

Is Open Data Ready for Use by Enterprises? Learnings from Corporate Registers

Pavel Krasikov^a, Timo Obrecht, Christine Legner^b and Markus Eurich^c
Department of Business and Economics, University of Lausanne, Lausanne, Switzerland

Keywords: Open Data, Corporate Registers, Open Corporate Data, Usability, Data Quality, Open Data Assessment.

Abstract: Open data initiatives have long focused on motivating governmental bodies to open up their data. The number of open datasets is growing steadily, but their adoption is still lagging behind. An increasing number of studies assess open data portals and open data quality to shed light on open data's current state. Since prior research addressed neither datasets' content, nor whether it met enterprises' data needs, our study aims to address this gap by investigating the extent to which open data is ready for use in the enterprise context. We focus on open corporate registers as an important segment of open government data with high relevance for enterprises. Our findings confirm that open datasets are heterogeneous in terms of access, licensing, and content, which makes them difficult to use in a business context. Our content analysis reveals that less than 50% of analyzed registers provide companies' full legal addresses, while only 10% note their contact information. We conclude that open data in corporate registers has limited use to its lack of required attributes and relevant business concepts for typical use cases.

1 INTRODUCTION

Open data can be defined as “data that is freely available, and can be used as well as republished by everyone without restrictions from copyright or patents” (Braunschweig, Eberius, Thiele, & Lehner, 2012). It is widely believed to have great business potential and to be linked to high expectations quantified to be as much as \$5.4 trillion (Manyika et al., 2013). Open data is therefore hailed as “a new goldmine” of business opportunities waiting to be unearthed (The Economist, 2013). According to previous research, open data could allow multiple industry sectors to benefit and prosper, including transportation, consumer products, electricity, oil and gas, healthcare, consumer finance, agriculture, urban development, and the social sector (Davies, Walker, Rubenstein, & Perini, 2019; Deloitte Analytics, 2012; Dinter & Kollwitz, 2016; Manyika et al., 2013; Publications Office of the EU, 2020).

Given governmental bodies' enforcing of open data provision and policies to improve its readiness and consumption have advanced significantly over

the last few years (European Data Portal, 2018). However, although the number of open datasets is growing steadily, their adoption is lagging behind (Publications Office of the EU, 2020). Application developers were the main users of the first wave of open data, achieving only modest success (Bizer, Heath, & Berners-Lee, 2009). Currently, the second wave is facilitating open data's wider adoption and using it to create added value (Puha, Rinciog, & Posea, 2018). Enterprises should, as part of this effort, use open data to improve their business processes and, ultimately, unveil new business opportunities. However, the severe challenges that enterprises face in order to find, access, select, and, finally, use open data make them reluctant to even try (Davies et al., 2019; Oliveira, Oliveira, Lima, & Lóscio, 2016). In fact, multiple studies have shown that users find open data's lack of transparency, unknown quality, and unclear licensing unsettling challenges (Janssen, Charalabidis, & Zuiderwijk, 2012; Martin, Foulonneau, Turki, & Ihadjadene, 2013).

^a <https://orcid.org/0000-0002-6427-7055>

^b <https://orcid.org/0000-0001-8891-3813>

^c <https://orcid.org/0000-0003-2850-4684>

Although an increasing number of studies have explored assessment methods of open data, most of them have primarily focused on the metadata level and proposed generic quality metrics. Enterprises' use of open data was not specifically investigated, despite its significant business potential. Consequently, this paper addresses the research question: "*Is open data ready for use by enterprises?*"

We answer this question by focusing on open corporate data (OCD), which is an important segment of open government data. OCD, which provides transparent and interoperable data about companies, has a high potential for reuse in a business setting (Koznov et al., 2016; Varytimou, Loutas, & Peristeras, 2015). We analyzed data from 20 open corporate registers (also known as business registries): first, by assessing the provided metadata and, second, by examining registered business entities' specific attributes. We compared the datasets' content with the typical attributes (intelligence and analytics, business processes, data management) that specific use cases require for an understanding of readiness for use. Our findings confirm that open datasets are heterogeneous in terms of access, licensing, and content, which make them difficult to use in a business context. In addition, our study shows that less than 50% of analyzed registers provide companies' full legal addresses, while only 10% note their contact information. We conclude that open corporate datasets have only limited use for typical use cases due to their lack of relevant business concepts. Our study thereby draws attention to the need for domain-specific semantic models that make open data more usable for enterprises.

The remainder of the paper is structured as follows: In Section 2, we review relevant literature on the barriers to open data adoption and assessment techniques, which clarifies the research gap. In Section 3, we explain the research methodology. Section 4 presents the study results in detail. We conclude with a discussion of our findings, the study's limitations, and provide an outlook on future research.

2 LITERATURE REVIEW

Open data initiatives focused on motivating governments to open their data for a long time (Zuiderwijk, Janssen, Choenni, Meijer, & Alibaks, 2012). Different organizations have started making their data available, but open data consumers still

experience difficulties with using open data. Consequently, researchers studied the barriers to open data adoption, proposing various assessment methods, which we review in this section.

2.1 Adoption Barriers of Open Data

Table 1 summarizes prior studies on the barriers to open data adoption. These studies integrate academic literature and practical insights, differentiating between open data consumption and supply (marked with an X in Table 1). They reveal that the barriers are related to the way open data is provided and to its condition, which make it difficult to use. Although the barriers are associated with either consumption or supply, there is a strong interdependency between the two: the way the data is published impacts how it is used. (Zuiderwijk et al., 2012, fig. 1).

Studies investigating open data provisioning identify several common issues: the risk of excessive costs, an unclear purpose, as well as litigation and differing licensing standards and documentation complicating open data suppliers' release process (Martin et al., 2013; Barry & Bannister, 2014; Conradie & Choenni, 2014; Beno, Figl, Umbrich, & Polleres, 2017). Studies addressing consumption barriers tend to emphasize the user perspective, claiming that these setbacks are not strictly technical (Beno et al., 2017; Martin et al., 2013; Zuiderwijk et al., 2012). Conversely, a lack of understanding of the contents and insufficient domain knowledge commonly hinder open data use (Beno et al., 2017; Janssen et al., 2012; Zuiderwijk et al., 2012). In fact, the absence of information describing an open dataset is often associated with poor metadata documentation (Zuiderwijk et al., 2012). The latter generally refers to technical barriers, demonstrating the interdependence of the impediments' consumption and supply sides.

The existing studies suggest that challenges with open data use relate mainly to three aspects: first, there is a lack of transparency about datasets' availability and their usefulness for the end user (Janssen et al., 2012). Second, open datasets' heterogeneity in terms of licensing conditions, available formats, and access to information complicates the integration efforts (Martin et al., 2013). Third, the quality of open data remains unknown and uncertain in terms of typical assessment criteria (Zuiderwijk et al., 2012). Finally, our review also points to a lack of research in the enterprise context, as only two studies examined enterprises as consumers of open data.

Table 1: Overview of prior literature on adoption barriers of open data.

Source and topic	Method	Adoption barriers	Open data	
			Consumption	Supply
(Janssen et al., 2012) Gap between the benefits of and barriers to open data adoption	Group session (n=9), findings were discussed during interviews (n= 14)	6 categories: institutional, task complexity, use and participation, legislation, information quality, technical. Categories are exemplified by a total of 57 examples of barriers	X (Generic)	X
(Zuiderwijk et al., 2012) Open data users' perspective on identified impediments	Literature review (n=37) Interviews (n=6) Workshops (n=4)	A total of 118 socio-technical impediments in 3 categories: data access, data use, and data deposition. 10 sub-categories: availability and access, findability, usability, understandability, quality, linking and combining data, comparability and compatibility, metadata, interaction with data provider, and opening and uploading	X (Generic)	X
(Martin et al., 2013) Risks for re-users of public data differ from those for open data providers	Analysis of open data platforms (n=3)	Typology of barriers comprising 7 categories: governance, economic issues, licenses and legal frameworks, data characteristics, metadata, access, and skills	X (Business)	X
(Conradie & Choenni, 2014) Release processes of government open data	Participatory action research: Exploratory workshop (n=5). Questionnaire answered by a consortium (n=14). Questionnaire answered by other civil servants (n=50). In-depth interviews (n=18). Workshop with data users (n=8). Plenary session discussion (n=21). Follow-up meeting with decision makers (n=2). Experiences with data release (n=4)	4 categories of barriers: fear of false conclusions, financial effects, opaque ownership and unknown data locations, and priority		X
(Barry & Bannister, 2014) Implications of opening up the data	Case studies (n=2), inductive approach to the analysis of collected data	6 types of barriers: economic, technical, cultural, legal, administrative, and task related. A total of 20 barriers to open data's release		X
(Beno et al., 2017) Practitioners using and providing open data in Austria	Literature review (n=17) Survey (n=110)	3 major groups: user specific, provider specific, and both users and providers with a total of 54 barriers	X (Enterprises, Academia, Public sector)	X

2.2 Open Data Assessment

The barriers to open data adoption have motivated researchers to investigate open data portals' assessment, focusing specifically on the data quality. Table 2 summarizes the ways prior studies assessed open data and categorizes two crucial aspects: (1)

whether the analysis only considered the metadata or the dataset content as well, and (2) the methods used. This summary allows us to conclude that the majority of the open data assessment studies focused almost exclusively on the metadata quality. Only three authors investigated the contents of the underlying datasets. As mentioned by Vetrò et al.

(2016), “poor data quality can be widespread, and potentially hamper an efficient reuse of open data.” Interestingly, the three studies propose generic quality assessment methods according to typical data quality dimensions, such as completeness, accuracy, or timeliness, but do not consider specific data requirements or the use contexts. This means that the reviewed literature largely ignores the actual user's perspective and the data domain knowledge, which has found to be crucial for overcoming the barriers (Section 2.1).

As a final point, open data’s usefulness is only partially addressed in the dataset assessment context. To our knowledge, only Osagie et al. (2017)

specifically address the usability of open data platforms’ features for specific use cases. In this regard, by omitting the datasets’ contents, the abovementioned assessment methods cannot address the usability aspect and its relevance.

2.3 Research Gap

The existence of barriers in the open data landscape shows a clear “lack of insight into the user’s perspective” (Janssen et al., 2012). There is a need to understand the particularities of open data access, publishing, licensing, and content, as well as the extent to which they meet the requirements for a

Table 2: Overview of open data assessment in literature.

Source	Scope	Method	Results
(Bogdanović-Dinić, Veljković, & Stoimenov, 2014)	Metadata	Case study: application of “data openness” model to 7 open data portals	Data openness index score based on eight open data principles (Open Government Working Group, 2007)
(Reiche, Höfig, & Schieferdecker, 2014)	Metadata	Case study: assessment of metadata quality of 10 open government data portals	Ranking of open data repositories with the average score computed by means of quality metrics
(Umbrich, Neumaier, & Polleres, 2015)	Metadata	Open data quality and monitoring assessment framework	Analysis of 82 CKAN portals by means of 6 quality dimensions
(Neumaier, Umbrich, & Polleres, 2016)	Metadata	Metadata quality assessment framework	Assessment of 260 open data portals to highlight common issues
(Vetrò et al., 2016)	Metadata and dataset	Quality framework supported by data quality models from the literature	Assessment of 11 datasets’ quality according to 6 dimensions and 14 metrics
(Máchová & Lněnička, 2017)	Metadata	Benchmarking framework for evaluating open data portals’ quality	Quality evaluation of 67 open data portals according to 12 general characteristics and 16 metrics
(Welle Donker & van Loenen, 2017)	Metadata	Holistic open data assessment framework with 3 main levels: open data supply, open data governance, and open data user characteristics.	Assessment of 20 “most wanted” datasets in Netherlands addressing open data quality on 3 levels
(Osagie et al., 2017)	Platform features	Usability evaluation with ROUTE-TO-PA and QUIN criteria	Scoring and testing of 4 functions in 5 datasets according to 12 usability criteria
(Bicevskis, Bicevska, Nikiforova, & Oditis, 2018)	Dataset	Three-part data quality model: definition of a data object, data object quality specifications, and implementation	Syntax analysis of data from 4 company registers for 11 attributes
(Kubler, Robert, Neumaier, Umbrich, & Le Traon, 2018)	Metadata	Open data portal quality (ODPQ) framework	Comparison of more than 250 open data portals according to 17 quality dimensions
(Stróżyńska et al., 2018)	Metadata	Quality-based selection, assessment, and retrieval method	Attribution of quality scores based on “ranking type Delphi” and 6 quality dimensions to 59 data sources
(Zhang, Indulska, & Sadiq, 2019)	Metadata and dataset	Design science research and a systematic approach to repurposed datasets’ quality	Discovery of data quality problems in 20 datasets using the LANG approach and according to 10 dimensions

specific usage scenario. Existing efforts regarding open data's use focus on open data platforms (Osagie et al., 2017) or refer to open data supply and its underlying technical impediments evoking users' behavioral intentions (Weerakkody, Irani, Kapoor, Sivarajah, & Dwivedi, 2017). Vetrò et al (2016) emphasize that barriers are therefore mostly studied on the platform level, rather than on the dataset level.

Moreover, the literature does not specifically cover open data's use in the business context. Users' perception of data availability (the way data is proposed and can be consumed), generally covers the usability concept's technical part (Osagie et al., 2017; Weerakkody et al., 2017), leaving a gap between the dataset's content and the user perspective. Furthermore, Welle Donker et al (2017) provided a more user-centric definition of open data's usability as "usable for the intended purpose of the user." In fact, being manageable is one of the indicators that the authors introduce in the same work, which implies that a "user should be able to use it (open data) for the goal the user had in mind" (Welle Donker & van Loenen, 2017).

The abovementioned gaps motivate our research aimed at answering the question whether open data is ready for use by enterprises.

3 RESEARCH METHODOLOGY

We address this research gap by focusing on open data in a specific domain and use context. We selected open corporate data (OCD), which is an important segment of open government data and has a confirmed reuse potential (Varytimou et al., 2015). OCD can be defined as data on companies that business registers, in keeping with local laws, usually collect. The resulting data is not only valuable for public authorities, but also for the businesses and researchers.

We conducted three types of research activities over a period of two years to understand how ready open corporate data is for use: a literature analysis to understand open data's current state and its adoption barriers; focus groups with practitioners to identify the use context; the in-depth assessment of open corporate sources and datasets.

Figure 1 summarizes the key phases of the research process (the numbers refer to the corresponding sections in this paper presenting the results).

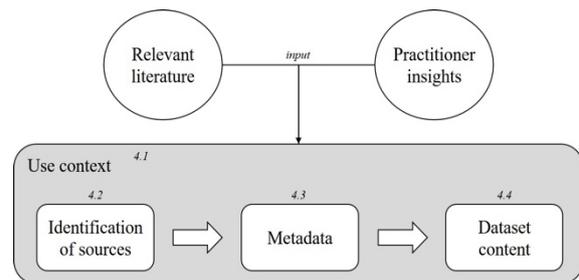


Figure 1: Research process.

Use Context: Open Data Domain and Use Cases

Focus groups (Bryman & Bell, 2007, p. 511; Creswell, 2009, p. 181) were formed with practitioners as a part of the broader analysis of open data use cases and addressed how OCD can be used in a business context. The focus group comprised seven Swiss-based data management experts representing transportation, consumption goods, and telecommunication industries. All the participants were knowledgeable about open data use cases and had been involved in the generation and documentation processes of OCD scenarios. The focus group first met during a Web conference during which it defined three high level use cases based on a structured use case generation framework (Krasikov, Harbich, Legner, & Eurich, 2019). Afterwards, the focus group met physically for a workshop that validated open corporate data use cases. Additional individual sessions were conducted with companies to refine the relevant use cases and obtain further insights. These focus group activities resulted in three core use case types and relevant business concepts (Section 4.1).

Identification of Data Sources and Datasets

So far, 223 countries have recorded 721 official corporate registers by following the global entity identifier foundation (GLEIF, 2017) procedure, which allows them to confirm their official status. However, only a small number of these registers are available as open data sources. For this paper, we selected 20 corporate registers that official state bodies provide and which are either available in full open access form or as open access with registration (Stróžyna et al., 2018). All of the registers are recorded within GLEIF and therefore have an assigned registry code. These registers cover open data initiatives in the United States, Europe, and other countries, but with different geographical granularity. Although we only wanted to consider registers that provide full access to the data (API or bulk download), we realized, during the course of

the analysis, that some only allow restricted access to the datasets, for example, the Indian, Danish, Belgian, Swiss, and Austrian business registers.

Metadata

As outlined in Section 2.2, most of the open data assessment methods focus on metadata. In fact, the primary insights into whether the desired data is usable or not are obtained through the metadata published at the source. We relied on previous literature (Table 1) when dealing with corporate registers and collected five categories of open data information: its identification, access, licensing, publisher, and basic information about the underlying datasets’ content. Two researchers collected and reconciled the metadata of the selected 20 corporate registers (see Appendix B).

Content Assessment of Open Corporate Datasets

Following the research process, a comprehensive content analysis of the corporate registers was undertaken to assess its readiness for use. Two researchers conducted a bottom-up analysis to understand the similarities between the attributes that the registers provide. Based on the focus group participants’ input, we examined the corporate registers to ascertain the presence of attributes related to the use cases’ relevant business concepts. Moreover, we took existing efforts regarding the OCD semantics’ standardization into consideration for this analysis. Section 4.1 elaborates these efforts. The understanding we gained of the existing domain ontologies and the content analysis of the relevant business registers allowed us to derive 21 typically used attributes (see Appendix A). We assessed all 20 registers on the basis of this attribute list.

4 RESULTS

This section compares the use context (4.1), i.e. the relevant use cases for open corporate data that we collected from practitioners, with the datasets that the corporate registers provided (4.2) in terms of metadata documentation (4.3) and its content (4.4).

4.1 Use Cases for Open Corporate Data

There is very little academic literature on OCD, which means that online sources are the main providers of insights. The working sessions described in Section 3 provided practitioners’ insights into how OCD can be used in the business environment. These sessions identified, discussed, and validated the following scenarios, as well as providing a summary of the relevant business concepts (see Table 3):

Intelligence and analytics: OCD can be used to gain insights into customers, partners, and competitors. Moreover, it is possible to identify a particular enterprise with a unique identifier, which helps to prevent confusion due to similar company names. OCD can also support investigations into corruption, abuse of power, and violations of cartel laws (Varytimou et al., 2015).

Business processes: OCD can be used in procurement processes to verify a given enterprise’s shipping or billing addresses. In turn, this leads to lower return rates and overall acceleration of procurement activities. Additionally, OCD helps to identify potential clients in particular industries and to target marketing campaigns at them. In this case, it is crucial to have up-to-date information about their activities and their initial contact information.

Table 3: Summary of the use cases for open corporate data.

Use case type	Relevant business concepts
1. Intelligence and analytics	Identification: Company Name, Identifier Organizational Information: Legal Form, Status, Date of Incorporation, Management Information, Financial Information, Number of Employees Address: Country, Post Code, Thoroughfare, Identifying Name Organizational and Management Information, Financial Statement, Number of Employees, Legal Form, Industry Classification Type, Incorporation Status
2. Business processes	Billing / Shipping Address: Country, Administrative area, Administrative Area, Locality, Post Code, Premise, Thoroughfare, Identifying Name Identification: Company Name Organizational Information: Status, Industry Classification Contact: Website, Postal Delivery Point, Phone Number, E-mail
3. Data management	Identification: Company Name, Identifier, Tax Number. (VAT) Address: Country, Administrative area, Locality, Post Code, Premise, Thoroughfare Organizational Information: Date of Incorporation, Status, Legal Form Contact: Website, Postal Delivery Point, Phone Number, E-mail

Data management: The maintenance of business partner data within a company’s IT systems is one example of an OCD application. OCD can help to ensure the data quality by removing duplicates, reconciling concepts representing the same real-world object, enriching the data with new entries, and ensuring its completeness and accuracy by adding up-to-date information from authoritative sources.

Each of these use cases can be related to business concepts, for example, the billing details for procurement process, which are similar to the attributes usually found in corporate registers (see Appendix A). It is interesting that “Address” is an overarching concept in all the use cases, while other concepts (identification numbers, organizational information, and contact details) are only relevant for selected use cases.

4.2 Identification of Data Sources

Corporate registers are usually assigned to a country or an administrative area and cover local business entities that need to undergo a local registration procedure. Aggregated lists of existing company registers are available online per country (Wikipedia, 2019), although there is no assessment process that confirms this sources’ accuracy. The abovementioned GLEIF has an attribution procedure by means of a legal entity identifier (LEI), and maintains a catalogue with accredited official business registers, which provides initial insights into the available OCD (GLEIF, 2018). Nevertheless, the register’s presence on this list does not guarantee that the provided data is open.

Our analysis considered 20 sources, i.e. business registers covering the United States, Europe, and other countries, but with different geographical granularity, as listed in Table 4.

4.3 Metadata Analysis

The analysis of the collected metadata provides first

insights into the sources (see Appendix B). The blue color denotes registers with open access, whereas those colored green refer to the ones with limited access to the download or a restrictive license policy. The registers present identification information regarding the relevant countries and registry codes, as well as the webpage to locate them.

Metadata regarding access revealed many interesting insights. CSV, JSON, and XML file formats, which are machine readable and suitable for processing, are of the most frequently used formats for downloadable business register files. Five registers required a login procedure in order to access the data, which also complicated the latter. In terms of licensing, most registers operated under an open license, a Creative Commons one or a national equivalent, whereas six registers provided access to the data without a license. Moreover, with the exception of one, all of the registers offered a free lookup service to query the data. Only 12 of the business registers provided a publishing date, which was after 2013. The data’s update frequency varied from daily or weekly to monthly or even yearly. Finally, these attributes’ importance should not be underestimated (Kampars, Zdravkovic, Stirna, & Grabis, 2020) as the enterprises’ specific needs drive this (see Section 4.1).

Metadata regarding the content revealed an important difference between the registers’ sizes, which ranged from 75,985 to 11,100,000 entry points. Their geographical coverage explains this, as larger registers have a national level of granularity (France, UK), while smaller ones cover states (US) or administrative areas. Almost all the registers are available in English, even though the country of origin has a different national language or more than one, which demonstrates the efforts taken to make data available. On average, the registers provide 33 distinct attributes. While this information allows first insights into the data, we, as shown in the following section, conducted a thorough analysis of the contents to specifically identify use cases.

Table 4: Analyzed corporate registers.

(1) Alaska Business Entity Register	(9) New York Business Entity Register	(14) Companies House UK
(2) Canada Corporate Register	(10) Norway Register of Business Enterprises	(15) Australian Business Register
(3) Colorado Business Entity Register	(11) Oregon Business Entity Register	(16) Indian Business Register
(4) Florida Business Entity Register	(12) Washington Business Entity Register	(17) Danish Company Register
(5) France Register of Companies	(13) Wyoming Business Entity Register	(18) KBO Central Belgium Company Database
(6) Iowa Business Entity Register		(19) Swiss UID Register
(7) Ireland Companies Register		(20) Austrian Corporate Register
(8) Japanese National Tax Agency		

Attribute / Register		Alaska	Canada	Colorado	Florida	France	Iowa	Ireland	Japan	New York	Norway	Oregon	Washington	Wyoming	UK	Australia	India	Denmark	Belgium	Switzerland	Austria	Σ (Attribute)	
ID	Company Name	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	20	
	Identifier	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	20	
	Tax N ^o . (VAT)		✓			✓			✓									✓		✓		5	
Address	Country	✓	✓	✓	✓	✓	✓	✓		✓	✓	✓	✓	✓	✓	✓	✓		✓	✓		17	
	Administrative Area	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	20	
	Locality	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	20	
	Post Code	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	20	
	Premise	✓		✓	✓	✓	✓	✓		✓					✓	✓	✓	✓	✓			13	
	Thoroughfare	✓	✓	✓	✓		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	19
	Identifying Name	✓		✓	✓		✓			✓		✓	✓	✓						✓			9
Organizational info	Legal Form	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓		✓	✓	✓	✓	✓	19	
	Type Classification (SIC)		✓			✓					✓		✓	✓	✓		✓	✓	✓	✓	✓	11	
	Status	✓	✓	✓	✓	✓		✓	✓		✓				✓	✓	✓	✓	✓	✓	✓	16	
	Date of Incorporation	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓				✓	✓	✓	✓	✓	✓	✓	18	
	Management Information		✓			✓											✓		✓		✓	5	
	Financial Information					✓												✓	✓		✓	4	
	Nr. of Employees					✓					✓								✓			3	
Contact	Website																		✓			1	
	Postal Delivery Point				✓	✓	✓			✓				✓	✓	✓			✓	✓		9	
	Phone Number																	✓	✓			2	
	E-Mail		✓			✓																2	
Σ (Register)		12	14	12	13	17	12	11	10	12	12	9	10	14	13	12	13	16	16	13	12		

Figure 2: Visualization of content analysis of open corporate registers.

4.4 Dataset Content Analysis

We conducted a content analysis for the previously identified registers with regards to the required attributes (see Table 3). Figure 2 summarizes the attribute’s presence in the dataset (with a green tick). On average, 12 of the 20 identified attributes were present; however, the French register shows 17 attributes present, followed by the Belgian and Swiss registers. Interestingly the US state registers do not provide the same attributes although they are part of the same country.

Companies’ address information and their identification concepts are present in the majority of the assessed corporate registers. “Address” is one of the central concepts for the considered use cases and available in all of the registers, with an exception of the “Premise” and “Identifying Name” attributes. Only seven registers provide all the address-related attributes, although the most evident attributes (“Administrative Area,” “Locality,” and “Post Code”) are all present.

The corporate registers present “Organizational information” only infrequently, with “Contact”

details appearing least. The intelligence and analytics use case suffers from a lack of “Organizational information” attributes. Even though a number of registers do contain management and financial information, this is too little to be useful. For instance, only the registers of Denmark, Austria, and the state of Iowa provide the full set of attributes in this category. Business processes-related use cases, i.e. marketing campaigns, suffer from a lack of contact information, which is also relatively scarce in all of the corporate registers. Data management use cases aim at maintaining the most accurate version of the data in the company’s internal systems; consequently, “Address” and “Identification” play a key role and are widely present in the registers.

5 CONCLUSION

Despite governments, NGOs, and companies’ enormous efforts to open their data and the open data movement’s decade of evolution, the condition and contents of the provided open data still do not

meet expectations. Open data is generally still assessed on the platform level and such assessments are almost exclusively aimed at the metadata quality.

We aimed to address this gap by means of a use case-driven analysis of open corporate registers, which considered both the metadata and the content. Our additional content analysis of 20 corporate registers revealed that open corporate datasets have limited use for typical use cases, because they do not cover relevant business concepts. Legally required information about companies, such as their addresses and identification, is mostly available, but not always complete, while many other attributes are only partially available.

Our study contributes to the emerging stream of research on the use of open data and addresses the “lack of insight into the user’s perspective,” which Janssen et al. (2012) mentioned. We propose a use case-driven approach comprising four steps: (1) the definition of the use context, (2) the identification of the open data sources, (3) a metadata analysis, and (4) a content analysis of the datasets. This approach goes beyond the existing assessments of open data quality by also integrating a content analysis. To the best of our knowledge, the conducted analysis is one of the first to provide insights into open data’s readiness for use from an enterprise perspective.

A limitation of this work is that, given the total number of existing business registers, the number of analyzed sources does not allow us to make conclusions about the domain as a whole. Nevertheless, the selected registers represent countries considered as advanced with regard to open data provision (Publications Office of the EU, 2020). As mentioned, we identified the use cases by means of a focus group, but others could be potentially discovered.

An implication from our study is that we find that the proposed open data assessment methods require amendments. Domain- and use case-specific content analysis need to complement these methods in order to assess open data’s usability. Future research in this direction could lead to the analysis being generalized into a method. As a first step, we plan to integrate the proposed approach further into the OCD domain by means of existing assessment methods, which we might generalize to other existing open domains by following this work’s usability assessment. Further links to open data quality should be explored with regard to usability (Bicevskis et al., 2018; Vetřo et al., 2016). In order to thoroughly address the data quality aspects, future research could embed the assessment techniques with metrics along the data quality dimensions in the content analysis part (e.g. timeliness, accuracy, and completeness).

Our study also underlines the need for domain ontologies, such as the euBusinessGraph (2019) common semantic model for company data, which could be a basis to provide more consistent and compatible open datasets across different open data portals and providers.

REFERENCES

- Barry, E., & Bannister, F. (2014). Barriers to open data release: A view from the top. *Information Polity*, 19, 129–152.
- Beno, M., Figl, K., Umbrich, J., & Polleres, A. (2017). Open Data Hopes and Fears: Determining the Barriers of Open Data. *2017 Conference for E-Democracy and Open Government (CeDEM)*, 69–81. Krems, Austria: IEEE.
- Bicevskis, J., Bicevska, Z., Nikiforova, A., & Oditis, I. (2018, September 26). *Data quality evaluation: A comparative analysis of company registers’ open data in four European countries*. 197–204.
- Bizer, C., Heath, T., & Berners-Lee, T. (2009). Linked Data—The Story So Far. *International Journal on Semantic Web and Information Systems*, 5, 1–22.
- Bogdanović-Dinić, S., Veljković, N., & Stoimenov, L. (2014). How Open Are Public Government Data? An Assessment of Seven Open Data Portals. In M. P. Rodríguez-Bolívar (Ed.), *Measuring E-government Efficiency* (Vol. 5, pp. 25–44). New York, NY: Springer New York.
- Braunschweig, K., Eberius, J., Thiele, M., & Lehner, W. (2012). *The State of Open Data. Limits of Current Open Data Platforms*.
- Bryman, A., & Bell, E. (2007). *Business Research Methods* (2nd ed.). Oxford, United Kingdom: Oxford University Press.
- Conradie, P., & Choenni, S. (2014). On the barriers for local government releasing open data. *Government Information Quarterly*, 31, S10–S17.
- Creswell, J. W. (2009). *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches* (3rd ed.). Thousand Oaks, CA: Sage Publications.
- Davies, T., Walker, S., Rubenstien, M., & Perini, F. (Eds.). (2019). *The State of Open Data: Histories and Horizons*. African Minds and International Development Research Centre.
- Deloitte Analytics. (2012). Open Data – Driving Growth, Ingenuity and Innovation. Retrieved July 9, 2019, from <https://www2.deloitte.com/content/dam/Deloitte/uk/Documents/deloitte-analytics/open-data-driving-growth-ingenuity-and-innovation.pdf>
- Dinter, B., & Kollwitz, C. (2016). Towards a Framework for Open Data Related Innovation Contests. *Proceedings of the 2016 Pre-ICIS SIGDSA/IFIP WG8.3 Symposium: Innovations in Data Analytics*, 13.
- euBusinessGraph. (2019). Ontology for Company Data. Retrieved November 15, 2019, from EuBusinessGraph website: <https://www.eubusinessgraph.eu/eubusinessgraph-ontology-for-company-data/>

- European Data Portal. (2018). Landscaping Method—Overview. Retrieved November 19, 2019, from https://www.europeandataportal.eu/sites/default/files/method-paper_insights-report_n4_2018.pdf
- GLEIF. (2017, September 1). Accreditation Process. Retrieved November 16, 2019, from GLEIF website: <https://www.gleif.org/en/about-lei/gleif-accreditation-of-lei-issuers/accreditation-process>
- GLEIF. (2018, December). GLEIF Registration Authorities List. Retrieved November 4, 2019, from GLEIF website: <https://www.gleif.org/en/about-lei/code-lists/gleif-registration-authorities-list>
- Janssen, M., Charalabidis, Y., & Zuiderwijk, A. (2012). Benefits, Adoption Barriers and Myths of Open Data and Open Government. *Information Systems Management*, 29, 258–268.
- Kampars, J., Zdravkovic, J., Stima, J., & Grabis, J. (2020). Extending organizational capabilities with Open Data to support sustainable and dynamic business ecosystems. *Software and Systems Modeling*, 19, 371–398.
- Koznov, D., Andreeva, O., Nikula, U., Maglyas, A., Muromtsev, D., & Radchenko, I. (2016). A Survey of Open Government Data in Russian Federation: *Proceedings of the 8th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management*, 173–180. Porto, Portugal: SCITEPRESS - Science and Technology Publications.
- Krasnikov, P., Harbich, M., Legner, C., & Eurich, M. (2019). *Open data use cases: Framework for the generation and documentation of open data use cases*. Retrieved from https://www.cc-cdq.ch/system/files/Open_data_use_cases_working_report_2019.pdf
- Kubler, S., Robert, J., Neumaier, S., Umbrich, J., & Le Traon, Y. (2018). Comparison of metadata quality in open data portals using the Analytic Hierarchy Process. *Government Information Quarterly*, 35, 13–29.
- Máchová, R., & Lněnička, M. (2017). Evaluating the Quality of Open Data Portals on the National Level. *Journal of Theoretical and Applied Electronic Commerce Research*, 12, 21–41.
- Manyika, J., Chui, M., Groves, P., Farrell, D., Van Kuiken, S., & Doshi, E. A. (2013). *Open data: Unlocking innovation and performance with liquid information*. McKinsey Global Institute.
- Martin, S., Foulonneau, M., Turki, S., & Ihadjadene, M. (2013). *Risk Analysis to Overcome Barriers to Open Data*. 11, 348–359.
- Neumaier, S., Umbrich, J., & Polleres, A. (2016). Automated Quality Assessment of Metadata across Open Data Portals. *Journal of Data and Information Quality*, 8, 1–29.
- Oliveira, M. I. S., Oliveira, L. E. R. de A., Lima, G. de F. A. B., & Lóscio, B. F. (2016). Enabling a Unified View of Open Data Catalogs: *Proceedings of the 18th International Conference on Enterprise Information Systems*, 230–239. Rome, Italy: SCITEPRESS - Science and Technology Publications.
- Open Government Working Group. (2007, December). The 8 Principles of Open Government Data (OpenGovData.org). Retrieved July 23, 2019, from <https://opengovdata.org/>
- Osagie, E., Waqar, M., Adebayo, S., Stasiewicz, A., Porwol, L., & Ojo, A. (2017). Usability Evaluation of an Open Data Platform. *Proceedings of the 18th Annual International Conference on Digital Government Research*, 495–504. New York, NY, USA: ACM.
- Publications Office of the EU. (2020). *Open data maturity report 2019*. Retrieved from https://op.europa.eu/publication/manifestation_identifier/PUB_OABE19001ENN
- Puha, A., Rinciog, O., & Posea, V. (2018). Enhancing Open Data Knowledge by Extracting Tabular Data from Text Images: *Proceedings of the 7th International Conference on Data Science, Technology and Applications*, 220–228. Porto, Portugal: SCITEPRESS - Science and Technology Publications.
- Reiche, K. J., Höfig, E., & Schieferdecker, I. (2014). Assessment and visualization of metadata quality for open government data. *Conference for E-Democracy and Open Government*, 335–346. Donau-Universität.
- Stróżyńska, M., Eiden, G., Abramowicz, W., Filipiak, D., Małyżko, J., & Węcel, K. (2018). A framework for the quality-based selection and retrieval of open data—A use case from the maritime domain. *Electronic Markets*, 28, 219–233.
- The Economist. (2013, May 18). A new goldmine. Retrieved November 19, 2019, from The Economist website: <https://www.economist.com/business/2013/05/18/a-new-goldmine>
- Umbrich, J., Neumaier, S., & Polleres, A. (2015). Quality Assessment and Evolution of Open Data Portals. *2015 3rd International Conference on Future Internet of Things and Cloud*, 404–411. Rome, Italy: IEEE.
- Varytimou, A., Loutas, N., & Peristeras, V. (2015). Towards Linked Open Business Registers: The Application of the Registered Organization Vocabulary in Greece. *International Journal on Semantic Web and Information Systems*, 11, 66–92.
- Vetò, A., Canova, L., Torchiano, M., Minotas, C. O., Iemma, R., & Morando, F. (2016). Open data quality measurement framework: Definition and application to Open Government Data. *Government Information Quarterly*, 33, 325–337.
- Weerakkody, V., Irani, Z., Kapoor, K., Sivarajah, U., & Dwivedi, Y. K. (2017). Open data and its usability: An empirical view from the Citizen’s perspective. *Information Systems Frontiers*, 19, 285–300.
- Welle Donker, F., & van Loenen, B. (2017). How to assess the success of the open data ecosystem? *International Journal of Digital Earth*, 10, 284–306.
- Wikipedia. (2019, October 19). List of company registers. Retrieved November 6, 2019, from Wikipedia website: https://en.wikipedia.org/w/index.php?title=List_of_company_registers&oldid=922064632
- Zhang, R., Indulska, M., & Sadiq, S. (2019). Discovering Data Quality Problems: The Case of Repurposed Data. *Business & Information Systems Engineering*, 61, 575–593.
- Zuiderwijk, A., Janssen, M., Choenni, S., Meijer, R., & Alibaks, R. S. (2012). *Socio-technical Impediments of Open Data*. 10, 17.

APPENDIX

Appendix A – Definition of Attributes

Source Information		
ID	Registry Code	Unique identification of legal entities by GLEIF (2018)
	Country	Defines a country to which the register refers
Access	Webpage	The webpage of the business register
	Resource Format	Describes the format of the published data, e.g. XML, JSON, CSV
	Access Login	Mentions whether access to the dataset requires an account
	Free lookup Service	Indicates whether the register has a free company lookup service
	License	License under which the data is provisioned
Publisher	Publisher	Entity responsible for providing the data
	Publishing Date	Date when the register originally published the dataset
	Update Cycle	Describes the frequency of the data update in days.
Content	Resource Language	Mentions the language(s) in which the data is published.
	Geographic Coverage	Defines the scope of the publishing institution, either on a state or national level.
	# of Diverse Attributes	Counts the different attributes that the register reports.
	#of Records	Estimate of the total number of entries in a register.
Content Information		
ID	Company Name	Defines the entity's name in a local language.
	Identifier	A unique identifier assigned to the relevant register.
	Tax № (VAT)	The tax number of the entity (VAT).
Address	Country	A geopolitical area, typically a nation
	Administrative Area	A top-level geographical or political area division in a country.
	Locality	A more granular level of an administrative area's geographical division.
	Post Code	A country-specific code for a certain address component.
	Premise	An area of land and its adjacent buildings.
	Thoroughfare	A form of the access route of the address: a street, road, avenue, etc.
	Organizational information	Identifying Name
Legal Form		The type of entity with respect to the local corporate law.
Type Classification (SIC)		Classification of entities and their respective industries.
Status		The entity's status, e.g. active, bankrupt.
Date of Incorporation		Date of the entry in the register.
Management Information		Information about the company's organizational structure.
Financial Information		Usually financial reports on corporate figures.
Number of Employees		The entity's number of employees.
Contact	Website	The entity's website.
	Postal Delivery Point	A single mailbox or other place at which postal mail is delivered.
	Phone Number	The entity's phone number.
	E-Mail	The e-mail address of the entity.

Appendix B – Metadata Documentation

	Alaska Business Entity Register	Canada Corporate Registry	Colorado Business Entity Register	Florida Business Entity Register	France Register of Companies	Iowa Business Entity Register	Ireland Companies Register	Japanese National Tax Agency	New York Business Entity Register	Norway Register of Business Enterprises	
	1	2	3	4	5	6	7	8	9	10	
ID	Registry Code	RA000594	RA000072	RA000599	RA000603	RA000189	RA000606	RA000402	RA000413	RA000628	RA000472
	Country	United States	Canada	United States	United States	France	United States	Ireland	Japan	United States	Norway
Access	Webpage	http://commerce.alaska.gov/CBP/DBDownloads/CorporationsDownload.CSV	https://open.canada.ca/data/en/dataset/0032e54-c5dd-4b66-99a0-320a7b5e99f2	https://data.colorado.gov/Business/Business-Entities-in-Colorado/4ykn-tg5h	http://dos.myflorida.com/sunbiz/other-services/data-downloads/corporate-data-file/	https://www.data.gouv.fr/en/datasets/base-sirene-des-entreprises-et-de-leurs-etablissements-	https://data.iowa.gov/Regulation/Active-Iowa-Business-Entities/cz5t-3qay	https://services.cro.ie/	http://www.houjin-bangou.nta.go.jp/pdf/download/zenken/	https://data.ny.gov/Economic-Development/Active-Corporations-Beginning-1800/n9v6-gdp6	http://data.breg.no/oppslag/enhetsregisteret/enheter.shtml
	Resource Format	CSV	XML	CSV, RDF, RSS, TSV, XML, REST	TXT	CSV	CSV, RDF, RSS, TSV, XML, SODA API	REST API	XML, CSV (Shift_JIS), CSV (Unicode)	CSV, RDF, RSS, TSV, XML	CSV, JSON, XML, REST API
	Access Login	no	no	no	no	no	no	no	no	no	no
	Free look up Service	available	available	available	available	not available	available	available	available	available	available
License	Open Government License	Open Government License	N/A	N/A	Open License V2.0	Creative Commons Attribution 4.0	Open License	Open License	Open Government License	Norwegian License for Open Government Data (NLOD)	
Publisher	Publisher	State of Alaska Department of commerce	Innovation, Science and Development Canada	Colorado Department of State	Division of Corporation Florida	National Institute of Statistics and Economic Studies	Secretary of State Iowa	Companies Registration Office Ireland	Financial Service Agency	New York Department of State	The Central Coordinating Register for Legal Entities
	Publishing Date	N/A	February 18, 2014	March 19, 2014	N/A	December 27, 2016	November 10, 2014	N/A	N/A	February 14, 2013	N/A
	Update Cycle	N/A	7d	1d	1d	1d	30d	N/A	30d	30d	N/A
Content	Resource Language	English	English, French	English	English	French	English	English	Japanese, English	English	Norwegian
	Geographic Coverage	State	National	State	State	National	State	National	National	State	National
	# of Records	75,985	930,000	1,716,403	N/A	11,100,000	230,117	N/A	N/A	2,776,167	1,103,302
	# of Diverse Attributes	35	25	35	45	118	19	18	19	30	42

	Oregon Business Entity Register	Washington Business Entity Register	Wyoming Business Entity Register	Companies House UK	Australian Business Register	Indian Business Register	Danish Company Register (CVR)	KBO Central Belgium Company Database	Swiss UID-Register	Austrian Corporate Register	
	11	12	13	14	15	16	17	18	19	20	
ID	Registry Code	RA000631	RA000641	RA000644	RA000585	RA000013	RA000394	RA000170	RA000025	RA000548	RA000017
	Country	United States	United States	United States	United Kingdom	Australia	India	Denmark	Belgium	Switzerland	Austria
Access	Webpage	http://sos.oregon.gov/business/Pages/temp-business-search.aspx	https://www.sos.wa.gov/cops/alldata.aspx	https://wyobiz.wy.gov/business/data-base.aspx	https://developer.companieshouse.gov.uk/api/docs/	https://data.gov.au/data/dataset/abn-bulk-extract	https://data.gov.in/catalog/company-master-data	http://datahub.virk.dk/dataset/systemtil-system-adgang-til-cvr-data	http://kbpub.economie.fgov.be/kpub/zoeknummerformhtml	https://www.uid.admin.ch/Search.aspx	https://www.justiz.gv.at/web2013/html/default/2c94848523082a60124066931e0860.de.html
	Resource Format	CSV, RDF, RSS, JSON, XML, SODA API	XML, JSON, Text	CSV	CSV, REST	XML, SOAP API	CSV	REST API	PDF	WebGUI	PDF
	Access Login	no	no	no	no	no	yes	yes	yes	no	yes
	Free look up Service	available	available	available	available	available	not available	available	available	available	available
License	N/A	N/A	N/A	Free, Open Government License v3.0	Creative Commons Attribution 3.0 Australia	National Data Sharing and Accessibility Policy (NDSAP) India	N/A	restricted to queries	restricted to queries	restricted access	
Publisher	Publisher	Secretary of State Oregon	Secretary of State Washington	Secretary of State Wyoming	Companies House UK	Australian Business Register	Ministry of Corporate Affairs India	Danish Business Authority	Ministry of Economy Belgium	Swiss Federal Statistical Office	Federal Ministry of Justice Austria
	Publishing Date	May 19, 2016	N/A	March 19, 2014	December 11, 2013	September 5, 2014	N/A	June 10, 2015	N/A	December 11, 2013	January 10, 2014
	Update Cycle	7d	1d	N/A	7d	1d	365d	1d	7d	1d	N/A
Content	Resource Language	English	English	English	English	English	English	Danish, English, Kalaallisut	English, French, Dutch, German	English, French, Italian, German	Deutsch
	Geographic Coverage	State	State	State	National	National	State	National	National	National	National
	# of Records	442,012	1,080,251	208,723	10,216,253	18,000,000	N/A	N/A	1,300,000	1,716,662	570'000
	# of Diverse Attributes	17	16	63	55	22	17	35	22	25	16