

Bachelor Thesis

Analyzing the Evolution of Open Data

Dominique Prinz

Date of Birth: 07.12.1997

Student ID: 11704419

Subject Area: Information Business

Studienkennzahl: 033 561

Supervisor: Univ.-Prof. Dr. Axel Polleres

Date of Submission: 15. April 2020

Department of Information Systems and Operations, Vienna University of Economics and Business, Welthandelsplatz 1, 1020 Vienna, Austria



Contents

1	Motivation	1
1.1	Problem Statement	2
1.2	Research Questions	3
1.3	Research Method	3
2	Definition of Terms	4
2.1	Open Data Portals	4
2.2	Monitoring Frameworks	5
2.3	Open Data and Metadata	5
2.4	RDF and Metadata Vocabularies	6
3	Summary of Background Literature	8
4	Related Work	9
5	Comparison of the Criteria Assessment for Monitoring Frameworks	11
5.1	The Open Data Barometer	13
5.1.1	Assessment Structure	15
5.1.2	Scoring	21
5.2	The Open Data Maturity	22
5.2.1	Assessment Structure	23
5.2.2	Scoring	24
5.3	The Open Data Portal Watch	28
5.3.1	Assessment Structure	28
5.4	Criteria Comparison between Monitoring Frameworks	32
6	Data Analysis	42
6.1	Evolution of Open Data in Europe - a Framework Dimension Overview	42
6.2	Comparative Comparison of Monitoring Metrics Results	45
6.3	Analysis Remarks and Findings	49
7	Limitations	50
8	Conclusion	51
9	Future Work	53
A	Unified Quality Metrics	57

List of Figures

1	Simplified RDF Graph	7
2	Assessment Structure ODB [15]	15
3	Scoring Criteria and Thresholds [15]	17
4	Dataset Clusters [15]	19
5	Scoring Distribution [24]	25
6	Open Data Portal Watch Dimensions 1 [21]	30
7	Open Data Portal Watch Dimensions 2 [21]	31
8	Framework (Metrics) Similarity Matrix	39
9	Average Assessment Points - European Countries ODB 15-16	42
10	Average Assessment Points per Dimension - European Countries ODB 15-17	44
11	ODM Average Percentage per Dimension - European Union 15-20	45
12	ODP Average Percentage per Dimension 16-19	46
13	ODB-ODP Metric Comparison	47
14	Selected ODP Metric Comparison	49
15	ODM-ODB Sub-Dimension Correlation	50
16	Open Data Barometer Metrics [16]	58
17	Open Data Maturity Metrics 2020 (1)	59
18	Open Data Maturity Metrics 2020 (2)	60
19	Open Data Maturity Metrics 2020 (3)	61
20	Open Data Maturity Metrics 2020 (4)	62
21	Open Data Maturity Metrics 2016 (1)	63
22	Open Data Maturity Metrics 2016 (2)	64
23	Open Data Portal Watch Metrics [21]	65

List of Tables

1	ODB Dataset Questions [15]	20
---	----------------------------	----

Abstract

Open Data has been a highly explored domain within research over the past years. With the released directives of the European Union on the re-use of Public Sector Information, European countries needed to provide Open Data for the public transparently. As a result, the development of Open Data portals in Europe started. This thesis is aimed to investigate and analyze the developments and evolution of Open Data in Europe within 2015-2020. More precisely, we present existing assessments of monitoring frameworks, created to capture trends and benchmark European countries and the quality of information they release on their national Open Data portals. We examined selected frameworks to find overlapping dimensions, sub-dimensions, or metrics that could potentially be compared to each other to either confirm observable trends or identify discrepancies between these metrics. The overall goal was to provide an in-depth overview on the different assessment processes to question the approaches, existing monitoring frameworks implemented to assess countries or governments and their associated Open Data portals. We used methodology sections and additional online published materials to describe the assessment processes of monitoring frameworks and to identify quantitative as well as qualitative metrics that could potentially confirm observable trends in Open Data. However, due to the constant developments of monitoring frameworks and their assessment processes, a detailed time series analysis was only possible for a limited amount of data. Furthermore, it was found that existing monitoring frameworks do not cover the (meta)data quality and data re-user perspective as much as they should and that the comparable metrics were not suitable to observe similar developments. Instead, it has become clear that there is a need for a constant global Open Data monitoring framework that rates the portals per se, and places great emphasis on the quality of published and linked data in the scoring and assessment process.

1 Motivation

Over the past few years, Open Data has been in the focus of researchers worldwide as there have been initiatives calling for open accessible data hubs that aim to collect and distribute available datasets. These data hubs are prevalent in the form of Open Data portals. As part of the strategy for shaping Europe's digital future, the European Commission (EC) is funding an Open Data harvesting portal since 2015, which is collecting Public Sector Information (also: government data) metadata across European countries. "The European Data Portal"¹ is the European Union's approach for a holistic integration of an Open Data infrastructure where metadata information, as well as APIs, are gathered and available through online resources ⁹.

By accumulating structured information about available datasets and providing access to these resources, the EC and the participating countries are pursuing sustainable developments in research as well as economic growth. Areas such as Artificial Intelligence (AI) and evidence-based decision-making in politics would highly benefit from a broad data infrastructure that enables instances to re-use highly relevant information ⁸. Hence, the need for Open Data portals became more crucial than ever, resulting in more and more countries implementing their own Open Data portals to drive further research and enable transparency on government data.

Monitoring Frameworks are developed to collect information from web portals and measure governments by their procedures of publishing data on the web. This processed information allows for later inspection and analysis of the activities happening on the so-called Open Data portals. Prominent examples for these frameworks are the Open Data Maturity² as well as the Open Data Barometer³. The former is part of the European Data Portal and funded by the European Union. The latter is initiated by the World Wide Web Foundation⁴ with support of the Omidyar Network⁵.

Due to the high priority the European Union puts on the development of Open Data, it is therefore all the more important to assess the results of the

¹European Data Portal <https://www.europeandataportal.eu/en>, accessed: 1.1.2021.

²Open Data Maturity <https://www.europeandataportal.eu/en/impact-studies/open-data-maturity>, accessed: 4.4.2021.

³Open Data Barometer <https://opendatabarometer.org/>, accessed: 4.4.2021.

⁴World Wide Web Foundation <https://webfoundation.org/>, accessed: 4.1.2021.

⁵Omidyar Network <https://omidyar.com/>, accessed: 4.1.2021.

measures taken to push Open Data within Europe.

1.1 Problem Statement

Open Data is highly available on the web, allowing for easy access of web-based resources by downloading datasets or via APIs. However, it is yet to determine if the emphasis on Open Data that some government programs provide has led to an actual move forward when it comes to publishing Open Data on the web. Especially the measures taken by the EC to expedite the Open Data movement with the funding of their data portal calls for an intermediate assessment of the actual quality of Open Data portal monitoring frameworks. In this way, we can assess the impact these measures have taken on European countries and their governments. To answer this question, we will conduct an analysis of the progress and evolution of Open Data monitoring frameworks over the past years across European Open Data portals. This will be done by an elaboration of the assessment processes of the frameworks and a comparison of metrics that are intersecting between them.

The already existing data provided by monitoring frameworks allow for an examination of the direction where Open Data is heading. However, the approaches of the selecting criteria for the assessment of the individual monitoring frameworks differ. Nevertheless, the metadata and qualitative data these frameworks are collecting contain valuable insights on Open Data portals and deliver key information when it comes to investigating the status quo of Open Data portals.

In order to derive reliable results from an analysis, it is necessary to gather as much information on Open Data portals as possible. In the past, Neumaier et al. [22] created a framework for the automated quality assessment of over 260 data portals called "Open Data Portal Watch". An in-depth analysis of data gathered by the "Open Data Portal Watch" framework will represent a solid foundation for the evaluation and will be the benchmark for comparing the information of further monitoring frameworks. Doing so, allows us to secure a broad scope of available data via different monitoring frameworks which will provide improved results and a way to answer the defined research questions, which will be derived based on the problem statement.

However, the amount and extent to which a monitoring framework assesses collected information is up to the organizations responsible for the development and administration of their monitoring frameworks.

As already mentioned, the Open Data Portal Watch framework is an automated quantitative assessment of metadata from Open Data portals, whereas other monitoring frameworks such as the Open Data Maturity also provide qualitative assessments based on surveys constructed by the initiators [5]. A categorization of quality metrics defined by the individual monitoring frameworks is indispensable to get an overview on the different procedures for assessments. Hence, they need to be compared on their methodological course of action in a first step.

1.2 Research Questions

Based on the above-described problem statement, there is one major question that will be addressed in this thesis.

- **Does Open Data actually gain momentum as some government programs promise?**

To assess this question, the research question is split up into two sub-problems represented by two specific questions:

- **How does the development of Open Data and Open Data quality differ across countries in Europe from 2015-2020?**
- **Can the assessments of qualitative monitoring frameworks be quantitatively confirmed based on quantitative monitoring of the metadata?**

1.3 Research Method

First, literature research was conducted, which is necessary for the definition of terms followed by a comparison of the quality assessment of popular Open Data monitoring frameworks. By utilizing empirically crawled metadata from the "Open Data Portal Watch" framework, we can evaluate the evolution of Open Data over the last couple of years for this particular framework. Furthermore, we will compare the outcomes of the Portal Watch metadata analysis to the historic data of the monitoring frameworks to allow for a more general assumption on the evolution. Therefore, it is indispensable to collect the data that is needed from the frameworks.

In the analysis of this metadata, the focus will be set primarily on a comparison between the data coming from the different EU countries. On a finer granularity, we would then also be able to identify high activity entities, that

emphasize the data collection and distribution to derive general conclusions. The data analysis is done mainly using Python⁶, a popular programming language that is best suited for data analysis. For frameworks that provide their assessment data only in the proprietary Microsoft Excel file format (.xlsx), we have done basic calculations of percentage changes for metrics, subtraction, addition and filtering, plus exporting of CSV files for Python within those Excel spreadsheets.

2 Definition of Terms

Before elaborating on the topics addressed in this thesis, the important expressions used in this thesis need to be defined. This will allow for a delimitation to possible related terms and prevent ambiguity. Hence, a detailed overview of the most important definitions is provided.

2.1 Open Data Portals

As briefly mentioned in the introduction, Open Data Portals are centralized hubs, intended to facilitate the exchange of all kinds of data. Open Data portals serve as data catalogs that index Public Sector Information on centralized web pages [21]. In this data catalogs users can explore data by browsing through categories, filtering, and downloading the actual data if it is linked correctly. Of course, the operators of these portals are responsible for the proper functioning and features of these portals. However, as the frameworks that will be compared amongst each other in this thesis are implemented by governments, we will focus on Open Data portals on national levels that provide primarily Public Sector Information.

Since the introduction of the European Data Portal⁷ in 2015 and the announcement of the Directive 2013/37 of the European Parliament and the European Council [10], the need for the development of Open Data portals had progressed rapidly. With the transposition into national law, this directive made it obligatory for EU member states to publish certain Public Sector Information.

Depending on the underlying software frameworks which are implemented for the respective portals to publish Open Data on the national data por-

⁶Python <https://www.python.org/>, accessed: 5.1.2021.

⁷European Data Portal <https://www.europeandataportal.eu/en>, accessed: 10.1.2021.

tals, we can distinguish between different software frameworks. Amongst these frameworks, CKAN is by far the most popular followed, by Socrata and OpenDataSoft [21]. The drawback of various types is that each individual software framework provides its own schema for metadata resulting in heterogeneous metadata expressions.

2.2 Monitoring Frameworks

Because Open Data portals are systems developed to enable the interchange and re-use of Public Sector Information to generate additional value for science, society, politics, and economy. Stakeholders might as well be interested in how and to what extent the Open Data portals serve their purposes.

Like in almost every other ICT system, one can derive specific metrics that allow for assessing the overall performance and actual condition of the system. A monitoring framework is intended to conduct a structured evaluation based on introduced metrics. Therefore, different approaches have been considered to benchmark the quality and evolution of Open Data on Open Data portals. There are qualitative approaches as well as mere quantitative assessments of different dimensions that can be derived from the key components of Open Data portals.

Depending on the type of assessment, different resources are used and serve as integral parts for the final validation of the quantitative results. Especially for automated quantitative assessments, the use of a portal's metadata is indispensable. Since metadata contains structured information on the actual data accessible via the portals, it is possible to develop automated assessment frameworks [21].

2.3 Open Data and Metadata

The "Open Definition"⁸, by the Open Knowledge Foundation⁹, serves as a guideline for (digital) content to be classified as "open". This definition does not only apply to source code but also to available information like data published at Open Data portals.

Data, which is often referred to as "knowledge", is classified as Open Data if it *"can be freely used, modified and shared by anyone for any propose"* [14].

⁸Open Definition <https://opendefinition.org/>

⁹Open Knowledge Foundation <https://okfn.org/>, accessed: 12.1.2021.

More precisely, in order to meet these requirements, there are four aspects that must be fulfilled by Open Data:

1. **Open License/Status:** The data needs to be published under an open license.
2. **Access:** The whole data (including publisher information) must be available online and ready to download.
3. **Machine Readability:** Open Data elements need to be usable for computers, i.e., the computer can process the data (e.g., images of relational tables cannot be automatically read and processed, relational tables in the form of comma-separated values (CSV) files, however, can be).
4. **Open Format:** The data must not be available in a proprietary format. Instead, it must be possible to access, discover and use it by utilizing open software tools. Common open formats are: CSV, JSON, PDF, PNG, etc.

If these requirements are met, we usually speak of Open Data. In order to describe characteristics of Open Data, structured information is used. This structured information is referred to as metadata. Depending on the type of data, metadata can contain different information. More specifically, metadata for Open Data should comprise the key information about data hosted on Open Data portals. Therefore, the license type, data format, publisher etc., are crucial parts of Open Data metadata.

2.4 RDF and Metadata Vocabularies

The Resource Description Framework (RDF), first published in 2004 [18], is a formal modeling language, which defines guidelines that enable the inter-linking of resources on the web. By using of syntax, that describes resources and relationships between data objects, data can be transformed into semantic data [11]. In this way, machines and people are able to discover complex structures between data.

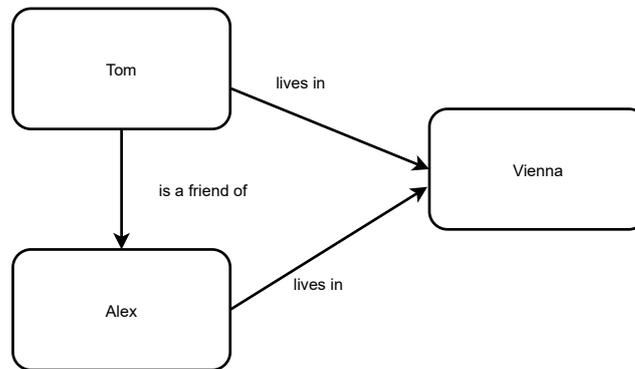
To give meaning to data, RDF makes use of three components:

1. **Subject** (e.g., Tom)
2. **Predicate** (e.g., lives in)

3. Object (e.g., Vienna)

These key components, which define properties of data objects, are the reason why RDF is also often referred to as triples. By defining triples, adhering to the RDF guidelines, we are able to portray a declarative fact about the world. Relationships between data subjects are established by linking the subject to an object, which again, is an instance of a subject. However, the prerequisite for this is a common vocabulary that is best suited to describe the elements of a domain. Figure 1 demonstrates how a simple knowledge graph is created by adding information between resources and using a common vocabulary for describing them.

Figure 1: Simplified RDF Graph



As a result, this technique cannot only be used to reference real-world objects. It is also used to establish knowledge between digital information and resources. The linking of data has already found many application examples. Up until 2002, many of them were rather experimental [13], but more and more initiatives have developed over time. Probably the most prominent example today is DBpedia¹⁰, which was first released in 2007. Originally, the idea behind DBpedia was to extract available data from Wikipedia infoboxes and publish it as RDF data to provide structured interlinked knowledge which can be accessed, queried and re-used for other purposes, free of charge [19]. The project evolved and included not only infobox data but also other information from online Wikipedia resources (references, page structure, whole text, etc.). By October 2016, DBpedia comprised 23 billion RDF triples [12].

The power of RDF relies upon the ontologies that are used when building the knowledge graphs, as the data providers need a homogenized approach

¹⁰DBpedia <https://www.dbpedia.org/>, accessed: 10.3.2021.

to define jointly used identifiers for resources and predicates. Ontologies are complex vocabularies that are specified to capture the relevant aspects of an area of application. They determine the set of terms that can be used for interlinking knowledge based on a common vocabulary. Because there are different use cases for RDF vocabularies to reproduce different kinds of data, there are different vocabularies that evolved over time.

Specifically for metadata, there is a RDF vocabulary called Data Catalog Vocabulary (DCAT)^[11], which is also a W3C^[12] recommendation and has the purpose to provide a metadata schema for the description of datasets on the web in the RDF format. The DCAT vocabulary is based upon the existing Dublin Core^[13] metadata vocabulary, i.e., it reuses some of the defined namespaces and classes as well as properties [1]. Therefore, it makes use of a predefined set of keywords that often differ from other metadata schemata exported by Open Data portals. These exports are provided by the underlying software frameworks such as CKAN^[14], Socrata^[15], and Opendatasoft^[16] [21] and therefore differ based on the underlying software framework.

However, as Neumaier [21] proposed, there is a way to homogenize those metadata schemata by mapping them to the DCAT vocabulary and adding additional quality-related keys in order to publish Open Data metadata as Linked Open Data.

3 Summary of Background Literature

As Open Data has been an important topic in the context of establishing the Web of (Linked) Data and the Semantic Web, there has been much effort to call for a standardized integration of Open Data. Tim Berners-Lee, for example, introduced a five-star rating schema that is encouraging data publishers to make their data available as Linked Open Data since this data is valuable for the society and people working with that data, as well as to further develop the Web of Linked Data [3].

¹¹DCAT <https://www.w3.org/TR/vocab-dcat-2/>, accessed: 3.3.2021.

¹²W3C <https://www.w3.org/>, accessed: 3.3.2021.

¹³Dublin Core <https://www.dublincore.org/specifications/dublin-core/dcmi-terms/>, accessed: 3.3.2021.

¹⁴CKAN <https://ckan.org/>, accessed: 13.2.2021.

¹⁵Socrata <https://dev.socrata.com/>, accessed: 13.2.2021.

¹⁶Opendatasoft <https://www.opendatasoft.com/>, accessed: 13.2.2021.

The transition from a "Web of documents" to an interconnected "Web of data" has been an anticipated goal of the World Wide Web Consortium (W3C)¹⁷ as highlighted by the Director, Tim Berners-Lee, since the beginning of the 21st century [4]. Making data not only available but also machine-readable to establish knowledge relations between data (using RDF) would further drive interchangeability of data between companies, software applications and other organizational communities. In the past, however, there have been major challenges identified for this adaption that yet need to be solved [23]. Nevertheless, there are some projects showcasing the potential of enriched Open Data, like the metadata monitoring framework called "Open Data Portal Watch" [21].

When establishing a monitoring framework that crawls information from data portals on the web, the underlying framework is dependent on the quality of corresponding metadata which is available at the portals. Due to the high number of data publishers and variety of different data sources at Open Data portals, the data quality can be influenced to a large extent. For this reason, Sebastian Neumaier [21] described 18 metrics together with five dimensions for the automated assessment of metadata quality for Open Data portals. This information can be used to gain insights into the actual quality of (meta)data provided by national Open Data portals.

Providing well-structured and complete metadata might have some disadvantages like the time-consuming aspect as well as the cost of updating and maintaining metadata. However, it offers much more advantages such as improved preservation, the process of analyzing structured information about data, and comparing the highly available information from the Web of Data [28]. For this reason, quantitative as well as qualitative methods of Open Data portal assessments are part of this thesis in order to analyze quantitative metadata metrics and qualitative information over several years.

4 Related Work

In the past, there have been efforts to evaluate the development of Open Data portals and their published data as well as to compare the portals or countries with each other. Besides the monitoring frameworks that we will cover in detail within this thesis, multiple approaches had been made to assess certain dimensions of Open Data portals, but they do come with certain

¹⁷W3C <https://www.w3.org/>, accessed: 15.1.2021.

limitations.

Máchová and Lněnička [20], for example, proposed an assessment for national Open Portals that is intended to cover the most relevant aspects of Open Data portals and also provide information on the actual data quality. Also, integrating the data user perspective and the overall data quality should be a crucial part of an Open Data assessment since the potential stakeholders such as the data re-users are dependent on the findability and subsequent exploitability of Open Data resources. This is in line with the findings of Alexopoulos et al. [2], in which they pointed out the importance of feedback mechanisms and the quality of Open Data for Open Data re-use and collaboration, which drives forward the Open Data movement. While the study of Máchová and Lněnička shows a holistic approach on how to assess Open Data it served more as a "prove-of-concept" to benchmark national Open Data portals. Moreover, the assessment of the Open Data portals was done by 10 postgraduate students and comprised a non-automated dataset evaluation which cannot display the quality of the whole data available at these portals and at the same time do not cover the needs of "real" data re-users. As a result, there is no data available in an annually prepared assessment that serves the purpose of drawing a general conclusion on the developments of Open Data.

Frederika Welle Donker and Bastiaan van Loenen [27] developed an assessment structure for the Open Data ecosystem of the Netherlands, which is based on the knowledge gained through investigating existing monitoring frameworks. Like the previously introduced work of Máchová and Lněnička, it also focuses heavily on the user perspective of Open Data. The assessment on the Open data supply, Open Data governance, and the user perspective dimensions are, unfortunately, not scalable as the assessment is based on an resource-intensive (time and interviewee) questionnaire, as well as desk research. Furthermore, like other assessment approaches, it only covers a fraction of national Open Data by including a selection (27) of datasets that are investigated during the assessment process.

There are also several other monitoring frameworks like the Global Open Data Index¹⁸ and the Open Data Monitor¹⁹. However, these frameworks are partially covering dimensions that cannot be compared to other monitoring

¹⁸Global Open Data Index <https://index.okfn.org/about/>, accessed: 18.1.2021.

¹⁹Open Data Monitor <https://www.opendatamonitor.eu/frontend/web/index.php?r=site%2Fabout>, accessed: 18.1.2021.

frameworks or offer only limited data because their assessments were discontinued at some point of time. Nevertheless, until now, there still needs to be work done to compare existing Open Data monitoring frameworks and validate whether the assessment is suitable to capture the direction where Open Data is heading.

5 Comparison of the Criteria Assessment for Monitoring Frameworks

In this section, we will elaborate on the structure of the assessment processes of individual frameworks which are implemented to monitor developments in the domain of Open Data. As we will see, these monitoring frameworks are operating on different dimensions and compute metrics whose similarities are not immediately apparent. Therefore, a first overview on what, or rather, which dimensions the different frameworks are examining will be the starting point of this section. Further, we will provide information on how the evaluation process is build up and how the final metrics are computed. The different frameworks will be covered independently and will comprise:

1. **The Open Data Barometer**²⁰
2. **The Open Data Maturity**²¹
3. **The Open Data Portal Watch**²²

First, we will provide a brief introduction for each framework and basic information on the framework providers as well as the period of time the framework assessed Open Data portals. Secondly, the frameworks will be split by their **assessment structure** and the **scoring** method which is applied to derive the final scores for the metrics that are used to benchmark governments and/or the respective Open Data portals. Lastly, the approach on how similarities between those frameworks could be identified will be presented together with a justification on the course of action.

²⁰Open Data Barometer <http://www.opendatabarometer.org>, accessed: 4.4.2021.

²¹Open Data Maturity <https://www.europeandataportal.eu/en/impact-studies/open-data-maturity>, accessed: 4.4.2021.

²²Open Data Portal Watch <https://data.wu.ac.at/portalwatch/about>, accessed: 4.4.2021.

In order to guarantee a uniform way to refer to metrics and dimensions of the monitoring frameworks, a simple notation is used to specify the frameworks indicators. Our figures [16](#) - [23](#) within the appendix show the mapping of indicators and metrics that are part of the monitoring frameworks. To distinguish between the metrics, we use indicators that are built as follows:

- a three-character prefix for the framework:
 - **ODB** = Open Data Barometer
 - **ODM** = Open Data Maturity
 - **ODP** = Open Data Portal Watch
- the prefix is, like the rest of the components, separated by a dot. Furthermore, the prefix is followed by the first four characters of the frameworks dimension it is a part of. Quantified metrics will be additionally indicated with "QN". The metrics obtained from the different monitoring frameworks will then be enumerated within the dimension, in the order we obtained them from the official documents of the frameworks.
 - e.g., **ODB.IMPL.1** indicates the first instance of the **Implementation** dimension of the **Open Data Barometer**. This is, see annex, figure [16](#):

"Does the data exist?"

For the Open Data Barometer, we extracted the questions out of the official methodology paper [16](#), whereas the Open Data Maturity metrics had been retrieved from two sources. We included only questions that are or were part of the framework's ranking system. Due to an overhaul of the ODM metrics in 2018, we used the Wayback machine^{[23](#)}, provided by the Internet Archive^{[24](#)}, to gather the scores before 2018. The Internet Archive is a web library that archives digital content (websites, videos, etc.). Despite using it, we were only able to source detailed scores for the 2016 edition^{[25](#)}. For the remaining editions before 2018, dashboards with the overall and sub-category scores are still provided on the official dashboard web page^{[26](#)}. For this thesis,

²³Wayback machine <https://archive.org/web/>, accessed: 19.1.2021.

²⁴Internet Archive <https://archive.org/about/>, accessed: 19.1.2021.

²⁵Detailed Scores ODM 2016 https://web.archive.org/web/20170717085028/https://www.europeandataportal.eu/sites/default/files/landscaping_2016_individual_scoring.xlsx, accessed: 19.1.2021.

²⁶Open Data Maturity dashboard <https://www.europeandataportal.eu/en/dashboard/2017>, accessed: 19.1.2021.

the detailed scores would be crucial since they comprise the detailed country scores for each metric. Those are needed for the comparison with metrics provided by other monitoring frameworks.

5.1 The Open Data Barometer

The elaboration in this section is based on the information available on the official website of the Open Data Barometer (ODB)²⁷ and the documents published for the latest edition of the study [17].

The Open Data Barometer is a monitoring project of the, in 2009 by Sir Tim Berners-Lee established, World Wide Web Foundation²⁸ in collaboration with the Omidyar Network²⁹. The intention of the developed monitoring framework is to provide insights on the global dispersion and adaptation of Open Data initiatives to evaluate the current state and generate assimilable results for governmental Open Data policies. By benchmarking governments on their practices to handle Open Data programs, they contribute actively to track the developments of Open Data.

The Open Data Barometer is one of the most comprehensive monitoring frameworks available for the assessment of governments with respect to their Open Data policies is the Open Data Barometer. Their latest report edition is called "Leader Edition"[17] and ranks 30 governments on their:

- *"Readiness for Open Data initiatives"*
- *"Implementation of Open Data programs"*
- *"Impact that Open Data is having on business, politics and civil society"*

Overall scores and the derived ranks for each government are based on a structured assessment of the government's Open Data practices.

The first publication of Open Data Barometer was in 2013. Since then, the assessment process had been iterated four times. In contrast to prior editions of the Open Data Barometer, the scale of the study for the "Leaders Edition" has dramatically changed. The Open Data Barometer assessed only

²⁷Open Data Barometer <http://www.opendatabarometer.org>, accessed: 4.4.2021.

²⁸World Wide Web Foundation <https://webfoundation.org/about/>, accessed: 24.1.2021.

²⁹Omidyar Network <https://omidyar.com/who-we-are/>, accessed: 24.1.2021.

governments who have adopted the Open Data Charter³⁰ or if being a G20 member, signed the G20 Anti-Corruption Open Data Principles, to investigate whether these governments actually fulfill their commitments. Ideally the report would cover all Open Data initiatives worldwide, however, the research process depends heavily on the available resources (time, staff). Prior to that, the ODB benchmarked between 77 and 115 countries for the fourth edition.

For the latest "Leaders Edition" the time covered by the study was 15 months, from July 2016 until September 2017. Final results were published in September 2018. However, in this thesis, especially the time frames, in which the assessments are conducted are important for us since we need a common baseline to compare the monitoring frameworks against each other. The timelines on which the information retrieval (i.e., expert survey and dataset assessment) were based for the editions are grouped as follows:

- **leaders edition:** July 2016 - September 2017³¹
- **fourth edition:** July 2015 - June 2016³²
- **third edition:** July 2014 - June 2015³³
- **second edition:** July 2013 - June 2014³⁴
- **first edition:** July 2012 - June 2013³⁵

So for the "regular" Open Data Barometer assessments each of the edition comprises a time period of 12 months. Although the "Leaders Edition" covers less countries and had some minor adaptations to the methodology, compared to the previous editions, the elaboration will be based on the assessment structure of the latest edition. In this way we can evaluate the most up to date approaches to benchmark Open Data initiatives.

³⁰Open Data Charter <https://opendatacharter.net/>, accessed: 1.2.2021.

³¹Methodology leaders edition <http://opendatabarometer.org/doc/leadersEdition/ODB-leadersEdition-Methodology.pdf>, accessed: 15.2.2021.

³²Methodology fourth edition <https://opendatabarometer.org/doc/4thEdition/ODB-4thEdition-Methodology.pdf>, accessed: 15.2.2021.

³³Methodology third edition <https://opendatabarometer.org/doc/3rdEdition/ODB-3rdEdition-Methodology.pdf>, accessed: 15.2.2021.

³⁴Methodology second edition <https://opendatabarometer.org/2ndEdition/about/method.html>, accessed: 15.2.2021

³⁵Methodology first edition <http://opendatabarometer.org/doc/1stEdition/Open-Data-Barometer-2013-Global-Report.pdf>, accessed: 15.2.2021.

5.1.1 Assessment Structure

The Barometer’s assessment of Open Data policies is divided into three dimensions, being the **Readiness**, **Implementation**, and **Impacts**. Each of which accounts for an equal contribution to the overall score, which is a maximum of 100 points. Those dimensions are grouped into sub-segments which are implemented for covering the relevant target groups (governments, entrepreneurs businesses, citizens civil society) and other factors (political, economic, social) that need to be considered for the sustainable evolution of Open Data. Figure 2 shows the Open Data Barometer’s structure of the assessment process and the information for the weighted measurement of the included factors. As a foundation for the calculation of the final government scores, the ODB used four types of data sources [15]:

Figure 2: Assessment Structure ODB [15]

Readiness (1/3) (Primary & secondary data)			
Government policies (¼)	Government action (¼)	Entrepreneurs & business (¼)	Citizens & civil society (¼)
Implementation (1/3) (Dataset assessments)			
Accountability dataset cluster (⅓)	Innovation dataset cluster (⅓)	Social policy dataset cluster (⅓)	
Impacts (1/3) (Primary data)			
Political (⅓)	Economic (⅓)	Social (⅓)	

1. Peer-reviewed expert survey:

For every government of the countries that are part of the study, there is a trained country researcher (expert) who is in charge of conducting the expert survey. In the expert survey, the country researcher needs to work on a questionnaire categorized by the dimensions **Readiness** and **Impacts**. By scoring points (0-10), providing justifications as well as a confidence level (0-100%) on each question of the questionnaire, the expert completes his assigned tasks. Throughout the research process, all country researchers are provided with a handbook that includes a structured guideline on how the survey needs to be conducted [16].

Because three distinct dimensions are observed in this framework which cover unequal characteristics of Open Data portals, the structures of the evaluation metrics - i.e., the questions that need to be answered - differ

as well. As a result, the ODB includes 30 different metrics/questions to assess the governments that are part of the study. In the **Readiness** dimension there are 14 questions covered. Six questions were developed for the **Impacts** dimension and ten checklist questions were assigned to the **Implementation** dimension.

Scoring the questions for the **Readiness** and **Impact** dimensions requires mainly desk research by the associated country researcher. This means that the expert is responsible for finding information and providing justifications by referencing sources that are needed to answer the question in the questionnaire. References for a question will mostly be found online and range from referencing to laws, regulations, interviews, official websites, third party surveys etc. A score for a particular question can range from 0 to 10 and is derived by comparing information collected through the desk research to an evidence and scoring criteria as well as the corresponding thresholds, which are used as reference values.

Incremental scoring guidelines and thresholds for the scores on each question are defined at the values 0, 3, 5 and 8, depending on the type of question asked. I.e., for every question there is an evidence and scoring criteria attached which comprises minimum standards that need to be fulfilled to achieve a corresponding score. The total scores are based on the reference scale ranging from 0-10 as well as the corresponding thresholds that allow for clear boundaries when scoring questions. These reference values and thresholds are defined for each question individually. Figure 3 shows an example of how these thresholds are structured. Additionally, the "Scoring" and "Source Guidance" attached to the question's thresholds support at getting a more accurate score and justification for the points given by the researcher. These guidelines provide assistance and examples on where the information needed is likely to be found.

An Example of a metric in the expert survey is **ODB.READ.1**: *"To what extent is there a well-defined open data policy and/or strategy in the country?"*. Figure 3 shows the scoring criteria and the thresholds for this question on which the country researcher builds upon his scoring. If the information about this policies or strategies are found and according to the defined standards in the research handbook referenced, the expert is able to register his score in the form.

Figure 3: Scoring Criteria and Thresholds [15]

Score > 0	There should be evidence of any official websites, government documents or guidelines referencing global open data practices in the country, although no formal open data policy or strategy may be yet in place.
Score > 3	There should be evidence of at least some national statements or guidelines on the publication of public sector information, even if just as part of the open government and transparency agenda or any other more general information management programme. A common definition of open data may still not be shared across the public sector and non-commercial restrictions or access fees may exist.
Score > 5	There should be evidence of a documented national open data policy or strategy that articulates processes, responsibilities, timelines and resources and a national institution or authority is in charge of its execution. There are general guidelines and standards for data publication covering different aspects such as specific datasets to be published, formats to be used, licensing to be applied, etc. Publication of raw machine readable data and adoption of data standards are clearly promoted.
Score > 8	There should be evidence of an active national open data strategy defined for a period of at least 2 years. The national data policy establishes a general right to reuse by means of an explicit 'open by default' statement and promotes standard licenses or terms of use to be adopted by the public sector bodies without any possible access and re-use restriction more than attribution and share-alike. General open data training and awareness programmes for civil servants are available to ensure they are capable of using open data effectively. The release of data is considered as part of the regular government performance indicators and progress reports are available.

Total scores of all questions are eventually reviewed by the ODB project management and quality assurance team and checked for conformance to the official researcher handbook guidelines. Ultimately, government officials or domain experts are invited to examine the assessment and may call for clarification, additional research or correction on the results which are subsequently forwarded to the country researcher. The expert will then re-check on his elaboration if there are any objections. This process will finally be repeated until the whole team and national counterparts approve the final scores.

2. Secondary data:

Additional secondary data is obtained from five independent surveys of third-party organizations, normalized before aggregation to the Readiness scores, and added to complement the expert survey questionnaire for a more precise measuring of the aspects for the dimension. As these are not covered in detail within the Methodology [15], the information was gathered from a Spreadsheet^[36] file which is provided by ODB on their website^[37] and contains all evaluated results and additional infor-

³⁶ODB Spreadsheet - Historical Data <https://docs.google.com/spreadsheets/d/1YbicyCIdnJjBTgQCN84YilqSyaW80yVHnALoPEj200I/edit#gid=1148338363>, accessed: 22.2.2021.

³⁷ODB - Get the Data <https://opendatabarometer.org/leadersedition/data/>

mation for the previous editions of the ODB.

These five third party surveys are:

- **World Economic Forum Global Information Technology Report** by the World Economic Forum³⁸ to measure to what extent the government has plans to utilize information and communications technology (ICT) in the future to improve the overall competitiveness.
- **World Economic Forum Global Competitiveness Index** also run by the World Economic Forum assesses the level of national companies on their general practices to adapt new technology.
- **UN E-Government Survey** by the United Nations³⁹ assesses the level of governmental online services (including e-services of national portals). The ODB uses this information as an indicator whether the government has the capabilities to provide and manage mature Open Data projects.
- **Proportion of individual using the Internet** by the International Telecommunications Union (ITU)⁴⁰ complements the "Entrepreneur business" segment in the Readiness dimension and should provide a measure on how much people use modern technology on a daily basis within a country.
- **Freedom House Political Freedoms and Civil Liberties Index** by the Freedom House⁴¹ is a value that indicates, based on the political rights and civil liberties in a country, whether it is considered a "free" country. This variable provides information about the extent to which citizens in a country experience restrictions by governments.

To answer the question of whether the additional surveys are suitable for the intended use would go beyond the scope of this work. However, as the publishers of the secondary data are renowned institutions, the quality of the data they contain will likely provide valuable information.

³⁸World Economic Forum <https://www.weforum.org/>, accessed: 24.2.2021.

³⁹United Nations <https://www.un.org/en/desa/>, accessed: 24.2.2021.

⁴⁰ITU <https://www.itu.int/en/about/Pages/default.aspx>, accessed: 24.2.2021.

⁴¹Freedom House <https://freedomhouse.org/>, accessed: 24.2.2021.

3. Dataset Assessments:

Implementation of Open Data policies is evaluated by a non-automated dataset assessment which is also within the field of responsibility of the country researcher.

Figure 4: Dataset Clusters [15]

Innovation	Social Policy	Accountability
<i>Data commonly used in open data applications by entrepreneurs, or with significant value to enterprise.</i>	<i>Data useful in planning, delivering and critiquing social policies & with the potential to support greater inclusion and empowerment.</i>	<i>Data central to holding governments and corporations to account. Based on the 'Accountability Stack'.</i>
Map Data, Public Transport Timetables, Crime Statistics, International Trade Data, Public Contracts.	Health Sector Performance, Primary or Secondary Education, Performance Data, National Environment Statistics, Detailed Census Data.	Land Ownership Data, Legislation, National Election Results, Detailed Government Budget, Detailed Government Spend, Company Register.

In the focus of this assessment are 15 categories of data rather than specific names of datasets as they might differ between countries and not all of them might actually exist for each country. A category of data can score up to 100 points. The 15 data categories are divided into groups of five and are classified based on their underlying content and for what purpose they can be used. The "Innovation" cluster for example covers the categories of data that are frequently used for the development of Open Data driven applications and other value-adding tools that can be created by organizations. Figure 4 shows the split between the categories of data. Table 1 contains the questions that we extracted from the research handbook [16] and include the modified indicators provided in figure 16.

All 100 points of the ten dataset questions are assigned based on binary ("Yes" or "No") responses, i.e., for each question the dataset scores either the full points or nothing at all, except for metric **ODB.IMPL.7**. For this question, points could be deducted if the last update of the dataset was a long time ago or if the update frequency of updates is considered too low. The official research handbook [16] provides minimum update frequencies for the respective dataset categories and the expert needs to provide the actual frequency in his answer but ultimately, the final judgment relies on the responsibility of the country

researcher. Sustainability in metric **ODB.IMPL.8** means that there is clear evidence at the portal that this dataset is kept regularly updated. Otherwise, there will not be any points assigned to this indicator.

Four questions (Indicators **ODB.IMPL. +: 3,4,5,6**) are accounting for 60% of the total score, each one accounting for 15 points due to a weighting which is intended to specifically rate the available data on its key characteristics of the Open Definition⁴² covered in chapter 2.3. This is done in order to draw a general conclusion on the Implementation status within a country. Part of this Open Definition is metric **ODB.IMPL.5** : *"Is the dataset available free of charge?"*, which indicates if anyone can freely access the data and use it for his/her purposes. Examples of the other 14 data categories can be observed in figure 4. Examples for these categories are: mapping data, land ownership data, government spend data, public transport timetable data, etc.

4. **Government self-assessment:** In the self-assessment, governments are invited to conduct a streamlined version of the expert survey questionnaire without the option to score based on personal estimation of the governmental contact. This represents valuable information for the research team as official government Open Data representatives provide the team with their own point of view and first-hand information on their Open Data policies. Furthermore, this information is used by the researchers during the peer-review phase to validate and compare the justifications brought by the country researcher.

Table 1: ODB Dataset Questions [15]

Indicator	Question	Points
ODB.IMPL.1	Does the data exist?	5
ODB.IMPL.2	Is it available online from government in any form?	10
ODB.IMPL.3	Is the dataset provided in machine-readable formats?	15
ODB.IMPL.4	Is the machine-readable data available in bulk?	15
ODB.IMPL.5	Is the dataset available free of charge?	15
ODB.IMPL.6	Is the data openly licensed?	15
ODB.IMPL.7	Is the dataset up to date?	10
ODB.IMPL.8	Is the publication of the dataset sustainable?	5
ODB.IMPL.9	Was it easy to find information about this dataset?	5
ODB.IMPL.10	Are (linked) data URIs provided for key elements of the data?	5

⁴²Open Definition <http://opendefinition.org/>, accessed: 27.2.2021.

5.1.2 Scoring

Computation of the final barometer scores is a straightforward process but with a few idiosyncrasies. First, one needs to distinguish between the questions answered by using the secondary data and the ones where evaluation is solely at the discretion of the expert, taking into account the established rules by ODB.

As the values for the metrics that require secondary data are measured on different scales, depending on the type of source, these need to be normalized before aggregation. However, this only affects the **Readiness** dimension, as these are merely part of this dimension. As for the other type of questions, each of them has a maximum score of 10 points.

The **Readiness** dimension consists of 14 questions, five of which are dependent on secondary data. The average scores for each of the dimensions are calculated by using a prior calculated average value for each of the sub-segments within the dimensions, observable in figure 2. This results in a maximum score of 100 for the first dimension.

As far as the **Implementation** goes, this dimension assesses the before mentioned dataset categories. For every dataset that is a part of the dataset assessment, there is a maximum score of 100 points that can be assigned by the expert. The named dataset categories are grouped into Accountability, Innovation and Social policy dataset clusters based on a "*qualitative analysis of the common ways in which these categories of data are used*" [15]. Finally, the averages within these clusters are computed after aggregating the total results. Each of the average cluster values makes up a third of the total score on the Implementation dimension.

The **Impacts** dimension features six questions that are again grouped into three sub-segments, two questions for each segment. Theoretically, for this dimension, a government can reach a maximum score of 60 (0-10 range per answer). However, the individual scores of the metrics of the Impacts dimension are rescaled to the same 0-100 range for the index calculation in the final step. This allows for the calculation of the overall index score of the country which is made upon the three dimensions.

In the last step, average scores for the dimensions are weighted by a third and aggregated which results in the overall country score. At the same time, this score is the basis for ranking the governments on an index value (final

ODB score) and comparing the scores amongst the participating countries.

5.2 The Open Data Maturity

The Open Data Maturity⁴³ report is a monitoring framework that is an integrated part of the European Data Portal (EDP)⁴⁴. The EDP is an initiative by the Publications Office of the European Union⁴⁵ and by the European Commission⁴⁶. Capgemini Consulting⁴⁷ is in charge of conducting the annual study for the EDP.

The Open Data Maturity study was first introduced to evaluate governments on the basis of their political strategies and leadership skills in relation to Open Data handling. For this purpose, an annual assessment was developed, which primarily deals with the guidelines, plans and implementations of governance practices for Open Data portals. This assessment covers Open Data initiatives at national and local level as well as the possible difficulties that portal providers are facing. Particular attention is paid to the development of the Open Data portals in order to selectively evaluate their features and the practices of the government with regard to their further development in order to create a basis for comparison and to evaluate the Open Data implementation at European level.

As the EDP is the official Open Data hub for European countries, the team of the European Data Portal is particularly interested to investigate whether the measures taken to develop Open Data further yield fruits. It is therefore important for them to quantify the state of Open Data within European Open Data portals on a regular basis in order to benchmark and adjust their efforts according to the findings that are the essential part of the Open Data Maturity study.

The report was first conducted in 2015, which was the year the European Data Portal was first online and it is since carried out on a yearly basis. By now, in February 2021, the report is in its sixth iteration, allowing for a

⁴³Open Data Maturity <https://www.europeandataportal.eu/en/impact-studies/open-data-maturity>, accessed: 27.2.2021.

⁴⁴European Data Portal <https://www.europeandataportal.eu/en>, accessed: 27.2.2021.

⁴⁵Publication Office of the European Union <https://op.europa.eu/en/home>, accessed: 27.2.2021.

⁴⁶European Commission <https://ec.europa.eu/>, accessed: 27.2.2021.

⁴⁷Capgemini Consulting <https://www.capgemini.com/>, accessed: 27.2.2021.

constant learning process and to further make improvements on the actual assessment process.

5.2.1 Assessment Structure

The information on the methodology and structure of the assessment process is obtained from the Method Paper [24], as well as the information available in the official 2020 report [5] of the Open Data Maturity study and will further be explained in this section.

In the latest version of the Open Data Maturity Report, the team behind the project defines four dimensions on which their research is based on. These dimensions are:

1. **Open Data Policy**
2. **Open Data Portal**
3. **Open Data Impact**
4. **Open Data Quality**

Prior to the report of 2018, the Open Data maturity was measured by two dimensions, the **Open Data Readiness** and **Portal Maturity** dimension. Since then, the scope of the analysis was expanded by two additional dimensions, **Open Data Impact** and **Open Data Quality**. By that, the European Data Portal expects to be able to measure, not only the efforts taken by governments to provide a mature data portal, but also the measures taken to secure the quality of the uploaded information. This is measured by assessing the provided metadata and quantifying the expected impact the available data has for society and businesses.

In the **Open Data Quality** dimension, the Maturity Report examines the activities the portal providers adopted to secure that the portal's metadata, current as well as historical, is up to date and accessible. Furthermore, it is checked whether there are guidelines that help data publishers in the process of publishing, choosing licenses, complying with metadata standards and monitoring the status quo of metadata quality. In contrast, the main purpose of the **Open Data Impact** dimension is to assess the emphasis of a portal provider to measure the actual impact that the availability of Open Data has on the social, environmental, economic and political areas in a certain country. This is mostly done by showcasing real-world examples

of actual projects that made use of Open Data provided by the Open Data portal. The EDP team also awards points in the scoring for the governmental support of initiatives that serve as an example for the re-use of portal provided Open Data, with the aim of adding social, environmental or economic added value.

The remaining dimensions **Open Data Policy** and **Open Data Portal** are a more generalized approach to judge how much emphasis national governments are showing in regard to Open Data and the implementation of their national and regional portals. **Open Data Policy** specifically investigates on the measures taken by the Open Data representatives to further enhance the supply and re-use of Open Data by defining strategies and policies which are then adopted at regional and local levels. Moreover, the existence of training programs for civil servants who work with Open Data are taken into account. Additionally, the **Open Data Portal** dimension focuses on the actual features national Open Data portals offer. For example, dashboards that allow data publishers to self-monitor their publishing behavior and usage as well as analytic tools for the portal providers are indispensable to reach a good maturity score. Furthermore, governments will be assigned points if they implemented monitoring tools that allow them to gain insights on the portal user behavior and adjust their portal features on this basis.

5.2.2 Scoring

The Maturity Report's scoring is done, similar to the Open Data Barometer's procedure, which allows quantifying qualitative information sourced from the answers given to a predefined questionnaire. In contrast to the Open Data Barometer, the Maturity Report does not rely on external researchers to handle the questionnaire. Instead, questionnaires are sent to national Open Data representatives that work in cooperation with the European Commission and the Public Sector Information Expert Group. However, according to the methodology [24] the results of these questionnaires are complemented with additional desk research. Further details on this complementary desk research were not provided in the methodology document, nor in the final report [25].

The four mentioned dimensions build the foundation of the assessment structure for the monitoring framework of the European Data Portal. A nation can reach a maximum of 2.600 points in the assessment process. Each of the dimensions is accounting for 650 out of 2.600 points of the scoring system and are thus weighted with 25% of the total number of points.

Like in the assessment of the Open Data Barometer, the dimensions are split up into at least three to a maximum of five sub-indices and metrics, respectively. These metrics are, again, split up into indicators which represent best case scenarios for a Open Data portal and for which a portal can reach a distinct number of points. The best-case scenario of a portal is a condition that is considered sufficient to derive a general assumption of the portals condition for each metric. This status is questioned by including a dichotomous "yes" or "no" question in the official questionnaire. If questions could be answered with "yes", this would imply a positive response and will lead to the allocation of the full points. Negative responses will lead to a score of zero for one metric. There are, however, also questions that ask for specific information on percentage values in a particular metric.

Figure 5: Scoring Distribution [24]

Dimension	Key metrics	Scoring	Weight
Open data policy		650	25%
	Policy framework	220	
	Governance of open data	220	
	Open data implementation	210	
Open data Portal		650	25%
	Portal features	240	
	Portal usage	150	
	Data provision	110	
	Portal sustainability	150	
Open data impact		650	25%
	Strategic awareness	140	
	Political impact	130	
	Social impact	120	
	Environmental impact	150	
	Economic impact	110	
Open Data Quality		650	25%
	Currency and completeness	150	
	Monitoring and measures	150	
	DCAT-AP compliance	180	
	Deployment quality and linked data	170	
Total		2600	100%

Figure 5 shows the distribution of the number of points for each dimension and corresponding sub-dimension. The questions (metrics) of the maturity report of 2020 are attached to the official country scores [48]. A good example for a question that specifically asks for information on a percentage rate is **ODM.QUAL.12**, which is one indicator of the key metric "Monitoring and

⁴⁸Download Country Scores 2020 https://www.europeandataportal.eu/sites/default/files/country_scores_2020.xlsx accessed: 5.3.2021.

measures":

"What percentage of the open data available on the national portal is accompanied by licensing information?"

To answer this question, however, a positive response to the question if the portal quantitatively monitors the metadata at all is a prerequisite. The associated "yes" or "no" question for this is:

"Do you monitor the quality of the metadata available on your portal?"

Along with the prior mentioned question, which expects a response given in percentage, thresholds (at certain percentage levels) are provided to allow for a distinction between the possible assigned points, depending on the specified percentage range. It is worth mentioning though that the differences specified in these thresholds are only five score points of the overall 650 for the **Open Data Quality** dimension. The boundaries between the predefined percentage ranges are divided into five groups. This means, for example, that if over 90% of all open data of a portal is provided with licenses, the full number of points (25) is assigned, but a portal that only has between 31 to 50% of licenses receives 10 points. This circumstance leads to the fact that a portal with a very good rate (above 90%), would have a plus of 2.3% for the **Open Data Quality** dimension, comparing to a portal which only accompanies 31-50% of their data with licenses. Despite the major difference in actual data quality, this difference in percentage would not be too significant in the overall rating, making this metric less important for the overall score of a country.

The "Open Definition"⁴⁹, which was covered in section 2.3, serves as a guideline for (digital) content to be classified as "open" or Open Data, respectively. As a matter of fact, Open Data is heavily dependent on open licenses, like for example a Creative Commons⁵⁰ license [14]. Therefore, not providing a license at all means to not fulfill the requirements for Open Data. As a result, this heavily impacts the quality of the metadata.

To assess the harvested metadata for Open Data portals, the consortium of the European Data Portal has developed a tool called "Metadata

⁴⁹Open Definition <https://opendefinition.org/od/2.1/en/>, accessed: 5.3.2021.

⁵⁰Creative Commons <https://creativecommons.org/licenses/?lang=en>, accessed: 5.3.2021.

Quality Assessment"⁵¹(MQA). It automates the process of delivering quantified metrics for metadata quality and provides helpful insights on the actual performance of Open Data portals for national portal providers. This information is quite similar to the one the Open Data Portal Watch⁵² framework provides but, at the time of this thesis, the monitoring does not exist for data before August 2020 and is only limited to countries within the EU. Therefore, a comparison with Open Data Portal Watch metrics would be limited to European countries and the short time frame the MQA covered. A detailed overview of the Portal Watch frameworks assessment process will follow in the next section.

Nevertheless, low coverage of licenses would imply an insufficient rating for the "Reusability" metric of the MQA. However, on Open Data portals, the number of data publishers might be fairly high due to the number of instances that might be assigned to upload and release Public Sector Information on Open Data portals. A bad rating on metadata would mean that the strategies for training publishers need to be reviewed and might need to be changed as well. It would therefore make sense to give greater weight to the influence of metadata quality in order to increase the quality of the data on the European portals in the long term.

Finally, for the Maturity Report assessment, the final scores for Open Data portals are computed by accumulating the number of points reached for each metric as well as weighting the total number of points for each dimension with 25%. By that, the team behind the study can quantify the results of their questionnaire and benchmark the predefined dimensions across the countries that are part of the study.

Remarks:

The metrics that are part of the **Open Data Quality** dimension, might not be appropriate to draw a general conclusion on the performance of a Open Data portal with respect to the actual metadata quality. Despite the fact that minor changes in the questionnaire took place in 2018, the overall methodology as well as the weighting of the dimensions and sub-dimensions persisted in order to guarantee comparability amongst the results of prior years. This allows for a better long-term judgment on how the actual scores

⁵¹Metadata Quality Assessment <https://www.europeandataportal.eu/mqa/?locale=en>, accessed: 5.3.2021.

⁵²Open Data Portal Watch <https://data.wu.ac.at/portalwatch/about>, accessed: 6.3.2021.

changed for each country over time. However, the team behind the Open Data Maturity Report has announced an overhaul of the questionnaire (for their 2021 edition), including the general assessment structure as well as the weighting and computation for the final metrics [25]. By the time the new study is completed, the provided information in this thesis will likely be outdated.

5.3 The Open Data Portal Watch

In the previous sections, two monitoring frameworks were covered that use qualitative or a mix between qualitative and quantitative approaches in the form of questionnaires and dataset assessments to derive a general conclusion on the development of Open Data in distinct countries. The next framework is developed in order to automate the process of assessing the quality of metadata. Whereas the Open Data Barometer and the Maturity index are primarily developed to serve as as-is information for portal managers and other instances that are responsible for the maintenance and authorities responsible for the developments of the portals, the Portal Watch framework monitors the quality of the metadata on Open Data portals and provides the accumulated information as RDF data. This gathered information can be utilized to analyze it on a finer granularity. The information in this section references the PhD dissertation of Dr. Sebastian Neumaier⁵³ [21]. Together with the researchers Univ.-Prof. Dr. Axel Polleres⁵⁴ and Dr. Jürgen Umbrich, Dr. Neumaier implemented the framework. Since the 24th calendar week of 2016, the monitoring of Open Data portals by the framework is active and up until now (February 2021) the project still provides data.

5.3.1 Assessment Structure

As already discussed in the introduction, the idea behind the Open Data Portal Watch framework was to showcase that Semantic Web standards and technologies could be used to further push the evolution of Open Data by implementing automated quality assessments of metadata. Therefore, the whole information gathering and evaluation process differs significantly from the previous monitoring frameworks. The developed framework consists of four essential blocks, "Input", "Analysis", "Backend" and the "Output".

⁵³Dr. Sebastian Neumaier <https://sebneumaier.wordpress.com/>, accessed: 7.3.2021.

⁵⁴Univ.-Prof. Dr. Axel Polleres <https://aic.ai.wu.ac.at/~polleres/>, accessed: 7.3.2021.

The Input includes metadata that is harvested from selected Open Data portals. These portals might use different software in the backend, and therefore Neumaier proposed a homogenized mapping of the different metadata schema (CKAN, OpenDataSoft, Socrata) to the DCAT⁵⁵ as well as to export and publish Schema.org⁵⁶ metadata descriptions on the web. DCAT is specifically developed to provide a metadata vocabulary in order to homogenize metadata schema obtained from different sources [1]. Additionally, there is other information stored in the backend database from accessing the web page as well as already processed data with information on the database.

The harvested data is subsequently analyzed. This includes a basic statistical metric computation for the extracted portal data as well as the mapping to the DCAT vocabulary and an "on the fly" computation of the quality metrics. The resulting datasets, which include the calculated metrics, are finally stored in the backend as well. A fraction of the final output is eventually published on the official website of the Open Data Portal Watch⁵⁷.

In the analysis, the framework covers five dimensions split up into 18 quality metrics. In the dissertation, these dimensions are defined as:

1. **Existence**
2. **Conformance**
3. **Retrievability**
4. **Accuracy**
5. **Open Data**

Existence is a dimension that is created in order to check the availability of certain DCAT metadata keys that are considered important for datasets on Open Data portals and to what degree these keys are included. **Conformance** comprises metrics, that validate a certain format for metadata keys, assuming that they exist. Figure 6 is showing the specified **Existence** and **Conformance** dimensions along with the defined metrics as well as the DCAT keys used for the computation of the final values.

⁵⁵DCAT vocabulary <https://www.w3.org/TR/vocab-dcat/>, accessed: 8.3.2021.

⁵⁶Schema.org <https://schema.org/>, accessed: 8.3.2021.

⁵⁷Open Data Portal Watch <https://data.wu.ac.at/portalwatch/>, accessed: 8.3.2021.

Figure 6: Open Data Portal Watch Dimensions 1 [21]

Metric		dcat:Dataset	dcat:Distribution
EXISTENCE			
<i>Existence of important information (i.e. exist certain metadata keys)</i>			
Access*	Is there access information for resources provided?		dcat:accessURL dcat:downloadURL
Discovery	Is information available that can help to discover/search datasets?	dct:title dct:description dcat:keyword	
Contact*	Existence of information that would allow to contact the dataset provider.	dcat:contactPoint dct:publisher	
Rights	Existence of information about the license of the dataset or resource.		dct:license
Preservation	Existence of information about format, size or update frequency of the resources	dct:accrualPeriod.	dct:format dcat:mediaType dcat:byteSize
Date	Existence of information about creation and modification date of metadata and resources respectively.	dct:issued dcat:modified	dct:issued dcat:modified
CONFORMANCE			
<i>Does information adhere to a certain format if it exist?</i>			
AccessURL*	Are the values of access properties valid HTTP URLs?		dcat:accessURL dcat:downloadURL
ContactEmail*	Are the values of contact properties valid emails?	dcat:contactPoint dct:publisher	
ContactURL*	Are the values of contact properties valid HTTP URLs?	dcat:contactPoint dct:publisher	
DateFormat	Is date information specified in a valid date format?	dct:issued dcat:modified	dct:issued dcat:modified
License	Can the license be mapped to the list of licenses reviewed by opendefinition.org?		dct:license
FileFormat	Is the specified file format or media type registered by IANA?		dct:format dcat:mediaType

Retrievability basically checks if the source can be accessed or downloaded by using the URL provided in the metadata. As far as the **Accuracy** is concerned, this dimension assesses if the resources that are available online are conform to the extracted metadata. However, as we will see in a later section, the metrics covered by **Accuracy** are not present in the data from 2016-2018 within the data dumps. This might be due to the fact that this data is gathered before the publication of the dissertation in 2019 by Neumaier [21] and the framework’s metrics/dimensions changed within this time frame. Finally, the **Open Data** dimension is implemented to evaluate if information on the resource adheres to conventions that make it possible to be classified as Open Data. For a more detailed representation of the DCAT keys that are present in the metrics, please take a look at figure 7.

Each of the quality metrics that are part of one of the defined dimensions is calculated using a Boolean evaluation function. Additionally, an aggregation function on DCAT metadata paths pointing to the corresponding value (leave), extracted by a preceding path selector function is subsequently deployed. In the case that a quality metric is defined by multiple input metadata instances, the aggregation is done by calculating the average over the

Figure 7: Open Data Portal Watch Dimensions 2 [21]

RETRIEVABILITY		
<i>Availability and retrievability of the metadata and data</i>		
Retrievable	Can the described resources be retrieved by an agent?	dcat:accessURL dcat:downloadURL
ACCURACY		
<i>Does information accurately describe the underlying resources?</i>		
FormatAccr	Is the specified file format accurate?	dct:format dcat:mediaType
SizeAccr	Is the specified file size accurate?	dcat:byteSize
OPEN DATA		
<i>Is the specified format and license information suitable to classify a dataset as open?</i>		
OpenFormat	Is the file format based on an open standard?	dct:format dcat:mediaType
MachineRead	Can the file format be considered as machine readable?	dct:format
OpenLicense	Is the used license conform to the open definition?	dct:license

result of applied path selector functions [21]. In this way, the computation of the metrics can be automated, and it also enables comparability for the extracted portal data with similar metrics from other monitoring frameworks or a comparison between different portals.

Remarks:

The necessary metadata for the analysis is publicly available in compressed (.tar.gz) folders at the official Portal Watch website⁵⁸. Available folders are provided for each crawl process separately. For every individual crawl there is a snapshot generated, named after the year (e.g., 21) and followed by the calendar week (e.g., 10) of which the snapshot was created. All files within the compressed folders are named after the respective Open Data portal and contain the collected metadata as well as the automated quality metrics. In this way, it is ensured that the data is properly structured and could be used for a time series analysis.

According to Neumaier [21], the portal crawls were first automated on a predefined, weekly time interval. Nevertheless, after gaining experience and gathering metadata for an ongoing period of time, the Portal Watch framework switched to less frequent time intervals that were intended to assess the data quality on an adapted interval, which is supposed to capture changes on Open Data portals on the basis of estimating/learning the change frequency of actual modifications on the portals themselves [26]. All archived files for

⁵⁸Portal Watch data dumps <https://data.wu.ac.at/portalwatch/data>, accessed: 16.3.2021.

the snapshots are stored in RDF, following a particular metadata schema which extends DCAT.

At the time of this thesis, however, the data within the data dumps were not complete. I.e., for the snapshots created in 2016, 2017 and 2018 there were only a fraction of portal data included within the data dumps. The crawled data for these time frames still have to be converted to be published in RDF format. However, without going into too much technical detail, transforming the data and publishing it as a part of this thesis would have exceeded the scope of this thesis. As a result, this thesis is limited to the data which was available in the data dumps.

5.4 Criteria Comparison between Monitoring Frameworks

For our analysis, it is necessary to identify categories, dimensions or metrics that allow us to compare the results between our chosen monitoring frameworks. The reference framework, on which we will specify intersecting metrics that allow investigating the Open Data Barometer's and Open Data Maturity's developments, is the Open Data Portal Watch framework. Additionally, the similarities between the Open Data Barometer and the Maturity Report will be covered. However, this comparison is not straightforward as the different frameworks tend to have different approaches on how to assess governments and their associated Open Data portals. Therefore, scoring and computation of metrics as well as the time covered by the studies differ a lot.

Whereas the Open Data Portal Watch is a purely automated quality assessment of Open Data metadata, the Open Data Maturity Report and the Open Data Barometer study are primarily survey-based, and the associated project teams also do their research and evaluation on different levels of dimensions. Additionally, the survey-based frameworks differ in terms of how and, above all, who evaluates the questionnaires. Only having one source of information like a questionnaire that is sent to government officials and consists of mainly "Yes" or "No" responses most likely will lead to different outcomes than independent research which sets thresholds and rules for scoring. That said, obviously different dimensions on which the monitoring frameworks are assessing Open Data portals and the governmental Open Data management lead to the fact that the frameworks cannot easily be compared with one another. As a result, we need to find similarities between the

particular metrics and questions on which the frameworks are based upon in order to explain different rankings and overall scores.

Another major aspect is the fact that the scope of the monitoring process is not as consistent as hoped. Of course, it is to be endorsed that the frameworks develop over time, allowing for a more precise outcome and rich information content if other and/or better metrics are introduced on how to best evaluate Open Data portals. This, however, can result in a restructuring of the whole assessment process due to the evolution of metrics, making it hard to draw long term conclusions on which dimensions and sub-categories progress had been made and on which there had not. If we have a look for example at the Open Data Maturity assessment overhaul since 2018 and compare the metrics that are listed in figure [17](#) - [20](#), then we can see that there are different dimensions which consist of modified metrics and are thus not directly comparable to prior editions of the monitoring process. Undoubtedly, to cover every single question (metric) that was or is used by one of the monitoring frameworks would not be beneficial because it would end up in a rather confusing elaboration on the actual metrics that are covered by the frameworks and disregard the overall goal of this thesis which is to analyze the evolution of Open Data.

In a first step, one needs to identify overlapping time frames based on the time under study (i.e., the time or current state of portal advancement on which the monitoring frameworks base their assessment on). This subsequently allows for a contrasting juxtaposition of data from the selected monitoring frameworks over time. Consequently, we will only use overlapping time frames for our data analysis.

In addition, we will present a selection of metrics from the Open Data Barometer and Open Data Maturity framework that cover aspects of a metadata assessment. This selection of metrics will be based on an examination to find indicators in the Open Data maturity and the Open Data Barometer framework that are suitable to match specific metadata quality metrics of Open Data Portal Watch. We will cover the framework's newest methodology approaches as well as those prior to incurred restructurings of the assessment processes to portrait the assessment processes in a holistic way and also understand how they changed over time. In a next step, the resulting selection of metrics will be grouped by the respective monitoring framework and accompanied by a justification why they are suitable to be compared to individual metadata metrics. Figure [8](#) provides an overview of the time frames, dimensions and similarities (metrics) that we identified for the com-

parison between the monitoring frameworks.

Comparison - Open Data Portal Watch & Open Data Barometer:

The relevant assessment editions of the Open Data Barometer for this comparison are:

1. **fourth edition:** July 2015 - June 2016
2. **leaders edition:** July 2016 - September 2017

The remaining editions are not overlapping with the data of the Portal Watch framework. In the second step, the metrics and corresponding dimensions that are suitable for the comparison need to be identified. Primarily, we will focus on questions of the questionnaire that overlap in terms of content or that, in any way, allow conclusions to be drawn about the quality of the metadata. For the ODB we identified two dimensions that include suitable metrics:

- **Readiness:** Within the Readiness dimension, the Open Data Barometer covers 14 questions, or rather, metrics. Out of these metrics just one question is related to the quality of metadata. In the methodology paper [15] this metric is indicated as **ODB.READ.2** and is as follows: *"To what extent is there a consistent (open) data management and publication approach?"*. At first, this might not directly be an obvious indicator of metadata quality. However, after studying the thresholds, the scoring and source guidance of the related question [16], we can observe that this metric covers the level of development of a government to manage and guarantee that data is published and accompanied by the information it needs to be classified as Open Data. Furthermore, the experts are especially encouraged to look for evidence on the adoption of metadata standards (e.g., DCAT, DCAT-AP). A government that reaches a high score on this question is likely to have a good management and training scheme for their data publishers to ensure a certain level of quality of their published materials. To investigate this hypothesis, we propose to compare this metric with the overall score of the **Existence** dimension of the Open Data Portal Watch, which checks for the existence of distinct metadata keys. These overall scores could indicate whether the training scheme had been successful and improved the completeness of the dataset information.

- **Implementation:** As already mentioned, the Implementation dimension consists of a manual dataset assessment which comprises 15 categories of data. These datasets are evaluated on the basis of ten predefined questions (points ranging from 1-10) shown in table 1 and have been retrieved from the working document of the historical scores of the governments that are part of the study⁵⁹.

A major part of these questions could be also answered by an automated assessment of the datasets if they are published on the Open Data portals. More precisely, we identified five out of ten concrete questions (**ODB.IMPL. + 1,2,3,5,6**) that could be covered by the metrics of the Open Data Portal Watch in a similar fashion. If datasets could be identified based on their resource identifier and publisher, we could use Open Data Portal Watch data to directly compare the metadata quality against the ODB assessment. However, this approach would require a certain amount of preprocessing to filter out the datasets from the data collections of the ODP. As a matter of fact, this would not be practically as there are thousands of datasets covered by the Portal Watch, depending on the size of the Open Data portal. These portals might have different naming conventions across the data publishers that are in charge of releasing Public Sector Information. But since the datasets, that are part of the ODB's non-automated assessment, are clustered into three clusters anyway, we can compare the overall metadata quality (summarized ODP metric values) to the suitable metrics from the ODB.

Metric **ODB.IMPL.1 & ODB.IMPL.2** could be answered by using the **ODP.EXIS.QN1** (Access) metric of the Open Data Portal Watch. With the existence of the access URL, it would be possible to confirm that this dataset is available online. Question **ODB.IMPL.3** of the dataset assessment could be replaced by the **ODP.OPEN.QN2** (Machine Read) metric, which checks if the provided format is machine-readable.

Additionally, indicator **ODP.CONF.QN1** (AccessURL) provides information if the access link is a valid HTTP URL and could be also used to verify the provided link. Question **ODB.IMPL.5 + ODB.IMPL.6**

⁵⁹Historical country scores <https://docs.google.com/spreadsheets/d/1YbicyCI dnJjBTgQCN84YilqSyaW80yVHnALoPEj200I/edit#gid=1240276003>, accessed: 20.3.2020

would be answered by using the **ODP.OPEN.QN.3** (OpenLicense) metric of the Portal Watch framework because it examines the provided license and checks whether to classify the license as open. An open license would, at the same time, imply that the dataset is available free of charge and vice versa.

Comparison - Open Data Portal Watch & Open Data Maturity:

For the Open Data Maturity framework, we need to distinguish between the dataset assessment prior to the overhaul in 2018 and the ones that have been conducted afterward. Before 2018 the assessment was based on two dimensions: **Open Data Readiness** and **Portal Maturity**. The Readiness dimension was further grouped into **Policy and Use** and the **Impact** category. That said, the Impact category is not comparable to the metrics of the Open Data Portal Watch framework. Here only the practices of a country are assessed in relation to their efforts to measure the impact of the provision of Open Data on politics, society and companies. Furthermore, the Open Data representatives of a country had to self-assess a level of impact that Open Data had on companies, society and politics. It was possible to choose between low, medium and high [6], which is not the most accurate way to measure the impact of Open Data. Since the Portal Watch monitoring is solely based on provided metadata keys, impact cannot be measured in any way.

For the remaining dimensions, there were four metrics that are suitable for comparison with Portal Watch. These are also extracted from the Maturity Report before the overhaul and also transformed to the uniform framework metrics, observable in figure [21] - [22] of the appendix as indicators **ODM.READ.22**, **ODM.READ.23**, **ODM.MATU.6** and **ODM.MATU.3**.

- **Open Data Readiness:** Metric **ODM.READ.22** and **ODM.READ.23** are part of the Readiness (Policy and Use) dimension and specifically asks if the data on the Portal is free of charge (**ODM.READ.22**) and if it is open licensed (**ODM.READ.23**). The Portal Watch metric **ODP.OPEN.QN3 (OpenLicense)** would answer both questions, i.e. if it is an open license, it would be at the same time free of charge.
- **Portal Maturity:** Metric **ODM.MATU.6** and **ODM.MATU.3** are part of the Portal Maturity dimension. In fact, they are part of the "Re-

usability of the data" category. In metric **ODM.MATU.6** the portal is able to score points depending on the proportion of machine-readable data which was sourced by the OpenDataMonitor⁶⁰. This metric can be compared with the information that is provided by **ODP.OPEN.QN2 (Machine Read)** metric of the Portal Watch framework. The use of an indicator based on the OpenDataMonitor seemed reasonable in 2015. However, we found out that this data is only available for the year 2015 and the project has not been continued since then. Unfortunately, we did not receive an answer to our request regarding the continuation of the OpenDataMonitor and possible access to historical data from the company behind the project. However, in the edition for 2016^[7] it was mentioned that due to the discontinuation of the OpenDataMonitor project after 2015, **ODM.MATU.6** was directly obtained from the countries if they had monitoring mechanisms in place.

Metric **ODM.MATU.3** checks whether the national portal offers the possibility to download datasets ^[7]. Although, the metrics of the Portal Watch framework can't provide information if all of the data is available as a bulk download, one of them actually can provide the information how much of the portals available datasets are retrievable through the download URL. For this information, we would need to sum up **ODP.RETR.QN1 (Retrievable)** metric for each data portal and calculate the average score (retrievable = 1, or not = 0) for the datasets to draw a conclusion on the level of availability for the downloadable datasets.

after the assessment overhaul in 2018: The ODM 2020 edition is used to find similarities between the ODP and ODM after the restructuring of the assessment process.

- **Open Data Quality:** Out of 163 indicators, see figure ^[17] - figure ^[20], only four can be compared to the Portal Watch metrics but they also do have their limitations regarding their meaningfulness. The concerned metrics are **ODM.QUAL.12**, **ODM.QUAL.23**, **ODM.QUAL.24** and **ODM.QUAL.28**.

ODM.QUAL.12 is stated as: "*What percentage of the open data available on the national portal is accompanied by licensing information?*" ^[25]. This information can exactly be answered by **ODP.EXIS.QN4**,

⁶⁰OpenDataMonitor <https://www.opendatamonitor.eu/frontend/web/index.php?r=dashboard%2Findex> accessed: 15.3.2021.

which is part of the existence dimension and tells whether the licensing information is present in the metadata. **ODM.QUAL.23** asks for the percentage of datasets at the national portal that provide a download-URL to obtain the data whereas **ODM.QUAL.24** is used to rate the percentage of metadata which includes information on where the actual data can be accessed. This simple metric could potentially be computed by the information the Open Data Portal Watch extracts. However, the current metric (**ODP.EXIS.QN1**), which either uses the DCAT key "accessURL" or "downloadURL" to specify if the access information for the resources is provided, needs only one of the metadata keys to be present in order for the metric to be classified as true [21]. As a result, to properly compare **ODM.QUAL.23** and **ODM.QUAL.23** to **ODP.EXIS.QN1**, we would need to define a new Portal Watch metric to automate this computation or use the available data in the data dumps to query the percentage of accessURL and downloadURL metadata keys separately.

ODM.QUAL.28 on the other hand, can be compared to **ODP.OPEN.QN3** of the Portal Watch's Open Data dimension. The metric **ODM.QUAL.28** from the Maturity Report is implemented in the questionnaire as: *"What percentage of datasets is made available under a standard open licence or an explicit custom open licence, in any data format including text documents?"*. Although the Portal Watch metric is a well-suited indicator in this regard, there could still be special national licenses at Open Data portals that might be conform to the Open Definition like those that are specified by the Open Knowledge Foundation [61]. But since not all of them are officially registered and listed on the website, they might not be present in the mapping list the Portal Watch framework uses for the calculation of **ODP.OPEN.QN3** [21] and therefore lead to different results.

Comparison - Open Data Barometer & Open Data Maturity:

For the Open Data Barometer and the Open Data Maturity Report, the similarities relate less to the individual metrics and more to the dimensions or sub-categories that the frameworks address. In terms of overlapping time frames, the editions of the Open Data Barometer range from 2012 to 2017. The time covered by the leaders edition was July 2016 - September 2017.

⁶¹List of conformant licenses: <http://opendefinition.org/licenses/>, accessed: 16.3.2020

Figure 8: Framework (Metrics) Similarity Matrix

	Open Data Barometer	Open Data Portal Watch	Open Data Maturity
Timeframes	<ul style="list-style-type: none"> • 2013 • 2014 • 2015 • 2016 • 2017-2018 (LE*) 	<ul style="list-style-type: none"> • 2016 • 2017 • 2018 • 2019 • 2020 • 2021 	<ul style="list-style-type: none"> • 2015 • 2016 • 2017 • 2018 • 2019 • 2020
Dimensions	<ul style="list-style-type: none"> • Readiness • Implementation • Impacts 	<ul style="list-style-type: none"> • Conformance • Retrievability • Accuracy • Open Data Existence 	<ul style="list-style-type: none"> • Open Data Readiness - Policy and Use • Portal Maturity • Open Data Readiness - Impact <hr/> <p>since 2018:</p> <ul style="list-style-type: none"> • Open Data Policy • Open Data Portal • Open Data Impact • Open Data Quality
Metrics	<ul style="list-style-type: none"> • ODB.READ.2 • ODB.IMPL.1 • ODB.IMPL.2 • ODB.IMPL.3 • ODB.IMPL.5 • ODB.IMPL.6 	<ul style="list-style-type: none"> • ODP.EXIS.QN1 • ODP.EXIS.QN4 • ODP.RETR.QN1 • ODP.OPEN.QN2 • ODP.OPEN.QN3 	<ul style="list-style-type: none"> • ODM.MATU.3 • ODM.MATU.6 • ODM.READ.22 • ODM.READ.23 <hr/> <p>since 2018:</p> <ul style="list-style-type: none"> • ODM.QUAL.12 • ODM.QUAL.23 • ODM.QUAL.24 • ODM.QUAL.28

*LE = Leaders Edition

Unfortunately, within the leaders edition the only European countries that were included were France, Germany, Italy, United Kingdom (UK), Ukraine and the transcontinental country Turkey. But since the Open Data Maturity covers only countries within the European Union, the comparison is limited to the scores of France, Germany, Italy and the former member UK. The 2015-2016 edition of the ODB would include the EU 27 countries, but 2015 is the only year in which no dashboard or downloadable online content is available for the Maturity Report. This means that no content would allow insights into the complete results for the individual countries and categories, let alone metrics. As a result, the data for the dashboard of 2016⁶² is used for comparison with the fourth edition of the ODB. In the comparison, we will cover and contrast the following dimensions and sub-categories of the Open Data Barometer to the assessment of the Open Data Maturity (prior to 2018):

- **Readiness:** Two of the ODB's Readiness sub-categories (Government policies and Government action) each consist of three metrics that examine similar topics like those that are addressed in subcategories of

⁶²ODM dashboard 2016 <https://www.europeandataportal.eu/en/dashboard/2016>, accessed: 20.3.2021.

the Maturity Report's Readiness dimension. However, the frameworks differ greatly in their assessment processes. While ODB defined a more generalized question about the extent to which there is an Open Data strategy/policy in a country - and asks the country researcher to rate them based on the provided thresholds within the handbook [16] - the questionnaire of the ODM Report assesses whether national portals are in a certain state or whether the Open Data managers have carried out certain activities within a year that ought to improve Open Data provision (e.g.: **ODB.READ.10** = "*Has your approach on how to open data changed since mid 2015?*"). A higher score for the sub-categories is therefore considered better. Consequently, the sub-category **Government policies** of the ODB assessment will be compared to the sub-category **Presence of Open Data Policy** of the ODM, observable in figures [21] and [22].

The **Extent of Coordination at national level** (ODM) will be compared to the scores of the **Government action** sub-category (ODB) to verify the developments that took place. This sub-category of the Open Data Maturity evaluates the countries according to their implementation of national data portals and assigns points to those countries that offer data portals also on a regional level. Similarly, the ODB rates the governments on their action to implement data portals at national levels. Higher scores would therefore imply a more mature evolution regarding the integration of data from local levels.

- **Impact:** Quantifying the impact of Open Data on different sections of the society is a challenging process [16]. The Open Data Barometer therefore adjusted their threshold schema so that it only offers three levels of differentiation. Depending on the sections of society (**Political, Social, Economic**) the ODB assigns zero points to metrics of the Impact dimension where there is no evidence that Open Data would improve certain aspects within those sections. Five out of ten points are assigned if there are two or more mentions in the media or credible online sources where open data is cited to have had an impact on one of the groups mentioned in the metrics in figure [16]. The full 10 points are assigned to those countries where Open Data had a significant impact on either politics, environmental/marginalized groups or on businesses and economics. Similarly, the Open Data Maturity includes the same sub-categories within their assessment process prior to 2018 and developed it further for the overhaul of the assessment - i.e., they have subdivided the metrics into a dedicated dimension (Open

Data Impact) and expanded the number of questions for them, see figure [17](#) - [18](#). However, due to the overlapping time frames only the metrics **ODM.READ.44** - **ODM.READ.52** will be used for the comparison with the ODB Impact dimension.

In order to check whether both the developments of the Open Data Barometer will correlate between the dimensions and sub-categories of the frameworks, respectively, we will review and compare the total scores for countries that are covered in both, the Open Data Barometer assessment as well as the Open Data Maturity Report.

Remarks:

It should be noted that we could identify less similarities with respect to the assessment process, the included countries, and overlapping time frames between the monitoring frameworks than expected. The frameworks were created to serve different purposes in the areas of Open Data. Additionally, the identified quantitative metrics of the Portal Watch framework address only a small part of the entire assessment process of the other frameworks.

While the Open Data Barometer is a framework that tries to capture the global developments of Open Data and provide a comprehensive benchmark for the global Open Data landscape, their approach is probably the most resource intensive. Their information acquisition is primarily based on an extensive desk research of the project team and the country researchers that are in charge to assess the national Open Data initiatives. This could be a possible reason for the fact that the last edition (Leaders edition) includes only a fraction (30 compared to 100+) of countries in the evaluation process and that there are no signs of a new edition since its publication in 2018.

The Open Data Maturity Report is a quite similar approach with respect to the included dimensions to the Open Data Barometer but does only include countries that are part of the European Union. Also, the assessment process does differ in terms of information retrieval (primarily desk research vs. external answered questionnaires). Furthermore, the ODM metrics are best suitable for assessing the current state of Open Data portals and their features, as well as the presence of government policies. For example, countries are assigned points if they provide a mechanism for portal users to rate datasets or if they generally monitor the metadata quality on their portals.

On the other side, the Open Data Portal Watch's automated quality

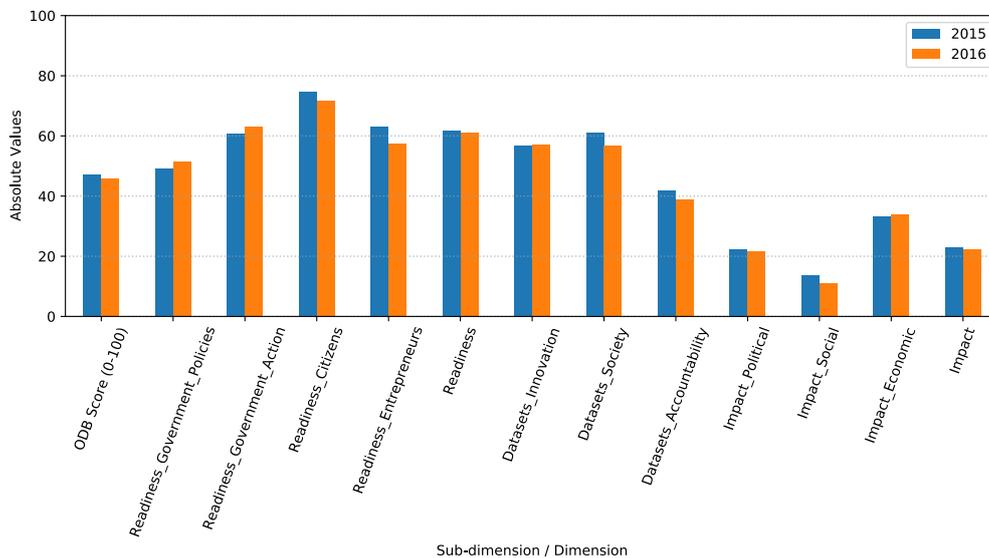
assessment of metadata provides valuable information on the actual content that is accessible at these portals. Instead of only assessing high value datasets within certain categories, the results of the Open Data Portal Watch allow for more detailed insights on the actual quality for the entire content that Open Data portals host.

6 Data Analysis

In this section, we present our approach to compare the frameworks dimensions and metrics. First, we will present a time series analysis on how the dimensions of the different monitoring frameworks changed over time. In a next step we will present the results of the metrics and dimensions that are suitable for a comparison in this context. Finally, we will show the differences and similarities that we have found based on the comparison.

6.1 Evolution of Open Data in Europe - a Framework Dimension Overview

Figure 9: Average Assessment Points - European Countries ODB 15-16



Open Data Barometer: After the year of 2014, only the scores of three editions of the Open Data Barometer assessment were available at the official

website⁶³. With the latest being the "leaders edition" and solely including five European countries in the assessment process, solely two editions (2015 2016) provided data for more than 20 European countries. The edition of 2015 covered 27 countries. In 2016, 34 European countries were included. Hence, for the analysis of the dimension evolution only the overlapping countries for both editions were used. Countries where only a part of the territory is located within the European continent (Russia and Turkey) were excluded from the overview. Furthermore, we limited the evaluation on overlapping countries. I.e., only European countries that were included in the 2015 and 2016 assessment were observed.

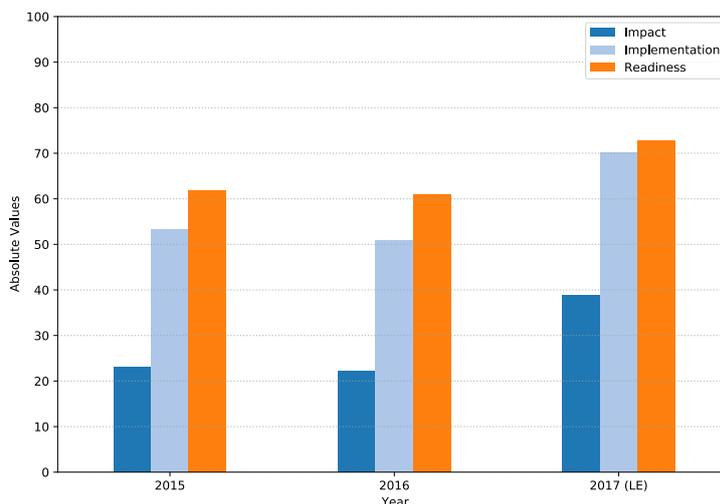
In figure 9, we plotted the average absolute scores of European countries for the ODB assessment of 2015 and 2016. The Results were grouped per sub-dimensions/dimensions and we also provided the average ODB score. The figure shows that in the majority of dimension and sub-dimension, there is a decline of the absolute scores. As a result, the overall ODB scores of European countries decreased by 1.37 points. Moreover, the biggest change from 2015 to 2016 happened within the dataset quality (**Implementation** dimension). On average, the dataset quality within Europe decreased by 2.4 points. Only policies and actions were better rated withing the **Readiness** dimension. Additionally, on average, there appeared to be a slightly higher economic impact of Open Data in European countries. In contrast, within figure 11 we see the average absolute values per dimension of five European countries in 2017 which adopted the G20 Anti-Corruption Open Data Principles or the Open Data Charter. Unsurprisingly, countries that commit to deliver certain Open Data policies, scored better on average. However, due to the limited scope regarding the number of countries that were covered within the leaders edition, the scores of 2017 were not suitable for a direct comparison with the scores of 2015 and 2016.

Open Data Maturity: We extracted the scores for the ODM from the dashboard⁶⁴ as well as the report of 2015⁶. The evolution of the dimensions that are part of the Open Data Maturity framework shows a different development than the ODB. From 2015 - 2017 there was a constant increase of the assessment scores per dimension. The dimension scores increased by

⁶³ODB Country Scores <https://docs.google.com/spreadsheets/d/1YbicyCI dnJjBTgQCN84YilqSyaW80yVHnALoPEj200I/edit#gid=1240276003>, accessed: 20.3.2021.

⁶⁴ODM dashboard <https://www.europeandataportal.eu/de/dashboard/2016>, accessed: 23.3.2021

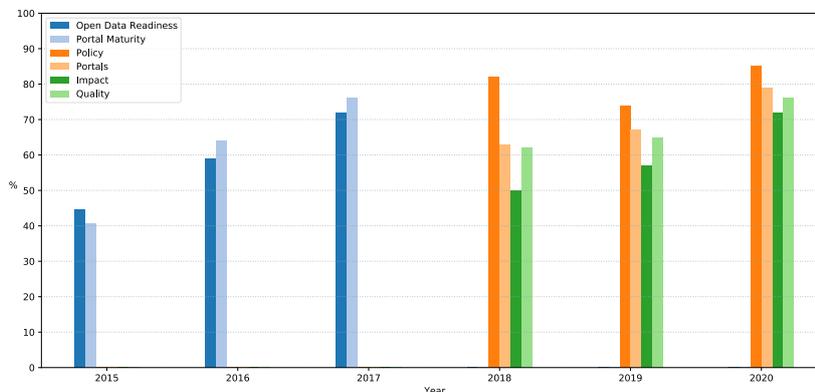
Figure 10: Average Assessment Points per Dimension - European Countries ODB 15-17



approximately 30 percent. After the assessment overhaul in 2018, the trend continued for most of the dimensions. Only in 2019 there was a decline in the Policy dimension. However, for the Open Data Maturity Report, the maximum achievable points of the dimensions changed from year to year because of the fact that the assessment metrics had been revised frequently. As a result, the comparability of the historical data has suffered, and percentage changes should be taken with a grain of salt. Nevertheless, the ODM dimensions show a clear upward trend which suggests that governments have been able to improve in every respect year on year.

Open Data Portal Watch: For the Open Data Portal Watch, we used the data from the dumps to calculate the average values for each dimension per year. The data comprises the national Open Data portals of Austria, Germany and the United Kingdom from 2016 to 2019. France is covered from 2018 to 2019 only. Greek, Hungary, Ireland, Iceland, Latvia and Slovakia are the remaining European countries from the Open Data Portal Watch data in this time frame and are only available for 2019. The data from 2020 to 2021 is, at the moment (April 2021), also listed at the website but unfortunately not accessible to download. As a result, the value for each dimension will be affected depending on the year in question due to the limited number of European countries that are covered by the data.

Figure 11: ODM Average Percentage per Dimension - European Union 15-20

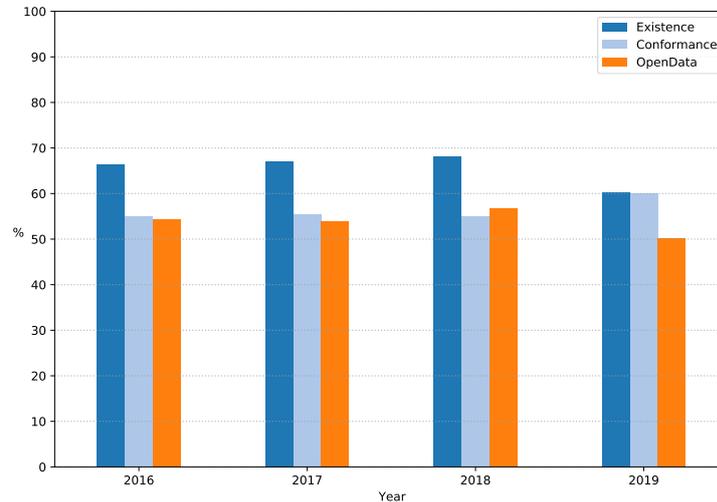


6.2 Comparative Comparison of Monitoring Metrics Results

Although the analysis is limited due to the amount of data available, the included countries of the Open Data Portal Watch are compared with those of the Open Data Barometer and of the Open Data Maturity Report to showcase whether or not the comparison provides valuable insights. Even the comparison of a small number of metrics can provide a good indication of the extent to which the developments in qualitative metrics are comparable with those in quantitative metrics. For this reason, this comparison is limited to the data from Austria, Germany and the UK, as these are already included in the Portal Watch data over a longer period of time (2016-2019) and therefore more meaningful results can be achieved.

The Open Data Barometer includes the non-automated dataset assessment on categorized high-value datasets and uses these results to rate the implementation status on Open Data portals. The identified ODP metrics for comparison in section 5.4 were used to compute the differences between those and the dataset assessment of the countries mentioned above to see whether these results can be quantitatively confirmed. The ODB assessment includes only a small part of Open Data datasets on the portals. Therefore, a deliberate comparison with the metadata quality for the entire datasets (from ODP) enables a more holistic view and reveals the actual implementation status. Hence, the average scores for the dataset categories, extracted

Figure 12: ODP Average Percentage per Dimension 16-19



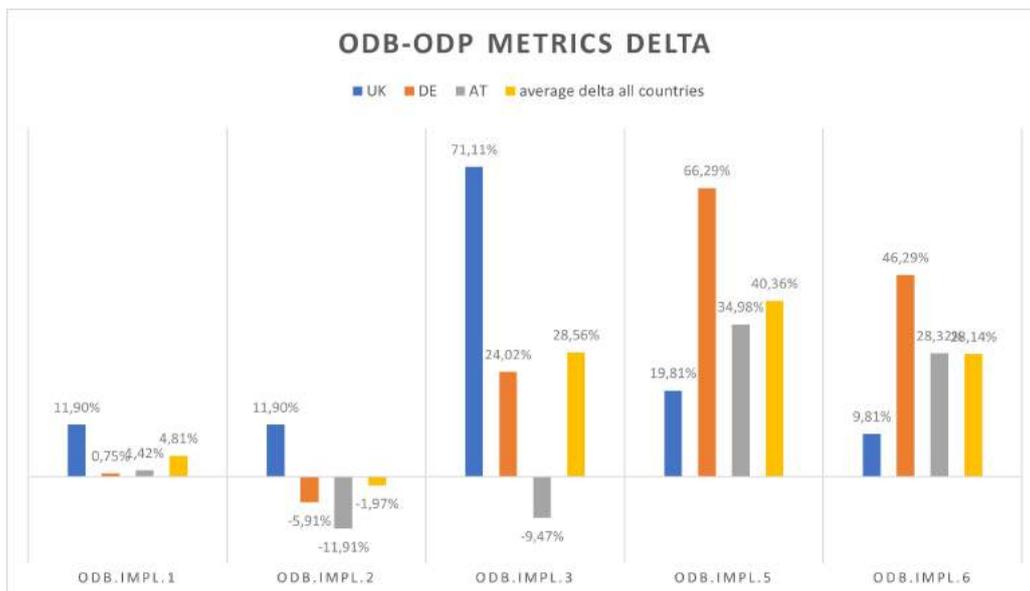
from the historical online data⁶⁵, were used to calculate the percentage of the maximum achievable points per metric. The percentages of achieved points were used to compare them to the ODP metrics, which gives information about the level of completeness, conformance and openness of the data. Deviations (in percentage) of the ODP metrics from those of the ODB assessment can be seen in figure 13. Unfortunately, the correlation between **ODB.READ.2** and the **Existence** metrics cannot be investigated due to the lack of data (only three countries and two years covered) which would be needed to compute reliable results.

When it comes to the Open Data Maturity Report, the quantitative metrics from the Open Data Portal Watch framework are not as suitable to confirm developments from the ODM as they are for the ODB assessment. This is primarily due to the fact that the questionnaire is mostly build upon closed questions (yes/no responses). Nevertheless, depending on the version of assessment process (after 2018 or prior), we identified certain metrics that examine similar, if not the same, characteristics like for example the percentage of datasets that are accompanied by an open license. Unfortunately, the ODM assessment that was elaborated on and on which basis we selected

⁶⁵ODB historical data <https://docs.google.com/spreadsheets/d/1YbicyCI dnJjBTgQCN84YilqSyaW80yVHnALoPEj200I/edit#gid=1240276003>, accessed: 25.3.2021

suitable metrics was the 2020 edition. As the ODM metrics differ over time, we used metrics that cover similar aspects like our identified ones but may be formulated slightly different. Additionally, the Portal Watch Data from 2016-2018 in the data dumps does not contain the metric for the "Retrievability" dimension. Therefore, metric **ODM.MATU.3** could not be compared with the **ODP.RETR.QN1** metric. Because only the ODP data from 2016-2019 was available in the data dumps, the Open Data Maturity scores from Austria, Germany and the United Kingdom (2016 and 2018) assessment editions were compared to the metrics from the Open Data Portal Watch data (2016 and 2018).

Figure 13: ODB-ODP Metric Comparison



- **Differences:**

When looking at the average differences of the **ODB.IMPL. + 3,5,6** metrics and the comparison metrics **ODP.OPEN.QN2, ODP.OPEN.QN3** we can see that these results differ substantially. Obviously, this is due to the fact that the ODB assessment covers only a fraction of the datasets and is likely to show differences when comparing these metrics

to overall metadata quality of the datasets.

Additionally, the majority of Open Data Maturity metrics that were identified for comparison (prior and after 2018), showed different results than the ones that were gathered by preparing the Open Data Portal Watch data for analysis. For the percentage of machine-readable data on the national Open Data portal, the national contact persons (AT, DE, UK) that were in charge to answer the questionnaire, rated the percentage of datasets that are provided in a machine readable format with 90% (expect UK in 2016). However, as figure 14 shows, the percentage of **ODP.OPEN.QN2** heavily differs to the provided answers which were extracted by the 2016⁶⁶ and 2018⁶⁷ edition. The same holds true for **ODP.OPEN.QN3**, which provides information on the percentage of datasets that are accompanied by a registered open license. Especially for the year 2018, the share of those datasets that were equipped with open licenses was very low for Austria and Germany. However, as the metrics differ enormously from the results of the UK data, there could have been an error in the crawling and transformation process of the ODP (e.g., metadata key not retrievable).

Lastly, the correlation between the sub-dimensions of the Open Data Barometer and the Open Data Maturity assessment in 2016 had been investigated. Only data (scores) for countries that was covered in both, ODM and ODB was used (24 European countries). The correlation coefficients are presented in figure 15. When comparing the identified sub-dimensions, mentioned in section 5.4, there are no sub-dimensions present which exceed the threshold of 0.75. Hence, we cannot see a strong positive nor negative correlation for the 2016 assessments. The strongest correlation coefficient is present between the Government Policies (ODB) and Presence of Open Data Policy (ODM) with a value of 0.62. However, this value and the remaining would not be sufficient to confirm a strong relation between the scores in the intersecting sub-dimensions.

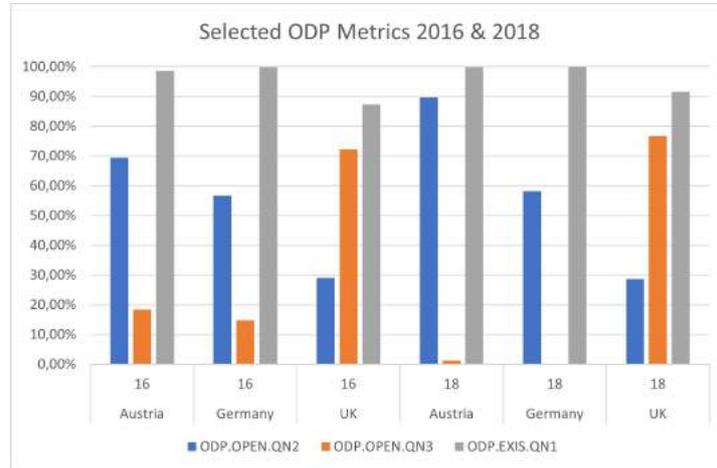
- **Similarities:**

For the Open Data Barometer, few similarities can be observed when

⁶⁶ODM 2016 scores https://web.archive.org/web/20170717085028/https://www.europeandataportal.eu/sites/default/files/landscaping_2016_individual_scoring.xlsx, accessed: 25.3.2021

⁶⁷ODM 2018 scores https://www.europeandataportal.eu/sites/default/files/country_scores_2018.xlsx, accessed: 25.3.2021

Figure 14: Selected ODP Metric Comparison



contrasting the results to those of the Open Data Portal Watch. Merely **ODB.IMPL 1+2** received good scoring results and at the same time the metrics of the Open Data Portal Watch showed similar results. However, these metrics only examine whether the datasets actually exist and are available online. Whilst these scores were contrasted to the **ODP.EXIS.QN.1** metric, which was used to rate how many of the datasets listed on the Open Data Portal actually contained access information to retrieve the data, these results do not fully have the same meaning.

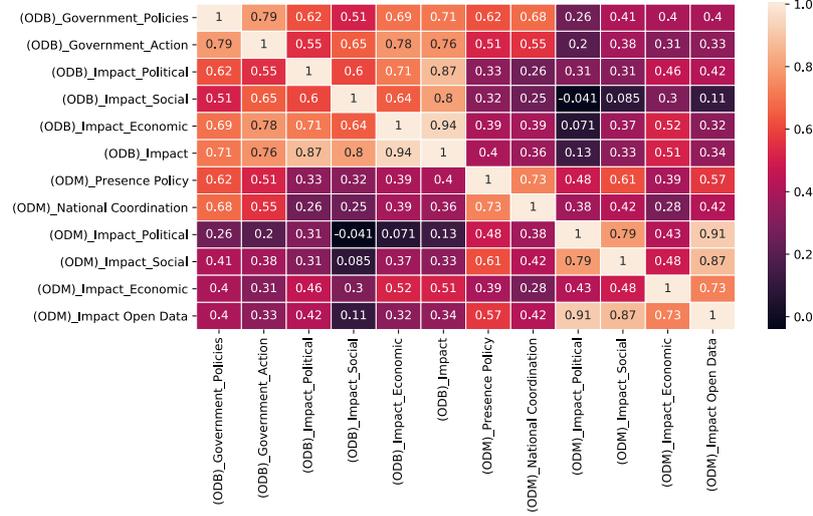
Only **ODM.QUAL.23** can be confirmed by an Open Data Portal Watch metric (**ODP.EXIS.QN1**). While this part of the Open Data Maturity assessment in 2018 rates the percentage of downloadable datasets at the national Open Data portal and therefore only has a maximum of 25 scoring points, the impact of this question on the overall score is very limited.

6.3 Analysis Remarks and Findings

The introduced Open Data Portal Watch metrics are not particularly suitable to compare with qualitative metrics of the frameworks that were covered within this thesis.

In the rare case that metrics cover the same characteristics (e.g., percentage of open licenses, or retrievable datasets), different results were observed.

Figure 15: ODM-ODB Sub-Dimension Correlation



This could be due to the fact that national Open Data providers do not have sufficient knowledge or resources to cover metadata quality monitoring, during the crawling process of the Open Data Portal Watch difficulties appeared or, for example in the case of ODB, only a fraction of datasets were used as a representation for metadata quality.

7 Limitations

One of the biggest limitations of this thesis is the lack of detailed data and especially qualitative metrics that were suitable for the comparison with quantitative metrics obtained from the Open Data Portal Watch framework. Even in the case where we found metrics that could be used for comparison, the historical values were not always available by the monitoring frameworks, despite the fact that detailed evaluation data of the remaining years had been provided. This was primarily an issue with the Maturity Report data of 2015 and 2017. In addition, the data in the data dumps were not complete, which limited the data analysis. Hence, a long-term comparison to analyze the evolution of these metrics and dimensions was only possible to a very limited extent.

Furthermore, the comparison of outcomes between the monitoring frameworks is limited to the overlapping time frames and metrics illustrated in figure 8. Additional intersecting dimensions or metrics would have led to

more meaningful analysis results. Moreover, the provided thresholds of the metrics scores made it impossible to precisely compare the quantitative metrics of the Open Data Portal Watch to confirm the observations of the other frameworks. I.e., the scores the governments received in the ODM or ODB were not directly comparable to the percentages resulting from the quantitative assessment of the Open Data Portal Watch.

Although the research on this topic was quite extensive, there may be global Open Data monitoring frameworks that were not mentioned during this thesis and actually covered some aspects of the suggestions that will be made based on the findings in the conclusion.

8 Conclusion

If Open Data has made a step forward and gained momentum within the last years cannot be answered in a generalized manner. This is due to the fact the available monitoring frameworks, have different approaches and cover different aspects of Open Data and a varying number of Open Data portals. Although this thesis is limited to the extent of covered countries, different trends were observable when looking at the developments of dimensions in the various monitoring frameworks.

The Open Data Maturity Report, for example, places more emphasis on the efforts and actual measures of the governments in relation to their Open Data programs. Furthermore, the features of a portal are particularly important when evaluating the countries. As the name "Open Data Maturity" already suggests, the framework determines how mature a national Open Data program is by benchmarking members of the European Union using the scores derived from the predefined questionnaire. **ODM.PORT.52**, which is stated as: *"Was there a user satisfaction survey concerning the national portal conducted in the past year?"*, might be a suitable metric to cover the efforts the Open Data providers place at data re-users but they can, by no means, be a measurement for the quality of relevant content (datasets) of those portals.

Other monitoring frameworks have appeared to discontinue their assessment processes, likely due to the resource intensive process. The Open Data Barometer for example has not launched a new edition since the publication of their leaders edition that comprised data from 2016-2017. Similarly, other frameworks that are more user-centered by focusing on dataset quality, like

the crowd-sourced Global Open Data Index⁶⁸ or the metadata-based OpenDataMonitor⁶⁹ have not released new editions of their assessments either.

As a result, what has been found is that there is a need for a more unified global Open Data monitoring framework which especially emphasizes on the data quality from a user perspective [2] [20]. This framework should:

- **Give more weight to the actual user-rated quality of datasets at Open Data portals when evaluating them** - by including average scores of a dataset rating scheme for example.
- **Comprise information of an automated metadata quality assessment** - to gain information about the quality of published metadata on the national portals and subsequently identify publishers that provide bad quality metadata. This information could be sourced by independent metadata monitoring frameworks like the Open Data Portal Watch to reduce resource expenditure.
- **Provide recommendations on how to improve the current state of national Open Data for countries with low ratings** - for example by referring to suitable training programs for data publishers to improve the quality of datasets or metadata to enhance reusability and findability of datasets.
- **Base the assessment on a constant selection of metrics for an ongoing period of time** - this ensures the comparability between different points of time and draw conclusions on long term developments.

This thesis highlights the need for a comprehensive framework. Similarly, this can be taken from an online article⁷⁰ by Open Data Watch. As an international non-profit organization of data experts that highly engages with companies managing and producing statistical data, they are highly familiar with the work in the Open Data domain. In this post they criticize the incompleteness of three assessment frameworks (Open Data Barometer, Open Data Index, and the World Bank's Statistical Capacity Indicator) covering relevant Open Data aspects such as country coverage, quality, and openness of data.

⁶⁸Global Open Data Index <https://index.okfn.org/about/>, accessed: 25.4.2021

⁶⁹OpenDataMonitor <https://www.opendatamonitor.eu/frontend/web/index.php?r=dashboard%2Findex>, accessed: 25.4.2021

⁷⁰Indexes of Data Quality and Openness: <https://opendatawatch.com/blog/indexes-of-data-quality-and-openness/>, accessed: 3.4.2021

9 Future Work

Work still needs to be done in the domain of Open Data. Firstly, the creation of a global monitoring framework for Open Data portals and the associated data quality. Given the framework comprises the suggestions made in the previous section, a holistic index on how well countries have implemented and are managing their Open Data initiatives would be provided. While certainly not all governments would be interested in collaborating or have sufficient resources to tackle the publication of Public Sector Information, the majority would highly benefit from a global monitoring project.

Secondly, long term tracking (e.g.: five years) of the developments by making use of a framework that does not, or at least, barely changes the metrics within the assessment process would enable researchers to conduct better time series analysis. As a consequence, these monitoring metrics could be used with a finer granularity to identify high- and low-quality data publishers and, if needed, initiate measures for quality improvement. This would benefit data re-users in particular and would almost certainly further the development of Open Data.

Lastly, responsible framework operators could revise their existing approaches on how to assess national Open Data portals. Because they often lack appropriate metrics to cover (meta)data quality, they have the opportunity to enrich their assessment processes and put more emphasis on provided content. For those frameworks where suitable metrics already exist, we found that low-quality metadata scores were not weighted enough in the calculation of total scores to ensure a significant difference from high-quality providers. As a result, the scoring systems should be revised to guarantee a more mature assessment process.

References

- [1] Riccardo Albertoni, David Browning, Simon Cox, Alejandra G. Beltran, Andrea Perego, and Peter Winstanley. Data catalog vocabulary (dcat) - version 2, 2020. <https://www.w3.org/TR/vocab-dcat/>, accessed: 11. February 2021, 17:10.
- [2] Charalampos Alexopoulos, Anneke Zuiderwijk, Yannis Charalabidis, E. Loukis, and Marijn Janssen. Designing a second generation of open data platforms: Integrating open data and social media. pages 230–241, 09 2014.
- [3] Tim Berners-Lee. Linked data, 2009. <https://www.w3.org/DesignIssues/LinkedData.html>, accessed: 5. December 2020, 10:32.
- [4] Tim Berners-Lee, James Hendler, and Ora Lassila. The semantic web. *Scientific American*, 284(5):34–43, 2001.
- [5] Marit Blank. Open data maturity report 2019, 12 2019. doi: 10.2830/073835, https://www.europeandataportal.eu/sites/default/files/edp_landscaping_insight_report_n6_2020.pdf, accessed: 1. January 2021, 14:00.
- [6] Wendy Carrara, Sander Fischer, and Eva van Steenbergem. Open data maturity report 2015, 01 2020. doi: 10.2830/23574, https://www.europeandataportal.eu/sites/default/files/edp_landscaping_insight_report_n1_-_final.pdf, accessed: 19. February 2021, 21:00.
- [7] Wendy Carrara, Cosmina Radu, and Heleen Vollers. Open data maturity report 2016, 01 2020. doi: 10.2830/41983, https://www.europeandataportal.eu/sites/default/files/edp_landscaping_insight_report_n2_2016.pdf, accessed: 24. February 2021, 16:00.
- [8] European Commission. Open data, 2020. <https://ec.europa.eu/digital-single-market/en/open-data>, accessed: 24. November 2020, 14:41.
- [9] European Commission. Open data portals, 2020. <https://ec.europa.eu/digital-single-market/en/open-data-portals>, accessed: 24. November 2020, 14:37.

- [10] Council of European Union. Directive 2013/37/eu of the european parliament and of the council of 26 june 2013 amending directive 2003/98/ec on the re-use of public sector information text with eea relevance, 2013. <https://eur-lex.europa.eu/eli/dir/2013/37/oj>.
- [11] Richard Cyganiak, David Wood, and Markus Lanthaler. Rdf 1.1 concepts and abstract syntax. w3c recommendation 25 february 2014. <http://www.w3.org/TR/rdf-concepts/>, 2014.
- [12] DBpedia. Dbpedia version 2016-10, 2017. <https://wiki.dbpedia.org/develop/datasets/dbpedia-version-2016-10>, accessed: 19. February 2021, 14:00.
- [13] Erik Duval, Wayne Hodgins, Stuart Sutton, and Stuart L Weibel. Metadata principles and practicalities. *D-lib Magazine*, 8(4):1–10, 2002. <http://www.dlib.org/dlib/april02/weibel/04weibel.html> accessed: 19. February 2021, 13:10.
- [14] Open Knowledge Foundation. Open definition 2.1, 2020. <https://opendefinition.org/od/2.1/en/>, accessed: 9. January 2021, 17:18.
- [15] World Wide Web Foundation. Open data barometer - leaders edition, odb methodology - v1.0., 2017. <http://opendatabarometer.org/doc/leadersEdition/ODB-leadersEdition-Methodology.pdf>, accessed: 5. January 2021, 13:20.
- [16] World Wide Web Foundation. Open data barometer - leaders edition, research handbook - v1.0, 2017. <https://opendatabarometer.org/doc/leadersEdition/ODB-leadersEdition-ResearchHandbook.pdf>, accessed: 5. January 2021, 13:20.
- [17] World Wide Web Foundation. Open data barometer - leaders edition, 2018. <https://webfoundation.org/research/open-data-barometer-leaders-edition/>, accessed: 3. January 2021, 15:30.
- [18] Graham Klyne, Jeremy J. Carroll, and Brian McBride. Resource description framework (rdf): Concepts and abstract syntax. w3c recommendation 10 february 2004. <http://www.w3.org/TR/rdf-concepts/>, 2004.
- [19] Jens Lehmann, Robert Isele, Max Jakob, Anja Jentzsch, Dimitris Kontokostas, Pablo Mendes, Sebastian Hellmann, Mohamed Morsey, Patrick Van Kleef, Sören Auer, and Christian Bizer. Dbpedia - a large-scale,

- multilingual knowledge base extracted from wikipedia. *Semantic Web Journal*, 6, 01 2014.
- [20] Renata Máchová and Martin Lněnička. Evaluating the quality of open data portals on the national level. *Journal of Theoretical and Applied Electronic Commerce Research*, 12(1):21–41, 2017.
- [21] Sebastian Neumaier. *Semantic enrichment of open data on the Web - or: how to build an open data knowledge graph*. Vienna, 2019.
- [22] Sebastian Neumaier, Jürgen Umbrich, and Axel Polleres. Automated quality assessment of metadata across open data portals. *Journal of Data and Information Quality*, 8, 10 2016.
- [23] Axel Polleres, Maulik Rajendra Kamdar, Javier David Fernández, Tania Tudorache, and Mark Alan Musen. A more decentralized vision for linked data. *Semantic Web*, (Preprint):1–13, 2020.
- [24] European Data Portal. Measuring open data maturity, 2020. https://www.europeandataportal.eu/sites/default/files/method-paper_insights-report_n6_2020.pdf, accessed: 5. February 2021, 13:00.
- [25] Laura van Knippenberg. Open data maturity report 2020, 12 2020. doi: 10.2830/619187, https://www.europeandataportal.eu/sites/default/files/edp_landscaping_insight_report_n6_2020.pdf, accessed: 19. February 2021, 20:00.
- [26] Thomas Weber, Johann Mitöhner, Sebastian Neumaier, and Axel Polleres. Odarchive – creating an archive for structured data from open data portals. In Jeff Z et al. Pan, editor, *The Semantic Web – ISWC 2020*, pages 311–327, Cham, 2020. Springer International Publishing.
- [27] Frederika Welle Donker and Bastiaan van Loenen. How to assess the success of the open data ecosystem? *International Journal of Digital Earth*, 10(3):284–306, 2016.
- [28] Anneke Zuiderwijk, Keith Jeffery, and Marijn Janssen. The potential of metadata for linked open data and its value for users and publishers. *Journal of e-Democracy and Open Government*, 4:222–244, 12 2012.

Appendix A Unified Quality Metrics

Figure 16: Open Data Barometer Metrics [16](#)

Open Data Barometer Qualitative Indicators

Dimension	Sub-Category	Indicator	Instance
Readiness	Government policies	ODB.READ.1	To what extent is there a well-defined open data policy and/or strategy in the country?
		ODB.READ.2	To what extent is there a consistent (open) data management and publication approach?
	Government action	ODB.READ.3	Importance of ICT to government vision of the future.
		ODB.READ.4	To what extent is there a well-resourced open government data initiative in the country?
		ODB.READ.5	Government online services index
	Citizen & Civil Society	ODB.READ.6	To what extent are city, regional and local governments running their own open data initiatives?
		ODB.READ.7	To what extent does the country have a functioning Right to Information law (RTI) / Freedom of Information (FoI) law?
		ODB.READ.8	To what extent is there a robust legal or regulatory framework for protection of personal data in the country?
		ODB.READ.9	Political Freedoms and Civil Liberties
		ODB.READ.10	To what extent are civil society and/or information technology professionals engaging with the government regarding open data?
	Entrepreneur & business	ODB.READ.11	To what extent is training about open data available for individuals or businesses wishing to increase their skills or build businesses to use (open) data?
		ODB.READ.12	To what extent is government directly supporting a culture of innovation with open data through competitions, grants or other support actions?
		ODB.READ.13	Firm-level technology absorption.
		ODB.READ.14	Internet users (per 100 people)
Implementation	Dataset assessment	ODB.IMPL.1	Does the data exist?
		ODB.IMPL.2	Is it available online from government in any form?
		ODB.IMPL.3	Is the dataset provided in machine-readable and reusable formats?
		ODB.IMPL.4	Is the machine-readable and reusable data available as a whole?
		ODB.IMPL.5	Is the dataset available free of charge?
		ODB.IMPL.6	Is the data openly licensed?
		ODB.IMPL.7	Is the dataset up to date?
		ODB.IMPL.8	Is the dataset being kept regularly updated?
		ODB.IMPL.9	Was it easy to find information about this dataset?
		ODB.IMPL.10	Are data identifiers provided for key elements in the dataset?
Impact	Political	ODB.IMPA.1	To what extent has open data had a noticeable impact on increasing government efficiency and effectiveness?
		ODB.IMPA.2	To what extent has open data had a noticeable impact on increasing transparency and accountability in the country?
	Social	ODB.IMPA.3	To what extent has open data had a noticeable impact on environmental sustainability in the country?
		ODB.IMPA.4	To what extent has open data had a noticeable impact on increasing the inclusion of marginalised groups in policy making and accessing government services?
	Economic	ODB.IMPA.5	To what extent has open data had a noticeable positive impact on the economy?
		ODB.IMPA.6	To what extent are entrepreneurs successfully using open data to build new businesses in the country?

Figure 17: Open Data Maturity Metrics 2020 (1)

OPEN DATA MATURITY ASSESSMENT 2020

Dimension	Sub-Category	Indicator	Instance
Open Data Policy	Policy framework	ODM.POLI.1	Is there an open data policy in your country?
		ODM.POLI.2	Is there an open data strategy in your country?
		ODM.POLI.3	Has this strategy/policy been updated in the past 24 months?
		ODM.POLI.4	Does the strategy/policy include an action plan with measures to be implemented in the open data field?
		ODM.POLI.5	Does the strategy/policy outline measures to incentivise the publication of and access to real-time or dynamic data?
		ODM.POLI.6	Does the strategy/policy outline measures to support the re-use of open data by the public sector?
		ODM.POLI.7	Does the strategy/policy outline measures to support the re-use of open data by the private sector?
		ODM.POLI.8	Does the strategy/policy mandate carrying out and maintaining a data inventory by public bodies, whether at national or local levels?
		ODM.POLI.9	If yes, do these data inventories also include the data collected by public bodies that cannot be published as open data?
		ODM.POLI.10	Have high-value domains and/or data sets been identified and prioritised for publication in your country?
	Governance of open data	ODM.POLI.11	Are there measures in place to assist other stakeholders' involvement in this prioritisation process?
		ODM.POLI.12	Is there a governance structure in place that enables the participation and/or inclusion of various open data stakeholders?
		ODM.POLI.13	Are the governance structure and its operating model published online and accessible to the public?
		ODM.POLI.14	Does the governance structure assist data providers with their open data publication process?
		ODM.POLI.15	Does the governance model include the appointment of official roles in civil service that are dedicated to open data (e.g., data officers / stewards)?
		ODM.POLI.16	Are all the open data initiatives (local/ regional/ national) facilitated and supported at the national level?
		ODM.POLI.17	What is the model used for governing open data in your country?
		ODM.POLI.18	Does the national open data policy incentivise and support open data initiatives at local or regional level?
		ODM.POLI.19	To what degree do local / regional public bodies conduct open data initiatives?
		ODM.POLI.20	Is there a regular exchange of knowledge or experiences between the different public sector bodies active in the open data field?
	Open data implementation	ODM.POLI.21	Is there a regular exchange of knowledge or experiences between public sector bodies and open data re-users?
		ODM.POLI.22	Are there annually held national, regional or local events (e.g. hackathons, conferences, users meet-ups) to promote open data in your country?
		ODM.POLI.23	Who organises most open data related events? Multiple answers are possible.
		ODM.POLI.24	Is there a guidebook at national level to assist data providers in their publication process?
		ODM.POLI.25	Are there data publication plans in place at national/regional/local or public body level?
		ODM.POLI.26	Are there any processes run at national level to ensure that the open data plans/strategy are implemented (e.g., monitoring)?
		ODM.POLI.27	If yes, would you describe the status of implementation as satisfactory/neutral/unsatisfactory?
		ODM.POLI.28	Are there any processes in place to assess if public sector bodies are charging for data above marginal cost?
		ODM.POLI.29	If yes, to what degree is data provided by public sector bodies free of charge?
		ODM.POLI.30	If not all datasets, how has this degree changed compared to the previous year?
		ODM.POLI.31	Are local/regional data sources discoverable via the national portal?
		ODM.POLI.32	If yes, to what degree are existing local/regional sources harvested?
		ODM.POLI.33	Are there activities conducted at national level to assist real-time and/or dynamic data holders in their publication process?
		ODM.POLI.34	Is there a professional development or training plan for civil servants working with data?
		ODM.POLI.35	If yes, do these training activities offer a publicly recognised certification and are they formally recognised as professional development training within the public bodies?



retrieved from: https://www.europeandataportal.eu/sites/default/files/country_scores_2020a.xlsx

Figure 18: Open Data Maturity Metrics 2020 (2)

Open Data Impact	
Strategic awareness	ODM.IMPA.1 Do you observe a trend in the public bodies in your country to increase measuring the re-use of their own data and of other public sector bodies' data?
	ODM.IMPA.2 Are there any processes run by public bodies to measure the re-use of their own data and of the data re-used from other public bodies (e.g., monitoring)?
	ODM.IMPA.3 Are there activities in place to support and incentivise public bodies in measuring the re-use of their own or other public bodies' open data (e.g., monitoring)?
	ODM.IMPA.4 Do you observe a trend in the public bodies in your country to increase measuring the re-use of their own data by the public?
	ODM.IMPA.5 Are there any processes run by public bodies to measure the re-use of their own data by the public (e.g., monitoring)? For example by web-analytics from the national portal.
	ODM.IMPA.6 Are there activities in place to support and incentivise public bodies in measuring the re-use of their own data by the public (e.g., monitoring)?
	ODM.IMPA.7 Do you have a definition of the impact of open data in your country?
	ODM.IMPA.8 If yes, do you have a methodology in place to measure the impact of open data?
	ODM.IMPA.9 Have you or other public bodies performed any activities in the past year to monitor the political impact of open data?
	ODM.IMPA.10 Has open data had a low/medium/high impact on increasing government efficiency, e.g. reducing operational costs? Please provide examples of how open data that has been used to develop applications on the topic.
Political impact	ODM.IMPA.11 Has open data had a low/medium/high impact on increasing government effectiveness, e.g. improving quality of service delivery? Please provide examples of how open data that has been used to develop applications on the topic.
	ODM.IMPA.12 Has open data had a low/medium/high impact on increasing transparency and accountability in your country? Please provide examples of how open data that has been used to develop applications on the topic.
	ODM.IMPA.13 Is open data used in policy-making processes in your country (i.e. are public administrations making use of the data as evidence to be included in their daily operations)?
	ODM.IMPA.14 Is open data used in decision-making processes in your country (i.e. are public administrations making use of the data as evidence to be included in their daily operations)?
	ODM.IMPA.15 Are there civil society initiatives that are open data driven and aim to tackle a problem identified in the political field?
	ODM.IMPA.16 Have there been any studies conducted in the past year that focus on assessing the political impact of open data?
	ODM.IMPA.17 Have you or other public sector stakeholders active in this field launched any activities in the past year to monitor the social impact of open data?
	ODM.IMPA.18 Has open data had a low/medium/high on increasing the inclusion of marginalised groups in society? Please provide examples of how open data that has been used to develop applications on the topic.
	ODM.IMPA.19 Has open data had a low/medium/high on raising awareness concerning housing in the city? Please provide examples of how open data that has been used to develop applications on the topic.
	ODM.IMPA.20 In your view and in light of your answers to the previous questions, has open data had a low/medium/high impact on society in your country?
Social impact	ODM.IMPA.21 Are there civil society initiatives that are open data driven and aim to tackle a problem identified in the social field?
	ODM.IMPA.22 Have there been any studies conducted in the past year that focus on assessing the social impact of open data?
	ODM.IMPA.23 Have you or other public sector stakeholders active in this field launched any activities in the past year to monitor the environmental impact of open data?
	ODM.IMPA.24 Has open data had a low/medium/high on raising awareness on the water and/or air quality in your country? Please provide examples of how open data that has been used to develop applications on the topic.
	ODM.IMPA.25 Has open data had a low/medium/high on raising awareness on the noise level in cities? Please provide examples of how open data that has been used to develop applications on the topic.
	ODM.IMPA.26 Has open data had a low/medium/high on dealing with waste management aspects? Please provide examples of how open data that has been used to develop applications on the topic.
	ODM.IMPA.27 Has open data had a low/medium/high on enabling more environmental-friendly transport systems in cities? Please provide examples of how open data that has been used to develop applications on the topic.
	ODM.IMPA.28 In your view and in light of your answers to the previous questions, has open data had a low/medium/high impact on the environment in the country?
	ODM.IMPA.29 Are there civil society initiatives that are open data driven and aim to tackle a problem identified in the environmental field?
	ODM.IMPA.30 Have there been any studies conducted/commissioned to assess the impact of open data in the environmental field or a sub-dimension of this field?
Environmental impact	ODM.IMPA.31 Have you or other public sector stakeholders active in this field launched any activities in the past year to monitor the economic impact of open data?
	ODM.IMPA.32 Has open data had a low/medium/high impact at macro-economic level in your country? Please provide examples of how open data that has been used to develop applications on the topic.
	ODM.IMPA.33 Has open data had a low/medium/high impact at micro-economic level in your country? Please provide examples of how open data that has been used to develop applications on the topic.
	ODM.IMPA.34 Has open data had a low/medium/high impact on economic benefits for public administrations in your country? Please provide examples of how open data that has been used to develop applications on the topic.
	ODM.IMPA.35 In your view and in light of your answers to the previous questions, has open data had a low/medium/high impact on the economy in the country?
	ODM.IMPA.36 Are there civil society initiatives that are open data driven and aim to tackle a problem identified in the economical field?
	ODM.IMPA.37 Have there been any studies conducted/commissioned to assess the impact of open data at micro- or macro-economic level in your country?

6

⁶retrieved from: https://www.europeandataportal.eu/sites/default/files/country_scores_2020a.xlsx

Figure 19: Open Data Maturity Metrics 2020 (3)

Open Data Portal		
Portal features	ODM.PORT.1 Is there a national portal in your country for making open data and PSI discoverable?	
	ODM.PORT.2 Does the national portal offer an advanced data search function (multiple field search, filter options etc.)?	
	ODM.PORT.3 Does the national portal offer the possibility for users to download data sets?	
	ODM.PORT.4 Does the national portal offer the possibility for users to search by file format?	
	ODM.PORT.5 Does the national portal offer the possibility for users to search by data domain?	
	ODM.PORT.6 Does the national portal offer a SPARQL search query feature?	
	ODM.PORT.7 Does the national portal offer a feedback mechanism at data set level?	
	ODM.PORT.8 Does the national portal offer a general feedback mechanism for users?	
	ODM.PORT.9 Does the national portal offer the possibility for users to request data sets?	
	ODM.PORT.10 If yes, what is the frequency of these requests?	
	ODM.PORT.11 Are these requests and their progress status presented in a transparent manner on the national portal?	
	ODM.PORT.12 Does the portal team monitor the extent to which these requests result in the publication of the requested data?	
	ODM.PORT.13 If yes, to what degree do these requests result in the publication of the requested data?	
	ODM.PORT.14 Does the national portal allow users to see what data exists but cannot be made available as open data?	
	ODM.PORT.15 Does the national portal offer the possibility for users to receive notifications when new data sets are available on the national portal (RSS, ATOM feeds, email notifications etc)?	
	ODM.PORT.16 Does the national portal provide a mechanism for users to rate data sets?	
	ODM.PORT.17 Does the national portal offer the possibility to link documentation and supporting materials to a given data set?	
	ODM.PORT.18 Does the national portal have a designated area to showcase use cases?	
	ODM.PORT.19 Does the national portal provide the possibility for users to submit their own use cases?	
	ODM.PORT.20 Does the national portal reference the data sets that the showcased use cases are based on?	
	ODM.PORT.21 Does the national portal include a discussion forum for users (whether data providers or re-users)?	
	ODM.PORT.22 Does the national portal offer a preview function for tabular data?	
	ODM.PORT.23 Does the national portal provide guidelines and tools for data publishers to improve the quality of their data publication?	
ODM.PORT.24 Does the national portal provide guidelines and tools for data publishers to improve the quality of their data publication?		
ODM.PORT.25 Is the national portal mobile responsive?		
ODM.PORT.26 Are log analytics performed on the portal to gain insights into its usage?		
ODM.PORT.27 If yes, are these insights (e.g., traffic and usage statistics) used to improve the portal?		
ODM.PORT.28 How many unique visitors visit the national portal on average per month?		
ODM.PORT.29 What is the typical profile of the portal visitor?		
ODM.PORT.30 Does this profile match the type of audience your national portal wants to cater to?		
ODM.PORT.31 What percentage of the visitors to the national portal is foreign?		
ODM.PORT.32 Do you monitor what keywords are used to search for data and content on the portal?		
ODM.PORT.33 Do you take measures to optimise the search and discoverability of content (data and editorial)?		
ODM.PORT.34 Do you monitor the most and least consulted pages?		
ODM.PORT.35 What are the top five data categories on the portal, with 1 being the most popular one?		
ODM.PORT.36 What data sets are most frequently consulted on the portal, with 1 being the most popular one?		
ODM.PORT.37 Is the metadata on your portal available in clear plain language as well to enable both humans and machines to read and understand it?		
ODM.PORT.38 Is the metadata describing the data sets accessible via a publicly available API?		
ODM.PORT.39 Do you run log analytics on the API usage?		
ODM.PORT.40 If yes, what percentage of outgoing portal traffic is generated by API usage only?		
ODM.PORT.41 Do all public sector data providers contribute data to the portal?		
ODM.PORT.42 If no, did you identify the data providers that are not yet publishing data on the national portal?		
ODM.PORT.43 Were there concrete actions taken to assist these data providers with their publication process?		
ODM.PORT.44 Does the national portal enable access to real-time or dynamic data?		
ODM.PORT.45 If yes, what percentage of metadata links to such data?		
ODM.PORT.46 Does the national portal provide a section where non-official data (not stemming from official sources, such as crowd-/ community-contributed data) can be published?		
ODM.PORT.47 Does the national portal have a strategy to ensure its sustainability?		
ODM.PORT.48 Does this strategy include a description of the portal's target audience and measures to reach this audience?		
ODM.PORT.49 Do you take actions to promote the national portal's activities and the available open data?		
ODM.PORT.50 Is your national portal active on social media?		
ODM.PORT.51 Are the portal's source code as well as relevant documentation and artifacts made available to the public?		
ODM.PORT.52 Was there a user satisfaction survey concerning the national portal conducted in the past year?		
ODM.PORT.53 Is there a process by which the portal is reviewed and improved regularly?		
ODM.PORT.54 If yes, what is the frequency of these reviews?		
ODM.PORT.55 Does the portal provide a monitoring of performance indexes in terms of the number of datasets published, the distribution across categories, number of visitors and how these changed over time?		
ODM.PORT.56 If yes, by what percentage has the number of datasets available on your portal increased compared to the same period last year?		
ODM.PORT.57 Does this monitoring allow data publishers to view the main performance indexes for their metadata or data featured on the national portal?		
Portal sustainability		
	Data provision	
Portal usage		



Figure 20: Open Data Maturity Metrics 2020 (4)

Open Data Quality	
Currency and completeness	ODM.QUAL.1 Is there a pre-defined approach to ensure that metadata is kept up-to-date?
	ODM.QUAL.2 What percentage of the metadata is obtained from the source automatically, rather than edited manually?
	ODM.QUAL.3 To what degree is the metadata describing the datasets available on the national portal updated within 1 day from the moment its primary source is updated?
	ODM.QUAL.4 Excluding the metadata that is updated within 1 week, to what degree is the metadata describing the datasets available on the national portal updated within 1 week from the moment its primary source is updated?
	ODM.QUAL.5 Excluding the metadata that is updated within 1 month, to what degree is the metadata describing the datasets available on the national portal updated within 1 month from the moment its source is updated?
	ODM.QUAL.6 To what degree do the datasets cover the full period from when they were first published until today? (for example, a dataset that started in collecting annual data since 2008, contains the data from 2008 until 2019 or if it is monthly data it contains data from 2008 until april 2020)
Monitoring and measures	ODM.QUAL.7 Do you monitor the quality of the metadata available on your portal?
	ODM.QUAL.8 Do you publish information on the quality of the metadata available on the portal?
	ODM.QUAL.9 Do you publish guidelines (e.g. written materials) and have tools in place, to assist publishers in choosing an appropriate licence for their data?
	ODM.QUAL.10 Did you develop your own open licence / licensing suite to foster the publication of open data in your country?
	ODM.QUAL.11 Do your open data publication/licensing guidelines provide recommendations for the use of Creative Commons (CC) licences or of your own licensing suite?
	ODM.QUAL.12 What percentage of the open data available on the national portal is accompanied by licensing information?
	ODM.QUAL.13 How has the percentage of data sets accompanied by licensing information changed compared to the same period last year?
	ODM.QUAL.14 Across all data sets you distribute, how many different licences are used on your portal?
	ODM.QUAL.15 Are there regular activities conducted, or mechanisms in place, to incentivise and / or assist data providers in the publication of data in machine-readable formats?
	ODM.QUAL.16 Are there regular activities conducted, or mechanisms in place, to incentivise and / or assist data providers in the publication of high-quality metadata?
	ODM.QUAL.17 Do you supply data providers with documentation on DCAT-AP (e.g. EDP factsheets, materials published on the EC websites such as the JoinUp platform, your own documentation)?
	ODM.QUAL.18 What is the percentage of metadata on your portal that is DCAT-AP compliant, in terms of mandatory classes? (agent, catalogue, data set, literal, resource)
	ODM.QUAL.19 What is the percentage of metadata on your portal that uses DCAT-AP recommended classes? (category, category scheme, distribution, licence document)
	ODM.QUAL.20 What is the percentage of metadata on your portal that uses DCAT-AP optional classes? (catalogue record, checksum, document, frequency)
	ODM.QUAL.21 Do you investigate the most common causes for the lack of DCAT-AP compliance?
	ODM.QUAL.22 If yes, what are the main causes for the lack of DCAT-AP compliance?
	ODM.QUAL.23 What is the percentage of data sets whose metadata provides a reference to where the data can be downloaded, or its API accessed ("download-URL" in the DCAT-AP specification)?
ODM.QUAL.24 What is the percentage of data sets whose metadata provides a reference to a web page from where the data can be accessed ("access-URL in the DCAT-AP specification)?	
ODM.QUAL.25 Is there a national extension of the DCAT-AP standard developed for your country?	
ODM.QUAL.26 Do you use a model (such as the 5-Star Open Data or FAIR) to assess the quality of deployment of data in your country?	
ODM.QUAL.27 If yes, please briefly describe.	
ODM.QUAL.28 Do you conduct activities to promote and familiarise data providers with ways to ensure higher quality data (such as promoting the model referenced in the previous question)?	
ODM.QUAL.29 What percentage of datasets is made available under a standard open licence or an explicit custom open licence, in any data format including text documents?	
ODM.QUAL.30 How has the percentage of data sets accompanied by licensing information changed compared to the same period last year?	
ODM.QUAL.31 Of the percentage indicated in the previous question (Question 28), what percentage of datasets is made available under a standard open licence or an explicit custom open licence, in a structured data format?	
ODM.QUAL.32 Of the percentage indicated in the previous question (Question 30), which part is also in an open and machine-readable format?	
ODM.QUAL.33 Of the percentage indicated in the previous question (Question 31), what percentage of datasets also consistently use Uniform Resource Identifiers?	
ODM.QUAL.34 Of the percentage indicated in the previous question (Question 32), what percentage of datasets also links to other renowned sources to provide additional context for the users, e.g. in a linked data fashion?	
Deployment quality and linked data	ODM.QUAL.35 Do you use a model (such as the 5-Star Open Data or FAIR) to assess the quality of deployment of data in your country?
	ODM.QUAL.36 If yes, please briefly describe.
	ODM.QUAL.37 Do you conduct activities to promote and familiarise data providers with ways to ensure higher quality data (such as promoting the model referenced in the previous question)?
	ODM.QUAL.38 What percentage of datasets is made available under a standard open licence or an explicit custom open licence, in any data format including text documents?
	ODM.QUAL.39 How has the percentage of data sets accompanied by licensing information changed compared to the same period last year?
	ODM.QUAL.40 Of the percentage indicated in the previous question (Question 28), what percentage of datasets is made available under a standard open licence or an explicit custom open licence, in a structured data format?
	ODM.QUAL.41 Of the percentage indicated in the previous question (Question 30), which part is also in an open and machine-readable format?
	ODM.QUAL.42 Of the percentage indicated in the previous question (Question 31), what percentage of datasets also consistently use Uniform Resource Identifiers?
	ODM.QUAL.43 Of the percentage indicated in the previous question (Question 32), what percentage of datasets also links to other renowned sources to provide additional context for the users, e.g. in a linked data fashion?



retrieved from: https://www.europeandataportal.eu/sites/default/files/country_scores_2020a.xlsx

Figure 21: Open Data Maturity Metrics 2016 (1)

OPEN DATA MATURITY ASSESSMENT 2016

Dimension	Sub-Category	Indicator	Instance
Open Data Readiness (Impact)	Presence of Open Data Policy	ODB-READ.1	Is there an (Open) Data policy in your country?
		ODB-READ.2	Are there policies supporting the re-use of Public Sector Information (within public administration, by the private sector)?
		ODB-READ.3	Is your Open Data Policy the same as your Public Sector Information Policy?
		ODB-READ.4	Was the (Open) Data policy in your country updated since April 2015?
		ODB-READ.5	Is there a national (Open) Data portal in your country?
		ODB-READ.6	Are there also regional, local portals?
		ODB-READ.7	Can all the different Public Sector Data Holders (ministries, departments, etc) upload data themselves to the portal?
		ODB-READ.8	What is the frequency in collecting the data from the relevant public sector data holders?
		ODB-READ.9	Is there a pre-defined approach to ensure the datasets are up-to-date?
		ODB-READ.10	Has your approach on how to open data changed since mid 2015?
		ODB-READ.11	Are there priority domains for the release of data?
		ODB-READ.12	Is Open Data used in decision making (i.e. are public administrations making use of the data as evidence to be included in evidence based policy making)?
		ODB-READ.13	Have there been national or regional events (e.g. hackathon, events) held annually to promote Open Data and PSI re-use (organized by whichever organization or group)?
		ODB-READ.14	Is there a national strategy in terms of Open Data for the next five years?
		ODB-READ.15	Is the revised PSI Directive transposed?
	Extent of Coordination at National Level	ODB-READ.16	Do you have national guidelines on the publication of Public Sector Information?
		ODB-READ.17	Do cities or regional governments run their own Open Data initiatives (e.g. dedicated data policies, portals, etc.)?
		ODB-READ.18	Would you describe the number of city/regional Open Data initiatives in your country as very few / some / many initiatives?
		ODB-READ.19	Are city/regional portals and datasets integrated into the national Open Data portal?
		ODB-READ.20	If yes, how many portals are integrated?
		ODB-READ.21	Is there a possibility for private companies and NGOs to share their datasets on the national portal?
		ODB-READ.22	Is all data available on the portal open licensed (i.e. open licence included in the metadata)?
		ODB-READ.23	Is there a national data policy that provides or stimulates the use of a standard licence (or suite of licences, for example Creative Commons licences)?
		ODB-READ.24	How many unique visitors go to your Portal every month?
		ODB-READ.25	Is your national Open Data portal accessible via a specific API?
	Licensing Norms	ODB-READ.26	What is the typical profile of your visitors?
		ODB-READ.27	What percentage of your visitors is foreign?
		ODB-READ.28	What proportion of traffic towards the portal is generated by API (i.e. Machine traffic)?
		ODB-READ.29	What are the top ten users/stakeholders of your portal (to the best of your knowledge)?
		ODB-READ.30	Which datasets are most often consulted? Please check the appropriate boxes from the 14 domains identified by the GB as shown in Annex 1: Domains listed in the GB Open Data Charter.
		ODB-READ.31	What are the top five domains, 1 being the most consulted?
		ODB-READ.32	What datasets are most often downloaded?
		ODB-READ.33	What are the least consulted domains, 1 being the least consulted?
		ODB-READ.34	What are the least downloaded datasets, 1 being the least downloaded?
		ODB-READ.35	What do you believe are the main reasons explaining the popularity?
	Use of Open Data	ODB-READ.36	What actions are you planning at national level to increase up-take of the data?
		ODB-READ.37	Would you be willing to share further information on traffic, statistics and high level profiles of your portal's users?
		ODB-READ.38	If yes, are you the relevant contact point to collect this information?
		ODB-READ.39	Did you see any changes in the re-use of data since mid 2015?
		ODB-READ.40	Did you launch activities to monitor the re-use of Open Data?
		ODB-READ.41	Do you support the re-use of Open Data?
		ODB-READ.42	Do you launch any specific communication activities to promote your Portal or Open Data in general?
		ODB-READ.43	Since June 2015, have you launched any activities to monitor the impact of Open Data?
		ODB-READ.44	Has Open Data had a low/medium/high impact on increasing government efficiency and effectiveness?
		ODB-READ.45	Has Open Data had a low/medium/high impact on increasing transparency and accountability in the country?
	Political Impact	ODB-READ.46	Has Open Data had a low/medium/high impact on environmental sustainability in the country?
		ODB-READ.47	Has Open Data had an impact on increasing the inclusion of marginalized groups in policy making and accessing government services?
		ODB-READ.48	Have there been macro-economic studies assessing the market value of Open Data (i.e. estimating the euro value of Open Data in your country)?
	Social Impact	ODB-READ.49	Have there been studies on assessing better service delivery for users of public services?
		ODB-READ.50	Are there any additional studies on the impact of Open Data in your country that you would like to share with the research team?
	Economic Impact	ODB-READ.51	What kind of funding model has been developed when setting up the portal?
		ODB-READ.52	



retrieved from: https://web.archive.org/web/20170717085028/https://www.europeandataportal.eu/sites/default/files/landscaping_2016_individual_scoring.xlsx

Figure 22: Open Data Maturity Metrics 2016 (2)

Portal Maturity		
Usability of the portal	ODM.MATU.1	Does your national Open Data portal offer a feedback mechanism on datasets?
	ODM.MATU.2	Does your national Open Data portal offer the possibility to contribute to datasets?
	ODM.MATU.3	Does your national Open Data portal offer the possibility to download datasets?
	ODM.MATU.4	Does your national Open Data portal offer the possibility to access datasets?
Re-usability of the portal	ODM.MATU.5	Does your national Open Data portal offer the possibility to download all datasets at once (in bulk, with one click)?
	ODM.MATU.6	What proportion of the data is available in machine readable format? [self assessment]
	ODM.MATU.7	Does your national Open Data portal offer the possibility to search on file format?
	ODM.MATU.8	Does your national Open Data portal include the possibility to request datasets?
	ODM.MATU.9	Does your national Open Data portal include a news section?
	ODM.MATU.10	Does your national Open Data portal include examples of the re-use of Open Data?
Spread of data across domains	ODM.MATU.11	Does your national Open Data portal offer the possibility to search on data domain?
	ODM.MATU.12	Does your national Open Data portal distinguish between more than 10 data domains?
	ODM.MATU.13	Does your national Open Data portal include datasets from more than five data publishers (departments, institutions)?

⁶

⁶retrieved from: https://web.archive.org/web/20170717085028/https://www.europeandataportal.eu/sites/default/files/landscaping_2016_individual_scoring.xlsx

Figure 23: Open Data Portal Watch Metrics [21]

Open Data Portal Watch Qualitative Indicators

Dimension	Sub-Category	Indicator	Instance
Existence	Access	ODP.EXIS.QN1	Is there access information for resources provided?
	Discovery	ODP.EXIS.QN2	Is information available that can help to discover/search datasets?
	Contact	ODP.EXIS.QN3	Existence of information that would allow to contact the dataset provider.
	Rights	ODP.EXIS.QN4	Existence of information about the license of the dataset or resource.
	Preservation	ODP.EXIS.QN5	Existence of information about the format, size or update frequency of the resources
	Date	ODP.EXIS.QN6	Existence of information about creation and modification date of metadata and resources respectively.
Conformance	Access URL	ODP.CONF.QN1	Are the values of access properties valid HTTP URLs?
	Contact Email	ODP.CONF.QN2	Are the values of contact properties valid emails?
	Contact URL	ODP.CONF.QN3	Are the values of contact properties valid HTTP URLs?
	Date Format	ODP.CONF.QN4	Is date information specified in a valid date format?
	License	ODP.CONF.QN5	Can the license be mapped to the list of licenses reviewed by opendefinition.org ?
	File Format	ODP.CONF.QN6	Is the specified file format or media type registered by IANA?
Retrievability	Retrievable	ODP.RETR.QN1	Can the described resources be retrieved by an agent?
Accuracy	Format Accuracy	ODP.ACCU.QN1	Is the specified file format accurate?
	Size Accuracy	ODP.ACCU.QN2	Is the specified file size accurate?
Open Data	Open Format	ODP.OPEN.QN1	Is the file format based on an open standard?
	Machine Read	ODP.OPEN.QN2	Can the file format be considered as machine readable?
	Open License	ODP.OPEN.QN3	Is the used license conform to the open definition?