OpenDataMonitor

D2.6 OPEN DATA STAKEHOLDER REQUIREMENT REPORT 2

PROJECT

Acronym:	**OpenDataMonitor**
Title:	Monitoring, Analysis and Visualisation of Open Data Catalogues, Hubs and Repositories
Coordinator:	SYNYO GmbH
Reference:	**611988**
Type:	Collaborative project
Programme:	FP7-ICT
Start:	November 2013
Duration:	24 months
Website:	http://project.opendatamonitor.eu
E-Mail:	office@opendatamonitor.eu

Consortium:

SYNYO GmbH, Research & Development Department, Austria, (SYNYO)

Open Data Institute, Research Department, UK, (ODI)

Athena Research and Innovation Center, IMIS, Greece, (ATHENA)

University of Southampton, Web and Internet Science Group, UK, (SOTON)

Potsdam eGovernment Competence Center, Research Department, Germany, (IFG.CC)

City of Munich, Department of Labor and Economic Development, Germany, (MUNICH)

Entidad Publica Empresarial Red.es, Shared Service Department, Spain, (RED.ES)

DELIVERABLE

Number:	**D2.6**
Title:	**Open Data Stakeholder Requirement Report 2**
Lead beneficiary:	IfG.CC
Work package:	WP2: Research Studies and Stakeholder Analysis
Dissemination level:	Public (PU)
Nature:	Report (RE)
Due date:	March 31, 2015
Submission date:	**March 30, 2015**
Authors:	**Sirko Hunnius**, IfG.CC **Grégoire Njacheun-Njanzoua**, IfG.CC **Elpida Prasopoulou,** ODI **Jack Hardinges,** ODI **Jamie Fawcett**, ODI **Tom Heath**, ODI
Contributors:	**Nacho Sanchez**, RED.ES **Sonia Castro**, RED.ES **Wolfgang Glock**, MUNICH **Bernhard Jäger**, SYNYO **Vassilis Kaffes**, ATHENA **Dimitrios Skoutas**, ATHENA
Reviewers:	**Peter Leitner**, SYNYO **Bernhard Jäger**, SYNYO

Acknowledgement: The OpenDataMonitor project is co-funded by the European Commission under the Seventh Framework Programme (FP7 2007-2013) under grant agreement number 611988.

MANAGEMENT SUMMARY

Goal and Approach

This report explores and analyses the roles, interests and understandings of the various stakeholders in the open data ecosystem. The goal is to identify patterns of understandings and interests in order to serve these stakeholders' information needs through the OpenDataMonitor platform. Stakeholders around open data can be found within public administrations (e.g. public managers, data generators, open data strategists) as well as outside public administrations (e.g. open data users, activists) and at the general policy level, the policy level of a specific instantiation of open data (e.g. a national or local project or catalogue) or at the operational level, a taxonomy developed in the first iteration of this report.

Method

The research in this report follows a mixed methods design consisting of a combination of qualitative and quantitative methods in order to capture and analyse in depth the practices of the various stakeholders in the open data ecosystem. The overall research design is organised around two major phases each exploring different aspects of the open data ecosystem. Phase I, presented in detail in a previous iteration of this report, aimed to identify the main stakeholders in the open data ecosystem as well as core issues related to the publication and use of open data. The second phase (II) explores in more depth the experiences of open data publishers and users across the European Union. Research covers all the stakeholder levels identified in D2.4 in a standardised survey with 243 completed questionnaires. It also focuses on a specific and very dynamic category of open data stakeholders: commercial organisations working with open data to create innovative products and services. These were investigated in a second standardised survey which reached 79 completed questionnaires as well as in-depth interviews for ten case studies.

Understandings and Interests

The results of our research in this report illustrate that overall, the various stakeholders' views are relatively aligned. Stakeholders across the board – whether within or outside public administration, whether they are active at the policy-level or the operational level of open data – attribute similarly high importance to aspects like e.g. open formats, open licenses, machine-readability and that data are made available free of charge. That data come with meaningful and high quality metadata are seen as only slightly less important. Interestingly, the amount of open data itself is considered the least important aspect among the surveyed properties. This technically-oriented understanding of open data is further substantiated by the findings of the second research strand. Both the survey and the follow up interviews show that timeliness and provenance of the data are quite important for

commercial stakeholders. Equally important are the licensing schemes attached to the open datasets.

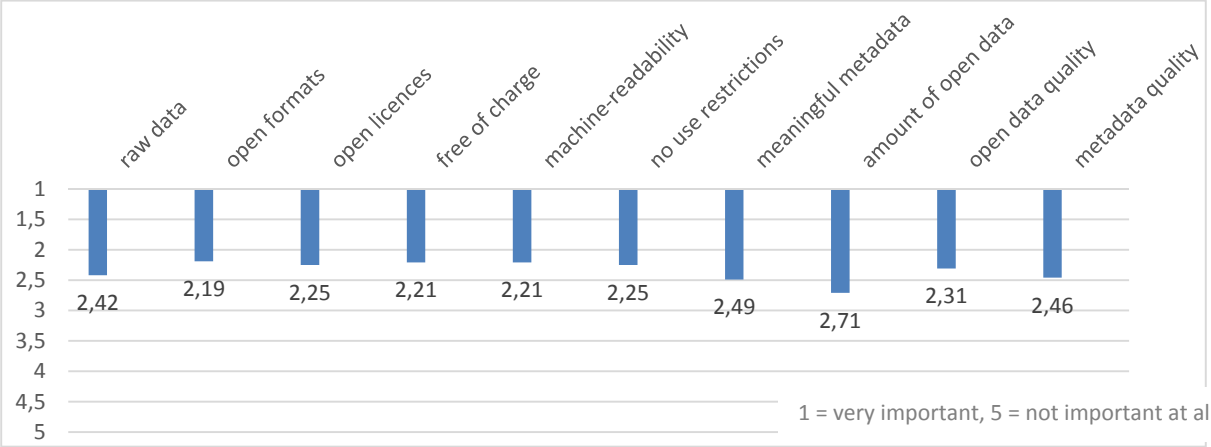

Figure 1 Importance of selected aspects for the success of open data

However, differences can be found in regard to what content interests different stakeholder groups. In general, “Transport and Traffic”-data, “Environment and Climate”-data, “Finance and Budget”-data and geographic data attract the highest interest. Also, stakeholders distinguish noticeably between the different kinds of open data, e.g. Sports-data being mentioned only twice among the 243 respondents, while Transport and Traffic data were mentioned 85 times. Thus, clear interests can be identified.

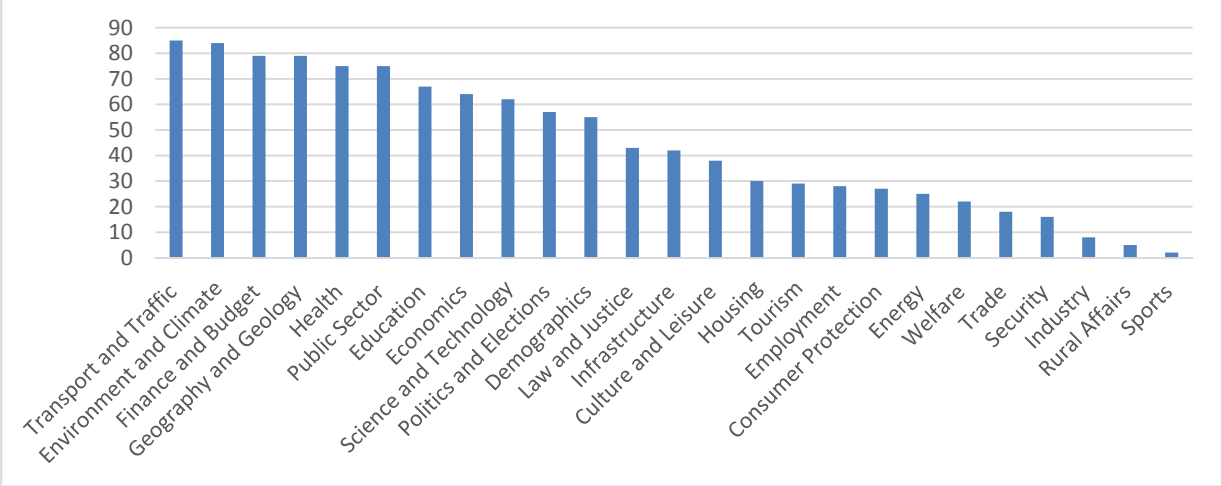

Figure 2 Relevant Content of Open Data

Furthermore, which content different stakeholder groups consider interesting varies in between these groups in an interesting way. While respondents who self-identify as open data activists are more concerned about data that is consistent with the Freedom of Information (FOI)/transparency-tradition, like e.g. Politics and Elections-data and Public Sector-data. These kinds of data rank

predominantly not highest among users. Even less important are these data seen among stakeholders in the public administration who appear to favour data that is less politicised.

Table 1 Frequency and Ranking of Relevant Content of Open Data by selected stakeholder groups

	Activists		Users		Public Administration	
Data category	frequency	rank	frequency	rank	frequency	rank
Politics and Elections	10.90%	1	5.78%	7	3.60%	12
Public Sector	10.26%	3	7.14%	2	6.00%	6
Finance and Budget	9.62%	4	5.78%	6	8.00%	4
Law and Justice	6.41%	5	5.10%	10	2.60%	15
Environment and Climate	3.85%	10	6.80%	5	9.60%	1
Geography and Geology	4.49%	8	4.76%	12	9.40%	2
Transportation and Traffic	5.13%	7	4.76%	11	8.40%	3
Science and Technology	3.21%	13	7.14%	4	5.60%	7

Stakeholders at the policy-level

Stakeholders at the policy-level of open data who are seen to most strongly support and push for open data are civic and political activists and activist groups in general. The group attributed with the second highest support are supra- and international bodies, like e.g. the European Commission, the World Bank and the OECD. Decision makers in politics and the higher echelons of public administration are seen to markedly lower support open data. Slightly less than 30 per cent of respondents see these two groups as pushing for open data. The group seen to provide the least support for open data are businesses and business associations whom only 15.5 per cent of respondents see as pushing for open data. This is in stark contrast to at least the demand and actual use of open data in the UK's ecosystem, which has been scrutinised in more detail for this report. One possible explanation is the small size of companies using open data, which in 70 per cent of the cases have fewer that ten employees and might thus be barely vocal. Journalists and media industry occupy a middle position between civic activists and commercial users, being seen by 38.7 per cent of respondents to support open data.

A large variety of stakeholders is involved in decisions about which data can be published as open data. Eight different stakeholder groups are involved in at least 10 per cent of the decisions, most frequently the administrative department that generates the data (%), an open data officer (42%), a data protection officer (36%) and the legal department (35%). Open data policies, prevalent in 56 per cent of the cases, exhibit a similarly large heterogeneity in regard to content and binding character.

Stakeholders at the operational level

At the operational level, a relatively small number of stakeholders is involved. Primarily the administrative department that generates and stores the data also prepares and uploads it as open data (63%), the ICT department being involved in about half of the cases. At least in a quarter of all instances exists the role of an open data officer who also supports the operational process. ICT apparently play an important role, supporting the operational process in some way in 57 per cent of the cases. At the simplest level, the software can support a user to manually export a data set from the data base, without any additional functionality. This function is provided in about one third of the cases. At a more sophisticated level, the software can support the manual export of data and by itself generate predefined metadata. In about one fifth of the cases the software offers this kind of functionality. The most sophisticated capability of a software is to generate and upload data as open data based on predefined rules. 29 per cent of respondents that the software they work with supports open data in this way, which could thus be characterized as open by design.

Regarding the use of open data, overall, 81 per cent of respondents have reportedly used an open data set in the past, despite the fact that only 26 per cent classified themselves as users. Across all stakeholder groups, more than 70 per cent say they have used an open data set. Particularly astonishing is the high number of stakeholders within the public administration report that they have used open data. Looking at the purposes respondents have used open data for, differences become apparent. Politicians almost exclusively and to a lesser extent public administrators predominantly used open data to play around and test it. However, this nevertheless means they have dipped into the details and thereby demonstrated interest and engagement. Even beyond, more than a third of public administrators also used the data to develop an application. Approximately half of the self-identified users and two thirds of external consultants used open data to develop an application. Business analytics and data journalism were far less common purposes of open data use. Nevertheless, 29 per cent and 27 per cent respectively of users also gave these reasons.

Examining the relevant prevalence of roles among companies in the UK open data ecosystem, we can easily distinguish two tiers of engagement in particular roles. A significant proportion of respondents (63.3%) are involved in processing open data, with similar numbers of respondents providing insights (59.5%) and developing products (57.0%) based on open data. Noticeably fewer respondents are providing infrastructure (41.8%) and publishing open data (40.5%). However, it is important to note that these values are still relatively higher than might be expected given the fact most focus on the commercial stakeholders as (re-)users of open data. For infrastructure this can perhaps be attributed to the necessity of having technical and data infrastructure in place before other aspects of the ecosystem can develop. Indeed many of the more established companies are involved in providing the technical infrastructure for publishing open data to the public sector, which evidently is required a priori of re-use. The high absolute proportion of companies publishing open data is perhaps

conversely attributable to more recent developments in open data and an indication of the direction a more mature commercial ecosystem might look like.

Analysing in more detail which attributes of open data commercial companies deem important, it is apparent that timeliness and licensing are the key elements that influence the decision of most companies to introduce open data in their operations. This is to be expected to some extent, as most of the products and services in the ecosystem (see also the case studies in the following section) rely on the timely publication of data from its publishers. Equally important is the licensing of open datasets.

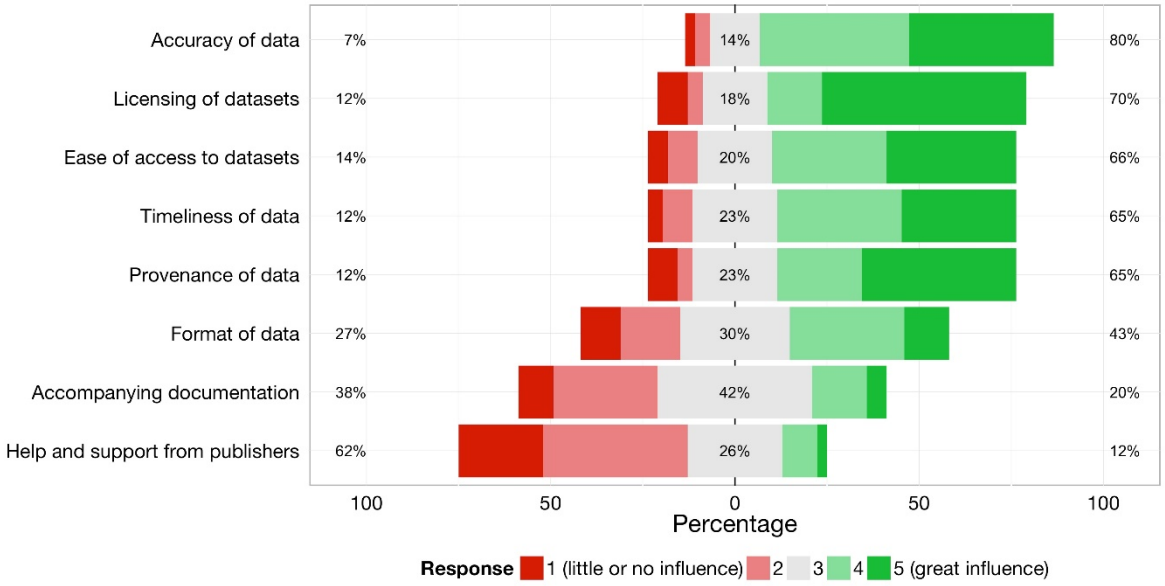

Figure 3 Open data attributes influencing the decision to use open data

Finally, data provenance is the third most important open data attribute as ranked by our respondents. This is another finding showcasing the importance of good quality data for the commercial ecosystem to flourish. Companies need to be able to understand how each open dataset is created, what each data field means and the type of the values so that they can confidently reuse the data in their products and services. However, most of the respondents did not feel that they need to establish closer collaboration with the publishers in order to solve some of the technical problems they encounter. It is our contention, that this is the case because most of our respondents are IT companies with significant experience in data operations.

Consequences

Altogether, the findings show that the OpenDataMonitor already covers the broad areas of interests of stakeholders in open data. However, since there is a high heterogeneity within the open data ecosystem, specialised interests might still be uncovered. Against this background, it has been

decided within the consortium to create mini user stories for general as well as specific roles and interests. This approach is in line with the agile development philosophy and also strongly connects the research and the development part of the project. User stories are generated on a constant basis, initially within the project and eventually in an open process. These stories are then prioritised and implemented. Thereby, the OpenDataMonitor platform will not only be able to inform broad and general interests, but also provide insights for specific and granular interests within the open data ecosystem.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

1. INTRODUCTION

1.1 Problem Statement and Research Questions

Open Data has for some years now caught attention among politicians and public managers inter alia. Promising great potentials regarding government transparency, democratic participation, public decision making, digital innovation, and economic growth (Deloitte 2012; Geiger and von Lucke 2012; McKinsey Global Institute 2013), the idea has attracted political support at all levels, globally (e.g. Group of Eight (G8) 2013), nationally (e.g. Bundesministerium des Innern 2014) and locally (e.g. Birmingham City Council 2013), generated government as well as private investment and stirred up implementation efforts by public administrations on a global scale (Davies 2013). Some scholars have documented ecosystems that have subsequently emerged around open data (Heimstädt, Saunderson, and Heath 2014). Within these ecosystems different elements intermingle, ranging from politicians, public administrations, civil society, private businesses and academia. Their understandings, interests and perspectives on open data are however not always clear and potentially manifold.

Different principles have been put forward by various actors to lay out the fundamentals of open data and constitute a unified understanding (Open Knowledge Foundation 2005; Sunlight Foundation 2010). Nevertheless, the adoption and implementation in different political and administrative traditions shows significant variety (Hunnius, Krieger, and Schuppan 2014; Perritt and Rustad 2000). That universal reform concepts are adapted during adoption is in itself relatively unsurprising (see e.g. Pollitt and Bouckaert 2011). However, when it comes to open data, the movement's success in terms of coverage, scalability, and sustainability seems to hinge in part on homogeneity (Hunnius, Krieger, and Schuppan 2014). This directs special attention to its unified understanding and thus to its intellectual origins.

Various barriers and impediments have been identified that are seen to hinder the progress and success of open data as it stands today (for an overview see OpenDataMonitor Project 2014b). These relate to the amount of available open data, the related metadata and their overall quality (e.g. Barry and Bannister 2014; Zuiderwijk et al. 2012). The amount of available open data is not an end in itself, but only an intermediate step. Instead, the success of open data hinges on the discovery of datasets by users that these deem relevant and useful. One way to facilitate the discovery of data sets is listing and describing data sets in data catalogues, an approach generally based on the concept of the semantic web (Shadbolt, Hall, and Berners-Lee 2006). Thus, particularly metadata and its quality are essential to discover data sets within a semantic web environment. The general idea underlying the semantic web is to describe its resources based on knowledge classification schema so machines are able to read them (Berners-Lee, Hendler, and Lassila 2001). This concept largely depends on shared ontologies and the validity of some broad-ranging epistemic presuppositions (Madsen 2013).

Setting aside the epistemic presuppositions, the question remains among whom ontologies shall be shared and accepted. This question points to the stakeholder groups and what their understandings and interests in open data are. Various stakeholders are subsumed in the open data movement and project different interests at open data. The origins of the open data movement can be traced back to two different sources: On the one hand is the Freedom of Information (FOI)/Transparency movement which is largely directed at democratic and civic engagement (Ackerman and Sandoval-Ballesteros 2006; Bovens 2002) and on the other hand the PSI industry (Janssen and Dumortier 2003) with its focus on innovation and economic growth. While the former is strongly focused on information access, unearthing troves of government and scrutinizing those in power, the latter stresses aspects like machine-readability and the rights to use the data without any restrictions, including commercial re-use. These two traditions seemingly amalgamate in open data (Janssen 2011). However, whether their interaction in the context of the open data movement leads to a shared understanding of what constitutes open data begs further consideration.

The way various stakeholder groups understand and engage with open data is critical for the Open Data Monitor project. The web platform built by the project automatically monitors the state of open data in the European Union conveying important information in the form of metrics and visualisations. To this end, it is important to capture the needs of all interested parties in the open data ecosystem in order to facilitate the evolution of the open data movement in the EU. In D2.6, we do this by surveying the views of open data stakeholders. We have developed a research design that establishes a holistic view of the ecosystem and its various stakeholders while it also captures and analyses the experiences of the business section of the ecosystem, which is currently a strong user, and to some extent publisher, of open data. Our aim is to explore two different, yet complementary, issues, which are at the core of this report: (1) gain a more refined understanding of the stakeholders within the open data ecosystem, which we split between policy and operational levels, following the taxonomy of D.2.4, and (2) bring to the fore their experiences and the issues they consider critical in their involvement with open data. Our final objective is to showcase the variety within the ecosystem while discussing the best way to serve the needs of these stakeholders.

This report is a second iteration, building on the qualitative fieldwork conducted for the first stakeholder report (OpenDataMonitor Project 2014b). The intention of this second iteration is therefore two-fold: It aims to validate the findings from the first phase on a larger European scale and at the same time refine the findings by collecting more detailed information.

1.2 Stakeholders, Stakeholder Interests and Requirements around Open Data

There is no taxonomy of stakeholders in open data or even an established methodology to identify and distinguish them. Therefore, we have previously developed a three-layer stakeholder distinction to identify stakeholders, their roles and interests. "At the core are the **stakeholders involved in the operational day-to-day activities around open data**. A middle layer consists of those **stakeholders who are involved at the policy-level of the open data case in point**. The outer layer consists of **stakeholders who are involved in open data at a more general, abstract level** that have an indirect effect on the open data project at hand." (OpenDataMonitor Project 2014b)

At which level stakeholders are involved does not directly translate into a specific understanding of open data or interests therein. However, it does play a role in how granular or general information they need about open data, what barriers they perceive as important and thus, what information they expect and require from the OpenDataMonitor. This translates into analyses over the aggregate, per geography and per catalogue on the one hand and on the other hand to those at the per dataset level (OpenDataMonitor Project 2014a). Besides the level of interest, the stakeholder role has some implications for whether someone is more interested in the technical aspects of open data (data formats, metadata standards etc.) or whether the focus lays on content-related aspects (demographic vs. geographic vs. budget data, update intervals etc.).

In this deliverable, we build on the stakeholder taxonomy (see Table 1.1), already presented in D2.4, constitutes a holistic depiction of the stakeholders in the open data ecosystem. More specifically, we delve in the commercial part of the ecosystem and present a more detailed view of the main stakeholders. The rationale for this approach lies in the strong orientation towards business innovation with open data characterising the commercial open data ecosystem. As companies experiment with open data they gather significant experiences regarding the opportunities and challenges accompanying the use of open data as a business resource. It is our contention that such experiences need to inform the development of the Open Data Monitor platform. It is for this reason, that we study in more detail commercial uses of open data identifying stakeholder groups and engaging with them through both qualitative and quantitative research to collect their views on open data. For convenience, our stakeholder taxonomy is presented below. It is based on the qualitative data collected and analysed for the first iteration of this deliverable, we have described generic interests and requirements and translated them into potentials for the OpenDataMonitor-platform in Table 1.1.

Table 1.1 Stakeholder Groups, Exemplified Requirements and ODM-Potential (OpenDataMonitor Project 2014b)

Stakeholder	Requirements	ODM-Potential
Policy Makers: Parliament, ministries pushing open data, coordination bodies for e-government and ICT, governance structures for cross-level collaboration in e-government and ICT	Understanding barriers to open data publication and use; understand, develop and enforce widely used standards (formats, structure, licenses etc.)	benchmark volume and sophistication of the published data as well as its use; highlight coverage of used standards; present usage of open data; metrics per geography (see D2.3)
Commercial User (Associations): corporate advocacy groups, business associations, media outlets, commercial data users	Detect high-value data sets with minimally and transparent strings attached; detect mashable, harmonised data sets on a large scale;	highlight high-value data sets (e. update frequency of a data set); map mashable content (congruent licenses, harmonised structure and vocabulary); highlight coverage of used standards (esp. licenses); metrics per geography and per data set
Civic Users: civic activists, civic advocacy groups, civic data users	Advocate the publication of and detect already published politically sensitive data sets (politico-administrative)	Highlight and compare sensitive data sets to advocate their publication in other locations; map mashable content; metrics per data set and per geography
Government bodies and associations: inter- and supra-national bodies and associations, coordinating bodies around ICT and e-government, public enterprise in charge of furthering the information society, network of smart cities, standardisation bodies	Advocate the publication of high-value data sets; benchmark volume and sophistication of the published data as well as its use to name and shame understand coverage of used standards to align with these; understand what constitutes high-value data sets to advocate	benchmark volume and sophistication of the published data as well as its use; highlight coverage of used standards; highlight high-value data sets; metrics per geography, catalogue and data set

	their publication	
Data Generators: data generating and (potentially) providing government bodies	Understand what constitutes a high-value data set in their professional domain; learn about standards in open data in general and their professional domain; understand how open data in their professional domain can be used	highlight high-value data sets by domain or topic; highlight coverage of used standards (licenses, structure and vocabulary) by domain or topic; highlight applications of open data by domain or topic
Technology providers and consultancies: Private technology consultancies, ICT vendors, (public) ICT service providers, Open data platform providers; applied research centres	Understand widely adopted technologies and standards to align with these	highlight coverage of used infrastructure, technology and standards (formats, licenses); metrics per geography and overall

2. METHODOLOGY

This research aims to broaden the understanding of the behaviours and interests of stakeholders in open data as well as the processes around open data both in public administration and in industry. To this end, research in deliverable D2.6 follows a mixed methods design consisting of a combination of qualitative and quantitative methods in order to capture and analyse in depth the practices of the various stakeholders in the open data ecosystem. The overall research design is organised around two major phases each exploring different aspects of the open data ecosystem. The aim is to understand the needs of the stakeholders working in the ecosystem in order to develop the appropriate tools to support their open data activities. Phase I, presented in detail in D2.4, aimed to identify the main stakeholders in the open data ecosystem as well as core issues related to the publication and use of open data. The second phase (II) explores in more depth the experiences of open data publishers and users across the European Union. Research covers all the stakeholder levels identified in D2.4. It also focuses on a specific and very dynamic category of open data stakeholders: commercial organisations working with open data to create innovative products and services.

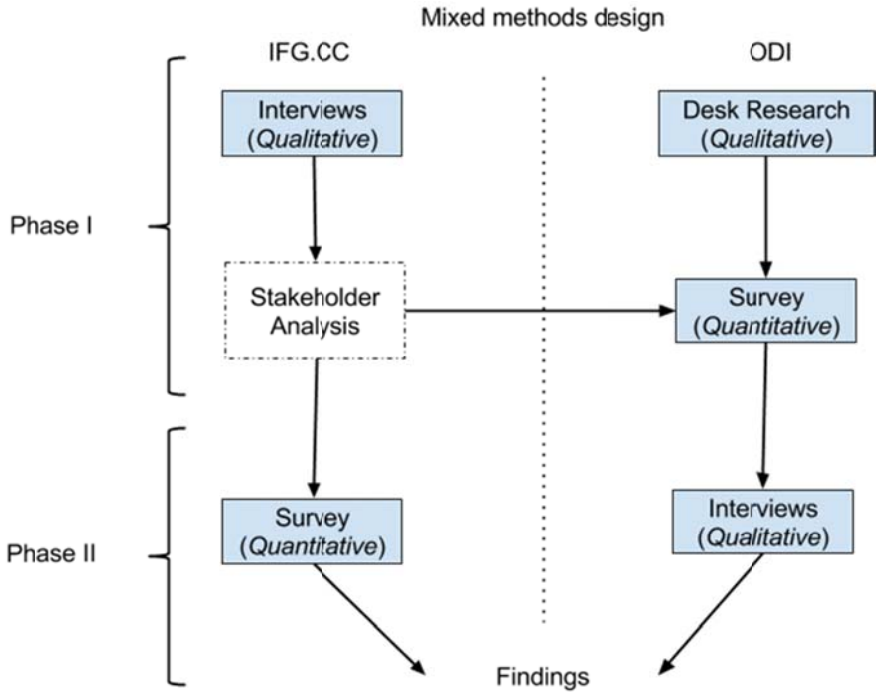

Figure 2.1 Schematic depiction of research design for T2.6

A mixed-methods research design is the appropriate choice to support the aforementioned objectives of research. It provides the necessary methodological tools to capture complex socio-technical phenomena without sidestepping the theoretical foundations of the research project (Hesse-Biber 2010). In the case of open data, the challenge rests on the emerging character of the

phenomenon under investigation (i.e. open data ecosystem) and the variety of stakeholders currently involved in open data activities. This means that we need to first establish reasonably solid boundaries for the ecosystem, understand the stakeholder groups populating the ecosystem, the various activities within it and finally capture their experiences from their everyday involvement with open data. The way the mixed-methods research design have been operationalized is described in the following sections.

2.1 Research strand I: Survey

Questionnaire Development

The survey builds on the identified stakeholder groups and their interests analysed in the first iteration of this deliverable as well as relevant categories derived from previous research. This comprises (1) the importance of the different aspects of data for the success of open data (e.g. open licenses, open formats, metadata), (2) the kinds of data they consider especially interesting and relevant as open data, (3) which groups support and push for open data, (4) the kind of relationship between users, activists and the public administration and (5) and a few questions to categorise participants (country of origin, industry they work in or level of government and policy field respectively etc.).

The questionnaire contains general questions all participants were asked as well as stakeholder specific questions that account for their respective context. For example, participants from within the public administration were asked specifically about the existence and content of an open data policy as well as operational questions regarding the preparation, generation, uploading and publication of open data. Users of open data were asked additional questions to gain a better understanding of how they use open data and for which purpose. Thus, in order to classify the stakeholders, participants were at the outset asked to assign themselves to one of the seven stakeholder groups (data generator, open data strategist, advisor on open data, politician, activist, user or consultant[1]).

The participants were asked general independent questions (without any conditionality on previous answers), e.g. “Which country are you from?”, and general conditional questions (independent from the stakeholder group they assigned themselves to, but dependent on a previous answer they gave). For example, after they had answered the general independent question whether they have ever personally used an open data set, those who answered “yes” were subsequently asked the general conditional question for which purpose they have used open data and a “No” answer did not

[1] These roles were briefly described to make them immediately accessible for participants (see Appendix).

introduce any follow-up-question. Furthermore, they were asked stakeholder-specific questions, depending on the stakeholder group they had assigned themselves to and stakeholder-specific conditional questions. For example after stakeholders from within the public administration affirmed that an open data policy exists in their organisation, they were subsequently asked for its specific content. Table 2.1 shows the structure of the survey.

Table 2.1 Questionnaire Structure of Survey I

Role in the context of open data	General independent questions	General conditional questions	Stakeholder-specific questions	Stakeholder-specific conditional questions	Total
open data generator in the public sector	6	3	7	4	20
open data / IT strategist within the public sector	6	3	7	4	20
advisor on open data in the public sector	6	3	7	4	20
politician	6	3	5	2	16
open data activist	6	3	3	1	13
external consultant on open data	6	3	3	1	13
open data user	6	3	2	1	12

Thereupon, a first draft of the questionnaire was tested with about ten participants. Participants in this pre-test were partners from within the project consortium as well as test participants recruited by partners. Subsequently, the questionnaire was adapted and questions refined, taking into account the appropriate formulation of questions, correct ordering of questions, correct scaling, as well as making sure it precisely captures the intended information, to accurately reflect the views and opinions of the participants.

Finally, the English questionnaire was translated into German, French and Spanish. The questionnaire was implemented with LimeSurvey to collect the required information in a data base, necessary for the empirical analysis. The survey started online on 2 December 2014 and ended on 14 January 2015.

Participant Selection, Recruitment and Participation

Participant selection and recruitment followed two strategies. One relied on contact data provided in open data portals across Europe and the other on word-of-mouth recommendation.

First, we downloaded the metadata of more than 10.000 open data sets stored in about 50 open data repositories of 18 European countries. We cleaned this metadata (deleting role-based addresses e.g. < contact@administration.com > and duplicates) for email addresses that we subsequently used to invite their owners to the survey. About 1.100 individuals were contacted in this way. Secondly, partners in the OpenDataMonitor project forwarded the survey invitation through their mailing lists, newsletters, posted Facebook messages and Twitter tweets via the OpenDataMonitor project. The included link to the survey was open to forward to those active in the open data realm, thereby profiting from the snowballing effect.

In total, 416 participants started the survey. 243 participants completed the survey to more than 95 per cent.

Table 2.2 Progress of Participation in the Survey

Progress of participation	No.
Started survey	416
Exited before page 2	66
Exited before page 4	81
Reached page 6 and/or last page	243

Considering how many respondents for each of the different stakeholder roles participated in the survey, results exhibit that every stakeholder group has at least 13 participants (Table 2.3). Open data users and open data generators represent approximately a quarter of all participants each. Since in the course of the analysis stakeholder groups will be combined in different constellations (e.g. stakeholder groups from within the public administration), samples become even larger. Politicians will receive only marginal attention in the analysis, because of their low response rate.

Table 2.3 Participants by Stakeholder Groups

Stakeholder Group	No. of Participants	Frequency
Open Data Activists	36	14.81%
Open Data Users	66	27.16%
Open Data Generators	65	26.75%
Open Data Strategists	27	11.11%
Advisors on Open Data	16	6.58%
External Consultant on Open Data	20	8.23%
Politicians	13	5.35%
Total	243	100.00%

Despite widespread efforts and various channels for distribution, the survey participants' countries of origin are strongly unevenly distributed. Participants from Germany, Spain and the United Kingdom together make up 63 per cent of all participants who completed the survey. Of those 243 participants, 25 come from countries outside Europe or non-EU European countries; two gave no country (Table 2.4).

Table 2.4 Countries with Highest Participation

Country of participants	Amounts
Germany	90
Spain	40
UK	24
North America (i.e. USA and Canada)	14
Netherlands	12
Italy	8

Therefore, the survey results cannot be seen as representative for the whole of Europe or open data in general. Especially for participants from the public administration there is a significant selection bias because of the means of recruitment. By recruiting participants who provided an open data set in an open data repository and correctly added metadata (at least name and e-mail address) we targeted participants who expose some level of interest and sophistication in open data, institutionally and personally. This seems nevertheless permissible, since the questionnaire presupposed a minimum of understanding of open data and the project's focus is on stakeholders within the open data ecosystem. The survey does not attempt to survey the general understanding of and proficiency in open data of the public administration.

Data Analysis

After the online survey had been closed, a csv-file was extracted from LimeSurvey. This dataset was then coded and finally analysed with the statistical analysis-software R. Besides basic descriptive statistics the analyses predominantly used were ANOVA-test (to analyse the differences and the variation of answers between groups of participants), X_2-test (to analyse the dependence of the responses between groups of participants, or to analyse the dependence between the responses in a group of participants) and correlation-test (to analyse the correlation of the quantitative answer between groups of participants).

2.2 Research strand II: Survey II and Interviews

Survey: Design, sampling and analysis

The first stage of our research design consists of a survey aiming to identify the commercial companies operating in the open data ecosystem, and establish a better understanding of the ways they are using open data for product and service innovation, but also the challenges they are facing in this endeavour. In preparation for the survey, which unlike other research designs (e.g. Nesta 2014) was not based on an open call for participation but purposive selection of respondents, we have mapped the commercial open data ecosystem in the UK and the open data companies we subsequently invited to do the survey. To this end, in this section we present in detail our approach in mapping the commercial open data ecosystem and explain how this activity fed into the survey and the interviews we conducted for D.2.6.

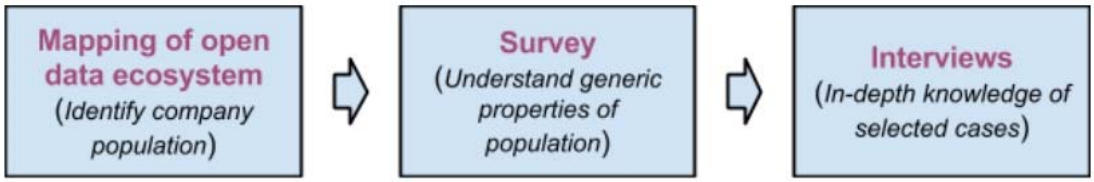

Figure 2.2 Research process for strand II

The mapping phase was completed within a two month period (July-August 2014). The breakdown of activities is as follows:

1. *Collection and aggregation of lists of open data companies*

   The data collection process was launched with extensive desk research. The aim was to identify available resources both in the ODI but also from partners and collaborators. We managed to procure a number of datasets of open data companies which we subsequently aggregated.

2. *Establishing data schema*

   The ensuing data schema captures detailed information for each company, its legal status, operations and work on open data. This allowed us to clean and homogenise the collected data and establish consistent information about the companies. It also allowed us to identify areas when we needed more information about the companies which informed the next stages of research.

3. *Establishment of initial inclusion criteria*

This stage served two main objectives: (a) curating the available data, and (b) developing a first set of operable criteria for the inclusion process. It also involved verifying/triangulating the data collected during the initial identification process given that it was aggregated through a number of disparate listings. More specifically, the purpose of this iteration was to:

a. Clearly identify that the company operates in the UK, resolving possible inaccuracies in the logging of the company in the initial sources of information
b. Record operational company data through the following variables. This data has been collected using OpenCorporates (www.opencorporates.com), an open database of company and corporate information.

Table 2.5: Classification criteria

No.	Variable	Source(s)
1	UK Company Number	OpenCorporates, Company Website
2	UK Office	Company Website
3	Trading Status	OpenCorporates, Company Website
4	UK Company Type	OpenCorporates

Company registration, operational status and geographical location constitute a solid set of inclusion criteria. The resulting list is inclusive and provides a good overview of the open data ecosystem in the UK.

4. *Data triangulation*

The data collected through the aforementioned process was triangulated through extensive desk research, largely using each company's website. This approach allowed us to verify our data and ensured we maintained a high level of confidence in the eligibility of the companies to be included. Through this process, we also added more companies working with open data that were not captured in the initial stages of research.

The outcome of this phase was a dataset comprising information about each company and a brief description of their work with open data. Following application of the inclusion criteria we identified 320 companies working with open data in the UK which constitutes our population for the survey.

Once the mapping of the ecosystem was completed and we had an initial understanding of its setup for the UK, we proceeded with the design of a survey which aimed to elicit more information on the

open data practices of the companies constituting the ecosystem. The survey was designed in August 2014 and launched in September 2014. Data collection lasted for two months until October 2014. The activities within these three months were the following:

1. *Design of questionnaire*

   In designing the questionnaire, we followed Couper (2011), in adopting a design that will capture the pertinent issues in the open data ecosystem while maximising response rates among very busy professionals. More specifically, the questions were designed to cover the various open data related activities within each company, their revenue streams, their pricing mechanisms, the datasets they use, the products and services they developed with open data and the challenges they face when using open data. The questions were procured after systematic survey of the open data literature (academic, government and industry) which identified the core themes discussed by the main open data stakeholders (identified in D.2.4). In order to refine these themes, we also conducted a small number of elite interviews (6 interviews ranging from 40 to 60 minutes) with open data experts which allowed us to verify and supplement the outcome of the literature survey but also get a sense of what are the core concerns on the ground. Finally, great detail has been placed on the format, graphic layout and question routing in the questionnaire, as suggested by Sanchez (1992), also in view of maximizing the response rate for the questionnaire.

2. *Trial period/Internal validation*

   Following the initial design of the questions, we invited 10 people to try the questionnaire and provide feedback on the questions and its general structures. These people included ODI employees and also startup companies, which are part of our targeted population, from ODI's incubation programme. Following the trial period, we incorporated the comments and created the final version of the questionnaire which was subsequently circulated to the open data ecosystem.

3. *Launching of survey*

   The questionnaire has been sent to all 320 companies identified during the mapping stage of research. We used MailChimp to send bulk invitations to the various groups of respondents (ODI members, unaffiliated). The survey was done online via GoogleForms integrated to ODI's website. The link to the survey has been included to the invitation email. There have been two calls for participation. The first call was launched on 23 September 2014. In tandem with the questionnaire a redacted version of the dataset has been made available via the ODI website to give companies the opportunity to see what information has been collected on them and provide us with feedback/corrections/additional information. The second call for

participation has been closed on the 28th of October. We had 88 responses out of the 320 companies identified from desk research as open data companies. This accounts for a 20% response rate. In order to capture companies that were not identified in desk research, we have also launched an open call via the @ODIHQ Twitter account. This procured us with 10 additional responses which went through due diligence to verify their work with open data.

Survey Data Analysis

Following the closure of the online survey, the 88 responses recorded by the Google Form were displayed as a Google Spreadsheet. The resulting dataset was cleaned and coded, which required the removal of duplicate and incomplete or inadequate responses. This left 79 valid responses which were directly downloaded into the statistical analysis-software R. The analysis consisted of basic descriptive statistics and visualisations.

Case studies: Interviewees selection and analysis

The fundamental objective of D2.6 is to understand the experiences, challenges and requirements of open data publishers and users – in this case with a focus on actors from the commercial sector. Additionally, we want to identify and understand innovative and high-value cases that can provide interesting insights into the way open data is adding value to the businesses. As open data and its usage is in its formative stages there cannot be a definitive or widely-adopted set of criteria to evaluate an open data company and predict the adoption of their products and services. To this end, case selection should be an inclusive process. However, at this stage, it is of great value to establish a rich overview of the various ways companies are experimenting with open data. To this end, there are a number of heuristic rules based upon experience in the field of open data that we can apply. These are the following:

- Companies covering as many industry sectors as possible to gain in depth knowledge of sector specific issues and the role of open data (We followed the sector classification used in the survey).
- Companies exemplifying the stakeholder roles in the commercial open data ecosystem identified in our stakeholder analysis as presented in section 1.2 of this deliverable. Additionally, we endeavoured to select companies with interesting, innovative products and services with the potential to grow in the near future (these are mainly startups)

We have identified ten companies which fulfil these criteria and interviewed their CEO's on their work with open data (see interview questionnaire in Appendix 3). The companies were selected out of the survey respondents in order to have the opportunity to triangulate the collected data and conduct a comprehensive analysis of the data collected through our mixed methods design. The

interviews were done in person or via Skype. They last from 30 to 60 minutes. The average duration was around 45 minutes. The companies are the following:

Table 2.6: Case studies - Companies classified per sector, data specialisation and stakeholder role

	Company Name	Sector	Data specialization
1	Adzuna	Data/Technology	Marketing/Recruitment
2	Spend Network	Finance & Investment	Finance
3	Geolytix	Research/Consulting	Retail/Geo-spatial
4	TransportAPI	Transportation	Transport
5	Swirrl	Data/Technology	-
6	Doorda	Data/Technology	Geo-spatial/Crime
7	OpenCorporates	Finance & Investment	Finance
8	Shoothill	Geo-spatial/Mapping	Environmental
9	SkillsRoute	Education	Education
10	FoodTrade	Food & Agriculture	Food/Allergens

Interview data analysis

The interviews were recorded and transcribed for analysis. The interview questionnaire was designed to serve as a coding scheme in the analysis phase (see Saldaña 2009). The questions followed the themes of the survey questionnaire. However, the interview was open ended with the aim to invite the interviewees to reflect on their experiences on using open data as a business resource. The interviews fed also directly to brief case studies that present the selected companies and provide a snapshot of their experiences with open data (see section 6.2 of this deliverable).

3. EMPIRICAL RESULTS: STAKEHOLDERS AND STAKEHOLDER ROLES

3.1 Stakeholders at the Policy Level

Stakeholders at the policy level play an influential role in the progress of open data (Zuiderwijk and Janssen 2014). They influence not only the general strategy in a jurisdiction or organisation on whether or not open data shall or must be published, but also shape decisions about which specific data sets can be published (Hunnius, Krieger, and Schuppan 2014). Therefore, it is important to understand which stakeholders at the policy level support and push for open data.

The group that is seen to most strongly support and push for open data are civic and political activists and activist groups in general. 79 per cent of respondents see this group as supporting open data which is relatively unsurprising, since they fulfil a lobbyist role for open data. The group attributed with the second highest support are supra- and international bodies, like e.g. the European Commission, the World Bank and the OECD. Decision makers in politics and the higher echelons of public administration are seen to markedly lower support open data. Slightly less than 30 per cent of respondents see these two groups as pushing for open data.

The group seen to provide the least support for open data are businesses and business associations. Only 15.5 per cent of respondents see potential commercial users of open data as pushing for open data. Journalists and media industry occupy a middle position between civic activists and commercial users, being seen by 38.7 per cent of respondents to support open data.

Table 3.1 Groups supporting and pushing for open data

Group supports and pushes for open data?	Yes	No	Total
politicians	28.57%	71.42%	100%
senior and mid-level public managers	31.51%	68.48%	100%
supra- and international bodies	52.52%	47.47%	100%
civic and political activists / activist groups	78.99%	21.00%	100%
journalists and media industry	38.65%	61.34%	100%
businesses and business associations	15.52%	84.45%	100%

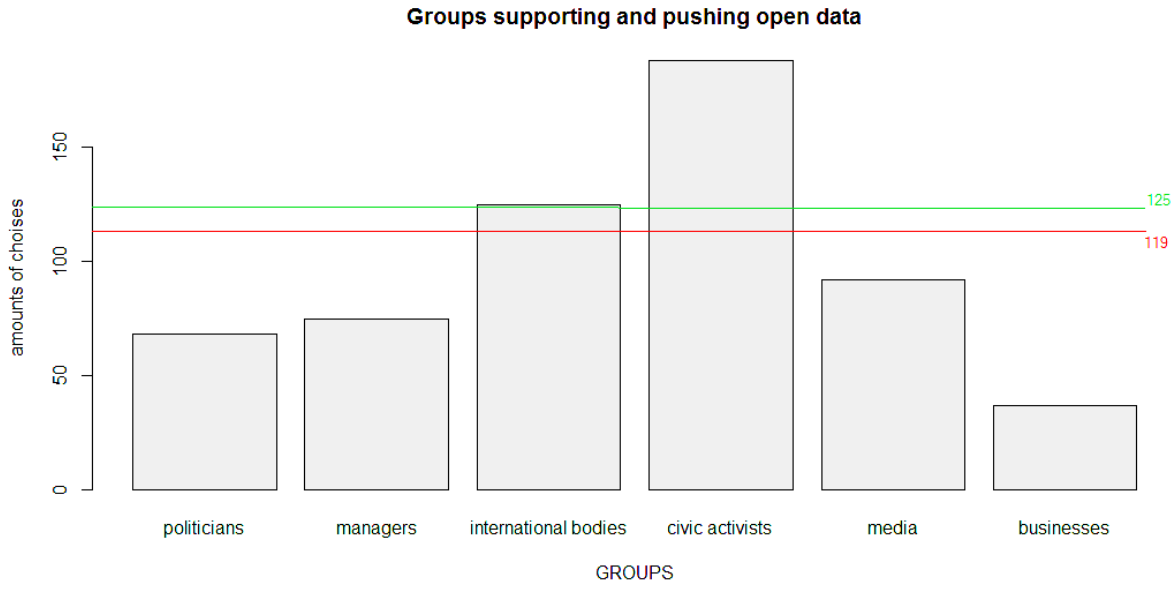

Figure 3.1 Groups supporting and pushing for open data

Apart from the general support and push for open data, stakeholders at the policy level are also involved in decisions whether a specific data set can be published. It is necessary to know who decides whether data can be published to better understand the organisational arrangements in which these decisions are made, the different rationalities involved and potentially explain different results.

Only in 19 per cent of the surveyed cases exists a general “open by default”-rule under which no general decision to publish is necessary. Still, this does not necessarily imply that nobody needs to be involved in a decision to publish a dataset and it is automatically uploaded. It might nevertheless be necessary to determine that no personal data are included, e.g. by a data protection officer.

The departments and roles most frequently involved in decisions about whether or not a data set can be published are the administrative department that generates the data and the open data officer. The administrative departments that generate the data are involved in 72 per cent of the cases, making them the most deeply involved group. In 42 per cent, an open data officer is involved in the decision, which implies that in nearly half of the cases such a position exists at all. The legal department, the ICT department and the data protection officer are each involved in approximately a third of the cases.

Decisions are rarely taken jointly together. In one fifth of the cases, joint committees with representatives from various departments from within the public administration make the decision whether or not to publish a data set. Even less common is involvement of stakeholders from outside the public administration in such committees and thus in decisions about open data. Such outside stakeholders can for example be representatives of business associations or activist groups. However, only in one tenth of the cases such mixed committees make decisions about open data.

Table 3.2 Stakeholders involved in decisions about open data

	Amount choices		
Involved in decisions whether data can be published as open data	**Yes**	**No**	**Frequencies "Yes"**
general "Open by default"-rule	19	81	19%
administrative department that generates data	72	28	72%
legal department	35	65	35%
ICT department	33	67	33%
open data officer	42	58	42%
committee internal	19	81	19%
committee mixed	10	90	10%
data protection officer	36	64	36%
politicians	19	81	19%
Total	285	615	

Overall, the results illustrate a large variety of organisational arrangements under which decisions about open data are made. Also, a relatively large number of stakeholders is involved in the decisions. No group can single-handedly decide to publish open data. Even though the mode of decision-making is unclear (majority rule, veto rights etc.), decisions with widely distributed decision rights in general potentially enable veto players and make passage less likely (Tsebelis 2002). This is especially true, if decision makers involved have different mind sets and interests (Tsebelis 1999 in regard to parliamentary decision making).

Besides the decisions about whether or not a data set can be published, actors at the policy level also shape the manner in which open data can be published, e.g. by mandating certain licenses, data formats etc. This influence can be exerted through open data policies. In 56 per cent of the cases, such an open data policy exists, in 44 per cent of the cases it does not.

Table 3.3 Existence of an open data policy

	Yes	No	NA
Amounts of answer	56	44	45

While at the local level, only 35 per cent report to have an open data policy, at the national level exists an open data policy in 62.5 per cent of the cases.

Table 3.4 Existence of open data policy by governmental level

	local / municipal	regional/state level	national	supra- / international
Yes	8	21	20	3
No	15	15	12	0

The higher the administrative level, the likelier it is that an administration has an open data policy, even though the numbers for the supra- and international level are too small to bear any meaning. The regional and state-level administration depict an inclined position with local administrations being less likely to have an open data policy and national levels being more to have an open data policy (chi square: 4.6193, p-value 0.0993).

Regarding the content of existing open data policies, they mostly contain some regulation about data formats, metadata formats and licenses (around 80 per cent for each). To a lesser degree, open data policies contain some statement about pricing (60 per cent), a general liability clause (55 per cent) or set a number of data sets that shall be published (52 per cent). Thus, open data policies cover a significant area of relevant aspects of open data.

Table 3.5 Content and binding character of open data policies

Aspect	Any rule	General recommendation	Binding rules	Non-existent
pricing	34	15	19	20
data formats	47	16	31	8
metadata formats	45	18	27	10
licenses	47	23	24	8
number of data sets	29	4	25	26
general liability clause	31	10	21	22

However, when it comes to setting binding rules, open data policies remain rather vague. On average, they contain binding rules only in 44 per cent of the cases for specific aspects of open data. Most prevalent are binding rules for data formats (55 per cent). A binding liability clause can be found in only 37.5 per cent of the surveyed cases. The pricing of open data is the aspect scarcest regulated with a binding rule (34 per cent). However, this might be attributable to the fact that open data is often understood as free of charge by definition.

The number of respondents in this subset remains too low to make any statements regarding the content of open data policies at different governmental levels or comparing different countries.

3.2 Stakeholders at the Operational Level

Stakeholders at the operational level are involved in extracting, preparing and making the data available as open data. Often, more than one department is involved in this process. In 63 per cent of the surveyed cases, the administrative department that generates the data is involved here. In about one half of the cases the ICT is involved. Open data officers are involved at the operational level only in about one quarter of the cases. Other stakeholders play a negligible role.

Table 3.6 Stakeholders involved at the operational level of open data

	Amount choices		
	Yes	No	Frequencies "Yes"
administrative	63	37	63%
ICT department	49	51	49%
open data officer	24	76	24%
other	7	97	6,7 %

Assuming that the administrative departments that generate the data rather have a professional perspective on the data than a technical assuming that different understandings of open data, it remains to be seen how profound an understanding they have of what is necessary to render open data easily usable or usable at all (see Chapter 4).

By definition, data that is published as open data is digital. Thus, besides the human actors involved in the operational process of open data generation, the ICT in which the data is processed and stored potentially play an important role in facilitating open data. In principle, ICT in the public sector are built to be closed and protect data from outside view, shielding citizens' private data and state secrets. Only in a limited number of administrative departments there is a history of sharing or selling data (e.g. geological and statistical data). Therefore, whether and to what extent the software supports open data is a relevant question.

Table 3.7 Software's capability to enable open data

Software supports generating open data	Yes	No	Don't know
absolute	56	30	12
Frequency	57,1%	30,6%	12,3%

In more than half of the surveyed cases, the software provides some sort of support to the public administrators who generate the data. About 30 per cent of the respondents need to rely solely on

their manual ability to somehow generate open data. 12 per cent were unaware of the software's capability.

At the simplest level, the software can support a user to manually export a data set from the data base, without any additional functionality. This function is provided in about one third of the cases. At a more sophisticated level, the software can support the manual export of data and by itself generate predefined metadata. In about one fifth of the cases the software offers this kind of functionality. The most sophisticated capability of a software is to generate and upload data as open data based on predefined rules. 29 per cent of respondents reported that the software they work with supports open data in this way.

Table 3.8 Extent of software support in generating open data

Software…	automatically uploads		supports manual export, automatically generates metadata		supports manual export	
	Yes	No*	Yes	No*	Yes	No*
Total	25	60	18	67	27	58
Frequencies	29,4%	70,6%	21,2%	78,8%	31,8%	68,2%

*including respondents who reported that the software does not provide any sort of support

This relatively large share of "open by design"-software might be attributable to the selection bias of the survey. The mode of recruitment for the survey favoured administrations who provide open data and at least filled out the metadata for "author e-mail" correctly. Thus, the results cannot be seen as representative in any way for public administration as a whole. Rather, it indicates that public administrations that already provide open data with at least minimally good metadata have to a relatively high proportion open by design-software.

4. EMPIRICAL RESULTS: STAKEHOLDERS' UNDERSTANDINGS OF OPEN DATA

4.1 Important Aspects of Open Data

Different definitions of open data have been put forward in recent years since the advent of the open data phenomenon (for an overview see Heimstädt, Saunderson, and Heath 2014). While these definitions differ in detail and breadth, there appears to be a general consensus on some constitutive characteristics of open data. Initially, aspects around the legal openness were emphasized, however, aspects of technical openness have since gained prominence (Bizer, Heath, and Berners-Lee 2009). Such aspects of legal openness are open licenses and that the use of the data is unrestricted. Technical openness concerns for example the machine-readability of the data and technical accessibility (open formats). Furthermore, usage costs (free of charge) are frequently deemed a core characteristic of open data.

In addition, in order to detect and use open data, it needs to be described with meaningful metadata of good quality. This aspect is of particular importance for the ODM-project. Therefore, such additional characteristics of the nature and shape of open data were included in the survey. Furthermore, during the preceding interviews within the ODM-project, public administrators often exhibited a special focus on the amount of open data. Apparently, an important point for comparison with other jurisdictions often was the number of data sets in a particular portal. Thus, whether the amount of open data is seen as a crucial success factor for open data was surveyed as well.

The results illustrate that in general all surveyed aspects are seen as highly and of relatively similar importance (Figure 4.1).

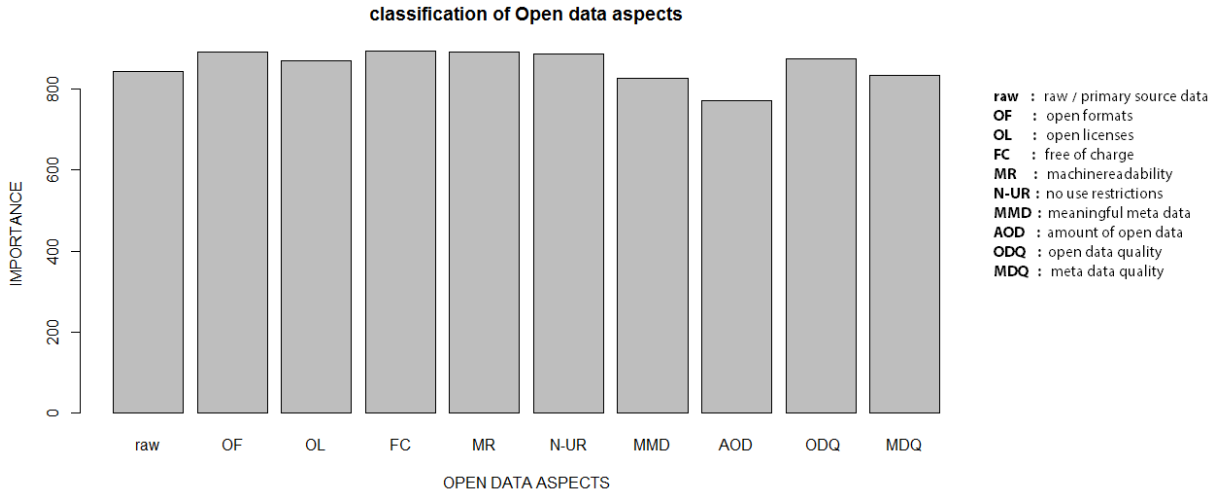

Figure 4.1 Importance of selected aspects for the success of open data

Open formats, open licenses, machine-readability and that data are provided free of charge are seen as paramount for the success of open data. Slightly less important are the quality of open data, the completeness and primacy of open data (raw data) and the metadata (quality and meaningfulness). However, the differences are minor. Only the amount of open data is deemed a little less important. Furthermore, the standard deviation is relatively high (Table 4.1). Thus, while the differences between the various aspects are small, the differences among respondents are rather high.

Table 4.1 Importance of selected aspects for the success of open data

	N	Mean	S.E. Mean	Standard Deviation	Variance	Median	Mode
raw data	235	2.42	.10	1.50	2.26	2	1
open formats	233	2.19	.10	1.53	2.33	1	1
open licences	232	2.25	.10	1.57	2.47	1	1
free of charge	235	2.21	.09	1.41	1.99	2	1
machine-readability	235	2.21	.10	1.58	2.51	1	1
no use restrictions	236	2.25	.09	1.36	1.84	2	1
meaningful metadata	235	2.49	.09	1.40	1.97	2	1
amount of open data	234	2.71	.07	1.14	1.30	3	2
open data quality	237	2.31	.09	1.46	2.12	2	1
metadata quality	235	2.46	.09	1.35	1.81	2	2

While the modal value for all but two aspects (metadata quality and amount of open data) is “1” (=very important), a significant number of respondents rated the very same aspects others saw as very important as not important at all (Table 4.2).

Table 4.2 Frequencies of ratings of aspects

	1 very important	2	3	4	5 not important at all	Total
raw	39,83 %	20,76 %	14,40 %	6,77 %	18,22 %	100%
open formats	50,63 %	18,29 %	8,08 %	5,53 %	17,44 %	100%
open licences	51,50 %	14,59 %	9,01 %	6,00 %	18,88 %	100%
free of charge	45,76 %	18,64 %	17,79 %	4,23 %	13,55 %	100%
machine-readability	53,38 %	14,83 %	7,62 %	4,23 %	19,91 %	100%
no use restrictions	40,50 %	24,89 %	13,50 %	10,97 %	10,12 %	100%
meaningful metadata	31,35 %	28,38 %	13,98 %	11,86 %	14,40 %	100%
amount of open data	14,46 %	31,48 %	31,06 %	14,04 %	8,93 %	100%
open data quality	40,33 %	27,73 %	5,88 %	11,34 %	14,70 %	100%
metadata quality	28,81 %	33,47 %	12,28 %	13,13 %	12,28 %	100%

In summary, while respondents on average deem most surveyed aspects of open data highly important, there is considerable variance for each aspect. Thus, there is no overall agreement on what is crucially important for the success of open data. Variance is particularly high in regard to the relevance of machine-readability and open licenses, two aspects seen as constitutive for open data in most definitions of what makes data open.

In how far these differences point to diverse understandings of open data among the various stakeholders, however, is so far not clear. During implementation of open data in multiple countries, public administrations have for various reasons opted to set rather vague standards in regard to licenses, formats etc. (e.g. Germany) while other countries have passed specific guidelines (e.g. Spain; for details see OpenDataMonitor Project 2014b). While in some instances, vague standards were seen as necessary to get administrators started in open data and raise acceptance within the public administration, it might as well be the case that the necessity for certain characteristics is not fully understood by public administrations.

Comparing the results between stakeholders within the public sector (generator, advisor, strategist combined) and those outside of the public sector (user, activists combined) can account for some of the variance. However, while the standard deviation is slightly lower among respondents from the public administration, it remains high or is even higher for the subset of users and activists. This means, this group attributes rather diverse importance to the various aspects. Again, this is especially true for the relevance of open licenses and machine-readability as well as open formats (Table 4.3).

Table 4.3 Importance of aspects of open data for stakeholders within and outside public administration

	Users and activists			Public sector personnel		
	Mean	Std. Deviation	Rank No.	Rank No.	Mean	Std. Deviation
raw data	2.56	1.61	7.	8.	2.18	1.31
open formats	2.35	1.65	2.	2.	1.94	1.36
open licences	2.47	1.75	5.	3.	1.98	1.39
free of charge	2.43	1.51	3.	3.	1.98	1.32
machine-readability	2.49	1.68	6.	1.	1.80	1.32
no use restrictions	2.33	1.41	1.	5.	2.07	1.23
meaningful metadata	2.66	1.46	8.	9.	2.28	1.30
amount of open data	2.70	1.21	10.	10.	2.71	1.02
open data quality	2.44	1.59	4.	5.	2.07	1.26
metadata quality	2.72	1.41	9.	7.	2.16	1.22

In general, public administrators attribute higher importance to the various aspects for the success of open data. Nevertheless, there is considerable alignment between these two groups, when ranking their preferences. The only remarkably different preferences are regarding the machine-readability of open data, which public administrators rank higher – even making it the most important aspect –, and that the data can be used unrestrictedly, what is paramount for users and activists, but only has a middle position among stakeholders in the public administration.

Drawing the distinction between stakeholders at the policy-level (activists, strategists, politicians, advisors combined) and those at the operational level (users and generators combined) it becomes apparent that is either no or no significant difference between what they see as fundamental for the success of open data (Table 4.4).

Table 4.4 Importance of aspects for open data for stakeholders at the policy-level and the operational level

	Policy level			Operational level		
	Mean	Std. Deviation	Rank No.	Rank No.	Mean	Std. Deviation
raw	2.36	1.51	6.	7.	2.43	1.48
open formats	2.27	1.62	2.	1.	2.10	1.46
open licences	2.36	1.69	6.	3.	2.15	1.51
free of charge	2.35	1.49	5.	1.	2.10	1.38
machine-readability	2.17	1.65	1.	4.	2.16	1.50
no use restrictions	2.27	1.48	2.	5.	2.19	1.24
meaningful metadata	2.45	1.40	9.	9.	2.48	1.39
amount of open data	2.58	0.96	10.	10.	2.76	1.20
open data quality	2.27	1.38	2.	6.	2.27	1.49
metadata quality	2.42	1.34	8.	8.	2.45	1.36

Interestingly, stakeholders at the policy-level rank aspects of technical openness as most relevant for the success of open data, machine-readability and open formats. Overall, however, the means for both groups are very close-by.

More pronounced are differences between respondents who have reported that they have actually used open data sets. However, even though both groups differ markedly in how they rate the relevance of the various surveyed aspects, non-users in generally simply assign greater importance to all aspects. This holds true systematically across all surveyed aspects of open data.

Some differences in how users and non-users rank the various aspects are remarkable nevertheless. Somehow counterintuitively, participants who have never used an open data set rank machine-readability highest. Strikingly, apart from the amount of available data, which is considered least relevant across all subsets, users of data sets see metadata as the least important aspect of open data. However, the difference is significant only for meaningful metadata (Sig. .004).

Table 4.5 Importance of aspects for open data for stakeholders have or never have used open data sets

	Personally used open data: yes			Personally used open data: no		
	Mean	Std. Deviation	Rank No.	Rank No.	Mean	Std. Deviation
raw	2.48	1.53	7.	9.	2.26	1.35
open formats	2.27	1.57	2.	3.	1.91	1.29
open licences	2.29	1.61	3.	7.	2.07	1.40
free of charge	2.26	1.43	1.	2.	1.89	1.20
machine-readability	2.34	1.64	5.	1.	1.73	1.21
no use restrictions	2.29	1.37	3.	7.	2.07	1.25
meaningful metadata	2.64	1.41	9.	3.	1.91	1.23
amount of open data	2.82	1.14	10.	10.	2.27	1.04
open data quality	2.43	1.49	6.	5.	1.95	1.28
metadata quality	2.60	1.36	8.	5.	1.95	1.18

4.2 Interesting and Relevant Data Content

Open Data is often characterised by rather formal attributes, such as machine-readability, an open license, open format etc. that were analysed in the previous sub-chapter. Its content remains so far rather underspecified and what stakeholders around open data view as relevant open data is not very well understood. This potentially leaves open data providers ambiguous about which data to make available first and users frustrated who do not have the data at hand they are looking for. The content of open data is potentially myriad, since it does not only cover all government activity (spending, public works, education etc.), but also government simply registering, monitoring and reporting on private activity (e.g. economy, sports, tourism).

Looking at the kinds of open data survey participants in general deemed most important it becomes apparent that for one it contains data that is to some extent already made available. Thus, what potentially insightful open data is seems influenced by what open data is made available. Existing open data shapes what open data is. Furthermore, it comprises a mix of data about government activity (Finance and Budget, Public Sector) as well as the government reporting about the world as it is (Environment and Climate, Geography and Geology, Demographics) and private activity (Economics, Tourism).

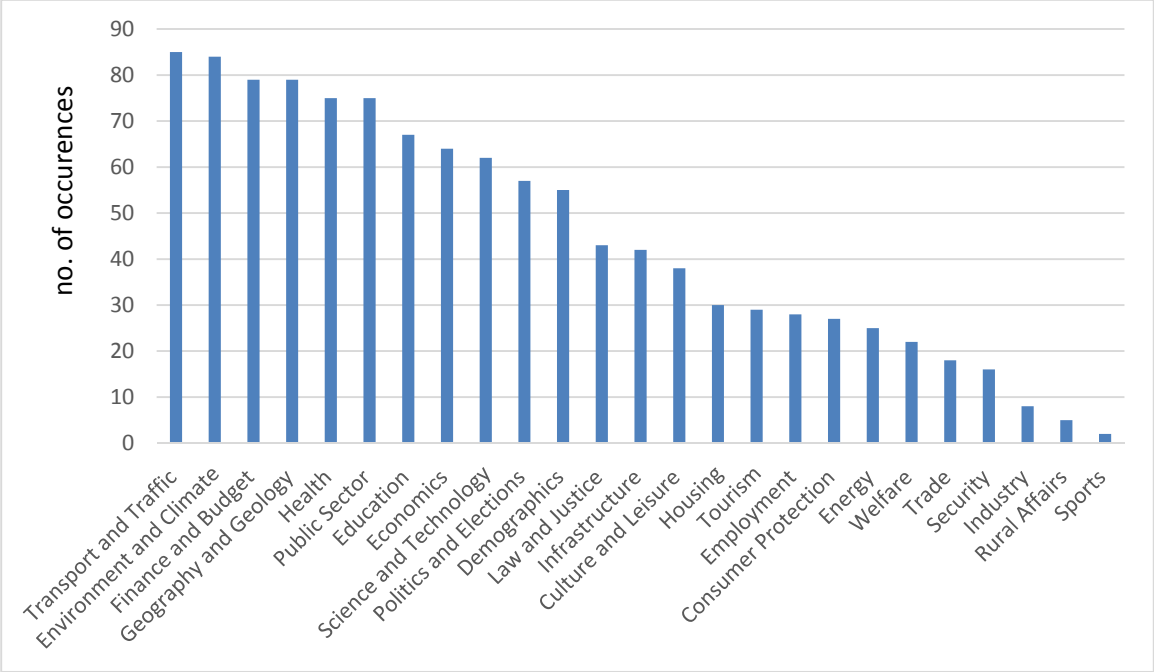

Figure 4.2 Relevant Content of Open Data

Scrutinising in detail which kinds of data the different stakeholder groups deem most useful illustrates that there are certain categories that are highly valued across the population, e.g. Finance and Budget (Table 4.6). However, there nevertheless is a large variety of different views. The data category mentioned most frequently (Transport and Traffic) has a frequency of only 7.62 per cent and no category remains unmentioned at all. Table 4.6 depicts that ten categories have a frequency above five per cent while six categories occur below two per cent with nine data categories in between.

There appears to be some alignment among stakeholders from within the public administration, with only minor differences between the strategic and the operational level in how they rank the different data categories.

Table 4.6 Relevant Content of Open Data by Stakeholder Groups

Data categ.	activist	advisor	external consultant	generator	politician	strategist	user	Total	Freq
Transport and Traffic	8	11	15	17	6	14	14	85	7,62%
Environment and Climate	6	7	6	30	4	11	20	84	7,53%
Finance and Budget	15	7	2	21	5	12	17	79	7,09%
Geography and Geology	7	8	9	25	2	14	14	79	7,09%
Health	9	6	9	14	2	11	24	75	6,73%
Public Sector	16	6	5	21	3	3	21	75	6,73%
Education	7	4	5	15	7	8	21	67	6,01%
Economics	17	5	8	11	2	5	16	64	5,74%
Science and Technology	5	2	7	21	1	5	21	62	5,56%
Politics and Elections	17	1	2	11	3	6	17	57	5,11%
Demographics	6	6	3	16	2	5	17	55	4,93%
Law and Justice	10	3	4	7	1	3	15	43	3,86%
Infrastructure	5	2	5	13	4	5	8	42	3,77%
Culture and Leisure	2	1	3	12	3	5	12	38	3,41%
Housing	3	1	5	8	8	2	3	30	2,69%
Tourism	0	1	5	9	2	5	7	29	2,60%
Employment	3	2	1	7	2	2	11	28	2,51%
Consumer Protection	3	0	2	8	4	4	6	27	2,42%
Energy	6	1	1	3	0	7	7	25	2,24%
Welfare	4	1	1	5	2	1	8	22	1,97%
Trade	4	2	0	4	0	3	5	18	1,61%
Security	0	1	2	4	1	3	5	16	1,43%
Industry	3	0	0	2	0	0	3	8	0,72%
Rural Affairs	0	2	0	1	1	0	1	5	0,45%
Sports	0	0	0	0	0	1	1	2	0,18%

Looking at the data categories the different stakeholder groups consider most important, it also becomes apparent that there is a considerably higher mutual understanding of what open data is actually about. The frequency is markedly higher, the top categories generally mentioned well above

ten per cent in each stakeholder group (Table 4.7). Especially among users of open data, however, the spread remains larger.

Table 4.7 Top Data Categories and Frequency by Stakeholder Groups

	Activists		Users		Generators		Advisors		Strategists	
Rank	Data Category	Freq.	Data Category	Freq.	Data Category	Freq.	Data Category	Freq.	Data Category	Freq.
1	Pol. and Elect.	10.90%	Health	8,16%	Env. and Climate	10.53%	Transp. and Traffic	13.75%	Transp. and Traffic	10.37%
2	Econ.	10.90%	Education	7.14%	Geogr. and Geology	8.77%	Geogr. and Geology	10.00%	Geogr. and Geology	10.37%
3	Public Sector	10.26%	Science and Tech.	7.14%	Science and Tech.	7.37%	Env. and Climate	8.75%	Finance and Budget	8.89%
4	Finance and Budget	9.62%	Public Sector	7.14%	Public Sector	7.37%	Finance and Budget	8.75%	Env. and Climate	8.15%
5	Law and Justice	6.41%	Env. and Climate	6.80%	Finance and Budget	7.37%	i.a. Demogr.	7.50%	Health	8.15%

The data categories activists consider the most relevant and interesting – politics and elections, economics, public sector, finance and budget and law and justice are predominantly related to the transparency/FOI-tradition of open data. From this perspective, open data is useful to scrutinise those in power make government activities themselves transparent. Thus, participants who self-identify as activists appear to come predominantly from this tradition. On the other hand, participants who self-identify as users consider data categories such as health, science and technology, education, public sector and environment and climate as most interesting. Although there is some overlap, these data are markedly less politicised. Similarly, public administrators rank environment and climate, geography and geology, transport and traffic, finance and budget highest.

Table 4.8 Frequency and Ranking of Relevant Content of Open Data by selected stakeholder groups

	Activists		Users		Public Administration	
Data category	frequency	rank	frequency	rank	frequency	rank
Politics and Elections	10.90%	1	5.78%	7	3.60%	12
Public Sector	10.26%	3	7.14%	2	6.00%	6
Finance and Budget	9.62%	4	5.78%	6	8.00%	4
Law and Justice	6.41%	5	5.10%	10	2.60%	15
Environment and Climate	3.85%	10	6.80%	5	9.60%	1
Geography and Geology	4.49%	8	4.76%	12	9.40%	2
Transportation and Traffic	5.13%	7	4.76%	11	8.40%	3
Science and Technology	3.21%	13	7.14%	4	5.60%	7

5. EMPIRICAL RESULTS: OPEN DATA USE

Open data users are a core element of the open data ecosystem, since they materialise the potential of open data, the provision of which is not an end in itself. Nevertheless, few specifics about open data users are known. Therefore, the actual use has also been surveyed, besides the self-classification as an open data user. Overall, 81 per cent of respondents have reportedly used an open data set in the past, despite the fact that only 26 per cent classified themselves as users. Across all stakeholder groups, more than 70 per cent say they have used an open data set. Particularly astonishing is the high number of stakeholders within the public administration report that they have used open data (Table 5.1). This might be due to the selection bias, because the survey targeted participants who are active in open data. Nevertheless, the large portion of those who have used an open data set shows a high level of interest and enthusiasm.

Table 5.1 Open Data Use by Stakeholder Group

Stakeholder Group	Yes	No	Frequencies Yes
activist	30	4	88.2%
advisor	10	4	71.4%
external consultant	18	2	90.0%
generator	40	17	70.2%
politician	10	2	83.3%
strategist	21	6	77.8%
user	55	9	86.0%

Looking at the purposes respondents have used open data for, differences become apparent (Table 5.2). Politicians almost exclusively and to a lesser extent public administrators predominantly used open data to play around and test it. However, this nevertheless means they have dipped into the details and thereby demonstrated interest and engagement. Even beyond, more than a third of public administrators also used the data to develop an application. Approximately half of the self-identified users and two thirds of external consultants used open data to develop an application. Business analytics and data journalism were far less common purposes of open data use. Nevertheless, 29 per cent and 27 per cent respectively of users also gave these reasons. The most common “other purpose” specified in a free text field was “research”.

Table 5.2 Purposes of Open Data Use

	to play		develop application		business analytics		data journalism		other purpose	
	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No
user	67%	33%	51%	49%	29%	71%	27%	73%	27%	73%
activist	83%	17%	43%	57%	17%	83%	17%	83%	33%	67%
external consultant	83%	17%	67%	33%	44%	56%	44%	56%	22%	78%
PA	80%	20%	39%	61%	16%	84%	13%	87%	33%	67%
politician	100	0%	0%	100%	0%	100%	10%	90%	40%	60%
Total	78%	22%	43%	57%	22%	78%	20%	80%	30%	70%
	183		186		184		188		188	

No significant differences can be observed regarding the origins of the data sets have been used. 70 per cent of respondents have used data sets from the national level, 59 per cent used data sets from the regional level and 55 per cent used local level open data. Only the supra- and international level is distinctly less frequently used, mentioned by only 37 per cent of respondents (Table 5.3).

Table 5.3 Origins of Open Data

	Yes	No	Total
local	55,3 %	44,7 %	100%
regional	59,3 %	40,7 %	100%
national	69,8 %	30,2 %	100%
supra- / international	37%	63%	100%

Most commonly, users do not solely use only one source of open data and do not specifically use data because it originates from an administrative level. The administrative level a data set originates from does not seem to be particularly relevant. The administrative level that makes data available is predominantly related to the task distribution across the vertical layers of government. A government organisation that is responsible for a certain task generates data in its course and could potentially make it available as open data. The task distribution and the administrative origin of open might thus be seen as negligible for open data is especially since data is increasingly federated to national and henceforth European open data portals. Nevertheless, the administrative level is relevant for open data, because the administrative origin of a data set shapes the technical and legal characteristics of the data.

6. COMMERCIAL USERS OF OPEN DATA: IN-DEPTH STUDY

6.1 Empirical Results: Survey

As explained in the introductory section of this deliverable, a very strong current within the open data movement promotes the use of open data by business organisations. The objective is to promote business innovation and ultimately boost economic growth through the creation of open data products and services. The companies using open data for such purposes are at the forefront of innovation. They experiment with existing datasets, develop their own open data resources and gradually build the necessary conditions for a generative ecosystem. In the process, they show what can be done with open data in a business setting while achieving economic, social and environmental impact. In this journey towards innovation and commercial success for their products and services, open data companies have a breadth of experiences to share about the business opportunities that can be found in the open data movement but also the challenges with using open data as a business resource. These are invaluable lessons that need to be captured in order to inform policy and guide the development of the necessary tools for promoting the next stages of open data for economic growth.

In the following sections, we present in detail, the characteristics of the commercial open data ecosystem in the UK. We show the composition of companies with the ecosystem, their primary activities and the ways they work with open data. We also present a series of case studies as examples of open data companies but also as a way to convey in more detail the experiences of the front-runners of the open data movement in the commercial world.

Overview of UK commercial ecosystem

In order to understand the commercial ecosystem, we first need to understand what types of companies are involved. This requires an examination of traditional business metrics such as size of the companies and the areas in which they operate. We can then examine the roles which these companies play in the ecosystem and build up profiles for open data commercial stakeholders. In this section, we will explore these facets using data collected from our survey of UK companies using open data. This will provide us with a good indication of the types of companies involved, although it cannot be considered to be representative of the entire UK ecosystem.

Survey responses indicate that engagement with open data is not limited to companies of a particular size. For example, a sizeable majority of surveyed companies (69.6%) can be defined as ‘micro’ enterprises – that is companies with fewer than 10 employees. However, larger enterprises are also well represented, with 7.6% of respondents being drawn from large or enterprise sized companies (251 or more employees). Due to the opportunistic sampling used to select survey

respondents it is not possible at this stage to meaningfully compare this distribution of companies sizes to those across the UK economy as a whole.

Table 6.1: Size of open data companies

	Fewer than 10 employees	10 - 50 employees	51 - 250 employees	251 - 1000 employees	More than 1000 employees
Count	55	14	4	2	4
Percentage of total	69.62	17.72	5.06	2.53	5.06

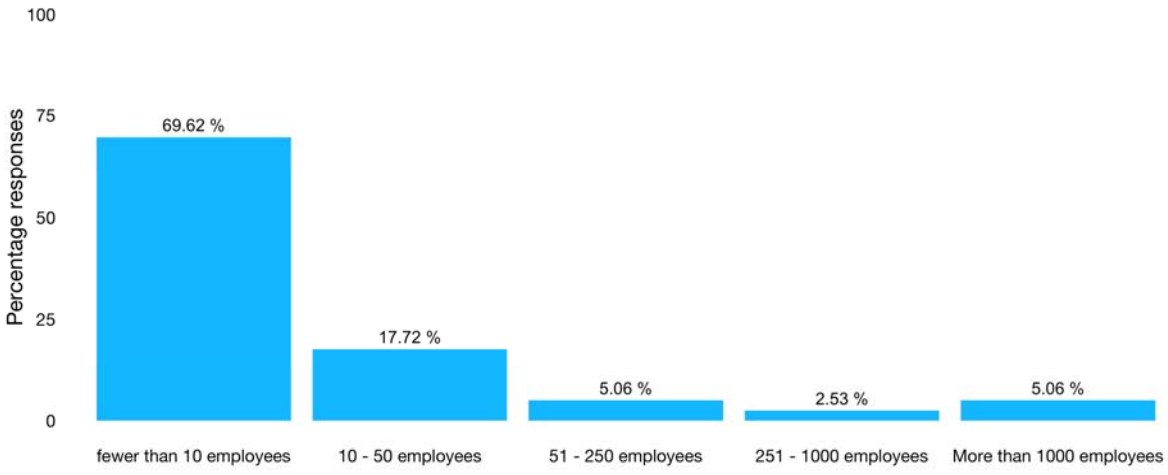

Figure 6.1: Percentage of total companies by size

When asked to specify their main area of business, the majority of respondents (58.4%) chose 'Data/Technology'. The most significant proportion of remaining respondents (33.1% of respondents who did not choose Data/Technology, 13.0% of all respondents) chose 'Research & Consulting' as their main area of business. These companies are primarily small startups or consultancies, using open data to create innovative services for their industries. The full distribution of primary business areas is shown below. However, it is worth noting that 'Finance & Investment', 'Healthcare', 'Housing/Real Estate' or 'Insurance' were not selected by any respondents as their main area of

business. This may reflect companies in these sectors being less willing to identify themselves as users of open data, by completing the survey or more general engagement with the open data community. Certainly anecdotal evidence suggests extensive use of open data in these sectors.

Responses gathered in subsequent interviews provide an alternative/additional explanation for this finding, namely that many companies view themselves mainly as data/technology companies specialised in different industry sectors, rather than identifying with e.g. finance or health care as their primary area of business. A related theme emerging from interviews was of companies not initially targeting a specific sector, but developing sector-related expertise during consultancy projects. There are a number of companies for which this is also true for their introduction to open data. Being brought into a project which required the use of open data has led them to becoming experts in open data solutions.

Table 6.2: Clasification of open data companies per industry sector

Sector	Count	Percentage
Data/Technology	45	58.44
Research & Consulting	10	12.99
Business & Legal Services	4	5.19
Geospatial/Mapping	4	5.19
Education	3	3.90
Energy	3	3.90
Transportation	3	3.90
Scientific Research	2	2.60
Environment & Weather	1	1.30
Food and Agriculture	1	1.30
Governance	1	1.30

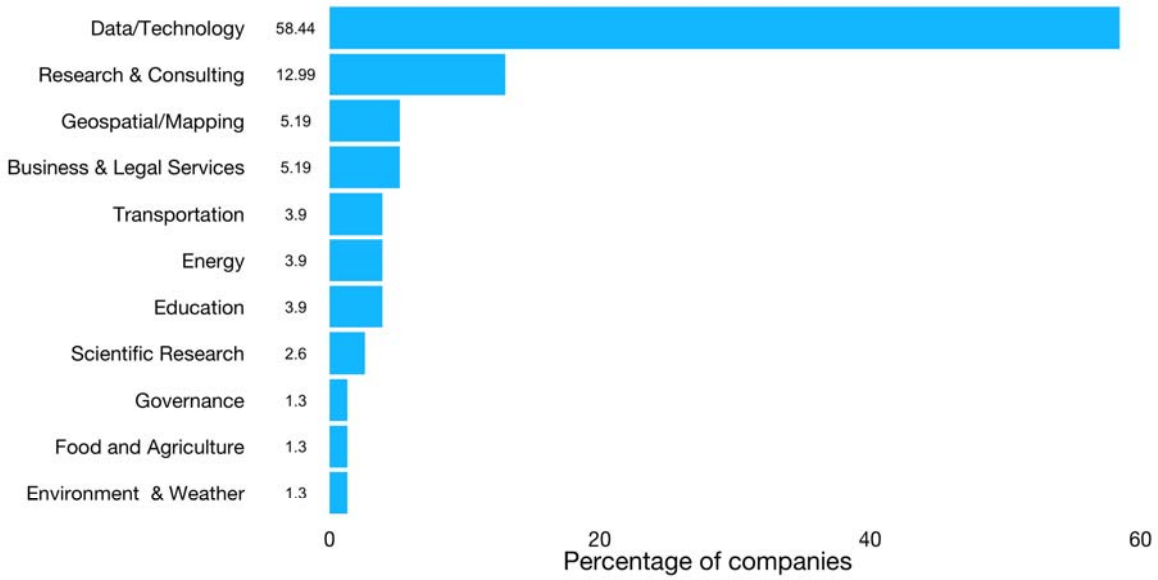

Figure 6.2: Company classification by area of business

To better understand these trends requires an understanding of the roles played by stakeholders in the UK commercial ecosystem. Given the lack of empirical work in the field of categorising commercial open data stakeholders, respondents were instead asked to express how their company is involved in using open data. They were allowed to choose from as many categories as they were involved with so that we could build up an understanding of how the different uses or types of engagement with open data are being combined by commercial entities to create sustainable business practices around open data. The key finding here is that our assumption that businesses were not limited to one role was well formulated. Over three quarters of companies (78.5%) are performing two or more of the five identified roles in the ecosystem. This may suggest that the various roles are complementary, and therefore one company can play several (something we examine later in this paper), or that the roles that each company plays in the ecosystem are not very clearly defined, perhaps due to the relatively low maturity of the field. This may also suggest that there is a high degree of experimentation with appropriate business models among the companies surveyed.

When we examine the relevant prevalence of roles we may perceive two tiers of engagement in particular roles, though these have not been tested for statistical significance. A large proportion of respondents (63.3%) are involved in processing open data, with similar numbers of respondents providing insights (59.5%) and developing products (57.0%) based on open data. Fewer respondents are providing infrastructure (41.8%) and publishing open data (40.5%). Regarding infrastructure, this may be a reflection of either a relatively immature field (i.e. more providers are yet to emerge) or, conversely, a recognition that infrastructure providers will likely always be less numerous than those exploiting the infrastructure itself. The relatively lower number of companies publishing open data

may reflect that the commercial rationale for publishing (as opposed to reusing) open data is not widely understood, however this hypothesis requires further research.

Table 6.3: How companies use open data

My company...	provides infrastructure for others to publish open data	processes open data	develops products based on open data	publishes open data	provides insights based on open data
Count	33	50	45	32	47
Percentage of total companies	41.77%	63.29%	56.96%	40.51%	59.49%

Given that the majority of respondents indicated that their companies were involved in performing more than role, analysing the minority who are only performing one role (21.5%) may offer further insights. For example, just over a third (35.3%) of those who indicated they were only involved in one role provide infrastructure. This may reflect the nature of providing infrastructure as being distinct from the other roles, which all involve direct interaction with open data itself. The effect being especially pronounced for companies providing technical infrastructure such as portals.

Table 6.4: When open data companies specialise in one role, which one is it?

My company...	provides infrastructure for others to publish open data	processes open data	develops products based on open data	publishes open data	provides insights based on open data	Total
Count	6	4	2	3	2	17
Percentage of only one role	35.29	23.53	11.76	17.65	11.76	100.00

A more comprehensive understanding of how companies position themselves in the ecosystem can be formed by examining where companies play more than one role at the same time. An important caveat being that these figures only reflect the co-occurrence of two roles and do not account for where the company plays more than two roles – a level of complexity beyond the scope of this analysis. Noteworthy co-occurrences are those between companies processing open data and those developing products (38.0%) and companies processing open data and providing insights (43.0%). Intuitively, this suggests that companies who develop products or provide insights are required to process relevant data before doing so. Interestingly a similar proportion of companies (36.7%) provide insights and also develop products, suggesting that these activities are complementary.

Table 6.5: When were each of the categories chosen together

	provides infrastructure for others to publish open data	processes open data	develops products based on open data	publishes open data	provides insights based on open data
provides infrastructure for others to publish open data	33* 41.77%**	18 22.78%	20 25.32%	16 20.25%	20 25.32%
processes open data	18 22.78%	50 63.29%	30 37.97%	24 30.38%	34 43.04%
develops products based on open data	20 25.32%	30 37.97%	45 56.96%	21 26.58%	29 36.71%
publishes open data	16 20.25%	24 30.38%	21 26.58%	32 40.51%	19 24.05%
provides insights based on open data	20 25.32%	34 43.04%	29 36.71%	19 24.05%	47 59.49%

*Count of respondents **Percentage of total respondents

Trends in data use

As may be expected, a large proportion of respondents (68.8%) are using open data provided by government. Almost half of the surveyed companies (49.4%) are using open data from other sources, including data provided by the private and third sectors, although only a tenth (10.4%) are using this data exclusively. A notable proportion of respondents (39.0%) are using data from both public and private sources. It is worth noting that a fifth of respondents (20.8%) answered that they were not using data from either source, which may indicate that they provide infrastructure or publish data and thus do not fall into the 'user' criteria.

Table 6.6: Use of government or other open data

		Use other data?		
		Yes	No	Row Total
Use government data?	Yes	30 38.96%	23 29.87%	53 68.83%
	No	8 10.39%	16 20.78%	24 31.17%
	Column Total	38 49.35%	39 50.65%	77* 100%

*The total responses analysed is 77 as two respondents failed to provide an answer to either one or both of the questions.

Survey responses indicate that there is wide variety in the types of data being used. The most prominent of these is geospatial/mapping which is being used by well over half of respondents (56.6%). Several other types of data are being used by more than a third of respondents: transportation (43.4%), environment (42.1%), demographics & social (39.5%) and business (38.2%). Many other types of data were well represented, as shown below, but it is also worth exploring the 'other' category where respondents listed data they are using which they did not feel was covered by the categories. These included: "3D", "Jobs data", "Internet of Things", "political" and "Crime".

Table 6.7: Types of open data used by companies

	Count	Percentages
Geospatial/Mapping	43	56.58
Transportation	33	43.42
Environment	32	42.11
Demographics & social	30	39.47
Business	29	38.16
Government operations	24	31.58
Education	23	30.26
Health/Healthcare	22	28.95
Economics	21	27.63
Energy	21	27.63
Weather	21	27.63
Housing	20	26.32
Consumer	16	21.05
Finance	14	18.42
Science and research	14	18.42
Agriculture & food	13	17.11
International/Global development	13	17.11
Public safety	11	14.47
Legal	9	11.84
Tourism	8	10.53
Manufacturing	5	6.58
Other	5	6.58

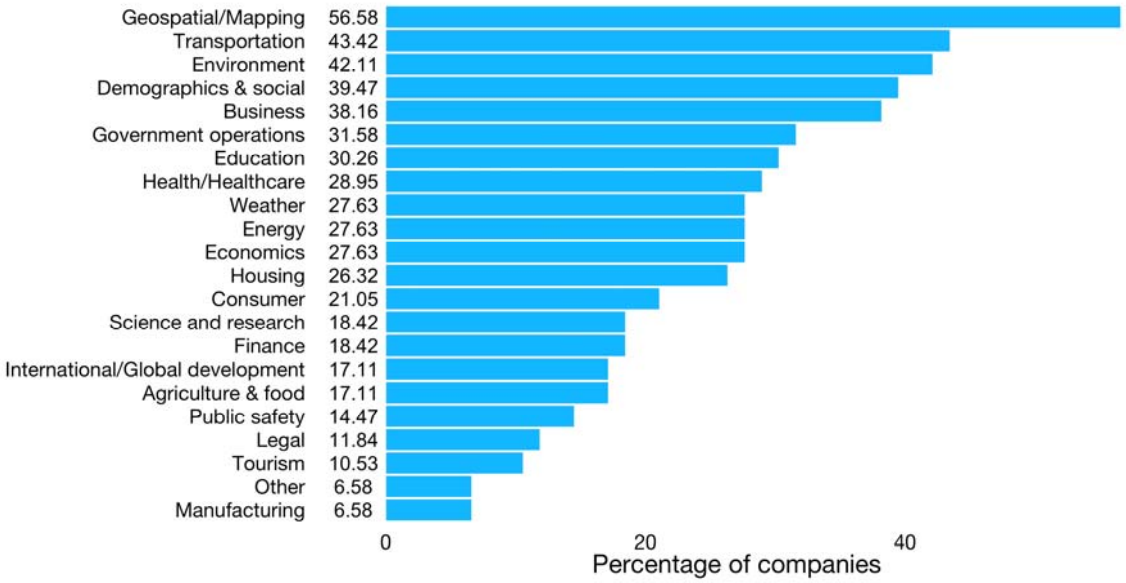

Figure 6.3: Types of data used by companies

Requirements of commercial stakeholders

Respondents were asked to rate the importance of various factors in their decision to use open data. Of particular interest is the relatively low importance of dataset documentation and support from publishers, particularly when contrasted with features such as accuracy, licensing, ease of access, timeliness and provenance.

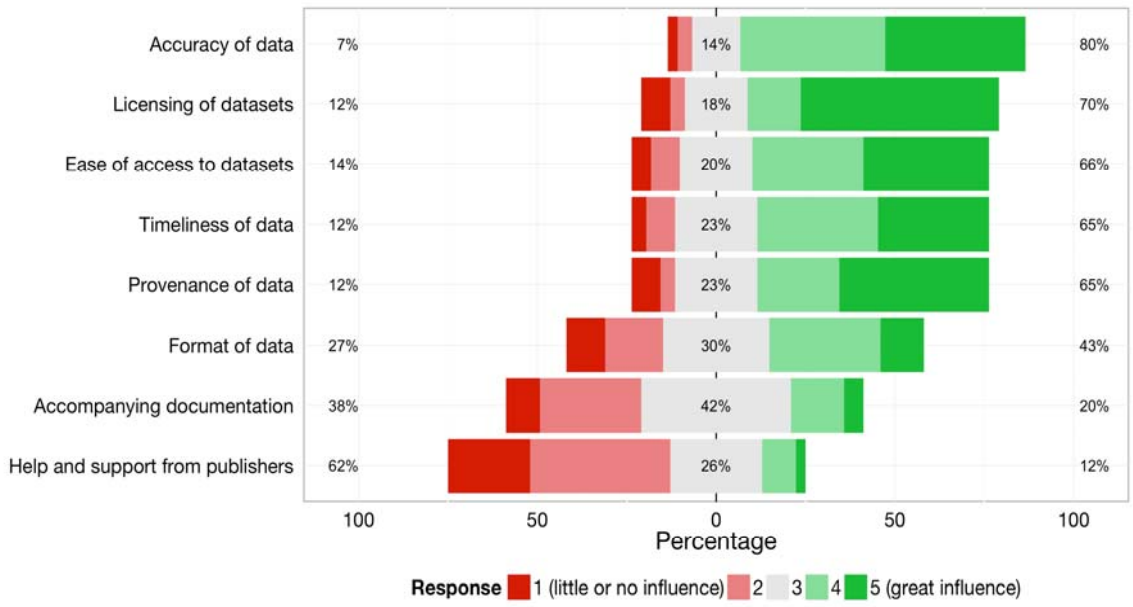

Figure 6.4: Open data attributes influencing the decision to use open data

Table 6.8: Open data attributes

	1 (little or no influence)	2	3	4	5 (great influence)
Accuracy of data	2 2.70%	3 4.05%	10 13.51%	30 40.54%	29 39.19%
Licensing of datasets	6 8.11%	3 4.05%	13 17.57%	11 14.86%	41 55.41%
Ease of access to datasets	4 5.41%	6 8.11%	15 20.27%	23 31.08%	26 35.14%
Provenance of data	6 8.11%	3 4.05%	17 22.97%	17 22.97%	31 41.89%
Timeliness of data	3 4.05%	6 8.11%	17 22.97%	25 33.78%	23 31.08%
Format of data	8 10.81%	12 16.22%	22 29.73%	23 31.08%	9 12.16%
Accompanying documentation	7 9.46%	21 28.38%	31 41.89%	11 14.86%	4 5.41%
Help and support from publishers	17 22.97%	29 39.19%	19 25.68%	7 9.46%	2 2.70%

This survey question also provided respondents with the opportunity to elaborate on their views through a free text box. Their answers led us to identify more data attributes important for commercial reusers of open data. They also provided additional insight into why some attributes are considered important. These answers are summarised in the table below.

Table 6.9: Additional open data attributes influencing business strategy

Attribute	Justification
Format	• Data need to be in a computer readable format. The format type does not matter. • Access to data via an API (of whatever sort) is usually a big advantage for us. • There are a number of reference datasets that we wish the UK public sector would provide in Linked Data form. If they did, we'd use them a lot. • Has to be processable using opensource tools • Has to be complete datasets i.e. not stuck behind a sparql endpoint
Accuracy	• There is little or no benefit to the user is they do not trust that the data is accurate and up to date.
Timeliness	• There is little or no benefit to the user is they do not trust that the data is accurate and up to date. • Ideally has scheduled updates (e.g. OpenStreetMap every 60 seconds)
Ease of access	• It is notable that public data stores are ordinarily presented as 'separate' from corporate web pages, such that open data of relevance to specific policy objectives and/or services do not appear in the same place to promote awareness of them and/or facilitate ease of access for non-technical users. • The current 'improvements' being implemented at government level has hindered much of our work due to data moving, data links often not working.

6.2 Empirical Results: Case Studies – Examples of open data companies

In order to complement these analyses, the following case studies are presented based on interviews with open data companies, selected using the criteria described above. For each company considered, a brief overview of their work is presented below in tandem with a discussion of their experiences using open data and their challenges working with a new and developing resource.

Adzuna: Using open data to modernise recruitment

Founded in 2011, Adzuna is a UK-based search engine for jobs. It aggregates over one million job adverts from a number of sources, including online job boards and vacancies advertised directly by employers. The search engine semantically analyses information from these adverts to

provide relevant results for jobseekers who can find jobs using a range of search options, and is now available in 11 countries.

Adzuna's database of job data is aggregated from several hundred different sources, through XML feeds or automated scraping. Each month, Adzuna combines open labour market statistics data from the UK Office of National Statistics (ONS) with their job ad data to generate statistics such as jobseekers per vacancy by location, which are reported in their monthly Job Market Report. Adzuna also provides an API for their data which is publicly available at no cost or usage limits; allowing other websites, businesses and applications to provide job searches and queries of the Adzuna database. An Adzuna executive says that '*We use open data ourselves... and it feels the right thing to do to be making data publicly available back out there where we can*.'

Adzuna faces two key challenges in using open data. Firstly, the timeliness of the ONS labour market statistics data often lags behind that of data from other sources, which is increasingly offered as live-feeds. Secondly, Adzuna would like to incorporate more open data into their database but have found that relevant datasets are not currently published openly. For example, the incorporation of detailed open company register data may allow for more job ads to be linked to more information on the hiring companies themselves.

Spend Network: Using open data to win public sector contracts

Spend Network is a publicly available repository of UK government transaction data. Founded in 2013, it compiles extensive spending, tender and contract data to allow users to find UK and EU transactions using various search options. Its users include individuals, businesses looking to compete for government contracts and government entities themselves seeking to identify suppliers.

The company's database is aggregated from around 330 different sources (including over 170 UK local councils) sourced through data portals, APIs, or through Freedom of Information requests. It also incorporates open data from OpenCorporates, an open database of company information. The Managing Director of Spend Network states that: *'We couldn't exist without open data. It is open data; our product is open data. We take open data as raw material and turn it into something else.'* In addition to its search tool, the company provides a subscriptive API for its database, allowing for use and reuse of the open data for journalism, academic research and other open data projects. Membership with different licensing terms is available for commercial use of the data. Spend Network offers detailed insights based upon the data, such as analytics and consultancy services for businesses, and also completes advisory services for government bodies on the process of making more open data available.

Spend Network experiences a number of challenges in its use of open data. Primarily, it finds that not all UK publicly-funded entities publish their data openly. A second issue encountered is a lack of standards in open data publishing. Many publishers maintain unique standardisation processes and some open data publishing standards vary within a government entity or by publishing intervals. This increases the level of work required by the company to aggregate the data and provide meaningful insight and use. Finally, incompleteness of open transaction data, such as incomplete Contract Award Notices, can also increase the difficulty of working with the data and may be indicative of low input or source quality.

Geolytix: Geospatial insights and decision-making with open data

Geolytix is a specialist geospatial data company founded in 2012. It provides geospatial data, analysis, training and consultancy services to organisations to embed insight into their processes and decision-making where location matters.

Geolytix offers a number of datasets through its GeoData products. These datasets incorporate open data from a range of sources, including Transport for London (TfL), the Land Registry, the Department for Education, the Department for Health and OpenStreetMap. GeoLytix incorporates this open data via two approaches; through adding auxiliary value to the existing datasets or by deriving new, novel datasets using the open data as an input. Geolytix subsequently releases a selection of its GeoData products as open datasets under open licensing. These include census data, a snapshot of postal sector boundaries, a subset of retail places data, workplace data and manually geocoded supermarket locations.

The primary challenge to Geolytix is the current supply of open data available in the UK. A number of datasets that are currently not released as open data may be of use to the company in the future, such as an open addresses dataset and detailed Companies House (company information) and Ordnance Survey (geospatial data). Geolytix is an interesting case as it publishes some of the datasets it creates for its clients as open data. These include government datasets, which the company aggregates and cleans for possible errors, or datasets created to map the retail industry. This is a unique practice for a small consultancy and shows that every company, irrespective of its size, can publish open data and make a substantial contribution to the open data ecosystem through improving the quality of the available open data. At the same time, it gradually familiarises big retail companies in the UK with the benefits of open data, paving the way for them to join the open data ecosystem.

Transport API: Creating a business ecosystem around transport open data

Transport API is a transport solutions platform founded in 2010 (as Placr). It provides API design and consultancy services focused on transport and location open data to developers and organisations. It seeks to create a single source of UK transport information by unifying timetables, routes, live running and performance history information of a wide range of transport types.

Transport API aggregates transport open data from a range of sources into a single API format to provide data-as-a-service (DaaS). These sources include Transport for London (TfL), the Department for Transport, Network Rail, Traveline and OpenStreetMap. Access to the API is based upon the usage rate required by the user, with free access provided for limited use and different commercial packages available to users with more advanced requirements. TranportAPI provides accurate and timely data from several sources, establishing itself as the go-to provider of transport data in the UK and developing a flourishing ecosystem of developers that work with the data to create apps and other services.

The company encounters two challenges when using open data. Primarily, Transport API would like to combine more open data with the data they already make use of, but have found that it is not made available openly. The Managing Director of the company says that *'there are still some interesting data sets we would like to get our hands on. Certain unreleased Oyster data would tell us in real time how busy the London transport network is.'* More broadly, Transport API seeks to demystify open data and embed it into society, to overcome the challenge of a lack of general public understanding; it is not currently widely understood that open data can be reused many times by anyone.

Swirrl: Creating a linked open data infrastructure for publishing organizations

Swirrl is a linked data solutions company founded in 2008. It specialises in making data easy to find, understand and reuse by creating solutions for organisations seeking to publish their data. In addition to open data publishing services, Swirrl offers custom features, apps and visualisations, training and consultancy services to its customers.

Swirrl's PublishMyData platform allows organisations to publish their open data on the web in a format that's easy to understand and machine readable so that users can access the data via a SPARQL endpoint and other APIs. Swirrl's CEO describes that, *'the value that we are offering to data owners is to help them to get their data used. That's the objective… We want to work with people who want to make their data used so that it can create value.'* Swirrl publishes linked

data to open standards from the World Wide Web Consortium (W3C), whereby 5-star open data exploits the strengths of the web to interconnect, contextualise and compare data. During the process of publishing datasets, Swirrl often incorporates UK government open data, including Ordnance Survey open data, ONS geography and census data, government organisation and time interval data, national public transport access nodes and Companies House URIs. Open data from DBPedia and Geonames is also sometimes used.

The company encounters a number challenges in using and publishing open data. Firstly, it is often difficult for Swirrl to impart the value of publishing open data to organisations. Concerns regarding exposure of the quality or mistakes in the data or the potential uses of the data are often raised by these customers. The linked data approach to publishing taken by Swirrl provides its own challenges with respect to audiences. Whilst the data must be machine readable and accessible, it must also be presented in a format for easy use and understanding by a wide range of public audiences, including casual browsers, developers and analysts, which requires specialist consideration and skill. Swirrl experiences further challenges in its incorporation of open data into the publishing process. Firstly, there are a number of datasets that are currently not released as open data which may be of use to the company. These include open addresses and Unique Property Reference Numbers datasets. In addition, Swirrl may benefit from the release of reference datasets by the UK public sector in linked data form; to allow for linkages with the data they publish.

Doorda: Using open data visualisations to understand neighbourhoods

Doorda is an interactive web service that promotes the use and understanding of data within local communities in the UK. Founded in 2013, it specialises in visualising a range of data onto an interactive map to encourage community engagement and describe what is happening on citizens' streets and in their towns or regions. Doorda also provides data aggregation and analytic services.

Doorda seeks to aggregate and visualise as much local open data as possible. It currently collects and combines open datasets from a number of sources; Ofsted, the Department for Education, the Department for Transportation, the Land Registry, the Food Standards Agency (FSA), the NHS and the Home Office. The open data available on Doorda can be used for a number of local purposes, such as searching for property and schools, identifying crime and antisocial behaviour, understanding restaurant hygiene and finding local news. Their aim is to aggregate available open data into a single source and establish its portal as a reference point for all citizens. Doorda's CEO states that, *'What (Doorda) does is make data accessible to people. Be you a business, be you an individual on the street; that's what we're going to do.'*

The company encounters three main challenges when using open data. Firstly, a lack of publishing standards between sources means that there are often changes and inconsistencies in a dataset from interval to interval. This can cause issues with aggregating and comparing the data. Variation in data publishing intervals is also an issue encountered; publishing intervals for the open data used by Doorda range from daily (e.g. FSA hygiene data) to quarterly (e.g. Ofsted reports) updates. This lack of constant or live-feed data across the board inhibits Doorda from providing a truly real-time map. Finally, a key challenge to Doorda is the limited release of open data by local authorities in the UK. More data held by these authorities, such as planning and building data, which is currently not released openly, could be used provide even more detailed descriptions of communities in the UK.

OpenCorporates: Combining transparency and entrepreneurial innovation through open data

OpenCorporates is an open database of company information founded in 2010. It aggregates data for over 84 million companies from 102 jurisdictions and aims to document a URL for every corporate entity in the world. Users can search the database for a particular company or browse companies by jurisdiction.

OpenCorporates incorporates data from a number of different sources. These include national company registers, government websites, national information centers, official filings, gazettes and data released under the Freedom of Information Act. OpenCorporates' CEO explains that, *'When you look at global government data that relates to companies, it is often unclear, incomplete, inaccurate or hasn't been kept up to date. We also found that the same companies appeared on different government registers without being linked, so there was duplication and valuable connections in the data that had been missed. We wanted to change that and we knew there would be a market for it.'* In addition to providing access to the database through search, the company has developed an API, which is licensed under the share-alike attribution Open Database Licence. Those who wish to use the data without the share-alike restrictions, or rate and/or volume limits, can do so under commercial licensing agreements. Users include individuals, governments and major financial institutions. OpenCorporates also provides its data, with support from The World Bank Institute, to the Open Company Data Index. The Index scores the access to company data and the ability to reuse that data across every country in the world.

In using open data, the company encounters two key challenges. Primarily, critical datasets related to companies, maintained by governments around the world, remain closed. The release of these as open data would enable OpenCorporates to build a further inclusive database of

companies. Secondly, the handling of such vast amounts of data has caused some technology scale challenges, which require skilled technical management.

Shoothill: Creating a flood warning system for the public with open data

Shoothill is a software development company founded in 2006. It specialises in data visualisation and online mapping, working with consumers and corporate clients to create visuals and other tools to convert geospatial and/or statistical data into dynamic and accessible formats.

Shoothill offers three mapping products based on open data provided by the UK Environment Agency. FloodAlerts is an online graphical representation of flood warnings, which provides localised updates to keep users informed about the potential flooding in their area. Shoothill's GaugeMap provides a live map of river levels based on data from each of the 2,400 river level monitoring gauges, which is updated every 15 minutes. In addition, each gauge is assigned a Twitter account for users to follow to provide snapshots of this data twice per day. Finally, Check My Flood Risk provides a map based on flood risk data, allowing users to visualise flood risks and calculate the risk of flooding to specific properties based on their location. Shoothill also provides APIs to allow developers access to live Environment Agency Flood Data, River & Tidal Levels, Groundwater, River Flow and 3-Day Flood Forecast datasets.

The company experiences two challenges in using open data. Firstly, the open data used by Shoothill requires expertise to work with it. Data such as live river levels require an understanding of the specialist publishing practices and data formats to which it conforms. Shoothill's Managing Director says that, *'The art of it is to make it (the data) simple… The general public out there are not interested in all the rubbish out there, they just want to get to the information they need to get to.'* Secondly, Shoothill believes that more open data could be incorporated into its products to provide further benefits to users, but finding and identifying useful datasets is difficult. For example, the Environment Agency maintains numerous other datasets that may be useful to the company if published as open data, but specifying which is a challenging process.

Mime Consulting: Providing job advice to students through open data

Mime Consulting is a specialist data consultancy company founded in 2007. It specialises in the use of data to aid decision making, offering a range of management information services, including data warehousing, analysis, visualisations and dashboard design. It also offers user-friendly software and website design to help automate data collection, analysis and reporting.

Mime Consulting's Skills Route tool combines students' chosen subjects, grades and location with open data to demonstrate personalised university, further education and employment options available in the future. Its projections are based on open data published by the Department for Education (the Level 3 value-added dataset [L3VA]), which captures the progression of students between Key Stage 4 education and the end of their Level 3 qualification (which includes A-Levels and vocational equivalents). It also uses other open data on higher education and apprenticeship providers to show students a wide range of further and higher education opportunities. It is designed to increase awareness of the range of educational options available and ensure young people make informed choices regarding their education.

The company encounters challenges in its use of open data. The L3VA dataset used in Skills Route, which shows progress of students and other aspects such as the impact of schools and teachers, is complex. This complexity requires specific expertise in order for the company to use it in a meaningful way. A second challenge is the timeliness of the data. As the dataset is released a year behind the timeframe upon which it is focused, the tool can not produce truly real-time information and projections. Mime Consulting's Director explains that, *'Datasets have risks in the way people interpret them. There is a lag, as it is information that happened a while ago, and there is a fear that schools could get quite defensive about it, so the timeliness of information is a problem. If data was available sooner, and the more it comes on stream, it would enable us to do better things.'*

FoodTrade: Using allergens regulation to create innovative open data products

Founded in 2012, FoodTrade is a global food network. It maps the sector's infrastructure system in order to make connections between businesses, organisations and individuals to allow them to find, buy or sell food. FoodTrade's CEO states that it aims to *'be a wiki of the food system; then anyone can use that data. And so the first step of innovating into the food system is removed*' in order to encourage the development of a fair, sustainable and local food system.

FoodTrade Menu (FTM) is an FIC regulation compliance tool launched in January 2015. The first version of the product helps restaurants and other unpackaged food retailers meet allergen regulations. As the next level EU meat sourcing regulations become active in 2015, the product will evolve to keep FTM customers compliant. FTM will integrate into FoodTrade's existing online platform and food search engine. Using the platform, the menus data that FTM customers input become 'shopping lists'; customers can discover local and sustainable suppliers, and suppliers can find new routes to market. FTM enables FoodTrade to transform open data published by the Food Standards Agency into an easy-to-use product that introduces small food retailers into the

data economy. This way, it can map the food system and start developing APIs on food recipes and other information.

The main challenge for FoodTrade is the proprietary nature of the most valuable datasets in the the food system. Currently, data concerning product ingredients is held by large food retailers which do not release it. There are only a few notable exceptions to this rule, with limited open datasets which often built by activists in the field. This means that there is limited transparency in the food system and as a result, limited space for transparency and innovation. Mapping local producers and promoting sustainability through open data is equally challenging as these are not familiar with IT enabled working practices. As a result, they have little motives for joining an online platform and offer information about their companies.

In the following table, we have summarised the main challenges as identified by the interviews we conducted with the aforementioned companies. They validate the results of the survey, as open data entrepreneurs also identify issues as timeliness as quite important for their businesses. However, they also pinpoint to a number of important issues related more to the ways the open data movement evolves but also the level of societal understanding and appreciation of the importance of open data in entrepreneurial activities.

Table 6.10: Summary of challenges faced by UK commercial stakeholders

No.	Challenge Name	Challenge Description
1	Timeliness	The open data currently used by commercial stakeholders is often not published quickly enough for their demands. This can mean that there are large variances in the timeliness of data when various datasets are aggregated. Lags in the time taken to publish data also inhibit these stakeholders from creating truly real-time products and services.
2	Complexity	The complexity of some open datasets present challenges to commercial stakeholders. Specialist expertise and skill is required to manage and draw meaning and use from large quantities of data. Publishing complex data also requires specialist consideration to allow for use and dissemination by a range of public audiences.
3	Incompleteness	The incompleteness of some open datasets used by commercial stakeholders can present challenges in aggregating and comparing it with other data. It may also indicate low input or source quality of the data, which can affect trust of the data and the veracity of stakeholders' products and services which use it.
4	Lack of open datasets	Many commercial stakeholders would like to incorporate more open data than they do at present. However, many find that desirable datasets, which may be of use in the future to improve or develop new products and services, are not published openly. There exists a particular demand for more open data to be published by government and publicly funded entities. The adoption of advanced publishing formats, such as linked data publishing, for some datasets may also benefit stakeholders.
5	Lack of publishing standards	A lack of widely adopted publishing standards for open data is often a challenge for commercial stakeholders. Different publishers maintain unique standardisation processes and some standards even vary within a government or publicly funded entity or by publishing intervals. Changes and inconsistencies in the publishing of data can cause issues when aggregating and comparing it, increasing the difficulty for stakeholders in developing meaningful insight and use.
6	Identifying and finding relevant datasets	Some commercial stakeholders are challenged by the identification of datasets that may be of use to them. In particular, stakeholders may be aware that government or publicly-funded entities maintain datasets, published openly or otherwise, that could be potentially be of use; but defining which is a challenging process. In addition, some more complex open datasets used can be difficult to locate and access.
7	Lack of commercial understanding	A lack of wide commercial understanding of open data can provide challenges to commercial stakeholders. Some stakeholders have difficulty in explaining their use of open data and its relevance to their products and services. Imparting the value of publishing open data to other commercial stakeholders is also challenging. Concerns regarding the subsequent uses of open data are sometimes raised by such stakeholders.

7. DISCUSSION

The two research strands, pursued in the context of D2.6 (see Section 2: Methodology), have brought to the fore interesting insights on how the various stakeholder groups understand open data and incorporate it (or not) in their work practices. These insights require integration and further consideration in order to establish a concise picture of the open data ecosystem. This is done in this section, where we discuss in depth the experiences of open data stakeholders. This approach is also in line with our research design which anticipates the integration of findings in the last stage of research.

From the survey on open data stakeholders, there are some general patterns regarding the actors involved at the policy and operational levels and the roles and attitudes they hold. The strongest open data supporters are seen outside the public sector, specifically in civic and political activist groups, in supra- and international bodies and to a lesser degree among journalists and media industry. Business organisations, in a variety of sectors (see section 6.1), are also strong open data advocates (Survey II). This is more pronounced among startups and SMEs which constitute the majority of the commercial open data ecosystem. Less than one third of respondents see politicians and the higher management in public administration as supportive for open data. Since these two groups play an important role in the general decisions about whether or not an administration launches an open data initiative, how many resources are allocated to that task and how strongly lower hierarchical level perceive leadership support for the task, could constitute an inhibitor for open data. Nonetheless, it can be argued that about one third see these two groups among the three most supportive of open data and thus relatively high for a fairly recent phenomenon which just starts to show concrete evidence of business innovation and economic impact.

In administrations that are active in open data, the governance structure is fairly diverse. Thus, decision rights are distributed differently and general rules, manifest in open data policies, vary. The most frequently involved stakeholders in decisions about publishing specific data are the administrative departments that generate the data. However, each of the eight different answering options offered in the survey were chosen by at least ten per cent of respondents. Thus, no stakeholder sub-group within the public administration can single-handedly decide which data to publish. Considering the variety of backgrounds, rationalities and interests e.g. legal departments, ICT departments and politicians have, illustrate the potentially contested nature of these decisions, especially if there is no strong leadership support. Open data policies in general provide some guidance and support, even though they are not yet widespread and often non-binding. However, especially the technical aspects (data and meta-data formats) are

more often regulated by mandatory rules compared with the legal aspects (e.g. licenses). Here, ODM might inform their further development and alignment. The operational level in comparison appears to be significantly less crowded. Only the administrative department that generates the data, the ICT department and in a quarter of the cases an open data officer are involved in preparing and uploading open data sets. Other stakeholders play a negligible role. For these tasks, ICTs in provide some basic support for open data. In about 30 per cent of the cases, the software might even be declared open by design. However, this large share might again be traced back to the fact that the survey targeted specifically administrations who are fairly advanced in open data.

Regarding the understanding of what constitutes open data across the various stakeholder groups, there seems to be relative homogeneity. Aspects from the Open Definition (Sunlight Foundation 2010), like e.g. open licenses and machine-readability, are considered highly important as well as the quality of open data and its meta data. Across the board, the sheer amount of open data is seen as relatively unimportant, thus these metrics (e.g. number of distributions, volume of distributions) appear of lower relevance. Instead, the quality, technical and legal aspects are fairly evenly considered more important. Interestingly, stakeholders from within the public administration in general attribute higher relevance to all aspects of open data than users and activists. This might mean that the rhetoric of advocacy groups caught on and fulfilled its goal in convincing open data providers that these aspects are highly important when in fact, users consider them less important. Differences between stakeholders at the policy-level and the operational level of open data are marginal and insignificant.

Even though the various surveyed aspects are all considered important, a relatively high variance can be observed, even within the different stakeholder groups. Thus, the stakeholder groups' views and understandings are not as distinct as the term often implies. In regard to what characteristics make data open, the stakeholder groups are not inwardly homogenous and outwardly different. This might mean that the stakeholder group definition developed is not robust or that the stakeholder concept might be inadequate. More adequately it would be to speak of roles, rather than stakeholders, because they combine various roles within one person (Dahrendorf 2006). Even the general and widely used differentiation between data providers and data users is questionable. As previously pointed out, individuals who use open data frequently engage in activist groups, lobby public administration to publish open data and at the same time consult them and educate public administrators on the details of open data (OpenDataMonitor Project 2014b). Thus, the roles they combine transcend formal and professional positions. This makes it particularly challenging to characterise them and pinpoint their interests. Especially true is this for users of open data. Here, variance is particularly high.

There appear to be very different needs, depending on factors so far not identified. Therefore, more exploratory qualitative research appears to be necessary in this realm. More robust is the distinction between those in the open data ecosystem who are more strongly influenced by its transparency/freedom of information (FOI) tradition (civic activists) and those more interested in re-using public sector information resources (PSI) (commercial re-use). This becomes obvious when looking in details at the kinds of data these two groups are most interested in. While the former are more focused on data revealing insights into the working and conduct of government itself, the latter focus on data the public administration reports about natural phenomena, society and commercial activity. However, civic activists and commercial users do not significantly disagree on what constitutes open data (e.g. machine-readability).

This technically-oriented understanding of open data is further substantiated by the findings of the second research strand. Both the survey and the follow up interviews show that timeliness and provenance of the data are quite important for commercial stakeholders. Equally important are the licensing schemes attached to the open datasets. This is to be expected to some extent, as most of the products and services in the ecosystem rely on the timely publication of data from its publishers, which may be indicative of a power imbalance in their favour them within the ecosystem. Data properties are one of the biggest challenges of the data economy, which operates under the assumption that the data embedded in such products is accurate and renewed in a timely fashion. This also holds true for the open data ecosystem since our findings show that there is great variance in the quality of datasets released as open. Many of our interviewees discussed extensively the importance of high quality data for the success of the open data ecosystem. Moreover, there seems to be a close link between the quality of open data publications and the degree of understanding of the value of open data in government organisations. The interviews showed that commercial stakeholders with successful products use open datasets that fulfil all the quality attributes identified in the survey (Q14). These usually come from government organisations with strong internal open data cultures, as indicated in the interviews, which understand the value of open data to the economy and the benefits it will offer if published with high quality standards. These findings constitute an important indication that when open data starts to be used to enable the creation of new products and services, its attributes become critical for commercial success. They also show that platforms like OpenDataMonitor are well positioned to solve such problems as they serve as signaling mechanisms for the quality of published open datasets and the publication practices supporting them. Finally, they may indicate a causal relationship between government open data practices in specific sectors and the potential for product innovation and business success in these sectors of the economy. The interviews indicate that government organisations working closely with

startups in the field receive direct feedback from the users of the open data they publish and adjust its quality accordingly. Therefore, it is safe to say that a shared understanding of what is good quality open data improves both publication and use and the creation of economic value in the long run.

Equally important is the licensing of open datasets. The legal context in which open data commercial stakeholders operate is critical both for their success and the promotion of a sharing culture within the broader open data ecosystem. Licenses are an integral part of what constitutes open data and therefore they should be firmly integrated in the business practices of both publishers and users of open data. Empirical research showed that this is indeed the case for most of the commercial stakeholders operating in the ecosystem. However, there are some notable exceptions which still equate “open” with “free” data; releasing data without appropriate licensing. Albeit a minority, these examples show that there is still need for cultural change within the commercial ecosystem in order to institute an understanding of open data that is conducive of economic growth. Finally, data provenance is an important attribute. Commercial stakeholders need to be able to understand how each open dataset is created, what each data field means and the type of the values it contains, so that they can confidently reuse the data in their products and services. This level of familiarity with open data may be easier to achieve for stakeholders that have direct working relationships with government organisations or have sector-specific experience and knowledge prior to engagement with open data within that domain. It is, however, critical for the creation of successful open data products and services. There were other open data attributes highlighted as important during empirical research (see section 7). However, the overall sense from the field is that as commercial stakeholders work with open data they develop a shared understanding of what open data is and how it can be used. Moreover, the creation of open data products and services allows specific open data attributes are identified as critical for business innovation. It is therefore very important to establish mechanisms for this information to be communicated to the entire open data ecosystem so that all stakeholders can improve their publication practices in order to harness innovation.

A final observation has to do with the variety of sectors that start using open data. These is one of the most important outcomes of the analysis as it constitutes a clear indication of the rapid evolution of the open data ecosystem. It can serve as a strong argument in favour of open data publication and use while successful companies with innovative products and services constitute examples of the potential of open data for business innovation. Foremost, the rapid growth of the ecosystem justifies the need for an infrastructure, like OpenDataMonitor, that will monitor the quality of open data and provide clear information for both publishers and users.

8. CONCLUSION: CONSEQUENCES FOR ODM

The various analyses have illustrated that there exist barely any robust distinctions between the various stakeholder groups. Neither at a general level (public vs. private sector, data generator vs. data user) nor at more fine-grained sub-level (infrastructure provider vs. data processor vs. product developer) do stakeholder groups differ significantly in regard to what they see as fundamental for open data (data providers consider e.g. open licenses and machine-readability as important or more important than users of open data) or in the way they engage with open data (generators use open data to build applications, users provide open data themselves). Thus, there might be an apparent alignment in what the majority of stakehodlers engaged with open data deem important and relevant to know. This concerns most of the surveyed aspects of open data, in particular licenses and rights of use, formats and machine-readability, metadata and its quality, only the amount of open data being slightly less important. Hence, the current plans for the OpenDataMonitor platform are well on track to cover the most important aspects stakeholders need and what to know about open data.

Nevertheless, despite this apparent alignment, there still exists a high variance which means there are considerable differences at a more fine-grained level. These differences cut across stakeholder groups, however, and illustrate different opinions within the groups themselves. This finding is in line with the insight that roles cut across stakeholder groups. This situation might be indicative of the still nascent nature of the open data ecosystem, in which roles and functions are still emerging and actors have not yet specialised. Thus, stakeholder groups in general do not constitute an appropriate starting point for structuring the OpenDataMonitor platform. Rather, the aspects themselves and more granular and specific interests shall be offered to more informed users on a self-selection basis.

A point where interests do differ is the content of open data. Here, significant differences exist between the stakeholder groups, especially between civic activists, application-oriented developers and public administration staff. The content-related dimension of open data is thus important. This aspect has been considered within the consortium and built into the harmonisation engine. The categorisation of data into topic areas thus serves the interests of a large portion of the stakeholders.

Altogether, the findings show that the OpenDataMonitor covers the broad areas of interests of stakeholders in open data. However, since there is a high heterogeneity within the open data ecosystem, specialised interests might still be uncovered. Against this background, it has been decided within the consortium to create mini user stories for general as well as specific roles and

interests which will become part of Deliverable 3.5. This approach is in line with the agile development philosophy (Paetsch, Eberlein, and Maurer 2003; Sy 2007) and also strongly connects the research and the development part of the project. User stories are generated on a constant basis, initially within the project and eventually in an open process. These stories are then prioritised and implemented. Thereby, the OpenDataMonitor platform will not only be able to inform broad and general interests, but also provide insights for specific and granular interests within the open data ecosystem.

REFERENCES

Ackerman, John M., and Irma E. Sandoval-Ballesteros. 2006. “The Global Explosion of Freedom of Information Laws.” *Administrative Law Review* 58(1): 85–130.

Barry, Emily, and Frank Bannister. 2014. “Barriers to Open Data Release: A View from the Top.” *Information Polity* 19: 129–52.

Berners-Lee, Tim, James Hendler, and Ora Lassila. 2001. “The Semantic Web.” *Scientific American* 284(5): 29–37.

Birmingham City Council. 2013. *Birmingham City Council Open Data Policy*. Birmingham.

Bizer, Christian, Tom Heath, and T Berners-Lee. 2009. “Linked Data-the Story so Far.” *International Journal on Semantic Web and Information Systems* 5(3): 1–22.

Bovens, Mark. 2002. “Information Rights: Citizenship in the Information Society.” *The Journal of Political Philosophy* 10(3): 317–41.

Bundesministerium des Innern. 2014. *Nationaler Aktionsplan Der Bundesregierung Zur Umsetzung Der Open-Data-Charta Der G8*.

Couper, Mick P. 2011. “The Future of Modes of Data Collection.” *Public Opinion Quarterly* 75(5): 889–908.

Dahrendorf, Ralf. 2006. *Homo Sociologicus. Ein Versuch Zur Geschichte, Bedeutung Und Kritik Der Kategorie Der Sozialen Rolle*. 16. Auflag. Wiesbaden: VS Verlag für Sozialwissenschaften.

Davies, Tim. 2013. *Open Data Barometer. 2013 Global Report*. London.

Deloitte. 2012. *Open Data. Driving Growth, Ingenuity and Innovation*. London.

Geiger, Christian P., and Jörn von Lucke. 2012. “Open Government and (Linked) (Open) (Government) (Data).” *eJournal of eDemocracy & Open Government* 4(2): 265–78.

Group of Eight (G8). 2013. *G8 Open Data Charter*. Lough Erne.

Heimstädt, Maximilian, Fredric Saunderson, and Tom Heath. 2014. “From Toddler to Teen: Growth of an Open Data Ecosystem.” *eJournal of eDemocracy & Open Government* 6(2): 123–35.

Hesse-Biber, S.N. 2010. *Mixed Methods Research: Merging Theory with Practice*. New York: Guilford Press.

Hunnius, Sirko, Bernhard Krieger, and Tino Schuppan. 2014. “Providing, Guarding, Shielding: Open Government Data in Spain and Germany.” In *2014 EGPA Annual Conference, 10-12 September 2014 in Speyer, Germany*,.

Janssen, Katleen. 2011. “The Influence of the PSI Directive on Open Government Data: An Overview of Recent Developments.” *Government Information Quarterly* 28(4): 446–56.

Janssen, Katleen, and Jos Dumortier. 2003. “Towards a European Framework for the Re-Use of Public Sector Information: A Long and Winding Road.” *International Journal of Law and Information Technology* 11(2): 184–201.

Madsen, Anders Koed. 2013. “Virtual Acts of Balance!: Virtual Technologies of Knowledge-Management as Co-Produced by Social Intentions and Technical Limitations.” *Electronic Journal of e-Government* 11(2): 183–97.

McKinsey Global Institute. 2013. *Open Data: Unlocking Innovation and Performance with Liquid Information*.

Nesta. 2014. *Digital Social Innovation: Interim Report*.

Open Knowledge Foundation. 2005. “Open Knowledge Definition Released.” http://blog.okfn.org/2005/10/19/open-knowledge-definition-released/ (January 19, 2015).

OpenDataMonitor Project. 2014a. *D2.3 Best Practice Visualisation, Dashboard and Key Figures Report*. Vienna.

———. 2014b. *D2.4 Open Data Stakeholder Requirement Report 1*. Vienna.

Paetsch, Frauke, Armin Eberlein, and Frank Maurer. 2003. “Requirements Engineering and Agile Software Development.” In *WET ICE 2003. Proceedings. 12th IEEE International Workshops on Enabling Technologies: Infrastructure for Collaborative Enterprises, 2003.*, , 1–6.

Perritt, Henry H, and Zachary Rustad. 2000. “Freedom of Information Spreads to Europe.” *Government Information Quarterly* 17(4): 403–17.

Pollitt, Christopher, and Geert Bouckaert. 2011. *Public Management Reform. A Comparative Analysis*. 3rd editio. Oxford: Oxford University Press Oxford.

Saldaña, J. 2009. *The Coding Manual for Qualitative Researchers*. London: Sage.

Sanchez, M.E. 1992. “Effects of Questionnaire Design on the Quality of Survey Data.” *Public Opinion Quarterly* 56(2): 206–17.

Shadbolt, Nigel, Wendy Hall, and Tim Berners-Lee. 2006. “The Semantic Web Revisited.” *Intelligent Systems, IEEE* 21(3): 96–101.

Sunlight Foundation. 2010. “Ten Principles for Opening Up Government Information.” (October 2007). http://assets.sunlightfoundation.com.s3.amazonaws.com/policy/papers/Ten Principles for Opening Up Government Data.pdf.

Sy, Desirée. 2007. “Adapting Usability Investigations for Agile User-Centered Design.” *Journal of Usability Studies* 2(3): 112–32.

Tsebelis, George. 1999. “Veto Players and Law Production in Parliamentary Democracies: An Empirical Analysis.” *American Political Science Review* 93(3): 591–608.

———. 2002. *Veto Players. How Political Institutions Work*. Princeton, N.J.: Princeton University Press.

Zuiderwijk, Anneke, and Marijn Janssen. 2014. “Open Data Policies, Their Implementation and Impact: A Framework for Comparison.” *Government Information Quarterly* 31(1): 17–29.

Zuiderwijk, Anneke, Marijn Janssen, Sunil Choenni, Ronald Meijer, and Roexsana Sheikh Alibaks. 2012. “Socio-Technical Impediments of Open Data.” *Electronic Journal of e-Government* 10(2): 156–72.

APPENDIX: DETAILED RESULTS TABLES

Table 0.1 Relevance of Aspects of Open Data by Policy-level vs. Operational Level

	Included	percent	Excluded	percent	Total		Mean	N	Std. Deviation
raw data									
policy-level	89	96.7391%	3	3.26087%	92	100%	2.36	89	1.51
operational level	126	96.1832%	5	3.81679%	131	100%	2.43	126	1.48
open formats									
policy-level	89	96.7391%	3	3.26087%	92	100%	2.27	89	1.62
operational level	124	94.6565%	7	5.34351%	131	100%	2.1	124	1.46
open licenses									
policy-level	89	96.7391%	3	3.26087%	92	100%	2.36	89	1.69
operational level	123	93.8931%	8	6.10687%	131	100%	2.15	123	1.51
free of charge									
policy-level	89	96.7391%	3	3.26087%	92	100%	2.35	89	1.49
operational level	127	96.9466%	4	3.05344%	131	100%	2.1	127	1.38
machine-readability									
policy-level	89	96.7391%	3	3.26087%	92	100%	2.17	89	1.65
operational level	127	96.9466%	4	3.05344%	131	100%	2.16	127	1.5
no use restrictions									
policy-level	89	96.7391%	3	3.26087%	92	100%	2.27	89	1.48
operational level	127	96.9466%	4	3.05344%	131	100%	2.19	127	1.24
meaningful metadata									
policy-level	89	96.7391%	3	3.26087%	92	100%	2.45	89	1.4
operational level	126	96.1832%	5	3.81679%	131	100%	2.48	126	1.39
amount of open data									
policy-level	89	96.7391%	3	3.26087%	92	100%	2.58	89	0.96
operational level	125	95.4198%	6	4.58015%	131	100%			
quality of open data									
policy-level	89	96.7391%	3	3.26087%	92	100%	2.27	89	1.38
operational level	128	97.7099%	3	2.29008%	131	100%	2.27	128	1.49
metadata quality									
policy-level	89	96.7391%	3	3.26087%	92	100%	2.42	89	1.34
operational level	127	96.9466%	4	3.05344%	131	100%	2.45	127	1.36

Table 0.2 Test of differences between the "classification of the aspects" and categories Policy-level (activists, strategists, politicians, advisors combined), Operational-level (users and generator) (ANOVA)

	ANOVA one-way test	Bartlett's test	Alternative Test for ANOVA (oav)
raw	0,740	0,857	0,739
OF	0,424	0,312	0,416
OL	0,344	0,255	0,335
FC	0,221	0,418	0,214
MR	0,960	0,337	0,959
N-UR	0,674	0,074	0,664
MMD	0,890	0,951	0,890
AOD	0,237	0,028	0,254
ODQ	0,985	0,442	0,985
MDQ	0,859	0,899	0,859

Table 0.3 Relevance of Aspects of Open Data by Users and non-Users of Open Data

	Included	percent	Excluded	percent	Total		Mean	N	Std. Deviation
raw data									
personally used: yes	182	98.91%	2	1.09%	184	100%	2.48	182	1.53
personally used: no	43	97.73%	1	2.27%	44	100%	2.26	43	1.35
open formats									
personally used: yes	181	98.37%	3	1.63%	184	100%	2.27	181	1.57
personally used: no	43	97.73%	1	2.27%	44	100%	1.91	43	1.29
open licenses									
personally used: yes	179	97.28%	5	2.72%	184	100%	2.29	179	1.61
personally used: no	44	100%	0	0%	44	100%	2.07	44	1.4
free of charge									
personally used: yes	182	98.91%	2	1.09%	184	100%	2.26	182	1.43
personally used: no	44	100%	0	0%	44	100%	1.89	44	1.2
machine-readability									
personally used: yes	181	98.37%	3	1.63%	184	100%	2.34	181	1.64
personally used: no	44	100%	0	0%	44	100%	1.73	44	1.21
no use restrictions									
personally used: yes	183	99.46%	1	0.54%	184	100%	2.29	183	1.37
personally used: no	44	100%	0	0%	44	100%	2.07	44	1.25
meaningful metadata									
personally used: yes	183	99.46%	1	0.54%	184	100%	2.64	183	1.41
personally used: no	43	97.73%	1	2.27%	44	100%	1.91	43	1.23
amount of open data									
personally used: yes	181	98.37%	3	1.63%	184	100%	2.82	181	1.14
personally used: no	44	100%	0	0%	44	100%	2.27	44	1.04
quality of open data									
personally used: yes	183	99.46%	1	0.54%	184	100%	2.43	183	1.49
personally used: no	44	100%	0	0%	44	100%	1.95	44	1.28
metadata quality									
personally used: yes	181	98.37%	3	1.63%	184	100%	2.6	181	1.36
personally used: no	44	100%	0	0%	44	100%	1.95	44	1.18

Table 0.4 Test of the differences between the "classification of the aspects" and "categories (Yes, personally used open data), (No, personally never used)" (ANOVA)

	ANOVA one-way test	Bartlett's test	Alternative Test for ANOVA (oav)
raw	0,347	0,309	0,382
OF	0,121	0,123	0,165
OL	0,364	0,280	0,400
FC	0,077	0,173	0,107
MR	**0,006**	**0,017**	**0,020**
N-UR	0,303	0,455	0,328
MMD	**0,001**	**0,277**	**0,002**
AOD	**0,003**	0,462	**0,004**
ODQ	0,037	0,208	0,054
MDQ	**0,002**	**0,260**	**0,004**

ANNEX I: SURVEY I QUESTIONNAIRE

Open Data Survey

Which country are you from?

Please choose **only one** of the following:

- Austria
- Belgium
- …
- Spain
- Sweden
- United Kingdom
- no-UE Europa
- Africa
- Asia
- Australia
- North America
- South America

What is your main role in the context of open data?

Choose one of the following answers. If you have numerous roles, please choose the one you consider primary. Please choose **only one** of the following:

- **open data generator in the public sector**: is personally involved in generating, preparing and / or uploading open data
- **open data user**: uses open data, no matter for what purpose (data journalism, app development, business analytics, research etc.)
- **open data activist**: is involved with open data activism, no matter for what purpose (commercial reuse, civic activism)
- **open data / IT strategist within the public sector**: responsible for the strategic development of open data, often in e-government, IT, open government and / or administrative modernisation
- **advisor on open data in the public sector**: is involved in decisions about publishing open data, e.g. legal counseling, technical support
- **external consultant on open data**: supports and consults the public administration in regard to open data, no matter if technical, legal or otherwise
- **politician**: is politically responsible for open data or involved with it, no matter if in the legislative or political executive branch

For the overall success of open data, how important are the following aspects:

Please classify from 1 to 5 (1 = very important... 5 = not important at all)

Please choose the appropriate response for each item:

	1	2	3	4	5
raw / primary source data					
open formats					
open licenses					
free of charge					
machine-readability					
no use restrictions					
meaningful metadata					
amount of open data					
open data quality					
metadata quality					

Which data do you consider especially interesting and relevant as open data?

Please choose the 5 categories you consider most relevant .
Double-click or drag-and-drop items in the left list to move them to the right.

- Culture and Leisure
- Consumer Protection
- Demographics
- Economics
- Education
- Employment
- Energy
- Environment and Climate
- Finance and Budget
- Geography and Geology
- Health
- Housing
- Industry
- Infrastructure
- Law and Justice
- Politics and Elections
- Public Sector
- Rural Affairs
- Science and Technology
- Security
- Sports
- Tourism
- Trade
- Transport and Traffic
- Welfare

In your opinion, the following groups support and push for open data:

Please select those three groups that you consider to must strongly support and push for open data.

Please select at most 3 answers
Please choose **all** that apply:

- politicians
- senior and mid-level public managers
- supra- and international bodies (e.g. European Commission, OECD, World Bank)
- civic and political activists / activist groups
- journalists and media industry
- businesses and business associations

Does in your administration or organisational unit exist an open data policy?

Only answer this question if the following conditions are met:
Answer was '**advisor on open data in the public sector** ' *or* ' **open data / IT strategist within the public sector** ' *or* ' **open data generator in the public sector**' at question '2 [Q1]' (What is your main role in the context of open data?)

Please choose **only one** of the following:

- ☐ Yes
- ☐ No

If yes, what does this policy contain?

Only answer this question if the following conditions are met:
Answer was 'Yes' at question '6 [Q2]' (Does in your administration or organisational unit exist an open data policy?)

Please choose the appropriate response for each item:

	general recommendation	binding rules	non-existent
pricing			
data formats			
metadata formats			
licenses			
number of data sets			
general liability clause			

Are you involved in drafting open data policies for the public administration?

Only answer this question if the following conditions are met: Answer was '**politician** ' *or* ' **external consultant on open data** ' at question '2 [Q1]' (What is your main role in the context of open data?)

Please choose **only one** of the following:

- Yes
- No

What should such a policy say?

Only answer this question if the following conditions are met:
Answer was 'Yes' at question '8 [QQ21]' (Are you involved in drafting open data policies for the public administration?)

Please choose the appropriate response for each item:

	general recommendation	binding rules	no regulation
pricing			
data formats			
metadata formats			
licenses			
number of data sets			
general liability clause			

Who is involved in the decision whether or not specific data can be published as open data?

Check all that apply

Only answer this question if the following conditions are met:
Answer was ' **advisor on open data in the public sector** ' *or* ' **open data / IT strategist within the public sector** ' *or* ' **open data generator in the public sector**' at question '2 [Q1]' (What is your main role in the context of open data?)

Please choose **all** that apply:

- General “Open by default”-clause
- administrative department that generates data
- legal department
- ICT department
- open data officer
- committee with representatives exclusively from within the public administration
- committee with representatives from within and outside the public administration (businesses, advocacy groups etc.)
- data protection / security officer
- politicians

Who should be involved in the decision whether or not specific data can be published as open data?

Check all that apply

Only answer this question if the following conditions are met:
Answer was ' **politician**' at question '2 [Q1]' (What is your main role in the context of open data?)

Please choose **all** that apply:

- general open by default clause
- administrative department that generates data
- legal department
- ICT department
- open data officer
- committee with representatives exclusively from within the public administration
- committee with representatives from within and outside the public administration (businesses, advocacy groups etc.)
- data protection / security officer
- politicians

Who actually prepares and uploads a specific data set?

Check all that apply

Only answer this question if the following conditions are met:
Answer was '**advisor on open data in the public sector** ' *or* ' **open data / IT strategist within the public sector** ' *or* ' **open data generator in the public sector** ' at question '2 [Q1]' (What is your main role in the context of open data?)

Please choose **all** that apply:

- administrative department or public organisation that generates data
- ICT department
- open data officer
- other

You chose "other", please specify your answer

Only answer this question if the following conditions are met:
Answer was 'other' at question '12 [Q4]' (Who actually prepares and uploads a specific data set?)

Please write your answer here:

Does the software (not the Office suite!) you work with enable you to generate open data?

Only answer this question if the following conditions are met:
Answer was '**advisor on open data in the public sector** ' *or* ' **open data / IT strategist within the public sector** ' *or* ' **open data generator in the public sector** ' at question '2 [Q1]' (What is your main role in the context of open data?)

Please choose **only one** of the following:

- yes
- no
- don't know

The software:

Check all that apply

Only answer this question if the following conditions are met:
Answer was 'yes' at question '14 [Q1]' (Does the software (not the Office suite!) you work with enable you to generate open data?)

Please choose **all** that apply:

- automatically uploads defined data sets to a public portal

- enables manually exporting data with automatically generated metadata
- enables manually exporting data without additional features

Have you ever had contact with users in the context of open data?

Only answer this question if the following conditions are met:
Answer was '**open data generator in the public sector** ' *or* ' **advisor on open data in the public sector** ' *or* ' **open data / IT strategist within the public sector** ' *or* '**politician** ' at question '2 [Q1]' (What is your main role in the context of open data?)

Please choose **only one** of the following:

- yes
- no

You were in contact with the user:

Check all that apply

Only answer this question if the following conditions are met:
Answer was 'yes' at question '16 [Q11]' (Have you ever had contact with users in the context of open data?)

Please choose **all** that apply:

- when specific data was requested
- to clarify questions around available data
- to discuss open data in general
- for any other reason

Have you ever contacted the public administration in the context of open data?

Only answer this question if the following conditions are met:
Answer was '**open data activist** ' *or* ' **open data user** ' at question '2 [Q1]' (What is your main role in the context of open data?)

Please choose **only one** of the following:

- yes
- no

For which reason?

Check all that apply

Only answer this question if the following conditions are met:
Answer was 'yes' at question '18 [Q12]' (Have you ever contacted the public administration in the context of open data?)

Please choose **all** that apply:

- to request specific data
- to clarify questions around available data
- to discuss open data in general
- for any other reason

Have you ever personally used an open data set?

Please choose **only one** of the following:

- yes
- no

You have personally used an open data set

Check all that apply

Only answer this question if the following conditions are met:
Answer was 'yes' at question '20 [Q2]' (Have you ever personally used an open data set?)

Please choose **all** that apply:

- to play around with / test it
- to build an application
- to carry out business analytics
- for data journalism
- for a different purpose

You chose "for a different purpose", please clarify this purpose

Only answer this question if the following conditions are met:
Answer was 'for a different purpose' at question '21 [Qqq]' (You have personally used an open data set)

Please write your answer here:

From which level(s) of Government do you use open data?

Check all that apply

Only answer this question if the following conditions are met:
Answer was 'yes' at question '20 [Q2]' (Have you ever personally used an open data set?)

Please choose **all** that apply:

- local / municipal
- regional / state level
- national

- supra- / international

With which level(s) of Government do you primarily engage?

Check all that apply

Only answer this question if the following conditions are met:
Answer was '**open data activist** ' at question '2 [Q1]' (What is your main role in the context of open data?)

Please choose **all** that apply:

- local / municipal
- regional / state level
- national
- supra- / international

In which policy field do you work?

Choose one of the following answers.
If more than one apply, choose the main policy field.

Only answer this question if the following conditions are met:
Answer was ' **advisor on open data in the public sector** ' *or* ' **open data / IT strategist within the public sector** ' *or* ' **open data generator in the public sector** ' at question '2 [Q1]' (What is your main role in the context of open data?)

Please choose **only one** of the following:

- Agriculture, Fisheries, Food, Energy and / or Natural Resources
- Business, Economy, Finance and / or Tax
- Culture, Education, Youth, Science and / or Technology
- Employment and / or Social Rights / Affairs
- Environment, Consumers and / or Health
- External Relations and / or Foreign Affairs
- Justice, Citizens' Rights, Home Affairs and / or General Administration
- Regions and / or Local Development
- Transport, Travel and / or Tourism

In which policy field are you politically active?

Choose one of the following answers.
If more than one apply, choose the main policy field.

Only answer this question if the following conditions are met:
Answer was '**politician** ' at question '2 [Q1]' (What is your main role in the context of open data?)

Please choose **only one** of the following:

- Agriculture, Fisheries, Food, Energy and / or Natural Resources
- Business, Economy, Finance and / or Tax
- Culture, Education, Youth, Science and / or Technology
- Employment and / or Social Rights / Affairs
- Environment, Consumers and / or Health
- External Relations and / or Foreign Affairs
- Justice, Citizens' Rights, Home Affairs and / or General Administration
- Regions and / or Local Development
- Transport, Travel and / or Tourism

In which industry do you work?

Choose one of the following answers. If more than one apply, choose the main industry (**NACE classification by Eurostat**).

Only answer this question if the following conditions are met:
Answer was ' **external consultant on open data** ' *or* ' **open data user** ' *or* ' **open data activist** ' at question '2 [Q1]' (What is your main role in the context of open data?)

Please choose **only one** of the following:

- Agriculture, forestry and fishing
- Mining and quarrying
- Manufacturing
- Electricity, gas, steam and air conditioning supply
- Water supply; sewerage, waste management and remediation activities
- Construction
- Wholesale and retail trade
- Transportation and storage services
- Accommodation and food service activities
- Information and communication
- …

At which level of Government do you work?

Only answer this question if the following conditions are met:
Answer was ' **advisor on open data in the public sector** ' *or* '**open data / IT strategist within the public sector** ' *or* ' **open data generator in the public sector** ' at question '2 [Q1]' (What is your main role in the context of open data?)

Please choose **only one** of the following:

- local / municipal
- regional / state level
- national
- supra- / international

Which level(s) of Government do you primarily consult on open data?

Check all that apply

Only answer this question if the following conditions are met:
Answer was '**external consultant on open data** ' at question '2 [Q1]' (What is your main role in the context of open data?)

Please choose **all** that apply:

- local / municipal
- regional / state level
- national
- supra- / international

At which level of government are you politically active?

Choose one of the following answers. If more than one apply,
choose the main level of government.

Only answer this question if the following conditions are met:
Answer was ' **politician**' at question '2 [Q1]' (What is your main role in the context of open data?)

Please choose **only one** of the following:

- local / municipal
- regional / state level
- national
- supra- / international

ANNEX II: SURVEY II QUESTIOINAIRE

The survey was created using Google forms. The questions in the order presented to the respondents are presented below.

Company Details

1. Company name

2. What is the size of your company?
   - fewer than 10 employees
   - 10 - 50 employees
   - 51 - 250 employees
   - 251 - 1000 employees
   - More than 1000 employees

3. Which category best describes your company's area of business?
   - Business & Legal Services
   - Data/Technology
   - Education
   - Energy
   - Environment & Weather
   - Finance & Investment
   - Food & Agriculture
   - Geospatial/Mapping
   - Governance
   - Healthcare
   - Housing/Real Estate
   - Insurance
   - Research & Consulting
   - Scientific Research
   - Transportation

4. Which of the following are significant sources of revenue for your company?

Tick all that apply.

   - Advertising
   - Consulting
   - Contributions/Donations
   - Data analysis for clients
   - Database licensing
   - Government contracts
   - Lead generation to other businesses
   - Membership fees

- Grants
- Software licensing
- Subscriptions
- User fees for web or mobile access
- Other: (free text answer)

Open Data Usage

5. How does your company currently use open data?

Tick all that apply.

- My company publishes open data
- My company provides infrastructure for others to publish open data (e.g. platforms, portals, data stores)
- My company processes open data (e.g aggregation, classification, anonymization, cleaning, refining, enriching)
- My company provides insights based on open data (e.g. analytics, visualisations)
- My company develops products based on open data (e.g. APIs, apps)
- Other: (free text answer)

6. What types of open data does your company use?

Tick all that apply.

- Agriculture & Food
- Business
- Consumer
- Demographics & Social
- Economics
- Education
- Energy
- Environment
- Finance
- Geospatial/Mapping
- Government operations
- Health/Healthcare
- Housing
- International/Global development
- Legal
- Manufacturing
- Science & Research
- Public safety
- Tourism

- Transportation
- Weather
- Other: (free text answer)

7. Does your company currently use open government datasets?
   - Yes
   - No
8. If yes, which open government datasets does it currently use?
   e.g. Transport for London ('station locations'), UK Hydrographic Office ('tidal predictions'), etc; please list all datasets or those of greatest significance/value to your company. (free text answer)
9. Does your company currently use other open datasets, such as those provided by businesses, charities, or community projects?
   e.g DBpedia, Geonames, Musicbrainz, etc; please list all datasets or those of greatest significance/value to your company.
   - Yes
   - No
10. If yes, which other open datasets does it currently use? (free text answer)

Product/Service Information

11. Please name and briefly describe your company's products and/or services that utilise open data. (free text answer)
12. Which pricing mechanism(s) does your company use for its open data products and/or services?
13. Tick all that apply
    - Provide unlimited free access to everyone
    - Provide limited free access to everyone (e.g. rate or volume limited)
    - Provide free access to only a subset of people
    - Provide only paid-for access
    - Other: (free text answer)
14. If your company has previously used open data in particular products and/or services but ceased to do so, could you please describe why? (free text answer)

Challenges in the use of Open Data

15. Please indicate the extent to which each of the following issues influence your company's decision to use open data:

	1 (little or no influence)	2	3	4	5 (great influence)
Provenance of data					
Licensing of datasets					
Accuracy of data					
Timeliness of data					
Ease of access to datasets					
Format of data					
Accompanying documentation					
Help and support from publisher					

16. If you wish to elaborate on any of these issues, please do so. (free text answer)

ANNEX III: INTERVIEW QUESTIONNAIRE

Open Data and innovation

1. What are your business incentives for publishing/using open data as a company (from a strategic point of view)?
2. What are the business opportunities when publishing/using open data?
   - Do you use open data to gain insights on your customer base?
   - Do you use open data to reach new market segments?
3. How central is open data in the development of products and services in your company?
4. What are the challenges when developing products and services based on open data?
   - Quality of datasets
   - Provenance of datasets
   - Continuity in the publication of datasets
5. How do you go about developing products with open data?
   - Was there a specific problem you were trying to solve?
   - Did you see potential for an innovative product or service in datasets released by government agencies?
6. Can you please explain to me the business rationale for developing a new line of products based on open data? How does it fit with your existing line of products and services? What was the rationale behind it?

Open data and growth

7. Do you have plans to scale up your open data product/services? If not, why?
8. If yes, how do foresee scaling up?
   - Introducing the product/service in overseas markets?
   - Replicating the service across other cities (UK and EU)?
   - Scaling the open data used/published by the company?
9. What are the challenges in scaling up with open data?
   - Do you require datasets that are difficult to find in other countries?
   - Are products/services at the city level challenging to scale?
   - Do you find difficulties to access markets abroad? If yes, for which reasons?

Open data and the future

10. Has your involvement with open data changed the philosophy and/or values of your business?
11. Will you diversify your service and product offerings based on open data?
12. How do you see the future of open data in your industry field?