used to provide a more detailed picture.

4.3.1. Search data

Figure 2 Questionnaire results - How data users look for data.

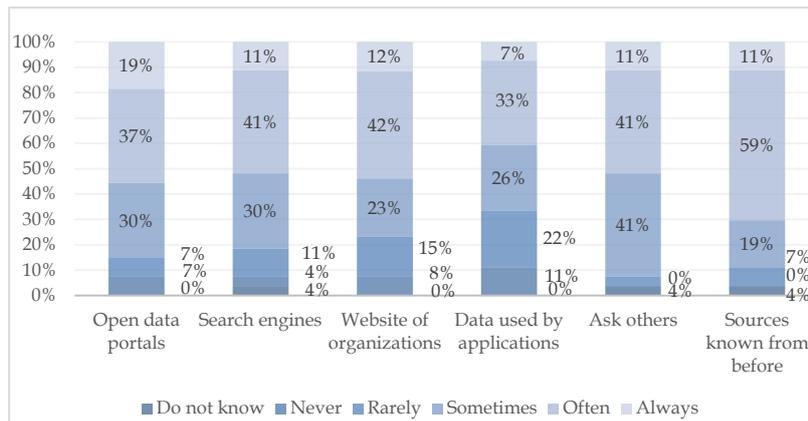


Figure 2 shows how the respondents use different approaches to find data. Most respondents use data sources they already know (70%). Official open data portals are also commonly used (56%). The Websites of data owners (54%) and search engines like Google (52%) are also popular, and the respondents also frequently ask each other about data sources (52%).

The interviews show that experienced and professional data users know the most common data sources. Users that are not very experienced quite often ask others for advice on where to find data, e.g., through an open data slack channel, and they also quite frequently use search engines.

Quite many respondents/interviewees were missing datasets that in fact are available. It is not always easy to find relevant data. Many interviewees explained that they did not know where to search, and the search terms used did not match the keywords and descriptions provided by the data publisher. Quite often, metadata and documentation are missing, of low quality or do not reflect the whole content of the dataset. Specific parts of the data content may not be addressed.

Interviewees requested one single entry point for all data of one kind at a national level, or preferably, for the whole of Europe. The situation varies between countries. In Norway, the availability of some types of governmental data in national portals is considered as quite good. One interviewee stated that in Germany, data are in general published for regions, and it is inconvenient for data users to visit all the different portals. In general, it is very challenging to find data from municipalities. Just a few of their datasets are published via national portals.

4.3.2. Criteria for selection of datasets

Figure 3 Questionnaire results – What criteria are important when you select and use data sources?

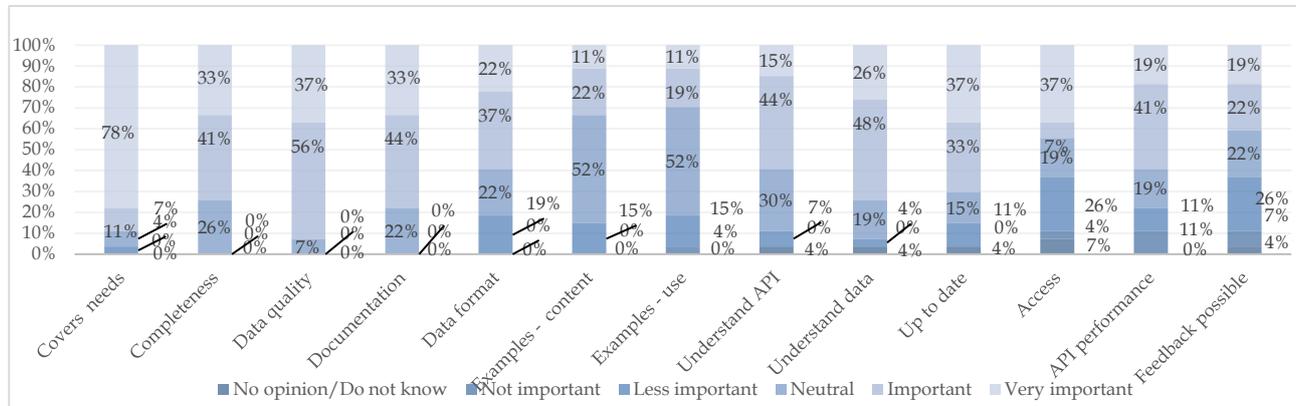


Figure 3 shows the questionnaire results with respect to what criteria the respondents emphasize when they select the datasets to use. The respondents reported that the data quality and the relevance of the data are the most important aspect the data users look for (93% and 89%), and completeness and up-to-date data were considered as important or very important by 74% and 70% of the respondents.

The interviewees explained that they must be able to quickly assess datasets to decide which they can use. It is however a challenge to judge the quality and relevance. They link relevance and quality to the completeness and freshness (up to date) of the data and claimed that old, outdated, and incomplete datasets in many cases may be just as bad as no data, and that data sources that are updated just once a year cannot be used. Address data should for example be used from official registers that are frequently updated. It was stated that data owners quite often do not have any incentives to update the data, and in many cases, they do not see the importance of updated data. It was addressed as a paradox that a one-time fulfillment of the Inspire Directive on the opening of public data in some cases is used as an excuse for no further actions since the Directive does not address the need for data updates. Some public data owners publish their data just once and do not have routines and mechanisms for provision of updated and quality assured data. It was observed that data owners that use their own open data in their own operations in general publish data of better quality than other data owners.

The documentation of the data content, and the ability to understand the APIs and data were considered as important or very important by 77%, 59% and 74% of the respondents. The provision of data content and use examples were however not that much emphasized (just 33% and 30%).

The interviewees stated that the quality and readability of the documentation, metadata and API documentation included, are crucial and affect the ability to understand the dataset. It was stated that the documentation should be in English since software developers and innovators are from different countries. The content of the data elements must be specified as well as the coverage (geographic area, time span, etc.). The assessment of a dataset is easier if representative data record examples are provided or if the whole dataset can be accessed and inspected.

The data formats used were considered as important or very important by 59% of the respondents. Some interviewees did however point out that metadata on the data format is useful but usually missing. Thus, it may be difficult to find datasets with a specific format. This is a problem since the data users quite often prefer specific formats and it may also be of interest to find data examples of a given format for training purposes. Some interviewees strongly preferred raw data rather than processed data. Raw data on earth observations may for example include clouds that are of interest to some data user, but the clouds may be removed from the processed data.

API performances are considered as important or very important by 60% of the respondents. The ability to provide feedback on the data was not considered as crucial at this early stage of the software development process. Just 41% of the questionnaire respondents considered this as important or very important. The access to the datasets through proper licenses was considered as important or very important by 44%. APIs and licenses are further addressed below.

4.3.3. Use of metadata

Figure 4 Questionnaire results – Which metadata are useful?

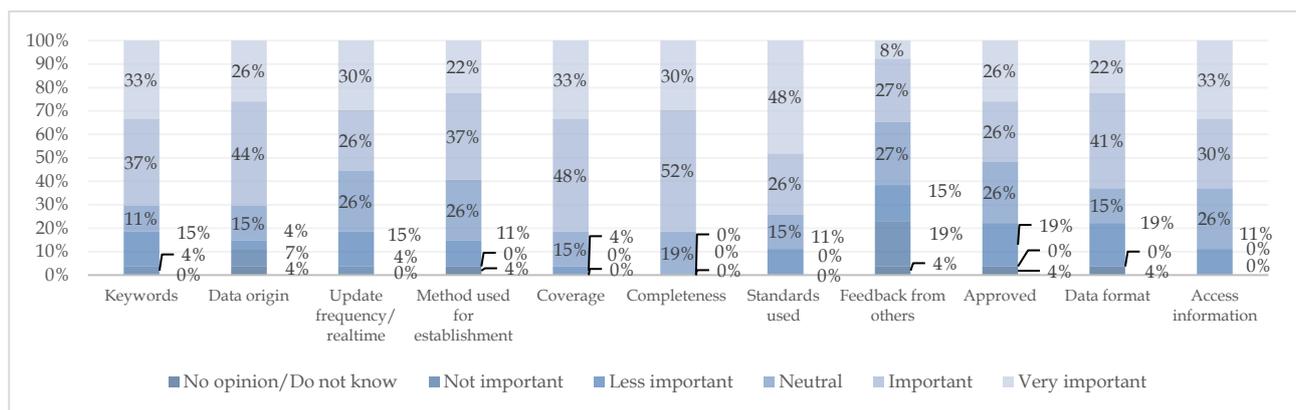


Figure 4 show the questionnaire results related to metadata. Focusing on the aspects that are considered as important or very important, the metadata on the completeness of the data (82%) and the coverage (81%) are in the lead. Coverage is about how well the dataset covers an area or time span, while completeness is the completeness within a coverage (e.g., all, the majority, or some/random samples). Metadata on the standards used to represent the data content is number three (74%). On the next places are keywords describing the dataset and the origin of the dataset (70% for both). More than half of the respondents considered the following metadata to be important or very important: the data format used (63%); access information such as licenses, availability, restrictions, and price (63%); the method used to collect the data (59%), the update frequency (56 %) and the approval of the data by a trusted party (52%). Third party feedback on the dataset is however not considered as important or very important (35%).

Many of the interviewees complained about missing metadata (e.g., on data origin and data format), metadata of low quality, and metadata that do not reflect the whole content of the dataset. This affects the ability to find relevant datasets and to assess datasets in an efficient way.

4.4. Access data

The access to open data is affected by technical solutions and restrictions as well as legal aspects like licenses. Both aspects will influence the innovation process.

4.4.1. Technical access

Figure 5 Questionnaire results – Are there restrictions to data access?

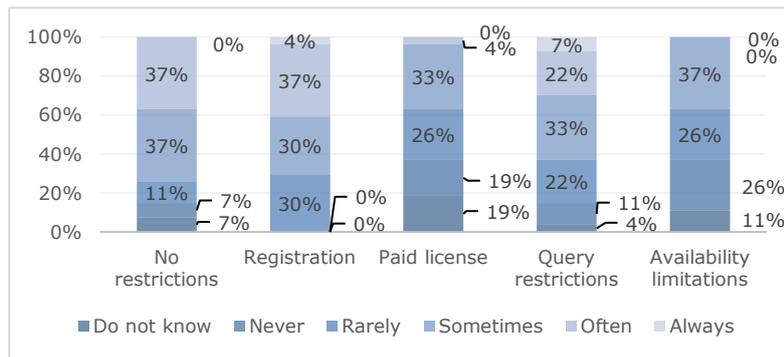


Figure 5 shows the questionnaire results on access restrictions. 41% of the respondents had always or often registered to get an access key to be able to access the data. 37% had however often experienced no access restrictions. 29 % had always or often experienced query restrictions, and additionally 33% had experienced restrictions sometimes. Data owners may for example have restrictions on the number of queries, and applications may be blocked out if they access data frequently. Availability limitations seems not to be very common, but 63% had experiences such limitations sometimes or rarely. Open data may for example not be available 24/7. Not many data users have experiences with paid licenses – probably because the survey focused on open data.

The interviewees pointed out that a registration and the assignment of access keys in some cases took several days and delayed the data assessment process. When professional data users have decided to use a dataset in a commercial service, the registration and use of access keys is however not considered to be a problem. The registration should however be on the company and not on an employee (person), which is common today. Such registrations of open data use arrange for communication between the data owner and the data user. The data user may provide information on how the dataset is used and wishes for future improvements of the dataset, and the data owner may provide information that is useful to the user, e.g., information on foreseen problems with the dataset, planned changes, and planned extensions of the datasets.

Several interviewees also reported problems with runtime access restrictions due to blocking, and that this might be discovered quite late in the system development process when several users are using the system simultaneously. In the worst case, such restrictions may put an end to innovations.

4.4.2. Licenses

The questionnaire did not focus on licenses, and as shown in Figure 2 and Figure 3, the access aspects and access information are in general not highly emphasized. The interviewees' view upon licenses varied a lot depending on their role.

Hobby developers and system developers just doing the programming had no or few concerns about licenses. Licenses were however one of the main concerns of interviewees with experience from or focus on commercialization. They stressed that data owners must emphasize on use of the right licenses. Many data owners are probably not aware of the importance of licenses and the effects they may have. The data users aiming for commercialization must be free to convey open data in value-added services without any restrictions. They can for example not commit to withdrawal of data that is provided to end users, as this may be technically impossible if for example the end users get data on DVDs for offline use.

The variety of license models was considered a problem. The administration of different conditions can be overwhelming when different datasets are combined into the same products, and when hundreds of data sources from many different countries are used. The national licenses for open access to open government data (e.g., NLOD⁸ in Norway) and the Creative Commons licenses⁹ CC0 and CC4 were considered as good. Some CC licenses have however put restrictions on commercialization. It is no problem to give attribution to the data owners, but unrestricted use of the data is a necessity.

It was said that if there is no open license, the data users must have contracts with the data owners. Some data users preferred to pay for datasets rather than having limitations in their use of data, and in some countries, they pay for datasets that are available for free in other countries. The prices are however to some extent set very high and do not always reflect the value of the dataset.

It was pointed out that the European ITS Directive states that transport related data must be open, and many countries have established portals for access to transport related datasets with CC licenses. The ITS Directive has however a self-declaration that by its vagueness causes uncertainty about the openness of the license and whether data users have the right to give away data, e.g., data that through processing and the combination with other data, have increased value.

4.5. Use data

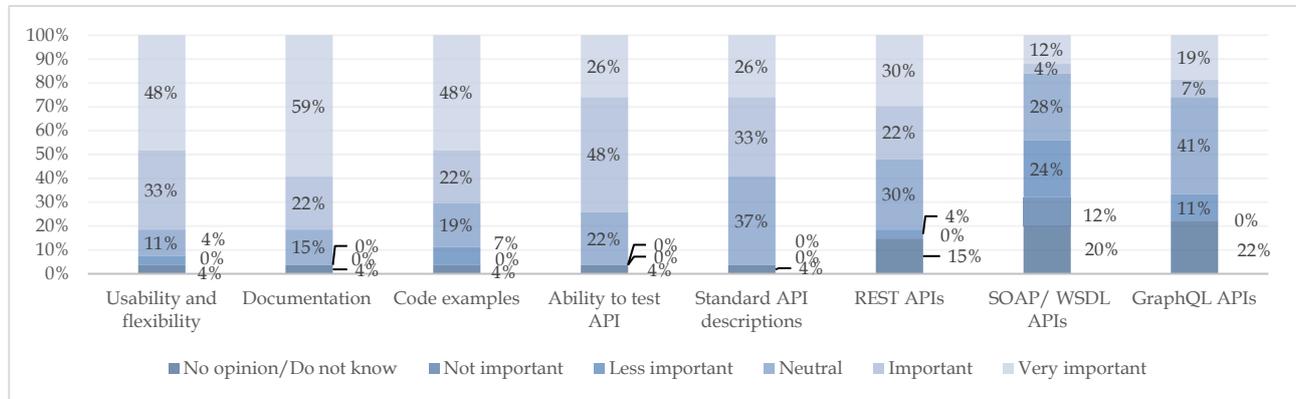
In the context of this paper, the use of data is twofold. The software developed must get the data needed, e.g., via APIs, and the actual data content must support the required functionality.

⁸ <https://data.norge.no/nlod/en/>

⁹ <https://creativecommons.org/licenses/>

4.5.1. APIs

Figure 6 Questionnaire result – What will make an API easy to use?



APIs are commonly used to access the data content. Figure 6 shows the questionnaire results addressing aspects that may make an API easy to use. 81% of the respondents reported that the usability and flexibility of the API are important or very important. This is about the need for data retrieval options that can be adapted to facilitate retrieval of the desired subset of data. The API documentation is equally emphasized (81%). Other important aspects are the ability to test the API (74%), the provision of code examples (70%), and use of standard API descriptions (59%). The respondents were also asked about which API type they preferred (REST API - 52%, SOAP/WSDL APIs - 16%, GraphQL APIs - 26%). Quite some respondents (28%-41%) were however, neutral, with respect to the API type.

One questionnaire respondent commented that the need for code examples and testing depends on the type of API used. For simple REST APIs, code examples and testing might not be required, but for more complex APIs (e.g., SOAP APIs), code examples and testing abilities are crucial since it is a lot of work to implement the API queries. A visualization of the data content returned by the APIs was also recommended, to support the understanding of what the API will return.

The interviewees explained that the technical solutions provided by the data owners are important for the usability of open data. The data users prefer datasets with easy access that do not require much programming and adaptation. Experienced users tend to prefer APIs. A few interviewees did however, prefer to download the whole dataset. Such an approach is however, just relevant for static data. APIs are required for real-time data and data that is frequently updated.

According to the interviewees, APIs are quite often not adapted to the user needs. An API may, for example, require data records to be requested one by one and not by selection criteria. APIs may also retrieve much more data than needed and cause a lot of work on data filtering. Thus, rather than one single API, there should be several optional APIs adapted to different data needs.

The interviewees stressed that APIs must be documented, preferably in the same way. Some had good experience with the standardized documentation supported by Swagger¹⁰. All parameters used in an API must be specified. API examples with links where the retrieved datasets can be inspected are also useful.

REST APIs are preferred by many but may deliver too much data and cause a lot of data filtering. Thus, some interviewees consider the GraphQL API a better option with support for queries for the retrieval of specific data elements. One interviewee also requested more frequent use of cross-origin resource sharing (CORS) to facilitate the direct use of open data from webpages.

4.5.2. Data content

The interviewees stressed that the quality of the dataset documentation is important when a dataset is used. It was also mentioned that quick start tutorials and examples showing the use of the dataset in different programming languages are useful. In comparison, the respondents considered the documentation of the data content important (77% in Figure 2) as well, but they did not emphasize the importance of data content examples (33% in Figure 2).

The interviewees have different preferences with respect to data formats. Many prefer JSON, CSV and XML dialects, but more specific formats for different data types were also mentioned. An unfamiliar or proprietary format is in many cases a reason for not using a dataset, and some interviewees complained about the need for programming to filter and transform the data content. Professional data users however, have data transformation tools.

Just one interviewee had experience with RDF¹¹, which is promoted as the solution for open, linked data. The interviewee did however, consider the quality to be too low, due to problems with the data links.

4.6. Innovation capacity

Open data is intended to fuel innovations. Thus, open data may be relevant when the innovation idea and the concept are developed, as described by the three first steps in the digital innovation process described by (Helmer et al., 2021) (see Figure 1). Thus, the interviewees were asked to describe how open data affect their innovation process.

4.6.1. Market analysis

Open data used in the market analysis does not necessarily have to be machine readable. Interviewees explained that they used statistics of all types, e.g., statistics on financial issues, socioeconomic issues, market developments, societal issues, etc. in analyses of potential user needs and market shares.

¹⁰ <https://swagger.io/>

¹¹ <https://www.w3.org/RDF/>

4.6.2. Idea and concept development

With respect to digital products, the knowledge of existing and new open datasets triggers ideas regarding both new products and improvements of existing products. Some interviewees got ideas from presentations given by data owners as well as examples of data use, tutorials, videos, etc. They reported that quite often, product ideas had been changed or skipped because data were missing. Some interviewees have in such cases established contact with data owners to influence their opening of relevant datasets or to get an agreement on access to the data.

The value of open data was illustrated by an interviewee from an international company. Usually, such companies first release new products in countries with large markets. Norway is a small market. However, due to the availability of good quality open governmental data with good coverage, some products were developed first and released for the Norwegian market.

All interviewees with commercial interests proclaimed that data from municipalities cannot be used, due to the limited market size of one municipality. The development of products for several municipalities is too demanding since the datasets are not harmonized and since the data are published via different channels.

Several interviewees have experienced severe problems such as loss of functionality and reduced quality of service, loss of customers and other problems due to unannounced changes in datasets and even removal of datasets. The interviewees requested announcements of both changes in datasets and future releases of new open datasets in advance to facilitate early planning, adaptations of existing products, and development of new concepts. Interviewees have also experienced that public data owners have killed new business opportunities by releasing unannounced products, based on their own data. In such cases, it is very difficult to compete with the products provided by the data owners.

To mitigate problems with data availability, data quality and API usability, some interviewees have frequent contact with data owners to ensure the availability of data that is useful in innovations. The data users provide feedback on existing data (data quality, etc.) and input on needs regarding data and APIs. Such dialogues are however not formalized, and it is mainly the professional and well-established companies and not small start-ups that have such dialogues with data owners.

5. Discussion and contribution

In this section, the findings in Chapter 4 are discussed and the knowledge gained is summarized. With a few exceptions, the context for the use of data and the discussions are for use of Norwegian data sources. The findings are however assumed to be of general value.

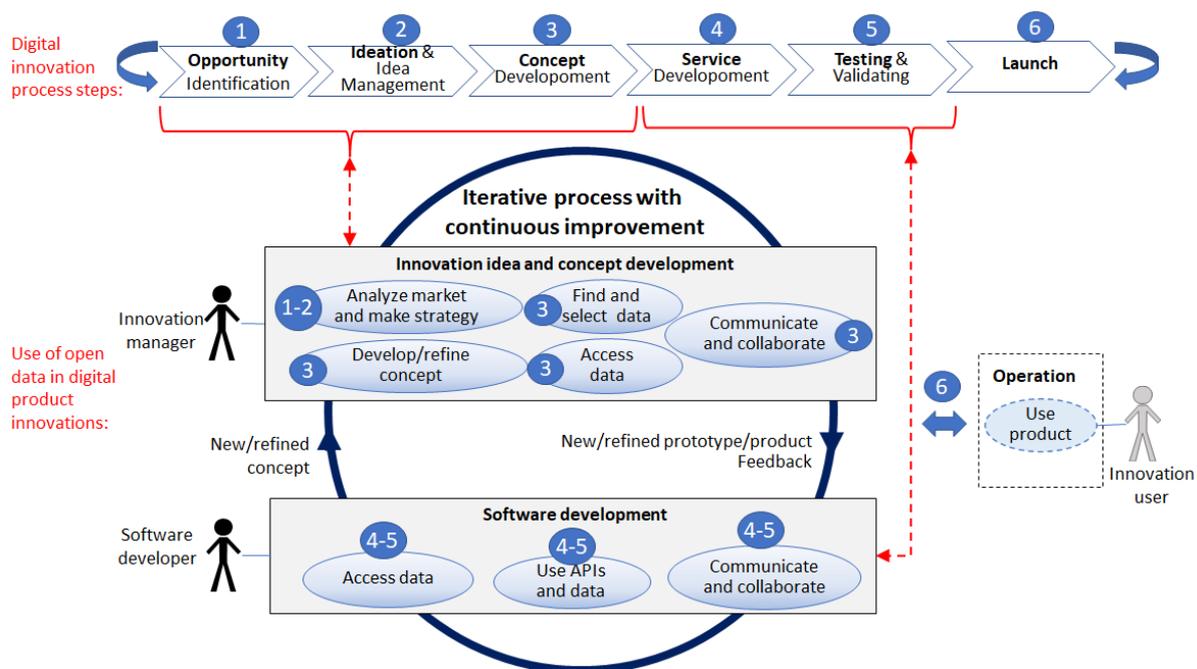
As the interviewees reflected upon their use of open data, we gained a deeper understanding of their knowledge. In some cases, it became clear that the self-assessment of their experience and insight, as described in 4.1, did not always match their real insight into open data issues. Thus, the background information cannot be used to weight the responses from the questionnaires, and we

have not analyzed the correlation between the self-reported experience and insight of the questionnaire respondents and their responses.

5.1. Open data in the digital innovation process

Based on the use cases created in the initial study and the results from the in-depth study described in Chapter 4, we suggest the overall use cases (the ovals) for use of open data in digital product innovations depicted in the lower part of Figure 7. The links towards the digital innovation process steps identified by (Helmer et al., 2021) are also illustrated (in the upper part of the figure). The process steps are further described in Chapter 2

Figure 7 Digital innovation process steps (Helmer et al., 2021) and use of open data.



The interviews contributed to the identification of two focus areas - innovation management and software development. Thus, we identify two generic roles that are non-overlapping and artificial archetypes - the innovation manager and the software developer. The innovation manager is responsible for the innovation idea and concept development. The software developer transforms the concept into a digital product through use of open data in the software development. Due to the iterative process, both roles cover both the initial round and the continuous improvement rounds done in later stages.

The roles are non-overlapping constructs. This makes it possible to define use cases with well-defined and non-overlapping responsibilities. In real life cases, the roles may be covered in different ways by one or more persons, and the same person may also take both roles. The latter is relevant in smaller companies. In larger companies, however, those doing the software development usually just cover the software developer role. (Smith & Sandberg, 2018) identified three archetypes with partly overlapping responsibilities - employees, entrepreneurs, and hobbyists. The employee is quite

like the software developer. The entrepreneur is like the innovation manager but may also cover parts of the software developer. The hobbyist has most in common with the software developer but must also cover parts of the innovation manager role. Dialogue with data publishers and license issues are for example not emphasized by the hobbyist.

The use cases illustrated as ovals in Figure 7 represent tasks carried out by the innovation manager and the software developer roles. The digital product operation and the innovation user are also included to provide a total picture but not addressed in the following. The numbered labels link the use cases to the associated process steps and show how the use of open data is connected to the digital product innovation process. Open data may affect all steps before launching.

Table 3 describes the use cases linked to the innovation manager and the software developer, and the steps in the innovation process corresponding to the use case described in the context of open data, based on the results in Chapter 4. The innovation manager is professional in finding and assessing the data usability. The innovation manager will not use datasets unless they have open licenses, and the licenses must be similar for all datasets used to simplify and ensure a proper handling of the conditions for use. In case of problems of any kind, the innovation manager may contact the data owner. Through communication and collaboration, they may handle issues regarding data access and quality, licenses, and service level. The innovation manager may also invite data owner/publisher to collaboration regarding API definitions. The actual collaboration on API definitions is however done by the software developer, since detailed technical knowledge might be required.

As for digital innovation in general, the process with the use cases is iterative and will include continuous improvements, bug fixing, extensions, and initiation of new innovations. The concept will guide the software development, and feedback from the software development is used to refine the concept. The software developer may for example detect that the data quality is too low. In such cases, data alternatives must be found, and the concept may have to be refined. As an alternative, actions may be taken to improve the data through communication and collaboration with the data publisher. Announcements or availability of new open data may also trigger new iterations.

Table 3 Use cases and roles with inks to digital innovation process steps

Use case	Software developer and steps	Innovation manager and steps
Analyze market and make strategy	NA	Opportunity (1): Open data, e.g., statistics, supports analysis of needs, markets, and opportunities. Ideation (2): The availability of open data, and possible combinations of data, may trigger overall ideas/strategy on what the innovation should address.
Access data	Service (4): Will in general use data selected by innovation manager, but may access additional data to investigate opportunities:	Concept (3): Relevant data is assessed: <ul style="list-style-type: none"> • For data assessment: Prefers no registration for quick access.

	<ul style="list-style-type: none"> • May register to access data but prefers that access is handled at company level. 	<ul style="list-style-type: none"> • When use of data is decided: Prefers registration to facilitate input from data publisher.
Find and select data	<p>NA</p> <p><i>Will use the data alternatives identified and approved by the innovation manager.</i></p> <p><i>No or little concern about licenses. Assumes they are handled at a company level.</i></p>	<p>Concept (3): Has knowledge on where and how to search. Quality and usability of the data, APIs and service level are assessed.</p> <ul style="list-style-type: none"> • Needs metadata and documentation for quick assessment of relevance. • License and service level may be showstoppers. • Prefers contracts/payment to unclear/not open licenses.
Develop/refine concept	<p>NA</p> <p><i>Concept is developed by innovation manager.</i></p>	<p>Concept (3): Data availability affects the concept. The use of data is planned. Modifies/refines concept if software developer reports problems and in case of new opportunities.</p>
Communicate and collaborate	<p>Service (4) / Testing (5): May in case of problems contact the data publisher or user communities regarding:</p> <ul style="list-style-type: none"> • Use of APIs • Use of data <p>May collaborate with data publisher on new API/API improvements (initiated by innovation manager).</p>	<p>Concept (3): May contact data owners/publishers to agree on:</p> <ul style="list-style-type: none"> • Access to un-published data • Licenses/contracts • Improved data quality/Service level • APIs modifications • Collaboration on API definitions <p>Opportunity (1)/Ideation (2) - for continual improvement: Needs early notifications on data updates, changes, planned products, etc.</p>
Use APIs and data	<p>Service (4) / Testing (5): Affected by:</p> <ul style="list-style-type: none"> • Data/API documentation • Data/API quality • Data/API usability • Service level <p>Feedback to innovation manager on problems with the selected data.</p>	<p>NA</p> <p><i>Will receive information from the software developer on data problems.</i></p>

5.2. Barriers towards use of open data in digital product innovation

Table 4 lists the topics where barriers are experienced when open data is used and the barriers, as reported in Chapter 4. The process steps affected are also indicated, based on the steps provided by (Helmer et al., 2021) (see the upper part of Figure 7). The two rightmost columns of the table

indicate which of the two roles, the Innovation manager (I) and the Software developer (S), that the barriers mainly affect, according to our analysis.

(Smith & Sandberg, 2018) also identify open data barriers, and our findings confirm their results. Compared to their work, we provide additional details on barriers regarding licenses, service level, transparency, documentation, API design, and data quality. We also identify new barriers regarding regulations, publishing channels and infrastructures, and standardization and harmonization.

Table 4 Barriers experienced in digital product innovation with open data and affected roles.

Topics	Process steps	Barriers experienced	I	S
Licenses	1 Opportunity 2 Ideation 3 Concept	No license specified. Datasets with no licenses cannot be used since there is no clear conditions regarding how the data can be used.	x	
		Many different licenses. The license management becomes very complex when datasets with different licenses are combined into one product.	x	
		Restrictions. Licenses put restrictions on how the data can be used. Such licenses may prohibit the use of the open data in digital products.	x	
Technical access	2 Ideation 3 Concept 4 Service	Delays. Registration causes delays during the assessment of datasets.	x	x
Service level	2 Ideation 3 Concept 4 Service 5 Testing	Limited availability. Data is not available 24/7.	x	
		Blocking. If frequent requests are performed, data access may get blocked without notice.		x
		API performance is too low. This may slow down the response time.	x	x
Transparency and predictability	1 Opportunity 2 Ideation 3 Concept	Data user is unknown. With no registration of data users, the data publisher cannot push information to data users.	x	
		No announcement of data changes. Changes in datasets and removal of datasets without any pre-announcements generate problems.	x	
		No announcement of new services. New services from data owners, based on their own open datasets, may compete with the products of innovators. Such services are not announced in advance.	x	
Communication and collaboration	1 Opportunity 2 Ideation 3 Concept 4 Service	No dialogue. Manual contact is required. Lack of mechanisms for feedback and dialogue on data needs and quality.	x	x
Publishing channel and infrastructure	1 Opportunity 2 Ideation 3 Concept 4 Service	Unknown channels. Potential users do not know where to search for data, and many users do not find the portals.	x	
		No single entry point. Many datasets are not registered in national portals, and there are several portals.	x	
		Data infrastructure limitations. Capacity and service levels are not sufficient.	x	x
Metadata	1 Opportunity 2 Ideation	Poor metadata quality. The quality of the metadata is in general low.	x	

	3 Concept 4 Service	Dataset assessment takes time. It takes time and may be difficult to understand the documentation. The data must be retrieved and inspected.	x	x
		Difficult to find data. The user must know exactly what to search for.	x	
		Quality assessments not supported. Metadata on provenance, coverage, update frequencies and service levels are quite often not provided.	x	
		Data selection not supported. Metadata on data format, use of standards, and license are quite often not supported.	x	x
Documentation	1 Opportunity 2 Ideation 3 Concept 4 Service	Insufficient API documentation. For complex APIs. Examples are needed.		x
		Insufficient dataset documentation. It is not always easy to understand the documentation, and the technical quality is not good enough. Examples are needed.	x	x
		Documentation language. English should be used to support all users.	x	x
API design	4 Service	APIs are too generic. APIs return too much data.		x
		APIs are not flexible. They are not adapted to user needs and may return records one by one instead of supporting selection criteria.		x
Data quality	3 Concept	Data is outdated. Data is published just once or not frequently updated.	x	
		Low coverage. Data coverage is not complete.	x	
Regulation	2 Ideation 3 Concept	Unclear regulations. The interpretation of European directives may be difficult. The need for data updates is not always emphasized.	x	
Harmonization and standardization	3 Concept 4 Service	Datasets are not harmonized. The same data is published in several different ways (e.g., data from municipalities).	x	x
		Not standardized metadata.	x	x
		Not standardized API documentation.		x

As indicated in Chapter 2, many barriers for use of open data in general are described. Our study links the barriers to the steps in the digital innovation process, and we also extend previous findings with details specific to digital product innovation. Poor metadata and data quality is previously addressed, e.g. by (C. Martin, 2014). Table 4 lists more specific barriers linked to data quality and to metadata for assessment of quality and data selection. Documentation and communication with data providers in general is addressed by (S. Martin et al., 2013) and (Zuiderwijk et al., 2012). The digital innovation process is highly sensitive to the technical quality of the documentation of both data and APIs. In addition, license and service level barriers may be showstoppers. During operation, the technical solutions for data provision and issues on performance are common problems. The lack of suitable APIs may also be a blocker. Eventually, the relevance of digital products may be reduced if the data is outdated, and the product may in the worst-case stop working in case of unannounced changes or removal of data.

It is also important to help the data user to discover relevant data. Thus, the search tools must be improved. One option may be to support semantic search for data, as suggested by (Jiang et al., 2019).

5.3. Data publishing recommendations to facilitate innovation

To increase the innovation capacity, the public sector must publish open data that supports innovation of new digital products. Table 5 provides recommendations on how the barriers in Table 3 should be addressed by data publishers, taking the needs of the roles in Table 4 into account. The recommendations should be used when the open data policies and strategies are established, and in the actual opening processes.

Table 5 Recommendations to be addressed by open data policies in the public sector.

Topics	Recommendations mitigating experienced barriers
Licenses	<ul style="list-style-type: none"> Use the same open license on all datasets – preferably national licenses for open data or Creative Common licenses supporting free use of data (CC0 or some variants of CC4).
Technical access	<ul style="list-style-type: none"> Offer two types of access: Access with no registration to facilitate quick data assessments; and user registration (companies and persons) to facilitate dialogue when data is used in operative services.
Regulation	<ul style="list-style-type: none"> European directive should emphasize quality requirements such as the need for relevant metadata, documentation, and frequent updates of the data content.
Service level	<ul style="list-style-type: none"> 24/7 operation with good response times and no blocking is needed. If this cannot be offered for free, a paid, reliable service is an alternative. Provide information on support mechanisms and point of contact as metadata. Publish service level agreements (SLAs).
Transparency and predictability	<ul style="list-style-type: none"> Provide overview of datasets that can be opened (if they are requested). Provide early announcements on new datasets and changes in datasets. Avoid unexpected competition. Make an open strategy on service offerings based on own data and provide early announcements on coming services.
Communication and collaboration	<ul style="list-style-type: none"> Provide mechanisms for interaction with data users in all stages of the innovation process – for transparency, feedback/requests from users with answers, and support from API/data experts. Establish procedures for the processing of input/feedback from users. Facilitate a community of data users and mechanisms for dialogues on data needs, data quality and APIs. Invite users to collaborate on data content/quality and API design.
Publishing channel and infrastructures	<ul style="list-style-type: none"> Publish metadata via one single national portal and arrange for the discovery of the data through metadata of good quality (see below).

	<ul style="list-style-type: none"> • Provide infrastructure for the publishing of the data content with sufficient scalability and service level.
Metadata	<ul style="list-style-type: none"> • Define guidelines on use of metadata (based on DCAT-AP) and emphasize the importance of metadata quality. Make certain metadata mandatory, e.g., keywords, update frequency, license, service level, provenance, data coverage (e.g., location and time span), and data format. • Provide metadata on point of contact for dialogue with dataset expert.
Documentation	<ul style="list-style-type: none"> • Define standards and guidelines for documentation of the data content and APIs. Emphasize the importance of documentation quality. Consider using standard API descriptions like OpenAPI/Swagger.
API design	<ul style="list-style-type: none"> • Define guidelines for API strategies emphasizing that several API alternatives should be provided to support different needs. • APIs must be of good quality (good performance, relevant, etc.).
Data quality	<ul style="list-style-type: none"> • Define data quality assurance procedures and checklists for data publishing. Use open data in own organization as an alternative to integrations between internal systems.
Harmonization and standardization	<ul style="list-style-type: none"> • Data from municipalities should be harmonized and if possible standardized. • Metadata standards such as DCAT-AP should be used. A more standardized terminology for metadata content should also be provided to ease and harmonize the provision of metadata content. • Open API/Swagger should be used to document APIs.

The recommendations partly overlap with recommendations provided by others, e.g. regarding metadata on coverage (Neumaier & Polleres, 2019) and the need for communication and user community (Benitez-Paez et al., 2018). Further details are however added, based on input from the interviews. Communication and collaboration with the data publishers are of special importance for professional users of open data. The data user should be encouraged to communicate how the dataset is used and wishes for future improvements of the dataset, the need for data quality improvements, API improvements, new datasets, etc. The data owner should provide information that is useful to the user, e.g., information on problems with the dataset, planned changes, planned extensions, etc. The public sector must also be transparent and communicate which data they have, plans for the opening of data, and which data that can be opened. The public sector must also be open about their own plans for new digital products, e.g., apps, based on their own data, to avoid unintended competition with innovators. Data users also like to discuss with other data users. Thus, the public sector should consider facilitating digital user communities.

Regarding regulation barriers, European directives should emphasize quality requirements such as the need for relevant metadata, documentation, and frequent updates of the data content.

5.4. Limitations

The inputs from respondents and interviewees reflect their perceived situation regarding the use of open data in a software development context. We have however, not verified their statements, so

we do, for example, not know the exact share of open datasets that require a registration by the users, and we have not monitored real data selection processes. Many of the detailed barriers reported, e.g., regarding metadata quality, are however confirmed by the overall findings in other studies.

The number of respondents and interviewees is not very high. Much of the information received, especially through the interviews, does however provide deep insight into their use of open data. The roles covered by the respondents and interviewees were broad, and some had a very broad experience. Together they highlighted the use of open data for different roles, and in different stages of the digital product innovation process.

Sometimes, the answers from the respondents and interviewees do not match. Some interviewees emphasized issues that are not ranked as the most important issues among the respondents, e.g., the access to API examples. One reason is probably the questions asked. The questionnaire asked for a ranking according to importance. The interviews talked more freely about their challenges and experiences.

6. Conclusion and further work

Existing literature addresses overall barriers and recommendations linked to open data. This paper takes this work further to a more detailed level, focusing on the use of open data in digital product innovation. An in-depth study was carried out through a questionnaire and semi structured interviews, with users of open data. The users were asked to provide input on which data they have used, which data they were missing, and their experiences with use of such data in the development of digital innovations. The overall findings are twofold: 1) The availability of open data affects the innovation capability, and 2) the way data is published affects the usability of the data and thereby also the innovation capacity.

As an answer to research question RQ1, the challenges and barriers with respect to the use of open data in digital product innovation are identified and linked to the digital innovation process. Data are published, but still more open data are requested, especially real-time data. It is a paradox that published data cannot be found, and that the data cannot be used due to access, service level and quality limitations. Many of the barriers are common for open data in general. Some barriers are however more specific to digital product innovation, e.g., the difficulty to combine data from datasets with different licenses into one product; challenges regarding quick assessments of datasets in the early stages of the innovation process; unannounced changes and removal of datasets causing problems for existing products; lack of metadata supporting quality assessments; the dependency on a high service level; the need for detailed, technical documentation of data formats, API and data content; and the lack of flexible APIs.

As an answer to research question RQ2, the paper discusses how to mitigate the challenges regarding the use of open data in digital product innovation, and a set of recommendations to the data publishers are provided. The data owners must take responsibility to ensure that data of good qual-

ity are published in a way that supports innovations. This includes use of licenses suited for commercial products, and metadata and documentation supporting quick assessments of datasets and use of the data in software development. There is also a need for more dialogue between data users and data owners. Data publishers must encourage and arrange for feedback on quality issues, data needs and APIs, and they must be transparent with respect to plans and strategies regarding data publishing and use of data in their own products.

The discussions of barriers and the recommendations provided by this paper target publishers of open data that are serious about supporting innovation. Future research should address how the recommendations should be followed in practice through an integration of data publishing in the working processes. The collaborations between data users and data publishers must also be addressed, e.g., mechanisms supporting the dialogue, and co-creation of innovation ideas to identify concrete data needs. Local data from municipalities are needed in many future digital product innovations for smart cities and societies. To arrange for sustainable service and product developments with large markets, the opening of local data must be coordinated. Innovators should get access to the same datasets via similar APIs from several municipalities.

References

- Abella, A., Ortiz-de-Urbina-Criado, M., & De-Pablos-Heredero, C. (2017). A model for the analysis of data-driven innovation and value generation in smart cities' ecosystems. *Cities*, 64, 47–53. <https://doi.org/10.1016/j.cities.2017.01.011>
- Benitez-Paez, F., Comber, A., Trilles, S., & Huerta, J. (2018). Creating a conceptual framework to improve the re-usability of open geographic data in cities. *Transactions in GIS*, 22(3), 806–822. <https://doi.org/10.1111/tgis.12449>
- Berends, J., Carrara, W., Engbers, W., & Vollers, H. (2020). *Re-using Open Data: A study on companies transforming Open Data into economic & societal value*. European Commission.
- Blank, M. (2019). *Open data maturity report 2019*. European Commission.
- Conradie, P., & Choenni, S. (2014). On the barriers for local government releasing open data. *ICEGOV 2012 Supplement*, 31, S10–S17. <https://doi.org/10.1016/j.giq.2014.01.003>
- Ferry, N., Omerovic, A., & Natvig, M. K. (2019). Towards Early Prototyping of Services based on Open Transport Data: A Feasibility Study. *Proceedings of the 9th International Conference on Cloud Computing and Services Science - Volume 1: CLOSER*, 257–262. <https://doi.org/10.5220/0007675402570262>

- Fichman, R. G., Dos Santos, B. L., & Zheng, Z. (Eric). (2014). Digital Innovation as a Fundamental and Powerful Concept in the Information Systems Curriculum. *MIS Quarterly*, 38(2), 329-A15. JSTOR.
- Folmer, E., Beek, W., Rietveld, L., Ronzhin, S., Geerling, R., & den Haan, D. (2019). Enhancing the usefulness of open governmental data with linked data viewing techniques. *Proceedings of the 52nd Hawaii International Conference on System Sciences*.
- Helmer, J., Huynh, T.-M.-T., Łobacz, K., Kör, B., & Wakkee, I. (2021). Innovating digitally for services: A review of innovation process literature focused on digital innovation and service innovation. *Knowledge-Based and Intelligent Information & Engineering Systems: Proceedings of the 25th International Conference KES2021*, 192, 2797–2806. <https://doi.org/10.1016/j.procs.2021.09.050>
- Huyer, E., & Knippenberg, L. van. (2020). *The Economic Impact of Open Data: Opportunities for Value Creation in Europe*. European Commission.
- Janssen, M., Charalabidis, Y., & Zuiderwijk, A. (2012). Benefits, Adoption Barriers and Myths of Open Data and Open Government. *Information Systems Management*, 29(4), 258–268. <https://doi.org/10.1080/10580530.2012.716740>
- Jiang, S., Hagelien, T. F., Natvig, M., & Li, J. (2019). Ontology-based semantic search for open government data. *2019 IEEE 13th International Conference on Semantic Computing (ICSC)*, 7–15.
- Krueger, R. A. (2014). *Focus groups: A practical guide for applied research*. Sage publications.
- Lnenicka, M., & Nikiforova, A. (2021). Transparency-by-design: What is the role of open data portals? *Telematics and Informatics*, 61, 101605. <https://doi.org/10.1016/j.tele.2021.101605>
- Martin, C. (2014). Barriers to the Open Government Data Agenda: Taking a Multi-Level Perspective. *Policy & Internet*, 6(3), 217–240. <https://doi.org/10.1002/1944-2866.POI367>
- Martin, S., Foulonneau, M., Turki, S., & Ihadjadene, M. (2013). Open data: Barriers, risks and opportunities. *2013 13th European Conference on EGovernment (ECEG)*, 301–318.

- Nambisan, S., Lyytinen, K., Majchrzak, A., & Song, M. (2017). Digital Innovation Management: Reinventing Innovation Management Research in a Digital World. *MIS Q.*, 41(1), 223–238.
- Neumaier, S., & Polleres, A. (2019). Enabling Spatio-Temporal Search in Open Data. *Journal of Web Semantics*, 55, 21–36.
- Sergey A. Yablonsky. (2019). Multidimensional Data-Driven Artificial Intelligence Innovation. *Technology Innovation Management Review*, 9(12). timreview.ca/article/1288
- Smith, G., & Sandberg, J. (2018). Barriers to innovating with open government data: Exploring experiences across service phases and user types. *Information Polity*, 23(3), 249–265. <https://doi.org/10.3233/IP-170045>
- Taylor, S. (2007). *The Official Introduction to the ITIL Service Lifecycle*. Stationery Office.
- Toots, M., McBride, K., Kalvet, T., & Krimmer, R. (2017). Open Data as Enabler of Public Service Co-creation: Exploring the Drivers and Barriers. *2017 Conference for E-Democracy and Open Government (CeDEM)*, 102–112. <https://doi.org/10.1109/CeDEM.2017.12>
- Union, E. (2010). Directive 2010/40/EU of the European Parliament and of the Council. *Official Journal of the European Union*, L 207.
- Union, E. (2019). Directive (EU) 2019/1024 on open data and the re-use of public sector information. *Official Journal of the European Union*, L 172.
- Yoo, Y., Henfridsson, O., & Lyytinen, K. (2010). Research Commentary: The New Organizing Logic of Digital Innovation: An Agenda for Information Systems Research. *Information Systems Research*, 21(4), 724–735. JSTOR.
- Zuiderwijk, A., Helbig, N., Gil-Garcia, J. R., & Janssen, M. (2014). *Innovation through Open Data-A Review of the State-of-the-Art and an Emerging Research Agenda: Guest Editors@ Introduction*.
- Zuiderwijk, A., Janssen, M., Choenni, S., Meijer, R., & Alibaks, R. S. (2012). Socio-technical impediments of open data. *Electronic Journal of E-Government*, 10(2), 156–172.

About the Author

Marit K. Natvig

Marit K. Natvig works as a senior researcher in SINTEF, Norway. She has a background from software engineering and her research areas are ICT for smart cities and societies, software architectures and ICT for multimodal and sustainable passenger and freight transport, open data and data sharing, and innovation based on shared data. She was the leader of a smart city research group in SINTEF for several years, and she has extensive experience as a leader and technical coordinator of national and European research and innovation projects in the areas of ICT for transport and open data.

Shanshan Jiang

Shanshan Jiang received her PhD degree from the Department of Telematics of Norwegian University of Science and Technology (NTNU) in 2008. She is currently a senior research scientist in SINTEF, Norway. Her research interests include software engineering and architecture, open data and data sharing, ICT for smart cities, and blockchain. She has extensive experience with EU and national research and innovation projects within the area of software engineering, open data, smart cities, and smart energy management. She has been leading several EU and national research and innovation projects within the open data and smart cities areas.

Erlend Stav

Erlend Stav is a senior research scientist at SINTEF, Norway. He received his Dr. Scient (PhD) in Computer and Information Science from NTNU in 2006. His research interests include open data, software architecture, quality of service, model driven development, tools development, and user centered design. He has extensive experience from national and European research projects and has applied his research in a variety of domains including smart cities, multimodal transport, avionics, health, assisted living, and fisheries.