



OpenGovIntelligence

Fostering Innovation and Creativity in Europe through Public
Administration Modernization towards Supplying and Exploiting
Linked Open Statistical Data

Deliverable 6.1

Data Management Plan

Editor(s):	Evangelos Kalampokis (CERTH)
Responsible Organisation:	CERTH
Version-Status:	V4.0
Submission date:	31/01/2019
Dissemination level:	CO

Deliverable factsheet

Project Number:	693849
Project Acronym:	OpenGovintelligence
Project Title:	Fostering Innovation and Creativity in Europe through Public Administration Modernization towards Supplying and Exploiting Linked Open Statistical Data

Title of Deliverable:	D6.1 – Data Management Plan
Work package:	WP6 – Project Management
Due date according to contract:	31/07/2016 31/01/2017 [first update] 31/07/2017 [second update] 31/01/2019 [third update]

Editor(s):	Evangelos Kalampokis (CERTH)
Contributor(s):	Jamie Whyte (TRAF) Trevor Alcorn (MI) Ramūnas Čižas (EL) Eleni Chaniotaki (MAREG) Paul Hermans (ProXML) Marijn Janssen (TUDelft) Robert Krimmer (TUT) Ricardo Matheus (TUDelft) Maarja Toots (TUT) Areti Karamanou (CERTH) Efthimios Tambouris (CERTH) Konstantinos Tarabanis (CERTH)
Reviewer(s):	-
Approved by:	All Partners

Abstract:	<p>This document provides the plan for managing the data generated and collected during the project. It covers: a) the handling of research data during and after the project, b) what data will be collected, processed or generated, c) what methodology and standards will be applied, d) whether data will be shared/made open and how and e) how data will be curated and preserved (including after the end of the project).</p> <p>The project in total involved twenty three data sets from eight different CSA domains including <i>Labour</i>, <i>Business statistics</i>, <i>Environment</i>, <i>Population and migration</i> and others. The data sets are provided in various formats (e.g. RDF, CSV), have various sizes (from 5kB to 127MB), and are originated from various sources (e.g. from the UK's National Statistics Excerpt, the Flemish government, or the Lithuanian Official Statistics Portal). Most data sets are openly provided to the public and published as linked data using the RDF Data Cube vocabulary facilitating their interoperability. Metadata of most data sets are modelled using standard vocabularies such as DCAT and PROV enhancing interoperability. Access to open data sets is given mainly through a HTML page, SPARQL endpoint, and/or a restful API. Regarding licensing, most data sets are provided under Open Government Licence 3.0 or publicly available data licences. The majority of open data sets will be preserved after the end of the project in various mediums including Open Science Framework and Zenodo. No resources were used to produce data sets. Moreover, the security of the data sets was ensured by using, for example, backup devices, and by backing up data in repositories like GitHub, Zenodo, and GitLab. Finally, regarding ethical aspects, for the data sets that include personal data, the consortium ensured that they comply with EU General Data Protection Regulation (GDPR) and also consent forms were provided to participants of interviews and surveys.</p>
Keyword List:	Project, management, data, plan

Effort of Participating Partners Consortium

	<i>Name</i>	<i>Short Name</i>	<i>Role</i>	<i>Person Months</i>
1.	Centre for Research & Technology - Hellas	CERTH	Participant	4
2.	Delft University of Technology	TU Delft	Participant	0.1
3.	National University of Ireland, Galway	NUIG	Leader	0.1
4.	Tallinn University of Technology	TUT	Participant	0.1
5.	ProXML bvba	ProXML	Participant	0.1
6.	Swirrl IT Limited	SWIRRL	Participant	0.1
7.	Trafford council	TRAF	Participant	0.1
8.	Flemish Government	VLO	Participant	0.1
9.	Ministry of Interior and Administrative Reconstruction	MAREG	Participant	0.1
10.	Ministry of Economic Affairs and Communication	MKM	Participant	0.1
11.	Marine Institute	MI	Participant	0.1
12.	Public Institution Enterprise Lithuania	EL	Participant	0.1

Revision History

<i>Version</i>	<i>Date</i>	<i>Revised by</i>	<i>Reason</i>
0.1	15/03/2016	CERTH	TOC
0.2	30/06/2016	CERTH	TOC updated
0.3	20/06/2016	ALL	Data management policy, input from pilots regarding data sets inserted
0.4	26/07/2016	CERTH	Introduction updated, abstract inserted
1.0	31/07/2016	CERTH	Final version
2.0 (1st Update)	25/01/2017	CERTH	Document updated to include new data set descriptions from the Trafford council (TRAF)
3.0	16/06/2017	CERTH	Updates according to the “Expert Report”
4.0	31/01/2019	CERTH	Final version

Statement of originality:

This deliverable contains original unpublished work except where clearly indicated otherwise. Acknowledgement of previously published material and of the work of others has been made through appropriate citation, quotation or both.

Table of Contents

DELIVERABLE FACTSHEET	2
EFFORT OF PARTICIPATING PARTNERS CONSORTIUM	4
REVISION HISTORY	5
TABLE OF CONTENTS	6
LIST OF TABLES	8
LIST OF ABBREVIATIONS.....	9
EXECUTIVE SUMMARY	10
1 INTRODUCTION	11
1.1 SCOPE.....	11
1.2 AUDIENCE.....	11
1.3 STRUCTURE.....	11
2 METHODOLOGY.....	12
3 DATA MANAGEMENT POLICY.....	13
3.1 DATA SUMMARY	13
3.2 FAIR DATA.....	13
3.2.1 <i>Making data findable, including provisions for metadata</i>	13
3.2.2 <i>Making data openly accessible</i>	15
3.2.3 <i>Making data interoperable</i>	16
3.2.4 <i>Increase data re-use</i>	17
3.3 ALLOCATION OF RESOURCES	20
3.4 DATA SECURITY	20
3.5 ETHICAL ASPECTS	20
3.6 OTHER ISSUES.....	20
3.7 OPEN ACCESS SCIENTIFIC PUBLICATIONS	20
4 DATA SETS ANALYSIS	21
4.1 DATA SUMMARY	21
4.2 FAIR DATA.....	22
4.2.1 <i>Making data findable, including provisions for metadata</i>	22
4.2.2 <i>Making data openly accessible</i>	24
4.2.3 <i>Making data interoperable</i>	26
4.2.4 <i>Increase data re-use</i>	29
4.3 ALLOCATION OF RESOURCES	35
4.4 DATA SECURITY	36
4.5 ETHICAL ASPECTS	40
4.6 OTHER ISSUES.....	40
4.7 OPEN ACCESS PUBLICATIONS	40
5 CONCLUSION	43
REFERENCES	44

APPENDIX 1: DMP TEMPLATE	45
APPENDIX 2: DATA SETS.....	47
<i>MARINE INSTITUTE DATA SETS</i>	<i>47</i>
<i>Oceanographic data on the weather and wave climate</i>	<i>47</i>
<i>Oceanographic data on the wave climate</i>	<i>49</i>
<i>TRAF DATA SETS.....</i>	<i>52</i>
<i>Claimant count.....</i>	<i>52</i>
<i>Claimant rate</i>	<i>53</i>
<i>Households with lone parent not in employment</i>	<i>55</i>
<i>Social rented households</i>	<i>56</i>
<i>Working age adults with no qualifications</i>	<i>58</i>
<i>Employees by broad industry group.....</i>	<i>59</i>
<i>Working age population</i>	<i>60</i>
<i>MAREG DATA SETS.....</i>	<i>62</i>
<i>Government Vehicles of the Greek Government Agency</i>	<i>62</i>
<i>FLEMISH GOVERNMENT DATA SETS</i>	<i>63</i>
<i>Air and Water Pollution Report Metadata.....</i>	<i>63</i>
<i>Air and Water Pollution Data.....</i>	<i>65</i>
<i>Company registry</i>	<i>66</i>
<i>ENTERPRISE LITHUANIA DATA SETS.....</i>	<i>67</i>
<i>Economic Activities in Lithuania.....</i>	<i>67</i>
<i>State Social Insurance - Sodra.....</i>	<i>69</i>
<i>Average earnings (monthly (2015))</i>	<i>71</i>
<i>Investment in tangible fixed assets at current prices (2015)</i>	<i>72</i>
<i>Foreign direct investment per capita at the end of the year (2014)</i>	<i>73</i>
<i>Economic entities in operation at the beginning of the year</i>	<i>75</i>
<i>Resident population at the beginning of the year (by age groups).....</i>	<i>76</i>
<i>Resident population at the beginning of the year (by gender)</i>	<i>78</i>
<i>Resident population at the beginning of the year (persons by municipality)</i>	<i>79</i>
<i>TUT DATA SETS.....</i>	<i>81</i>
<i>Interviews' and online survey's data.....</i>	<i>81</i>

List of Tables

TABLE 1 DATA QUALITY DIMENSIONS AND RECOMMENDATIONS 19

TABLE 2. DATA SETS AND CSA DOMAINS 22

TABLE 3 DATA SETS AND METADATA 23

TABLE 4. DATA SETS, ACCESS POLICY AND SHARING MEDIUM 24

TABLE 5. DATA SETS, STANDARDS AND METADATA 27

TABLE 6 DATA SETS AND LICENSES..... 29

TABLE 7 DATA SETS AND QUALITY ASSURANCE DESCRIPTIONS..... 30

TABLE 8 DATA SETS AND PRESERVATION DURATION 34

TABLE 9 DATA SETS AND COST ESTIMATION..... 35

TABLE 10 DATA SETS AND DATA SECURITY 36

TABLE 11 OPEN ACCESS PUBLICATIONS OF THE PROJECT..... 40

List of Abbreviations

The following table presents the acronyms used in the deliverable in alphabetical order.

<i>Abbreviation</i>	<i>Description</i>
DMP	Data Management Plan
DCAT	Data Catalogue Vocabulary
RDF	Resource Description Standard
WP	Work Package
PA	Public Administration
LOSD	Linked Open Statistical Data
ICT	Information and Communications Technology
EU	European Commission
CSA	Classification of Statistical Activities
GDPR	General Data Protection Regulation
FAIR	findable, accessible, interoperable and reusable

Executive Summary

This is the fourth submission of deliverable “D6.1 Data Management Plan” of the OpenGovIntelligence Project. The OpenGovIntelligence project aims at stimulating sustainable economic growth in Europe through fostering innovation in society and enterprises. Towards this end, OpenGovIntelligence suggests a holistic approach for the modernization of Public Administration (PA) by exploiting Linked Open Statistical Data (LOSD) technologies. This includes new business processes, policies, and tools that will enable the active participation of the society and enterprises in data sharing and in the co-production of innovative data-driven public services.

This deliverable documents activities that are performed in WP6 “Project Management”. WP6 intends to:

- Perform strategic and day-to-day administrative, financial, scientific and technical management of the project.
- Ensure the sound management of project activities and the fulfilment of project objectives.
- Monitor resource usage, budget allocation and project cash flow.
- Ensure efficient communication within the consortium and reassure effective liaison with the EC, other projects, communities and other bodies as required.
- Deal with all ethical issues that might emerge during the project.

The purpose of this document is to provide the plan for managing the data generated and collected during the project; The Data Management Plan. Specifically, the DMP describes the data management life cycle for all datasets to be collected, processed and/or generated by a research project. It covers:

- the handling of research data during and after the project
- what data will be collected, processed or generated.
- what methodology and standards will be applied.
- whether data will be shared/made open and how
- how data will be curated and preserved (including after the end of the project)

Following the EU’s guidelines regarding the DMP, this document was updated – three times during the project lifetime (in the form of deliverables). The consortium also complies with the EU GDPR regulation when storing data.

The DMP involves twenty three data sets from eight different CSA domains including *Labour*, *Business statistics*, *Environment*, *Population and migration* and others. Most data sets are openly provided to the public and also most data sets are published as linked data using the RDF Data Cube vocabulary. Metadata are modelled using standard vocabularies such as DCAT and PROV enhancing interoperability. All data sets will be preserved after the end of the project in various mediums including Open Science Framework and Zenodo.

1 Introduction

This document is the fourth version of OpenGovIntelligence Data Management Plan (DMP). The consortium is required to create the DMP because the OpenGovIntelligence project participates in the Open Research Data pilot. The DMP describes the data management life cycle for all datasets to be collected, processed and/or generated by a research project. It covers:

- the handling of research data during and after the project
- what data will be collected, processed or generated.
- what methodology and standards will be applied.
- whether data will be shared/made open and how
- how data will be curated and preserved (including after the end of the project)

This version of the DMP takes into account the latest version (v3.0) of the EU's guidelines for the DMP (European Commission, 2016b) that helps Horizon 2020 beneficiaries make their research data findable, accessible, interoperable and reusable.

1.1 Scope

The present document is the Deliverable 6.1 "D6.1 – Data Management Plant" (henceforth referred to as D6.1) of the OpenGovIntelligence project. The main objective of D6.1 is to provide the plan for managing the data generated and collected during the project.

According to the EU's guidelines regarding the DMP (European Commission, 2016), the document may be updated - if appropriate - during the project lifetime (in the form of deliverables).

1.2 Audience

The intended audience for this document is the OpenGovIntelligence consortium and the European Commission.

1.3 Structure

The structure of the document is as follows:

- Section 2 presents the methodology that the consortium used to create the DMP.
- Section 3 presents the policy used by the consortium of the OpenGovIntelligence project plans regarding the data sets collected, processed and/or generated within the project.
- Section 4 presents the analysis of the data sets included in the DMP.
- Section 5 draws conclusion and sets future goals.

2 Methodology

The Data Management Plan regards all the data sets that will be collected, processed and/or generated within the project. The methodology the consortium follows to create the DMP is as follows:

1. Create a data management policy. To this end, we describe a) the items that the EU proposes to address for each data set and b) the strategy that is used by the consortium to address each item. The items were also used to create a DMP template which was sent to the partners of the consortium in order to fill it in with information for each relative data set.
2. Analyse the completed by the project's partners DMP templates.

3 Data management policy

The DMP takes into account the latest version (v3.0) of the EU's guidelines for the DMP (European Commission, 2016b) that helps Horizon 2020 beneficiaries make their research data findable, accessible, interoperable and reusable. According to EU's guidelines, this document should provide answers to questions related to:

1. Data summary (e.g. purpose of the data, types and formats, expected size etc.)
2. FAIR Data
 - a. Making data findable, including provisions for metadata
 - b. Making data openly accessible
 - c. Making data interoperable
 - d. Increase data re-use
3. Allocation of resources
4. Data security
5. Ethical aspects
6. Other issues

In addition, partners have created a policy about facilitating the access to open access publications produced throughout the project's lifecycles and ensure their long time preservation.

For the data sets that include personal data, the consortium ensured that they comply with EU General Data Protection Regulation (GDPR). GDPR is EU's legal framework on privacy and personal data protection that replaces the Directive 95/46/EC (the 'Data Protection Directive'). More details can be found at deliverable D6.5 Ethics policy.

The consortium developed a number of strategies that were followed to address the elements proposed by the EU's guidelines for the DMP. In this section, we provide a detailed description of these strategies used to address each item of the above list.

3.1 Data summary

Each data set that was collected, processed or generated within the project is accompanied by a description called data summary. Data summary provides information regarding:

- The **purpose** of data and its relation to the project's objectives
- A description of the **type(s)** and **format(s)** of data
- The expected **size** of data
- A description of the **origin** of data (in case the data set is collected)
- A description of the **utility** of data

3.2 FAIR Data

3.2.1 Making data findable, including provisions for metadata

In order to make data findable, a policy has been developed for the (i) data set reference and name, and (ii) the metadata that will describe the data set.

3.2.1.1 Data set reference and name

In order to be able to distinguish and easily identify data sets, each data set is assigned with a unique name. This name can be also used as the identifier of the data sets.

In order to design the data set names, we use the following practice:

1. Each data set name consists of *four* different parts separated with a dot: *CSACode.CountryCode.PartnerName.DatasetName*, where
 - a. The *CSACode* part describes the domain that the data set fits. We use the Classification of Statistical Activities (CSA)¹ in order to represent different domains. Specifically, CSA uses five domains to classify the statistical activities undertaken by national and international statistical organizations: (a) Demographic and social statistics, Economic statistics, (b) Economic statistics, (c) Environment and multi-domain statistics, (d) Methodology for data collection, processing dissemination and analysis, and (e) Strategic managerial issues of official statistics. Each domain is then further divided into a number of subdomains. To represent the theme in the name of the data sets, we use the number that corresponds to each sub-domain of CSA. For example, a data set that comes from the education domain will have *1.3* as theme which corresponds to “Education” in CSA.
 - b. The *CountryCode* part represents the country associated with the data set using ISO Alpha-2 country codes:
 - i. BE for Belgium
 - ii. EE for Estonia
 - iii. GR for Greece
 - iv. IE for Ireland
 - v. LT for Lithuania
 - vi. NL for the Netherlands
 - vii. UK for the United Kingdom
 - c. The *PartnerName* part represents the name of the organisation (e.g. data owner, data custodian) associated with the data set:
 - i. MarineInstitute for the Marine Institue
 - ii. CERTH for the Center for Research and Technology, Hellas
 - iii. TUDelft for the Technische Universiteit Delft
 - iv. NUIG for the National University of Ireland, Galway
 - v. TUT for Tallinn University of Technology
 - vi. SWIRRL for Swirrl IT Limited
 - vii. TRAF for Trafford council
 - viii. VLO for the Flemish Government
 - ix. MAREG for the Ministry of Interior and Administrative Reconstruction
 - x. MKM for the Ministry of Economic Affairs and Communication

1

http://ec.europa.eu/eurostat/ramon/nomenclatures/index.cfm?TargetUrl=DSP_GEN_DESC_VIEW_N_OHDR&StrNom=CSA&StrLanguageCode=EN

- xi. EnterpriseLithuania for the Public Institution Enterprise Lithuania
- d. The *DatasetName* represents the full name of the data set.

An example of a data set's name could be the following:

2.4.4.GR.MAREG.GovernmentVehicles

The above name indicates that the data set describes Government Vehicles from the Ministry of Interior and Administrative Reconstruction in Greece and that the data set regards the 2.4.4 (Economic Statistics→Sectoral Statistics→Transport) domain.

In addition, only the data sets from the Marine Institute will have a second name. This is because (a) these data sets are very environmental specific and may not be fully described by CSA and (b) the Marine Institute has already assigned to them a name in a previous effort so we decided to keep those names too. The practice used to design these data sets' names is the following:

1. Each data set name consists of *five* different parts separated with a dot *INSPIRECode.INSPIREDataSpecification.CountryCode.PartnerName.DatasetName*, where
 - a. *INSPIRECode* consist of two letters for the appropriate INSPIRE theme the data set fits.
 - b. The *INSPIREDataSpecification* is the name of the data set to be portrayed for INSPIRE European data harmonisation and this comes from INSPIRE Data Specifications.
 - c. The *CountryCode* refers to the country associated with the data set using ISO Alpha-2 country codes.
 - d. The *PartnerName* is the full name of the organisation (e.g. data owner, data custodian) associated with the data set.
 - e. The *DatasetName* is the full name of the data set.

An example of such a data set's name is the following:

OF.PointTimeSeriesObservation.IE.MarineInstitute.IrishWeatherBuoyNetwork.Wave

The above name means that the 'Irish Weather Buoy Network' 'Wave' data set from the Marine Institute in Ireland is a PointTimeSeriesObservation in the inspire theme 'Oceanographic geographical features'.

3.2.1.2 Metadata

The data sets in the project are accompanied by proper metadata in order to be easier to find them. Metadata standards were used whenever possible. Metadata of a data set may include, for example, the publisher of the data set, the date of its last update and others. In case the data sets are published as linked data, standard vocabularies can be used (e.g. DCAT) for modelling metadata (see Section 3.2.3 for more details).

3.2.2 Making data openly accessible

This field describes the policies used in the project towards enhancing the open accessibility of the data sets. In particular, for each data set, information regarding whether the data set is made openly available as default is provided. For data sets that can't be shared, legal and contractual reasons are

provided. In addition, for open data sets, information regarding the medium that hosts data (e.g. repositories), the methods and software tools required to access data and other relevant information are also provided.

Partners agreed to upload, when possible, their open data sets in Zenodo (or other similar repository that does not require registering), in order to facilitate their easy access and long term preservation.

3.2.3 Making data interoperable

Interoperability of data sets facilitates data exchange and re-usability between researchers and organisations. To this end, standard vocabularies and technologies were used to describe the data as well as the metadata of the data sets. In OpenGovIntelligence, we identify two types of data sets: (a) linked data sets and (b) other data sets.

In the first category, data sets were published following Tim Berners-Lee's five star deployment scheme² that by default promotes interoperability. To this end, two types of standard vocabularies can mainly be used: (a) the Resource Description Standard (RDF)³ and (b) the RDF data cube vocabulary⁴. Additional vocabularies include the Registered Organization Vocabulary⁵ (regorg) for describing organizations and the ISA Programme Location Core Vocabulary⁶ (Icon) for the description of names.

RDF is a W3C standard model since 1999 for interchanging data on the Web. RDF models data as triples (subject-predicate-object). At the same time, the RDF data cube vocabulary is also a W3C standard since 2015 that is used to model statistical data as linked data cubes using RDF. Specifically, the RDF data cube vocabulary allows the modelling of statistical data as cubes using dimensions, attributes and measures. A measure represents the phenomenon that is being observed in the data set. A dimension defines what an observation in the data set applies to (e.g. time, area). Attributes are used to specify units of measure, scaling factors and other.

Metadata are crucial for the long-term use of the data sets. As a result it is advised to accompany (whenever possible) linked data sets with a set of metadata. To this end, existing can be used vocabularies to describe metadata such as the Data Catalogue Vocabulary (DCAT)⁷ and the PROV⁸ ontology.

DCAT is a W3C vocabulary that facilitates the interoperability between data catalogs published on the Web. Properties of DCAT that can be used to describe open data sets in the project include:

- *dct:title* to give a title to the data set
- *dct:description* to provide a brief description for the data set
- *dct:issued* to provide the date of the publication of the data set
- *dct:language* for the language of the data set

² <https://5stardata.info/en/>

³ <https://www.w3.org/RDF/>

⁴ <https://www.w3.org/TR/vocab-data-cube/>

⁵ <https://www.w3.org/TR/vocab-regorg/>

⁶ <https://www.w3.org/ns/locn>

⁷ <https://www.w3.org/TR/vocab-dcat>

⁸ <https://www.w3.org/TR/prov-o/>

- *dct:publisher* for the entity that publishes the data set
- *dcat:keyword* to provide a keyword that describes the data set

At the same time, the PROV ontology is also a W3C standard that can be used to provide provenance descriptions to the data sets. In this context, the PROV ontology can be used to model the activities, entities and actors involved in the data set production process. Specifically, according to the PROV ontology:

- *prov:Entity* can be used to describe a physical, digital, conceptual, or other kind of thing. A *prov:Entity*
- *prov:Activity* represents an activity that occurs over a period of time and acts upon or with entities; it may include consuming, processing, transforming, modifying, relocating, using, or generating entities
- *prov:Agent* is something that bears some form of responsibility for an activity taking place, for the existence of an entity, or for another agent's activity

Finally, in the second category comply all the data sets published in other formats (e.g. excel, csv, pdf, txt etc.). These data sets will also be accompanied, in some cases, by metadata.

3.2.4 Increase data re-use

The re-usability of data sets is enhanced in the project mainly through (i) clarifying licenses that will allow the widest possible re-usability, (ii) ensuring the quality of the data, and (iii) ensuring long preservation of data sets.

3.2.4.1 Data licensing

Data sets are accompanied by a suitable license that grant permissions to a second party to access and re-use data on condition that certain terms are met. Standard licenses that can be used include (Ball, 2014):

Creative commons. This kind of licenses is good for very simple and factual data sets and also data sets that are to be used automatically. There are six main Creative Commons licenses and have different values for certain conditions such as the *Attribute* condition, the *Non-Commercial* condition, the *Share Alike* condition and the *No Derivatives* condition:

- Attribution (CC BY)⁹;
- Attribution Share Alike (CC BY-SA)¹⁰;
- Attribution No Derivatives (CC BY-ND)¹¹;
- Attribution Non-Commercial (CC BY-NC)¹²;
- Attribution Non-Commercial Share Alike (CC BY-NC-SA)¹³;

⁹ <http://creativecommons.org/licenses/by/4.0>

¹⁰ <http://creativecommons.org/licenses/by-sa/4.0>

¹¹ <http://creativecommons.org/licenses/by-nd/4.0>

¹² <http://creativecommons.org/licenses/by-nc/4.0>

- Attribution Non-Commercial No Derivatives (CC BY-NC-ND)¹⁴.

The version 1 licences are very strict as permit derivations with the exact same version 1 licence. The version 2 licences permits to use a later version or a different port of the same license in derivations. The No Derivatives condition in the version 3 licences and earlier means that the licensee is forbidden from altering, transforming or building upon the work. The version 4 condition is more flexible: it allows these things for private use, but prevents the licensee from sharing the derivations.

Open data commons. It includes three licenses having some of the character of the Creative Commons licenses, but designed mainly for databases:

- The Public Domain Dedication and License (PDDL)¹⁵ that places the data(base) in the public domain (waiving all rights).
- The Attribution License (ODC-By)¹⁶ that is intended to allow users to freely share, modify, and use a database subject only to a number of attribution requirements.
- The Open Database License (ODC-ODbL)¹⁷, which is a copyleft (or "share alike") license agreement intended to allow users to freely share, modify, and use a database while maintaining this same freedom for others.

Open/Non-Commercial Government License. It was developed by the National Archives to enable public sector's information providers to license the use and re-use of their Information under a common non-commercial licence.

Public domain. When a work is in the public domain, it is free for use by anyone for any purpose without restriction under copyright law. Public domain is the purest form of open/free, since no one owns or controls the material in any way¹⁸.

3.2.4.2 Quality assurance

To ensure data quality, a number of open data quality dimensions (Dekkers et al., 2013) can be used including:

- Accuracy: the extent to which data correctly represents the characteristics of the real-world object, situation or event.
- Consistency: the extent to which data does not contain contradictions that would make its use difficult or impossible.
- Availability: the extent to which data can be accessed; this also includes the long-term persistence of data.
- Completeness: the extent to which data includes the data items or data points that are necessary to support the application for which it is intended.

¹³ <http://creativecommons.org/licenses/by-nc-sa/4.0>

¹⁴ <http://creativecommons.org/licenses/by-nc-nd/4.0>

¹⁵ <https://opendatacommons.org/licenses/pddl/index.html>

¹⁶ <https://opendatacommons.org/licenses/by/index.html>

¹⁷ <https://opendatacommons.org/licenses/odbl/index.html>

¹⁸ https://wiki.creativecommons.org/wiki/Public_domain

- Conformance: the extent to which data follows a set of explicit rules or standards for capture, publication and description.
- Credibility: the extent to which data is based on trustworthy sources or delivered by trusted organisations.
- Processability: the extent to which data can be understood and handled by automated processes.
- Relevance: the extent to which data contains the necessary information to support the application.
- Timeliness: the extent to which it correctly reflects the current state of the entity or event and the extent to which the data (in its latest version) is made available without unnecessary delay .

Table 1 presents a number of recommendations for each data quality dimension that will be followed by the project partners to ensure data quality.

Table 1 Data quality dimensions and recommendations

Data quality dimension	Recommendation
Accuracy	<ul style="list-style-type: none"> • The accuracy of data will be balanced against the cost in the context of the application to ensure that data is good enough for the intended use. • The partners will commit and invest in procedures and tools to maintain accuracy.
Consistency	<ul style="list-style-type: none"> • All data will be processed before publication to detect conflicting statements and other errors (in particular if data is aggregated from different sources).
Availability	<ul style="list-style-type: none"> • (Linked Open Data) Best practices will be followed for the assignment and maintenance of URIs. • Partners will ensure that the responsibility for the maintenance of data is clearly assigned in the organisation.
Completeness	<ul style="list-style-type: none"> • The capture and publication process will be designed to include the necessary data points. • The update mechanisms will be monitored on a continuous basis.
Conformance	<ul style="list-style-type: none"> • The most used standards in the domain that is most relevant for the data will be applied. • (Linked Open Data) Local vocabularies will be defined if no standard is available, and these vocabularies will be published according to best practices (e.g. dereferenceable URIs).
Credibility	<ul style="list-style-type: none"> • Data will be based on sources that can be trusted or on explicit Service Level Agreements where possible and appropriate. • Appropriate attributions will be made so that re-users can determine

	whether or not they can trust the data.
Processability	<ul style="list-style-type: none"> • (Linked Open Data) Identify the source of terminology and codes used in the data in machine-readable manner. • Recommendations for syntax of data given in common standards and application profiles will be applied.
Relevance	<ul style="list-style-type: none"> • Coverage and granularity of data will be matched to its intended use within constraints of available time and money. • Potential future usages of the data will be also considered.
Timeliness	<ul style="list-style-type: none"> • The update frequency of data will be adapted to the nature of the data and its intended use. • The processes and tools will be in place to support the updating.

3.2.4.3 Preservation time

The partners should decide for how long data will be preserved and remain re-usable after the end of the project. Partners also agreed to upload, when possible, their open data sets in Zenodo (or other similar repository that does not require registering), in order to facilitate their easy access and long term preservation.

3.3 Allocation of resources

The partners describe all costs associated with making their data sets FAIR. Detailed descriptions are given about how they covered these costs.

3.4 Data security

Partners describe the provisions that are in place for data security such as data recovery and secure storage of sensitive data. Partners also describe where (e.g. in certified repositories) data are stored to ensure long term preservation and curation.

3.5 Ethical aspects

Ethical or legal issues that can have impact on data sharing.

3.6 Other issues

Other issues may include other national, sectorial, departmental procedures for data management.

3.7 Open access scientific publications

All open access scientific publications that describe the results of the project are uploaded on Zenodo in order to facilitate their easy access and long time preservation.

4 Data sets analysis

The final version of the DMP includes twenty three data sets. This section presents the results of the analysis of these data sets related to the policies described in the previous section.

4.1 Data summary

Data sets used in the OGI project have various purposes. Specifically:

- two of them refer to real time oceanographic data on weather and wave climate collected from the Irish moored Weather Buoy network of stations in the North Atlantic Ocean,
- seven to labour data of the Department of Work and Pensions partner (including, for example, the number of working age adults with no qualifications, and the number of employees by broad industry group),
- one to the government vehicles used by Greek Government Agencies,
- two to air and water pollution data of companies and organisations of the Flemish government,
- one is a company registry of the Environment Department (LNE) of the Flemish government,
- six to economic activities (e.g. average company earnings and investments) at the municipality level in Lithuania,
- three to demographic distributions at the municipality level in Lithuania, and
- one to the aggregated results of the interviews and online surveys of the project.

Regarding the types and formats of the data sets:

- twenty are stored as RDF and/or ttl,
- two as CSV, JSON, and
- one as XLSX, CSV

Regarding the size to the data sets:

- nine of them have size 5-82kB,
- eight of them have size 0.5 – 28.5 MB,
- four of them have size 50 -127MB, and
- two of them are measured in rdf thousands of triples

Regarding the origin of the data:

- one comes from the Irish Weather Buoy Network
- one comes from the Irish Wave Buoy Network
- seven come from the Office for National Statistics. Excerpt in the UK
- one from the Greek Government Agencies
- two from the Flemish government,
- one from the federal Belgian government
- one from the Lithuanian State Food and Veterinary Service
- one from the Lithuanian Open Corporate Data (Sodra)
- seven from the Lithuanian Official Statistics Portal

- one from the of Tasks 1.2 and 1.4 of WP1 of the OGI project.

Finally, data sets are useful to the following entities:

- Renewable energy community
- the public (e.g. for tourism and leisure activities)
- businesses
- Governments and other public bodies
- Policy makers
- Researchers
- Meteorological agencies (e.g. to inform weather forecast models)

4.2 FAIR Data

4.2.1 Making data findable, including provisions for metadata

4.2.1.1 Data set reference and name

Table 2 presents the twenty three data sets recorded by the partners of the consortium along with their CSA domain. Data sets come from eight different domains. Specifically, six data sets come from the *Labour* domain, six from the *Business statistics* domain, four data sets come from the *Environment* domain, three from the *Population and migration* domain, one from the *Social protection* domain, one from the *Transport* domain, one from the *Banking, insurance, financial statistics* domain, and one from the *Data sources* domain.

Table 2. Data sets and CSA domains

Data set	CSA Domain
3.1.IE.MarineInstitute.IrishWeatherWaveCimate	Environment
3.1.IE.MarineInstitute.IrishWaveClimate	Environment
1.2.UK.TRAF.ClaimantCount	Labour
1.2.UK.TRAF.ClaimantRate	Labour
1.2.UK.TRAF.HouseholdsLoneParentNotInEmployment	Labour
1.6.UK.TRAF.SocialRentedHouseholds	Social protection
1.2.UK.TRAF.WorkingAgeAdultsNoQualifications	Labour
1.2.UK.TRAF.EmployeesBroadIndustryGroup	Labour
1.2.UK.TRAF.WorkingAgePopulation	Labour
2.4.4.GR.MAREG.GovernmentVehicles	Transport
3.1.BE.VLO.AirWaterPollutionMetadata	Environment
3.1.BE.VLO.AirWaterPollutionData	Environment
2.3.BE.VLO.CompanyRegistry	Business Statistics

2.4.6.LT.EnterpriseLithuania.StateSocialInsurance	Banking, insurance, financial statistics
2.3.LT.EnterpriseLithuania.EconomicActivities	Business statistics
2.3.LT.EnterpriseLithuania.AverageEarnings2015	Business statistics
2.3.LT.EnterpriseLithuania.Investments2015	Business statistics
2.3.LT.EnterpriseLithuania.ForeignInvestments2014	Business statistics
2.3.LT.EnterpriseLithuania.EconomicEntities2015	Business statistics
1.1.LT.EnterpriseLithuania.ResidentPopulationAgeGroups	Population and migration
1.1.LT.EnterpriseLithuania.ResidentPopulationGender	Population and migration
1.1.LT.EnterpriseLithuania.ResidentPopulation	Population and migration
4.3.TUT. InterviewsSurveyDataOGI	Data sources

4.2.1.2 Metadata

Table 3 presents a brief description of the metadata used for each data set. As the table presents, six data sets are using standard vocabularies for the metadata. Specifically, two of the data sets are using ISO 19139, and others are using standard vocabularies like ORG and Dublin Core. In general, common metadata used in the data sets include the title of the data set, its publisher, its license, next and/or last update.

Table 3 Data sets and Metadata

Data set	Metadata
3.1.IE.MarineInstitute.IrishWeatherWaveClimate	Coming from ISO 19139
3.1.IE.MarineInstitute.IrishWaveClimate	Coming from ISO 19139
1.2.UK.TRAF.ClaimantCount	Publisher, licence, next update
1.2.UK.TRAF.ClaimantRate	Publisher, licence, next update
1.2.UK.TRAF.HouseholdsLoneParentNotInEmployment	Publisher, licence, next update
1.6.UK.TRAF.SocialRentedHouseholds	Publisher, licence, next update
1.2.UK.TRAF.WorkingAgeAdultsNoQualifications	Publisher, licence, next update
1.2.UK.TRAF.EmployeesBroadIndustryGroup	Publisher, licence, next update
1.2.UK.TRAF.WorkingAgePopulation	Publisher, licence, next update
2.4.4.GR.MAREG.GovernmentVehicles	Title, Publisher, license, last update
3.1.BE.VLO.AirWaterPollutionMetadata	Coming from Dublin Core
3.1.BE.VLO.AirWaterPollutionData	Coming from RDF Data Cube, SKOS
2.3.BE.VLO.CompanyRegistry	Coming from ORG vocabulary

2.4.6.LT.EnterpriseLithuania.StateSocialInsurance	Last update, Title
2.3.LT.EnterpriseLithuania.EconomicActivities	Last update, Title
2.3.LT.EnterpriseLithuania.AverageEarnings2015	Last update, Title
2.3.LT.EnterpriseLithuania.Investments2015	Last update, Title
2.3.LT.EnterpriseLithuania.ForeignInvestments2014	Last update, Title
2.3.LT.EnterpriseLithuania.EconomicEntities2015	Last update, Title
1.1.LT.EnterpriseLithuania.ResidentPopulationAgeGroups	Last update, Title
1.1.LT.EnterpriseLithuania.ResidentPopulationGender	Last update, Title
1.1.LT.EnterpriseLithuania.ResidentPopulation	Last update, Title
4.3.TUT. InterviewsSurveyDataOGI	N/A

4.2.2 Making data openly accessible

Table 4 describes the access policy (i.e. open or restricted) and sharing medium for each data set. Specifically, 19 data sets are open, one data set is open but only in an aggregated version, and three data sets are restricted. From the open data sets, the majority can be accessed using a HTML page, SPARQL endpoint, and/or a restful API. Finally seventeen open data sets are also provided through Open Science Framework or Zenodo repositories.

Table 4. Data sets, access policy and sharing medium

Data set	Access Policy	Sharing medium
3.1.IE.MarineInstitute.IrishWeatherWaveClimate	Open	1. HTML page, 2. restful API
3.1.IE.MarineInstitute.IrishWaveClimate	Open	1. HTML page, 2. restful API
1.2.UK.TRAF.ClaimantCount	Open	1. HTML page, 2. SPARQL Endpoint, 3. Open Science Framework repository
1.2.UK.TRAF.ClaimantRate	Open	1. HTML page, 2. SPARQL Endpoint, 3. Open Science Framework repository
1.2.UK.TRAF.HouseholdsLoneParentNotInEmployment	Open	1. HTML page, 2. SPARQL Endpoint, 3. Open Science Framework

		repository
1.6.UK.TRAF.SocialRentedHouseholds	Open	1. HTML page, 2. SPARQL Endpoint, 3. Open Science Framework repository
1.2.UK.TRAF.WorkingAgeAdultsNoQualifications	Open	1. HTML page, 2. SPARQL Endpoint, 3. Open Science Framework repository
1.2.UK.TRAF.EmployeesBroadIndustryGroup	Open	1. HTML page, 2. SPARQL Endpoint, 3. Open Science Framework repository
1.2.UK.TRAF.WorkingAgePopulation	Open	1. HTML page, 2. SPARQL Endpoint, 3. Open Science Framework repository
2.4.4.GR.MAREG.GovernmentVehicles	Open (aggregated version)	1. Zenodo
3.1.BE.VLO.AirWaterPollutionMetadata	Restricted	-
3.1.BE.VLO.AirWaterPollutionData	Restricted	-
2.3.BE.VLO.CompanyRegistry	Open	1. SPARQL Endpoint
2.4.6.LT.EnterpriseLithuania.StateSocialInsurance	Open	1. A central storage RDF endpoint. 2. A GraphQL based API implementation. 3. Zenodo
2.3.LT.EnterpriseLithuania.EconomicActivities	Open	1. A central storage RDF endpoint. 2. A GraphQL based API implementation. 3. Zenodo
2.3.LT.EnterpriseLithuania.AverageEarnings2015	Open	1. A central storage RDF endpoint. 2. A GraphQL based API implementation. 3. Zenodo

2.3.LT.EnterpriseLithuania.Investments2015	Open	<ol style="list-style-type: none"> 1. A central storage RDF endpoint. 2. A GraphQL based API implementation. 3. Zenodo
2.3.LT.EnterpriseLithuania.ForeignInvestments2014	Open	<ol style="list-style-type: none"> 1. A central storage RDF endpoint. 2. A GraphQL based API implementation. 3. Zenodo
2.3.LT.EnterpriseLithuania.EconomicEntities2015	Open	<ol style="list-style-type: none"> 1. A central storage RDF endpoint. 2. A GraphQL based API implementation.
1.1.LT.EnterpriseLithuania.ResidentPopulationAgeGroups	Open	<ol style="list-style-type: none"> 1. A central storage RDF endpoint. 2. A GraphQL based API implementation. 3. Zenodo
1.1.LT.EnterpriseLithuania.ResidentPopulationGender	Open	<ol style="list-style-type: none"> 1. A central storage RDF endpoint. 2. A GraphQL based API implementation. 3. Zenodo
1.1.LT.EnterpriseLithuania.ResidentPopulation	Open	<ol style="list-style-type: none"> 1. A central storage RDF endpoint. 2. A GraphQL based API implementation. 3. Zenodo
4.3.TUT. InterviewsSurveyDataOGI	Restricted	-

4.2.3 Making data interoperable

Table 5 presents information regarding the standards that are used to model data sets (as linked data or not) and also the standards that are used to model their metadata (if any). Most data sets (nineteen out of twenty three) were published as linked data using the RDF data cube vocabulary. The rest of the data sets use standards such as CSV, XLS, JSON. In addition, from the data sets that were published as linked data, sixteen were accompanied by W3C Tabular Data and Metadata, two

by metadata following the ISO19139 and the BODC unit vocabulary, and the rest of them by Dublin Core, ORG and/or SKOS metadata.

Table 5. Data sets, standards and metadata

Data set	Data standard	Linked Data standard	Metadata standard
3.1.IE.MarineInstitute.IrishWeatherWaveClimate	Climate and Forecast standard names, CSV, JSON	-	ISO19139, BODC unit vocabulary
3.1.IE.MarineInstitute.IrishWaveClimate	Climate and Forecast standard names, CSV, JSON	-	ISO19139, BODC unit vocabulary
1.2.UK.TRAF.ClaimantCount	-	RDF data cube vocabulary	W3C Tabular Data and Metadata
1.2.UK.TRAF.ClaimantRate	-	RDF data cube vocabulary	W3C Tabular Data and Metadata
1.2.UK.TRAF.HouseholdsLoneParentNotInEmployment	-	RDF data cube vocabulary	W3C Tabular Data and Metadata
1.6.UK.TRAF.SocialRentedHouseholds	-	RDF data cube vocabulary	W3C Tabular Data and Metadata
1.2.UK.TRAF.WorkingAgeAdultsNoQualifications	-	RDF data cube vocabulary	W3C Tabular Data and Metadata
1.2.UK.TRAF.EmployeesBroadIndustryGroup	-	RDF data cube vocabulary	W3C Tabular Data and Metadata
1.2.UK.TRAF.WorkingAgePopulation	-	RDF data cube vocabulary	W3C Tabular Data and Metadata
2.4.4.GR.MAREG.GovernmentVehicles	-	RDF data cube	• RDF Data Cube for

		vocabulary	data <ul style="list-style-type: none"> • Dublin Core, SKOS, ORG for metadata
3.1.BE.VLO.AirWaterPollutionMetadata	-	-	Dublin Core
3.1.BE.VLO.AirWaterPollutionData	-	RDF data cube vocabulary	SKOS
2.3.BE.VLO.CompanyRegistry	-	-	ORG vocabulary
2.4.6.LT.EnterpriseLithuania.StateSocialInsurance	-	RDF data cube vocabulary	W3C Tabular Data and Metadata
2.3.LT.EnterpriseLithuania.EconomicActivities	-	RDF data cube vocabulary	W3C Tabular Data and Metadata
2.3.LT.EnterpriseLithuania.AverageEarnings2015	-	RDF data cube vocabulary	W3C Tabular Data and Metadata
2.3.LT.EnterpriseLithuania.Investments2015	-	RDF data cube vocabulary	W3C Tabular Data and Metadata
2.3.LT.EnterpriseLithuania.ForeignInvestments2014	-	RDF data cube vocabulary	W3C Tabular Data and Metadata
2.3.LT.EnterpriseLithuania.EconomicEntities2015	-	RDF data cube vocabulary	W3C Tabular Data and Metadata
1.1.LT.EnterpriseLithuania.ResidentPopulationAgeGroups	-	RDF data cube vocabulary	W3C Tabular Data and Metadata
1.1.LT.EnterpriseLithuania.ResidentPopulationGender	-	RDF data cube vocabulary	W3C Tabular Data and Metadata
1.1.LT.EnterpriseLithuania.ResidentPopulation	-	RDF data cube	W3C Tabular Data and

		vocabulary	Metadata
4.3.TUT. InterviewsSurveyDataOGI	XLSX, CSV	-	-

4.2.4 Increase data re-use

4.2.4.1 Data licensing

Table 6 presents the licenses used by the data sets of the project. Specifically, three data sets are provided under Creative Common CC-BY 4.0, seven data sets under Open Government Licence 3.0, nine data sets are provided under publicly available data licence, one under Flemish Open Data License 1, and one does not have a license. Finally the license of two data sets of the VLO partner has not been defined yet.

Table 6 Data sets and licenses

Data set	License
3.1.IE.MarineInstitute.IrishWeatherWaveCimate	Creative Common CC-BY 4.0
3.1.IE.MarineInstitute.IrishWaveClimate	Creative Common CC-BY 4.0
1.2.UK.TRAF.ClaimantCount	Open Government Licence 3.0
1.2.UK.TRAF.ClaimantRate	Open Government Licence 3.0
1.2.UK.TRAF.HouseholdsLoneParentNotInEmployment	Open Government Licence 3.0
1.6.UK.TRAF.SocialRentedHouseholds	Open Government Licence 3.0
1.2.UK.TRAF.WorkingAgeAdultsNoQualifications	Open Government Licence 3.0
1.2.UK.TRAF.EmployeesBroadIndustryGroup	Open Government Licence 3.0
1.2.UK.TRAF.WorkingAgePopulation	Open Government Licence 3.0
2.4.4.GR.MAREG.GovernmentVehicles	Creative Common CC-BY 4.0
3.1.BE.VLO.AirWaterPollutionMetadata	N/A
3.1.BE.VLO.AirWaterPollutionData	N/A
2.3.BE.VLO.CompanyRegistry	Flemish Open Data License 1
2.4.6.LT.EnterpriseLithuania.StateSocialInsurance	Publicly available data
2.3.LT.EnterpriseLithuania.EconomicActivities	Publicly available data
2.3.LT.EnterpriseLithuania.AverageEarnings2015	Publicly available data
2.3.LT.EnterpriseLithuania.Investments2015	Publicly available data
2.3.LT.EnterpriseLithuania.ForeignInvestments2014	Publicly available data
2.3.LT.EnterpriseLithuania.EconomicEntities2015	Publicly available data
1.1.LT.EnterpriseLithuania.ResidentPopulationAgeGro	Publicly available data

ups	
1.1.LT.EnterpriseLithuania.ResidentPopulationGender	Publicly available data
1.1.LT.EnterpriseLithuania.ResidentPopulation	Publicly available data
4.3.TUT. InterviewsSurveyDataOGI	-

4.2.4.2 Quality assurance

Table 7 describes how the quality of data sets was assured.

Table 7 Data sets and quality assurance descriptions

Data set	Quality Assurance
3.1.IE.MarineInstitute.IrishWeatherWaveClimate	<ul style="list-style-type: none"> • Data is monitored and assessed at key points by an Operational System Monitoring Component. • Email and SMS alerts are supported, to notify of any issues encountered • Checks include: Wave buoy location validation, parameter range limit checks (e.g. wave height & temperature), and expert visual assessment.
3.1.IE.MarineInstitute.IrishWaveClimate	<ul style="list-style-type: none"> • Data is monitored and assessed at key points by an Operational System Monitoring Component. • Email and SMS alerts are supported, to notify of any issues encountered • Checks include: Wave buoy location validation, parameter range limit checks (e.g. wave height & temperature), and expert visual assessment.
1.2.UK.TRAF.ClaimantCount	Data is an excerpt from Office for National Statistics sources which is subject to their quality assurance processes
1.2.UK.TRAF.ClaimantRate	Data is an excerpt from Office for National Statistics sources which is subject to their quality assurance processes

1.2.UK.TRAF.HouseholdsLoneParentNotInEmployment	Data is an excerpt from Office for National Statistics sources which is subject to their quality assurance processes
1.6.UK.TRAF.SocialRentedHouseholds	Data is an excerpt from Office for National Statistics sources which is subject to their quality assurance processes
1.2.UK.TRAF.WorkingAgeAdultsNoQualifications	Data is an excerpt from Office for National Statistics sources which is subject to their quality assurance processes
1.2.UK.TRAF.EmployeesBroadIndustryGroup	Data is an excerpt from Office for National Statistics sources which is subject to their quality assurance processes
1.2.UK.TRAF.WorkingAgePopulation	Data is an excerpt from Office for National Statistics sources which is subject to their quality assurance processes
2.4.4.GR.MAREG.GovernmentVehicles	The quality assurance process is based on a data management tool created within OGI project, which Government Agencies will use to confirm or update their fleet data.
3.1.BE.VLO.AirWaterPollutionMetadata	During conversion to RDF, values are normalized, and a lot of string values are reconciled into unique identifiers coming from controlled and centrally managed terminology vocabularies
3.1.BE.VLO.AirWaterPollutionData	During conversion to RDF, values are normalized. A lookup process is in place for replacing strings into unique identifiers.
2.3.BE.VLO.CompanyRegistry	The Federal data are validated and enhanced on the Flemish level and then once again on the Departmental level.
2.4.6.LT.EnterpriseLithuania.StateSocialInsurance	<ul style="list-style-type: none"> Data is an excerpt from the State Food and Veterinary Service sources which is subject to their quality

	<p>assurance processes.</p> <ul style="list-style-type: none"> • During the data cleansing process, well documented processes were run in order to preserve data quality. • During the data enrichment process new data was generated using BING reverse geocoding API to represent geographical distributed information. The quality of this generated data is subject to this service.
2.3.LT.EnterpriseLithuania.EconomicActivities	<ul style="list-style-type: none"> • Data is an excerpt from the Open Corporate Data sources which is subject to their quality assurance processes. • During the data cleansing process, well documented processes were run to preserve data quality.
2.3.LT.EnterpriseLithuania.AverageEarnings2015	<ul style="list-style-type: none"> • Data is an excerpt from the Lithuanian Official Statistics Portal sources which is subject to their quality assurance processes. • During the data cleansing process, well documented processes were run to preserve data quality.
2.3.LT.EnterpriseLithuania.Investments2015	<ul style="list-style-type: none"> • Data is an excerpt from the Lithuanian Official Statistics Portal sources which is subject to their quality assurance processes. • During the data cleansing process, well documented processes were run to preserve data quality.
2.3.LT.EnterpriseLithuania.ForeignInvestments2014	<ul style="list-style-type: none"> • Data is an excerpt from the Lithuanian Official Statistics Portal sources which is subject to their quality assurance processes. • During the data cleansing process, well documented processes were run to preserve data quality.

2.3.LT.EnterpriseLithuania.EconomicEntities2015	<ul style="list-style-type: none"> • Data is an excerpt from the Lithuanian Official Statistics Portal sources which is subject to their quality assurance processes. • During the data cleansing process, well documented processes were run to preserve data quality.
1.1.LT.EnterpriseLithuania.ResidentPopulationAgeGroups	<ul style="list-style-type: none"> • Data is an excerpt from the Lithuanian Official Statistics Portal sources which is subject to their quality assurance processes. • During the data cleansing process, well documented processes were run to preserve data quality.
1.1.LT.EnterpriseLithuania.ResidentPopulationGender	<ul style="list-style-type: none"> • Data is an excerpt from the Lithuanian Official Statistics Portal sources which is subject to their quality assurance processes. • During the data cleansing process, well documented processes were run to preserve data quality.
1.1.LT.EnterpriseLithuania.ResidentPopulation	<ul style="list-style-type: none"> • Data is an excerpt from the Lithuanian Official Statistics Portal sources which is subject to their quality assurance processes. • During the data cleansing process, well documented processes were run to preserve data quality.
4.3.TUT. InterviewsSurveyDataOGI	N/A

4.2.4.3 Preservation time

Long-term preservation of data sets of the project are ensured by archiving them for a number of years after the end of the project. Table 8 presents the data sets of the project along with their preservation duration and the preservation medium. Specifically, most of the data sets (20 out of 23) will be preserved for an unlimited period of time after the end of the project. In addition, two of the data sets will be preserved for at least five years after the end of the project. In addition, the majority of the data sets (17 out of 23) are already preserved in Open Science Framework or Zenodo in order to ensure long time preservation.

Table 8 Data sets and preservation duration

Data set	Preservation duration	Preservation Medium
3.1.IE.MarineInstitute.IrishWeatherWaveClimate	Unlimited	-
3.1.IE.MarineInstitute.IrishWaveClimate	Unlimited	-
1.2.UK.TRAF.ClaimantCount	Unlimited	Open Science Framework
1.2.UK.TRAF.ClaimantRate	Unlimited	Open Science Framework
1.2.UK.TRAF.HouseholdsLoneParentNotInEmployment	Unlimited	Open Science Framework
1.6.UK.TRAF.SocialRentedHouseholds	Unlimited	Open Science Framework
1.2.UK.TRAF.WorkingAgeAdultsNoQualifications	Unlimited	Open Science Framework
1.2.UK.TRAF.EmployeesBroadIndustryGroup	Unlimited	Open Science Framework
1.2.UK.TRAF.WorkingAgePopulation	Unlimited	Open Science Framework
2.4.4.GR.MAREG.GovernmentVehicles	At least 5 years	Zenodo
3.1.BE.VLO.AirWaterPollutionMetadata	Unlimited	-
3.1.BE.VLO.AirWaterPollutionData	Unlimited	-
2.3.BE.VLO.CompanyRegistry	N/A	-
2.4.6.LT.EnterpriseLithuania.StateSocialInsurance	Unlimited	Zenodo
2.3.LT.EnterpriseLithuania.EconomicActivities	Unlimited	Zenodo
2.3.LT.EnterpriseLithuania.AverageEarnings2015	Unlimited	Zenodo
2.3.LT.EnterpriseLithuania.Investments2015	Unlimited	Zenodo
2.3.LT.EnterpriseLithuania.ForeignInvestments2014	Unlimited	Zenodo
2.3.LT.EnterpriseLithuania.EconomicEntities2015	Unlimited	Zenodo
1.1.LT.EnterpriseLithuania.ResidentPopulationAgeGroups	Unlimited	Zenodo

1.1.LT.EnterpriseLithuania.ResidentPopulationGender	Unlimited	Zenodo
1.1.LT.EnterpriseLithuania.ResidentPopulation	Unlimited	Zenodo
4.3.TUT. InterviewsSurveyDataOGI	At least five years	-

4.3 Allocation of resources

According to **Σφάλμα! Το αρχείο προέλευσης της αναφοράς δεν βρέθηκε.**, none of the project's data sets required allocating resources of the project in order to make data FAIR.

Table 9 Data sets and Cost estimation

Data set	Costs
3.1.IE.MarineInstitute.IrishWeatherWaveCimate	No additional costs
3.1.IE.MarineInstitute.IrishWaveClimate	No additional costs
1.2.UK.TRAF.ClaimantCount	No additional costs
1.2.UK.TRAF.ClaimantRate	No additional costs
1.2.UK.TRAF.HouseholdsLoneParentNotInEmployment	No additional costs
1.6.UK.TRAF.SocialRentedHouseholds	No additional costs
1.2.UK.TRAF.WorkingAgeAdultsNoQualifications	No additional costs
1.2.UK.TRAF.EmployeesBroadIndustryGroup	No additional costs
1.2.UK.TRAF.WorkingAgePopulation	No additional costs
2.4.4.GR.MAREG.GovernmentVehicles	No additional costs
3.1.BE.VLO.AirWaterPollutionMetadata	No additional costs
3.1.BE.VLO.AirWaterPollutionData	No additional costs
2.3.BE.VLO.CompanyRegistry	No additional costs
2.4.6.LT.EnterpriseLithuania.StateSocialInsurance	No additional costs
2.3.LT.EnterpriseLithuania.EconomicActivities	No additional costs
2.3.LT.EnterpriseLithuania.AverageEarnings2015	No additional costs
2.3.LT.EnterpriseLithuania.Investments2015	No additional costs
2.3.LT.EnterpriseLithuania.ForeignInvestments2014	No additional costs
2.3.LT.EnterpriseLithuania.EconomicEntities2015	No additional costs
1.1.LT.EnterpriseLithuania.ResidentPopulationAgeGroups	No additional costs

1.1.LT.EnterpriseLithuania.ResidentPopulationGender	No additional costs
1.1.LT.EnterpriseLithuania.ResidentPopulation	No additional costs
4.3.TUT. InterviewsSurveyDataOGI	No additional costs

4.4 Data security

Table 10 presents how partners ensured data security for their data sets.

Table 10 Data sets and data security

Data set	Data security
3.1.IE.MarineInstitute.IrishWeatherWaveCimate	<ul style="list-style-type: none"> • Data are managed by the Irish National Marine Data Centre. • The Marine Institute ERDDAP server is hosted on a backend SQL Server database with standard SQL Server backup devices used to protect data in case of technical issues.
3.1.IE.MarineInstitute.IrishWaveClimate	<ul style="list-style-type: none"> • Data are managed by the Irish National Marine Data Centre. • The Marine Institute ERDDAP server is hosted on a backend SQL Server database with standard SQL Server backup devices used to protect data in case of technical issues.
1.2.UK.TRAF.ClaimantCount	The data are backed up in a dedicated project repository on GitHub as well as being stored as linked open statistical data on SWIRRL's instance of the Publish My Data platform, GM DataStore.
1.2.UK.TRAF.ClaimantRate	The data are backed up in a dedicated project repository on GitHub as well as being stored as linked open statistical data on SWIRRL's instance of the Publish My Data platform, GM DataStore.
1.2.UK.TRAF.HouseholdsLoneParentNotInEmployment	The data are backed up in a

	dedicated project repository on GitHub as well as being stored as linked open statistical data on SWIRRL's instance of the Publish My Data platform, GM DataStore.
1.6.UK.TRAF.SocialRentedHouseholds	The data are backed up in a dedicated project repository on GitHub as well as being stored as linked open statistical data on SWIRRL's instance of the Publish My Data platform, GM DataStore.
1.2.UK.TRAF.WorkingAgeAdultsNoQualifications	The data are backed up in a dedicated project repository on GitHub as well as being stored as linked open statistical data on SWIRRL's instance of the Publish My Data platform, GM DataStore.
1.2.UK.TRAF.EmployeesBroadIndustryGroup	The data are backed up in a dedicated project repository on GitHub as well as being stored as linked open statistical data on SWIRRL's instance of the Publish My Data platform, GM DataStore.
1.2.UK.TRAF.WorkingAgePopulation	The data are backed up in a dedicated project repository on GitHub as well as being stored as linked open statistical data on SWIRRL's instance of the Publish My Data platform, GM DataStore.
2.4.4.GR.MAREG.GovernmentVehicles	The data are stored in Zenodo which has all the necessary procedures for long lasting preservation and curation in place
3.1.BE.VLO.AirWaterPollutionMetadata	The data are kept in an archival system, which has all the necessary procedures for access control, long-lasting preservation (backup and restore) and rapid redeployment in place
3.1.BE.VLO.AirWaterPollutionData	The data are kept in an archival system, which has all the necessary

	procedures for access control, long-lasting preservation (backup and restore) and rapid redeployment in place
2.3.BE.VLO.CompanyRegistry	This is part of the central ICT infrastructure of the department where all measures for access control, high availability, backup and restore, quick deployment are in place. The triple store has 2 entry points: one internal with all records available, and an external one where privacy related data are not exposed since filtered out in the source database (PostgresQL) view
2.4.6.LT.EnterpriseLithuania.StateSocialInsurance	RDF data has been backed up in an internal GitLab repository and stored as open statistical data in the mentioned engine. The persistence of the data is ensured by NUIG backup mechanisms.
2.3.LT.EnterpriseLithuania.EconomicActivities	RDF data has been backed up in an internal GitLab repository and stored as open statistical data in the mentioned engine. The persistence of the data is ensured by NUIG backup mechanisms.
2.3.LT.EnterpriseLithuania.AverageEarnings2015	RDF data has been backed up in an internal GitLab repository and stored as open statistical data in the mentioned engine. The persistence of the data is ensured by NUIG backup mechanisms.
2.3.LT.EnterpriseLithuania.Investments2015	RDF data has been backed up in an internal GitLab repository and stored as open statistical data in the mentioned engine. The persistence of the data is ensured by NUIG backup mechanisms.
2.3.LT.EnterpriseLithuania.ForeignInvestments2014	RDF data has been backed up in an

	internal GitLab repository and stored as open statistical data in the mentioned engine. The persistence of the data is ensured by NUIG backup mechanisms.
2.3.LT.EnterpriseLithuania.EconomicEntities2015	RDF data has been backed up in an internal GitLab repository and stored as open statistical data in the mentioned engine. The persistence of the data is ensured by NUIG backup mechanisms.
1.1.LT.EnterpriseLithuania.ResidentPopulationAgeGroups	RDF data has been backed up in an internal GitLab repository and stored as open statistical data in the mentioned engine. The persistence of the data is ensured by NUIG backup mechanisms.
1.1.LT.EnterpriseLithuania.ResidentPopulationGender	RDF data has been backed up in an internal GitLab repository and stored as open statistical data in the mentioned engine. The persistence of the data is ensured by NUIG backup mechanisms.
1.1.LT.EnterpriseLithuania.ResidentPopulation	RDF data has been backed up in an internal GitLab repository and stored as open statistical data in the mentioned engine. The persistence of the data is ensured by NUIG backup mechanisms.
4.3.TUT. InterviewsSurveyDataOGI	<p>The original data will be stored by the Tallinn University of Technology in a secure server in Estonia, which only the research team has access to. The key allowing re-identification will be stored in an encrypted form only until necessary for the analysis, and will be deleted after the project's end.</p> <p>The anonymized data will be stored on a secure server of the Tallinn University of Technology up to 5</p>

4.5 Ethical aspects

For the data sets that include personal data, the consortium ensured that they comply with EU General Data Protection Regulation (GDPR). GDPR is EU's legal framework on privacy and personal data protection that replaces the Directive 95/46/EC (the 'Data Protection Directive'). More details can be found in deliverable D6.5 - Ethics policy. In addition, all partners who were required to apply for ethical approval from an REC, have been granted approval for current research activities. More details can be found again in deliverable D6.5- Ethics policy.

Finally, the participants of interviews and online surveys (e.g. to collect the data of dataset 4.3.TUT. InterviewsSurveyDataOGI) were provided with information regarding (1) the procedures that will be followed, (2) the purpose of the research (3) the benefit for participants (4) any risks to the participant, (5) the opportunity to ask questions, (6) the opportunity to avoid to complete some tasks, (7) the opportunity to withdraw at any time, (8) mechanisms for protection of and limitations to privacy and confidentiality, (9) voluntariness of participation, (10) possibility of future uses (secondary uses) of data collected, (11) storage and destruction of data collected before starting with the study. To this end, a sample consent form was created and can be found at Appendix E of the deliverable D6.5- Ethics policy.

4.6 Other issues

The partners of the project have also studied and followed related local rules, national legislations, and European directives. These are presented in deliverable D6.5- Ethics policy.

4.7 Open access publications

Table 11 Open Access publications of the project

Title	Zenodo URL	Partner	License
ICT Tools for Creating, Expanding, and Exploiting Statistical Linked Open Data	https://zenodo.org/record/2542713	CERTH	Creative Commons Attribution 4.0 International
Linked Open Cube Analytics Systems: Potential and Challenges	https://zenodo.org/record/2542719	CERTH	Creative Commons Attribution 4.0 International
Exploiting Linked Statistical Data in Public Administration: The Case of the Greek Ministry of Administrative Reconstruction	https://zenodo.org/record/2542756	CERTH	Creative Commons Attribution 4.0 International
Visualizing Linked Open Statistical Data to Support Public Administration	https://zenodo.org/record/2551073	CERTH	Creative Commons Attribution 4.0 International

Towards a Linked Open Statistical Data Innovation Ecosystem	https://zenodo.org/record/2551104	CERTH	Creative Commons Attribution 4.0 International
Open Statistics: The Rise of a New Era for Open Data?	https://zenodo.org/record/2542728	CERTH	Creative Commons Attribution 4.0 International
Theory and Practice of Linked Open Statistical Data	https://zenodo.org/record/2543480	CERTH	Creative Commons Attribution 4.0 International
Methods and Tools for Publishing and Reusing Linked Open Statistical Data	https://zenodo.org/record/2543478	CERTH	Creative Commons Attribution 4.0 International
Facilitating the exploitation of Linked Open Statistical Data: JSON-QB API requirements and design criteria	https://zenodo.org/record/2543624	CERTH	Creative Commons Attribution 4.0 International
Linked data cubes: Research results so far	https://zenodo.org/record/2543632	CERTH	Creative Commons Attribution 4.0 International
Open Statistical Data: Potential and Challenges	https://zenodo.org/record/2551006	CERTH	Creative Commons Attribution 4.0 International
Open Data as Enabler of Public Service Co-creation: Exploring the Drivers and Barriers	https://zenodo.org/record/2537031	TUT	Creative Commons Attribution 4.0 International
How does open government data driven co-creation occur? Six factors and a 'perfect storm'; insights from Chicago's food inspection forecasting model	https://zenodo.org/record/2537035	TUT	Creative Commons Attribution 4.0 International
Co-creating an Open Government Data Driven Public Service: The Case of Chicago's Food Inspection Forecasting Model	https://zenodo.org/record/2537025	TUT	Creative Commons Attribution 4.0 International
The Role of Linked Open Statistical Data in Public Service Co-Creation	https://zenodo.org/record/2537044	TUT	Creative Commons Attribution 4.0 International
Leader in e-Government, Laggard in Open Data: Exploring the Case of Estonia	https://zenodo.org/record/2537046	TUT	Creative Commons Attribution 4.0 International

Open Government Data Driven Co-Creation: Moving Towards Citizen-Government Collaboration	https://zenodo.org/record/2537057	TUT	Creative Commons Attribution 4.0 International
Is Government Welfare Able to Change? Analysing Efforts to Co-create an Improved Social Welfare System through Taking Advantage of a Collaborative Economy	https://zenodo.org/record/2537062	TUT	Creative Commons Attribution 4.0 International
Government as a Platform: Exploiting Open Government Data to Drive Public Service Co-Creation	https://zenodo.org/record/2537053	TUT	Creative Commons Attribution 4.0 International
A Framework for Data-Driven Public Service Co-Production	https://zenodo.org/record/2537039	TUT	Creative Commons Attribution 4.0 International
Co-creating an Open Government Data Driven Public Service: The Case of Chicago's Food Inspection Forecasting Model	https://zenodo.org/record/2537029	TUT	Creative Commons Attribution 4.0 International
Using Linked Statistical Data to Improve Marine Search and Rescue Operations in Ireland	https://zenodo.org/record/2551464	NUIG	Creative Commons Attribution 4.0 International

5 Conclusion

The purpose of this document was to provide the plan for managing the data generated and collected during the project; The Data Management Plan. Specifically, the DMP described the data management life cycle for all data sets to be collected, processed and/or generated by a research project. It covered:

- the handling of research data during and after the project
- what data will be collected, processed or generated.
- what methodology and standards will be applied.
- whether data will be shared/made open and how
- how data will be curated and preserved

Following the EU's guidelines regarding the DMP, this document was updated - if appropriate - during the project lifetime (in the form of deliverables).

The final version of the DMP involves twenty three data sets from eight different CSA domains including *Labour*, *Business statistics*, *Environment*, *Population and migration* and others. The data sets are provided in various formats (e.g. RDF, CSV), have various sizes (from 5kB to 127MB), and are originated from various sources (e.g. from the UK's National Statistics Excerpt, the Flemish government, or the Lithuanian Official Statistics Portal). Most data sets are openly provided to the public and published as linked data using the RDF Data Cube vocabulary facilitating their interoperability. Metadata of most data sets are modelled using standard vocabularies such as DCAT and PROV enhancing interoperability. Access to open data sets is given mainly through a HTML page, SPARQL endpoint, and/or a restful API. Regarding licensing, most data sets are provided under Open Government Licence 3.0 or publicly available data licences. The majority of open data sets will be preserved after the end of the project in various mediums including Open Science Framework and Zenodo. No resources were used to produce data sets. Moreover, the security of the data sets was ensured by using, for example, backup devices, and by backing up data in repositories like GitHub, Zenodo, and GitLab. Finally, regarding ethical aspects, for the data sets that include personal data, the consortium ensured that they comply with EU General Data Protection Regulation (GDPR) and also consent forms were provided to participants of interviews and surveys.

References

Ball, A. (2014). 'How to License Research Data'. DCC How-to Guides. Edinburgh: Digital Curation Centre. Available online: <http://www.dcc.ac.uk/resources/how-guides>

Dekkers, M., Loutas, N., De Keyzer M., and Goedertier, S. Open data and metadata quality. (2013). Available: https://joinup.ec.europa.eu/sites/default/files/D2.1.1%20Training%20Module%202.2%20Open%20Data%20Quality_v0.09_EN.pdf [12/06/2017]

European Commission (2016a), Guidelines on Data Management in Horizon 2020 version 2.1.

Available at

http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf [12/07/2016]

European Commission (2016b), Guidelines on Data Management in Horizon 2020 version 3.0.

Available at

http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf [12/06/2017]

Appendix 1: DMP Template

Data Summary	Purpose of data collection	Description of the purpose of the data collection and its relation to the project's objectives.
	Data type(s) and format	Description of the type and format of the data.
	Expected size	What is the estimated expected volume of data?
	Origin of data	Description of the origin of data (in case the data is collected). e.g. Data were be collected through a survey with key actors ... The questionnaires were disseminated to at least 20 recipients.
	Data utility	Describe to whom might data be useful e.g. Data will be useful for public authorities & policy makers in order to....
Findable data	Data set reference and name	Identifier for the data set to be produced (to be filled in by CERTH)
	Metadata	What metadata will be crated? What standards will be used? In case metadata standard do not exist in your discipline, please outline what type of metadata will be created and how.
Openly accessible data	Open access	<p>Will the data set be made openly available as the default? If the data set cannot be shared (or need to be shared under restrictions), explain why, clearly separating legal and contractual reasons from voluntary restrictions.</p> <p>Describe how will the data be made accessible (e.g. by deposition in a repository).</p> <p>What methods or software tools are needed to access the data? Is documentation about the software needed to access the data included? Is it possible to include the relevant software (e.g. in open source code)? Where will the data and associated metadata, documentation and code be deposited? Preference should be given to certified repositories (e.g. Zenodo) which support open access where possible. Please also add instructions on how to gain access to restricted data.</p>

Interoperable data	Interoperability and metadata	<p><i>Are the data interoperable, that is allowing data exchange and re-use between researchers, institutions, organisations, countries, etc. (i.e. adhering to standards for formats, as much as possible compliant with available (open) software applications, and in particular facilitating re-combinations with different data sets from different origins)?</i></p> <p><i>What data and metadata vocabularies, standards or methodologies are followed to make data interoperable? Are standard vocabularies used to allow inter-disciplinary interoperability? In case it is unavoidable to use uncommon or generate project specific ontologies or vocabularies, please provide mappings to more commonly used ontologies?</i></p>
Re-usable data	Data licensing	<i>How will the data be licensed to permit the widest re-use possible?</i>
	Quality assurance	<i>Describe the quality assurance process.</i>
	Preservation time	<i>Describe how long it is intended that the data remains re-usable.</i>
Allocation of resources	Cost estimation	<i>What are the costs for making data FAIR? How will these be covered? Note that costs related to open access to research data are eligible as part of the Horizon 2020 grant (if compliant with the Grant Agreement conditions).</i>
Secure Data	Data security	<p><i>What provisions are in place for data security (including data recovery as well as secure storage and transfer of sensitive data)?</i></p> <p><i>Is the data safely stored in certified repositories for long term preservation and curation?</i></p>

Appendix 2: Data sets

Marine Institute data sets

Oceanographic data on the weather and wave climate

Data Summary	Purpose of data collection	Real time oceanographic data on the weather and wave climate collected from the Irish moored Weather Buoy network of stations in the North Atlantic Ocean.
	Data type(s) and format	Point time-series observation data on weather and wave conditions are presently available in a number of common open formats including CSV (comma-separated value ASCII text table) and JSON (Java Script Object Notation).
	Expected size	Growing data set which currently is 50-55 MB in size as a CSV file.
	Origin of data	The Irish Weather Buoy Network moored surface metocean buoys collect data on weather and wave behaviour including wave height (in metres), wave period (in seconds), mean wave direction (degreeTrue) and maximum wave height (known as Hmax) on an hourly basis and this is disseminated via the Marine Institute ERRDAP data server. Information is collected on a frequency of an hourly basis on a continuous data collection basis when sensors are working correctly and in deployment.
	Data utility	<p>These weather buoy point time-series metocean observation data are useful to:</p> <ul style="list-style-type: none"> • Renewable energy community when considering developing a wave energy converter and requirements for access to real-time data on waves in order to predict the performance of devices and monitor safety once deployed in the environment. • Public – real-time information for tourism and leisure activities. • Government – marine spatial planning activities. • Search and Rescue – real-time information for of conditions. • Meteorological agencies – inform weather forecast models.
Findable data	Data set reference and name	<ol style="list-style-type: none"> 1. 3.1.IE.MarineInstitute.IrishWeatherWaveClimate 2. OF.PointTimeSeriesObservation.IE.MarineInstitute.IrishWeatherWaveClimate

	Metadata	<p>ISO 19139 metadata describing the observation data is hosted on the Irish Spatial Data Exchange and Ireland's Open Data Catalogue data.gov.ie.</p> <p>http://www.isde.ie/#/464864ce-319b-4645-b4c9-f408d656fd76</p>
Openly accessible data	Open access	<p>The Marine Institute hosts the Irish National Marine Data Centre (NMDC), which is the accredited National Oceanographic Data Centre (NODC) within the UNESCO IOC-IODE network of NODCs. These data are managed by the NMDC.</p> <p>The Marine Institute presently publish point time-series wave data via an ERDDAP data server under a Creative Common CC-BY 4.0 open licence. Users who require this data simply access the http link provided and choose the variables of interest (e.g. wave) to fit their requirements.</p> <p>http://erddap.marine.ie/erddap/tabledap/IWBNetwork.html</p> <p>The MI ERDDAP data server provides data access through a restful API.</p>
Interoperable data	Interoperability and metadata	<p>The data are interoperable as the metadata are marked up with machine-readable community vocabularies including the Climate and Forecast standard names for description of the parameters and the BODC unit vocabulary. The metadata are ISO19139 compliant and the data are available through a RESTful API with a selection of open data formats to choose from.</p>
Re-usable data	Data licensing	<p>The Marine Institute presently publish point time-series metocean data via an ERDDAP data server under a Creative Common CC-BY 4.0 open licence. Users who require this data simply access the http link provided and choose the variables of interest (e.g. wave) to fit their requirements.</p> <p>http://erddap.marine.ie/erddap/tabledap/IWBNetwork.html</p>

	Quality assurance	Irish Wave Buoy Network data is acquired, processed and stored by a central data acquisition system, operated and maintained by the Marine Institute. This system includes an Operational System Monitoring Component allowing the flow of real-time data to be monitored and assessed at key points throughout the system. Email and SMS alerts are supported, to notify the team of any issues encountered during data processing. In addition to real-time monitoring, data quality assurance routines are regularly applied to data post collection and quality flags recorded for processed data. Checks include: Wave buoy location validation, parameter range limit checks (e.g. wave height & temperature), and expert visual assessment.
	Preservation time	The data are being collected on an ongoing basis as part of a growing time-series. These data are to be kept indefinitely as a valuable time-series. Weather Buoy data are a core Marine Institute data asset of continuous data collection since 2001 when the first moored metocean buoy was deployed. The network now includes 5 metocean buoys.
Allocation of resources	Cost estimation	n/a already FAIR
Secure Data	Data security	These data are managed by the Irish National Marine Data Centre. The Marine Institute ERDDAP server is hosted on a backend SQL Server database with standard SQL Server backup devices used to protect data in case of technical issues.

Oceanographic data on the wave climate

Data Summary	Purpose of data collection	Real time oceanographic data on the wave climate collected from the Irish moored Wave Buoy network of stations in the North Atlantic Ocean.
	Data type(s) and format	Point time-series observation data on wave conditions are presently available in a number of common open formats including CSV (comma-separated value ASCII text table) and JSON (Java Script Object Notation).
	Expected size	Growing data set which currently is 65-70 MB in size as a CSV file.

	Origin of data	The Irish Wave Buoy Network moored surface wave buoys collect data on wave behaviour including wave height (in metres), wave period (in seconds), mean wave direction (degreeTrue) and maximum wave height (known as Hmax) on a half hourly basis and this is disseminated via the Marine Institute ERRDAP data server. Information is collected on a continuous data collection basis when sensors are working correctly and in deployment.
	Data utility	<p>These wave buoy point time-series observation data are useful to:</p> <ul style="list-style-type: none"> • Renewable energy community when considering developing a wave energy converter and requirements for access to real-time data on waves in order to predict the performance of devices and monitor safety once deployed in the environment. • Public – real-time information for tourism and leisure activities. • Government – marine spatial planning activities. • Search and Rescue – real-time information for of conditions. • Meteorological agencies – inform weather forecast models.
Findable data	Data set reference and name	<ol style="list-style-type: none"> 1. 3.1.IE.MarineInstitute.IrishWaveClimate 2. OF.PointTimeSeriesObservation.IE.MarineInstitute.IrishWaveClimate
	Metadata	<p>ISO 19139 metadata describing the observation data is hosted on the Irish Spatial Data Exchange and Ireland's Open Data Catalogue data.gov.ie.</p> <p>http://www.isde.ie/#/55eb27e0-2fc3-4dab-9963-a99d12402a9e</p>

Openly accessible data	Open access	<p>The Marine Institute hosts the Irish National Marine Data Centre (NMDC), which is the accredited National Oceanographic Data Centre (NODC) within the UNESCO IOC-IODE network of NODCs. These data are managed by the NMDC.</p> <p>The Marine Institute presently publish point time-series wave data via an ERDDAP data server under a Creative Common CC-BY 4.0 open licence. Users who require this data simply access the http link provided and choose the variables of interest (e.g. wave) to fit their requirements.</p> <p>http://erddap.marine.ie/erddap/tabledap/IWaveBNetwork.html</p> <p>The MI ERDDAP data server provides data access through a restful API.</p>
Interoperable data	Interoperability and metadata	<p>The data are interoperable as the metadata are marked up with machine-readable community vocabularies including the Climate and Forecast standard names for description of the parameters and the BODC unit vocabulary. The metadata are ISO19139 compliant and the data are available through a RESTful API with a selection of open data formats to choose from.</p>
Re-usable data	Data licensing	<p>The Marine Institute presently publish point time-series wave data via an ERDDAP data server under a Creative Common CC-BY 4.0 open licence. Users who require this data simply access the http link provided and choose the variables of interest (e.g. wave height) to fit their requirements.</p> <p>http://erddap.marine.ie/erddap/tabledap/IWBNetwork.html</p>
	Quality assurance	<p>Irish Weather Buoy Network data is acquired, processed and stored by a central data acquisition system, operated and maintained by the Marine Institute. This system includes an Operational System Monitoring Component allowing the flow of real-time data to be monitored and assessed at key points throughout the system. Email and SMS alerts are supported, to notify the team of any issues encountered during data processing. In addition to real-time monitoring, data quality assurance routines are regularly applied to data post collection and quality flags recorded for processed data. Checks include: Physical limits validation, wind and sea temperature spike checks and expert visual assessment.</p>

	Preservation time	The data are being collected on an ongoing basis as part of a growing time-series. These data are to be kept indefinitely as a valuable time-series. Wave Buoy data is a core Marine Institute data asset of continuous data collection since 2008 when the first moored surface buoy was deployed to a network of 4 buoys now deployed and observing the wave climate of the ocean at the designated ¼ scale and full scale test facilities.
Allocation of resources	Cost estimation	n/a already FAIR
Secure Data	Data security	These data are managed by the Irish National Marine Data Centre. The Marine Institute ERDDAP server is hosted on a backend SQL Server database with standard SQL Server backup devices used to protect data in case of technical issues.

TRAF data sets

Claimant count

Data Summary	Purpose of data collection	Key data set for Department of Work and Pensions partner.
	Data type(s) and format	The data are extracted in CSV format and have been converted to RDF using table2qb, part of the OGI Toolkit.
	Expected size	1,673 observations, 5.67 MB in .ttl format
	Origin of data	Office for National Statistics. Excerpt extracted from NOMIS platform: https://www.nomisweb.co.uk/datasets/ucjsa
	Data utility	Department of Work and Pensions and worklessness leads across Greater Manchester's councils
Findable	Data set reference and name	1.2.UK.TRAF.ClaimantCount

	Metadata	Publisher, licence, next update.
Openly accessible	Open access	The updated data set is stored in linked open data format on the GM DataStore: http://gmdatastore.org.uk/data/claimant-count . An archived copy is available on the Open Science Framework project site: https://osf.io/pa4rf/
Interoperable data	Interoperability and metadata	The RDF data are fully interoperable according to the W3C Tabular Data and Metadata on the Web as outlined in the guidance for table2qb.
Re-usable data	Data licensing	The data are re-usable under the terms of the <i>Open Government Licence 3.0</i> .
	Quality assurance	The data is an excerpt from Office for National Statistics sources which is subject to their quality assurance processes.
	Preservation time	Unlimited.
Allocation of resources	Cost estimation	Data storage on the GM DataStore is subject to the pricing terms determined by SWIRRL. However, this facility has been provided for the purposes of the project. Data storage on the Open Science Framework is free.
Secure Data	Data security	Information about data security on the Open Science Framework is provided here: http://help.osf.io/m/security/l/524224-faqs-security . The data are backed up in a dedicated project repository on GitHub as well as being stored as linked open statistical data on SWIRRL's instance of the Publish My Data platform, GM DataStore.

Claimant rate

Data Summ	Purpose of data collection	Key data set for Department of Work and Pensions partner.
------------------	-----------------------------------	---

	Data type(s) and format	The data are extracted in CSV format and have been converted to RDF using table2qb, part of the OGI Toolkit.
	Expected size	8,362 observations, 28.5 MB in .ttl format
	Origin of data	Office for National Statistics. Excerpt extracted from NOMIS platform: https://www.nomisweb.co.uk/datasets/ucjsa
	Data utility	Department of Work and Pensions and worklessness leads across Greater Manchester's councils
Findable data	Data set reference and name	1.2.UK.TRAF.ClaimantRate
	Metadata	Publisher, license, next update.
Openly accessible	Open access	The updated data set is stored in linked open data format on the GM DataStore: http://gmdatastore.org.uk/data/claimant-rate . An archived copy is available on the Open Science Framework project site: https://osf.io/pa4rf/ .
Interoperable data	Interoperability and metadata	The RDF data are fully interoperable according to the W3C Tabular Data and Metadata on the Web as outlined in the guidance for table2qb.
Re-usable data	Data licensing	The data are re-usable under the terms of the <i>Open Government License 3.0</i> .
	Quality assurance	The data is an excerpt from Office for National Statistics sources which is subject to their quality assurance processes.
	Preservation time	Unlimited.
Allocation of resources	Cost estimation	Data storage on the GM DataStore is subject to the pricing terms determined by SWIRRL. However, this facility has been provided for the purposes of the project. Data storage on the Open Science Framework is free.

Secure Data	Data security	<p>Information about data security on the Open Science Framework is provided here: http://help.osf.io/m/security/l/524224-faqs-security.</p> <p>The data are backed up in a dedicated project repository on GitHub as well as being stored as linked open statistical data on SWIRRL's instance of the Publish My Data platform, GM DataStore.</p>
--------------------	----------------------	---

Households with lone parent not in employment

Data Summary	Purpose of data collection	Key data set for Department of Work and Pensions partner.
	Data type(s) and format	The data are extracted in CSV format and have been converted to RDF using table2qb, part of the OGI Toolkit.
	Expected size	1,899 observations, 6.44 MB in .ttl format
	Origin of data	Office for National Statistics. Excerpt extracted from NOMIS platform: https://www.nomisweb.co.uk/census/2011/KS107EW
	Data utility	Department of Work and Pensions and worklessness leads across Greater Manchester's councils
Findable data	Data set reference and name	1.2.UK.TRAF.HouseholdsLoneParentNotInEmployment
	Metadata	Publisher, licence, next update.
Openly accessible data	Open access	<p>The updated data set is stored in linked open data format on the GM DataStore: http://gmdatastore.org.uk/data/households-with-lone-parent-not-in-employment. An archived copy is available on the Open Science Framework project site: https://osf.io/pa4rf/.</p>

Interoperable data	Interoperability and metadata	The RDF data are fully interoperable according to the W3C Tabular Data and Metadata on the Web as outlined in the guidance for table2qb.
Re-usable data	Data licensing	The data are re-usable under the terms of the <i>Open Government Licence 3.0</i> .
	Quality assurance	The data is an excerpt from Office for National Statistics sources which is subject to their quality assurance processes.
	Preservation time	Unlimited.
Allocation of resources	Cost estimation	Data storage on the GM DataStore is subject to the pricing terms determined by SWIRRL. However, this facility has been provided for the purposes of the project. Data storage on the Open Science Framework is free.
Secure Data	Data security	<p>Information about data security on the Open Science Framework is provided here: http://help.osf.io/m/security/l/524224-faqs-security.</p> <p>The data are backed up in a dedicated project repository on GitHub as well as being stored as linked open statistical data on SWIRRL's instance of the Publish My Data platform, GM DataStore.</p>

Social rented households

Data Summary	Purpose of data collection	Key data set for Department of Work and Pensions partner.
	Data type(s) and format	The data are extracted in CSV format and have been converted to RDF using table2qb, part of the OGI Toolkit.
	Expected size	1,899 observations, 5.72 MB in .ttl format
	Origin of data	Office for National Statistics. Excerpt extracted from NOMIS platform: https://www.nomisweb.co.uk/census/2011/ks402ew .

	Data utility	Department of Work and Pensions and worklessness leads across Greater Manchester's councils
Findable data	Data set reference and name	1.6.UK.TRAF.SocialRentedHouseholds
	Metadata	Publisher, licence, next update.
Openly accessible data	Open access	The updated data set is stored in linked open data format on the GM DataStore: http://gmdatastore.org.uk/data/social-rented-households . An archived copy is available on the Open Science Framework project site: https://osf.io/pa4rf/ .
Interoperable data	Interoperability and metadata	The RDF data are fully interoperable according to the W3C Tabular Data and Metadata on the Web as outlined in the guidance for table2qb.
Re-usable data	Data licensing	The data are re-usable under the terms of the <i>Open Government Licence 3.0</i> .
	Quality assurance	The data is an excerpt from Office for National Statistics sources which is subject to their quality assurance processes.
	Preservation time	Unlimited.
Allocation of resources	Cost estimation	Data storage on the GM DataStore is subject to the pricing terms determined by SWIRRL. However, this facility has been provided for the purposes of the project. Data storage on the Open Science Framework is free.
Secure Data	Data security	Information about data security on the Open Science Framework is provided here: http://help.osf.io/m/security/l/524224-faqs-security . The data are backed up in a dedicated project repository on GitHub as well as being stored as linked open statistical data on SWIRRL's instance of the Publish My Data platform, GM DataStore.

Working age adults with no qualifications

Data Summary	Purpose of data collection	Key data set for Department of Work and Pensions partner.
	Data type(s) and format	The data are extracted in CSV format and have been converted to RDF using table2qb, part of the OGI Toolkit.
	Expected size	1,899 observations, 6.23 MB in .ttl format
	Origin of data	Office for National Statistics. Excerpt extracted from NOMIS platform: https://www.nomisweb.co.uk/census/2011/lc5601ew
	Data utility	Department of Work and Pensions and worklessness leads across Greater Manchester's councils
Findable data	Data set reference and name	1.2.UK.TRAF.WorkingAgeAdultsNoQualifications
	Metadata	Publisher, licence, next update.
Openly accessible	Open access	The updated data set is stored in linked open data format on the GM DataStore: http://gmdatastore.org.uk/data/working-age-adults-with-no-qualifications . An archived copy is available on the Open Science Framework project site: https://osf.io/pa4rf/ .
Interoperable data	Interoperability and metadata	The RDF data are fully interoperable according to the W3C Tabular Data and Metadata on the Web as outlined in the guidance for table2qb.
Re-usable data	Data licensing	The data are re-usable under the terms of the <i>Open Government Licence 3.0</i> .
	Quality assurance	The data is an excerpt from Office for National Statistics sources which is subject to their quality assurance processes.
	Preservation time	Unlimited.

Allocation of resources	Cost estimation	Data storage on the GM DataStore is subject to the pricing terms determined by SWIRRL. However, this facility has been provided for the purposes of the project. Data storage on the Open Science Framework is free.
Secure Data	Data security	Information about data security on the Open Science Framework is provided here: http://help.osf.io/m/security/l/524224-faqs-security The data are backed up in a dedicated project repository on GitHub as well as being stored as linked open statistical data on SWIRRL's instance of the Publish My Data platform, GM DataStore.

Employees by broad industry group

Data Summary	Purpose of data collection	Key data set for Department of Work and Pensions partner.
	Data type(s) and format	The data are extracted in CSV format and have been converted to RDF using table2qb, part of the OGI Toolkit.
	Expected size	3,870 observations, 15.8 MB in .ttl format
	Origin of data	Office for National Statistics. Excerpt extracted from: https://www.ons.gov.uk/employmentandlabourmarket/peopleinwork/employmentandemployeetypes/bulletins/businessregisterandemploymentsurveybresprovisionalresults/provisionalresults2017revisedresults2016
	Data utility	Department of Work and Pensions and worklessness leads across Greater Manchester's councils
Findable data	Data set reference and name	1.2.UK.TRAF.EmployeesBroadIndustryGroup
	Metadata	Publisher, licence, next update.

Openly accessible	Open access	The updated data set is stored in linked open data format on the GM DataStore: http://gmdatastore.org.uk/data/working-age-population . An archived copy is available on the Open Science Framework project site: https://osf.io/pa4rf/ .
Interoperable data	Interoperability and metadata	The RDF data are fully interoperable according to the W3C Tabular Data and Metadata on the Web as outlined in the guidance for table2qb.
Re-usable data	Data licensing	The data are re-usable under the terms of the <i>Open Government Licence 3.0</i> .
	Quality assurance	The data is an excerpt from Office for National Statistics sources which is subject to their quality assurance processes.
	Preservation time	Unlimited.
Allocation of resources	Cost estimation	Data storage on the GM DataStore is subject to the pricing terms determined by SWIRRL. However, this facility has been provided for the purposes of the project. Data storage on the Open Science Framework is free.
Secure Data	Data security	Information about data security on the Open Science Framework is provided here: http://help.osf.io/m/security/I/524224-faqs-security . The data are backed up in a dedicated project repository on GitHub as well as being stored as linked open statistical data on SWIRRL's instance of the Publish My Data platform, GM DataStore.

Working age population

Data Summary	Purpose of data collection	Key data set for Department of Work and Pensions partner.
	Data type(s) and format	The data are extracted in CSV format and have been converted to RDF using table2qb, part of the OGI Toolkit.

	Expected size	226 observations, 689 KB in .ttl format
	Origin of data	Office for National Statistics. Excerpt extracted from: https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates/datasets/wardlevelmidyearpopulationestimatesexperimental .
	Data utility	Department of Work and Pensions and worklessness leads across Greater Manchester's councils
Findable data	Data set reference and name	1.2.UK.TRAF.WorkingAgePopulation
	Metadata	Publisher, licence, next update.
Openly accessible	Open access	The updated data set is stored in linked open data format on the GM DataStore: http://gmdatastore.org.uk/data/working-age-population . An archived copy is available on the Open Science Framework project site: https://osf.io/pa4rf/ .
Interoperable data	Interoperability and metadata	The RDF data are fully interoperable according to the W3C Tabular Data and Metadata on the Web as outlined in the guidance for table2qb.
Re-usable data	Data licensing	The data are re-usable under the terms of the <i>Open Government Licence 3.0</i> .
	Quality assurance	The data is an excerpt from Office for National Statistics sources which is subject to their quality assurance processes.
	Preservation time	Unlimited.
Allocation of resources	Cost estimation	Data storage on the GM DataStore is subject to the pricing terms determined by SWIRRL. However, this facility has been provided for the purposes of the project. Data storage on the Open Science Framework is free.

Secure Data	Data security	<p>Information about data security on the Open Science Framework is provided here: http://help.osf.io/m/security/l/524224-faqs-security.</p> <p>The data are backed up in a dedicated project repository on GitHub as well as being stored as linked open statistical data on SWIRRL's instance of the Publish My Data platform, GM DataStore.</p>
-------------	---------------	---

MAREG data sets

Government Vehicles of the Greek Government Agency

Data Summary	Purpose of data collection	Descriptive data about the fleet of Government Vehicles used by each Greek Government Agency.
	Data type(s) and format	The raw data was originally acquired in .csv format. The aggregated data was converted to .ttl using the table2qb app, which was created by the OGI project.
	Expected size	The Government Vehicles raw data set contains 4.332 entries and the size of the .csv file is 1.2 MB. The aggregated data set contains 894 observations 455 KB in .ttl format.
	Origin of data	Government Vehicles data were collected by MAREG with the contribution of Greek Government Agencies sending spreadsheets with data about their fleet. These data were then compared with data extracted from the Information System of the Greek Ministry of Transport.
	Data utility	<p>These data are useful to:</p> <ul style="list-style-type: none"> • MAREG to facilitate the internal decision making process regarding Government Vehicles • Decentralized Administrations having the responsibility of Government Vehicles of all Government Agencies operating within their geographic area • Government Agencies that want to gain insight of their fleet
Findable data	Data set reference and name	2.4.4.GR.MAREG.GovernmentVehicles
	Metadata	Title, Publisher, license, last update

Openly accessible data	Open access	An aggregated version of the data set is hosted in Zenodo repository (url: https://zenodo.org/record/2551095)
Interoperable data	Interoperability and metadata	Data and metadata have been created using standard vocabularies thus ensuring interoperability. <ul style="list-style-type: none"> Standards used for the data: RDF Data Cube Vocabulary Standards used for metadata: Dublin Core, SKOS, ORG vocabulary
Re-usable data	Data licensing	Data are published under the Creative Common CC-BY 4.0 license.
	Quality assurance	The quality assurance process is based on a data management tool created within OGI project, which Government Agencies will use to confirm or update their fleet data. Access to the tool is provided through the web from the following url: http://wapps.islab.uom.gr:9090 .
	Preservation time	At least 5 years in Zenodo
Allocation of resources	Cost estimation	None
Secure Data	Data security	The data are stored in Zenodo which has all the necessary procedures for long lasting preservation and curation in place.

Flemish Government data sets

Air and Water Pollution Report Metadata

Data Summary	Purpose of data collection	Companies and organizations must notify the Flemish government which waste materials they have produced or emitted in the past year. These reports contain next to the waste related data themselves (measurements) all kinds of metadata to allow efficient archival and retrieval.
	Data type(s) and format	The report's metadata are available in XML files and are archived in the department's archival system (DSPACE).

		These metadata are converted into RDF by the project and loaded in a triple store.
	Expected size	Final size covering 2011 to 2016: 2169613 triples on 157596 reports.
	Origin of data	The data have been collected by a dedicated web application called IMJV. See https://imjv.milieuinfo.be/
	Data utility	The data are useful to the department. They allow the department to archive, manage and retrieve the reports in a more efficient way.
Findable data	Data set reference and name	3.1.BE.VLO.AirWaterPollutionMetadata
	Metadata	These metadata on the documents are using the Dublin Core metadata element set. (http://dublincore.org/specifications/).
Openly accessible	Open access	The discussion on opening up the data to the larger public is still ongoing in the department since these are data as reported by the companies or organizations as such and are not validated or curated (yet).
Interoperable data	Interoperability metadata and	The metadata are completely standards based e.g. Dublin Core and hence interoperable without any mapping or conversion needed.
Re-usable data	Data licensing	Still under discussion (<i>cf. supra</i>)
	Quality assurance	During conversion to RDF, values are normalized, and a lot of string values are reconciled into unique identifiers coming from controlled and centrally managed terminology vocabularies.
	Preservation time	The data are already in an archival system, which has all the necessary procedures for long lasting preservation in place. When storing the data, the owners made sure that they adhere to the EU GDPR.

Allocation of resources	Cost estimation	Not applicable.
Secure Data	Data security	The data are kept in an archival system, which has all the necessary procedures for access control, long-lasting preservation (backup and restore) and rapid redeployment in place.

Air and Water Pollution Data

Data Summary	Purpose of data collection	<p>Companies and organizations must notify the Flemish government which waste materials they have produced or emitted in the past year.</p> <p>These reports contain the aggregated measurements/observations of their air/water pollution.</p> <p>These are the core statistical data to work with.</p>
	Data type(s) and format	<p>The emission data are available in XML files and archived in an archival system (DSPACE).</p> <p>These data are converted into RDF during the project and loaded in a triple store.</p>
	Expected size	Final size covering 2011 to 2016: 25434467 triples on 2828123 entities.
	Origin of data	The data have been collected by a dedicated web application called IMJV. (<i>cf. supra</i>)
	Data utility	The data are useful to the department to be able to generate analytics on the data as starting point for better planning of inspections, for validation and policy building.
Findable data	Data set reference and name	3.1.BE.VLO.AirWaterPollutionData
	Metadata	<p>The observations are encoded according to the RDF Data Cube vocabulary (https://www.w3.org/TR/vocab-data-cube/) also taking into account the best practices as established by the project (https://islab-uom.github.io/qbBestPractices/).</p> <p>Controlled vocabularies are encoded using SKOS (https://www.w3.org/2004/02/skos/).</p>

Openly accessible	Open access	The discussion on opening up the data to the larger public is still ongoing in the department since these are data as reported by the companies or organizations and are not validated or curated (yet).
Interoperable data	Interoperability metadata and	The observations are encoded according to the RDF Data Cube vocabulary. Controlled vocabularies are encoded using SKOS. For other supporting classes where no standard was available, own vocabularies have been developed compliant with the existing XML data model. The department is seeking to standardize those on the Flemish level after remodeling the data model.
Re-usable data	Data licensing	Still under discussion (<i>cf. supra</i>)
	Quality assurance	During conversion to RDF, values are normalized. A lookup process is in place for replacing strings into unique identifiers.
	Preservation time	The data are already in an archival system, which has all the necessary procedures for long lasting preservation in place.
Allocation of resources	Cost estimation	Not applicable.
Secure Data	Data security	The data are kept in an archival system, which has all the necessary procedures for access control, long-lasting preservation (backup and restore) and quick redeployment in place.

Company registry

Data Summary	Purpose of data collection	The Environment Department (LNE) of the Flemish government keeps company (and their sites) information in a PostgreSQL DB for their application needs.
	Data type(s) and format	PostgreSQLdatabase. These data are virtualised as a SPARQL endpoint by the project.
	Expected size	Actual size: 96875 database rows
	Origin of data	Data coming from the federal Belgian government (KBO), enhanced by the central Flemish Government and then extended by the Department itself with domain specific fields.
	Data utility	These are base registry data.
File	Data set reference and	2.3.BE.VLO.CompanyRegistry

	name	
	Metadata	The virtual SPARQL endpoint maps these to the W3C's ORG vocab (https://www.w3.org/TR/2014/REC-vocab-org-20140116/).
Openly accessible	Open access	The registry data are being opened by the Flemish Government.
Interoperable data	Interoperability metadata and	These data are using the ORG vocabulary and completely in line with the Belgian standards.
Re-usable data	Data licensing	Open (Flemish Open Data License 1)
	Quality assurance	The Federal data are validated and enhanced on the Flemish level and then once again on the Departmental level.
	Preservation time	This is a base registry. As long as the department exists it will be a central data set.
Allocation of resources	Cost estimation	Not Applicable
Secure Data	Data security	This is part of the central ICT infrastructure of the department where all measures for access control, high availability, backup and restore, quick deployment are in place. The triple store has 2 entry points: one internal with all records available, and an external one where privacy related data are not exposed since filtered out in the source database (PostgreSQL) view.

Enterprise Lithuania data sets

Economic Activities in Lithuania

Data Summary	Purpose of data collection	Understanding the geographical distribution of different economic activities done in Lithuania.
	Data type(s) and	Data is presented in XML format and then converted to CSV and

	format	RDF using specific scripts and tools developed for this purpose.
	Expected size	Raw CSV file size is 2.97 MB
	Origin of data	Data set was downloaded from the State Food and Veterinary Service webpage: http://vmvt.lt/
	Data utility	Data will be useful for citizens or investors aiming at opening new businesses in Lithuania as it will provide insights for areas with less competition.
Findable data	Data set reference and name	2.3.LT.EnterpriseLithuania.EconomicActivities
	Metadata	Last update, Title.
Openly accessible data	Open access	<p>Data set will be openly available through two different mechanisms:</p> <ul style="list-style-type: none"> a) A central storage RDF endpoint. Knowledge in SPARQL query language will be required by the interested parties to access this data. URL: http://vmogi01.deri.ie:8000/sparql b) A GraphQL based API implementation. Data accessibility will be enhanced by providing access with a widely known standard such as JSON. <p>Data are also be available for download in the following Zenodo repository: https://zenodo.org/record/2551455</p>
Interoperable data	Interoperability and metadata	The RDF data are fully interoperable according to the W3C Tabular Data and Metadata on the Web as outlined in the guidance for the tool table2qb used along this project.
Re- use	Data licensing	Publicly available data

	Quality assurance	<p>First, data is an excerpt from the State Food and Veterinary Service sources which is subject to their quality assurance processes.</p> <p>During the data cleansing process, well documented processes were run in order to preserve data quality.</p> <p>During the data enrichment process new data was generated using BING reverse geocoding API (https://msdn.microsoft.com/en-us/library/ff701714.aspx#Examples) to represent geographical distributed information. The quality of this generated data is subject to this service.</p>
	Preservation time	Unlimited.
Allocation of resources	Cost estimation	OpenLink Virtuoso data storage engine is free subjected to the use of Insight Centre's server infrastructure. The last one is provided during the OGI project lifecycle. No future plans for maintenance are defined.
Secure Data	Data security	<p>Information about data security on the OpenLink Virtuoso platform can be found here: http://vos.openlinksw.com/owiki/wiki/VOS/VirtRDFGraphsSecurity.</p> <p>RDF data has been backed up in an internal GitLab repository and stored as open statistical data in the mentioned engine. The persistence of the data is ensured by NUIG backup mechanisms.</p>

State Social Insurance - Sodra

Data Summary	Purpose of data collection	Understanding the insurances issued at the municipality level can provide insights about how many companies are settled in each municipality.
	Data type(s) and format	Data is presented CSV format and converted to RDF using specific tools developed for this purpose.
	Expected size	Raw CSV file size is 127 MB
	Origin of data	Data set was downloaded from the Open Corporate Data (Sodra) webpage: http://atvira.sodra.lt/imones/rinkiniai/index.html
	Data utility	Data will be useful for citizens or investors aiming at opening new businesses in Lithuania.
Findable	Data set reference and name	2.4.6.LT.EnterpriseLithuania.StateSocialInsurance

	Metadata	Last update, Title.
Openly accessible data	Open access	<p>Data set will be openly available through two different mechanisms:</p> <ul style="list-style-type: none"> a) A central storage RDF endpoint. Knowledge in SPARQL query language will be required by the interested parties to access this data. URL: http://vmogi01.deri.ie:8000/sparql b) A GraphQL based API implementation. Data accessibility will be enhanced by providing access with a widely known standard such as JSON. <p>Data are also be available for download in the following Zenodo repository: https://zenodo.org/record/2551455</p>
Interoperable data	Interoperability and metadata	The RDF data are fully interoperable according to the W3C Tabular Data and Metadata on the Web as outlined in the guidance for the tool table2qb used along this project.
Re-usable data	Data licensing	Publicly available data
	Quality assurance	<p>First, data is an excerpt from the Open Corporate Data (Sodra, http://atvira.sodra.lt/imonos/rinkiniai/index.html) sources which is subject to their quality assurance processes.</p> <p>During the data cleansing process, well documented processes were run to preserve data quality.</p>
	Preservation time	Unlimited.
Allocation of resources	Cost estimation	OpenLink Virtuoso data storage engine is free subjected to the use of Insight Centre's server infrastructure. The last one is provided during the OGI project lifecycle. No future plans for maintenance are defined.
Secure Data	Data security	<p>Information about data security on the OpenLink Virtuoso platform can be found here: http://vos.openlinksw.com/owiki/wiki/VOS/VirtRDFGraphsSecurity</p> <p>RDF data has been backed up in an internal GitLab repository and stored as open statistical data in the mentioned engine. The persistence of the data is ensured by NUIG backup mechanisms.</p>

Average earnings (monthly (2015))

Data Summary	Purpose of data collection	Understanding the average earnings at the municipality level.
	Data type(s) and format	Data is presented in CSV format and then converted to RDF using specific scripts and tools developed for this purpose.
	Expected size	Raw CSV file size is 5 KB
	Origin of data	Data set was downloaded from the Official Statistics Portal https://osp.stat.gov.lt/
	Data utility	Data will be useful for citizens or investors aiming at opening new businesses in Lithuania as it will provide insights for areas with higher income.
Findable data	Data set reference and name	2.3.LT.EnterpriseLithuania.AverageEarnings2015
	Metadata	Last update, Title.
Openly accessible data	Open access	<p>Data set will be openly available through two different mechanisms:</p> <ul style="list-style-type: none"> a) A central storage RDF endpoint. Knowledge in SPARQL query language will be required by the interested parties to access this data. URL: http://vmogi01.deri.ie:8000/sparql b) A GraphQL based API implementation. Data accessibility will be enhanced by providing access with a widely known standard such as JSON. <p>Data are also be available for download in the following Zenodo repository: https://zenodo.org/record/2551455</p>
Interoperable data	Interoperability and metadata	The RDF data are fully interoperable according to the W3C Tabular Data and Metadata on the Web as outlined in the guidance for the tool table2qb used along this project.
Re- use	Data licensing	Publicly available data

	Quality assurance	First, data is an excerpt from the Lithuanian Official Statistics Portal sources which is subject to their quality assurance processes. During the data cleansing process, well documented processes were run to preserve data quality.
	Preservation time	Unlimited.
Allocation of resources	Cost estimation	OpenLink Virtuoso data storage engine is free subjected to the use of Insight Centre's server infrastructure. The last one is provided during the OGI project lifecycle. No future plans for maintenance are defined.
Secure Data	Data security	Information about data security on the OpenLink Virtuoso platform can be found here: http://vos.openlinksw.com/owiki/wiki/VOS/VirtRDFGraphsSecurity RDF data has been backed up in an internal GitLab repository and stored as open statistical data in the mentioned engine. The persistence of the data is ensured by NUIG backup mechanisms.

Investment in tangible fixed assets at current prices (2015)

Data Summary	Purpose of data collection	Understanding the investment at the municipality level.
	Data type(s) and format	Data is presented in CSV format and then converted to RDF using specific scripts and tools developed for this purpose.
	Expected size	Raw CSV file size is 4 KB
	Origin of data	Data set was downloaded from the Official Statistics Portal https://osp.stat.gov.lt/
	Data utility	Data will be useful for citizens or investors aiming at opening new businesses in Lithuania as it will provide insights for areas with higher investment.
Findable data	Data set reference and name	2.3.LT.EnterpriseLithuania.Investments2015
	Metadata	Last update, Title.

Openly accessible data	Open access	<p>Data set will be openly available through two different mechanisms:</p> <ul style="list-style-type: none"> a) A central storage RDF endpoint. Knowledge in SPARQL query language will be required by the interested parties to access this data. URL: http://vmogi01.deri.ie:8000/sparql b) A GraphQL based API implementation. Data accessibility will be enhanced by providing access with a widely known standard such as JSON. <p>Data are also be available for download in the following Zenodo repository: https://zenodo.org/record/2551455</p>
Interoperable data	Interoperability and metadata	The RDF data are fully interoperable according to the W3C Tabular Data and Metadata on the Web as outlined in the guidance for the tool table2qb used along this project.
Re-usable data	Data licensing	Publicly available data
	Quality assurance	<p>First, data is an excerpt from the Lithuanian Official Statistics Portal sources which is subject to their quality assurance processes.</p> <p>During the data cleansing process, well documented processes were run to preserve data quality.</p>
	Preservation time	Unlimited.
Allocation of resources	Cost estimation	OpenLink Virtuoso data storage engine is free subjected to the use of Insight Centre's server infrastructure. The last one is provided during the OGI project lifecycle. No future plans for maintenance are defined.
Secure Data	Data security	<p>Information about data security on the OpenLink Virtuoso platform can be found here: http://vos.openlinksw.com/owiki/wiki/VOS/VirtRDFGraphsSecurity</p> <p>RDF data has been backed up in an internal GitLab repository and stored as open statistical data in the mentioned engine. The persistence of the data is ensured by NUIG backup mechanisms.</p>

Foreign direct investment per capita at the end of the year (2014)

Data	Purpose of data collection	Understanding the foreign investment at the municipality level.
------	----------------------------	---

	Data type(s) and format	Data is presented in CSV format and then converted to RDF using specific scripts and tools developed for this purpose.
	Expected size	Raw CSV file size is 7 KB
	Origin of data	Data set was downloaded from the Official Statistics Portal https://osp.stat.gov.lt/
	Data utility	Data will be useful for citizens or investors aiming at opening new businesses in Lithuania as it will provide insights for areas with higher foreign investment.
Findable data	Data set reference and name	2.3.LT.EnterpriseLithuania.ForeignInvestments2014
	Metadata	Last update, Title.
Openly accessible data	Open access	<p>Data set will be openly available through two different mechanisms:</p> <ul style="list-style-type: none"> a) A central storage RDF endpoint. Knowledge in SPARQL query language will be required by the interested parties to access this data. URL: http://vmogi01.deri.ie:8000/sparql b) A GraphQL based API implementation. Data accessibility will be enhanced by providing access with a widely known standard such as JSON. <p>Data are also be available for download in the following Zenodo repository: https://zenodo.org/record/2551455</p>
Interoperable data	Interoperability and metadata	The RDF data are fully interoperable according to the W3C Tabular Data and Metadata on the Web as outlined in the guidance for the tool table2qb used along this project.
Re-usable data	Data licensing	Publicly available data
	Quality assurance	<p>First, data is an excerpt from the Lithuanian Official Statistics Portal sources which is subject to their quality assurance processes.</p> <p>During the data cleansing process, well documented processes were run to preserve data quality.</p>
	Preservation time	Unlimited.

Allocation of resources	Cost estimation	OpenLink Virtuoso data storage engine is free subjected to the use of Insight Centre's server infrastructure. The last one is provided during the OGI project lifecycle. No future plans for maintenance are defined.
Secure Data	Data security	Information about data security on the OpenLink Virtuoso platform can be found here: http://vos.openlinksw.com/owiki/wiki/VOS/VirtRDFGraphsSecurity RDF data has been backed up in an internal GitLab repository and stored as open statistical data in the mentioned engine. The persistence of the data is ensured by NUIG backup mechanisms.

Economic entities in operation at the beginning of the year

Data Summary	Purpose of data collection	Understanding the economic entities in operation at the municipality level.
	Data type(s) and format	Data is presented in CSV format and then converted to RDF using specific scripts and tools developed for this purpose.
	Expected size	Raw CSV file size is 7 KB
	Origin of data	Data set was downloaded from the Official Statistics Portal https://osp.stat.gov.lt/
	Data utility	Data will be useful for citizens or investors aiming at opening new businesses in Lithuania as it will provide insights for areas with more economic activities.
Findable data	Data set reference and name	2.3.LT.EnterpriseLithuania.EconomicEntities2015
	Metadata	Last update, Title.

Openly accessible data	Open access	<p>Data set will be openly available through two different mechanisms:</p> <ul style="list-style-type: none"> a) A central storage RDF endpoint. Knowledge in SPARQL query language will be required by the interested parties to access this data. URL: http://vmogi01.deri.ie:8000/sparql b) A GraphQL based API implementation. Data accessibility will be enhanced by providing access with a widely known standard such as JSON. <p>Data are also be available for download in the following Zenodo repository: https://zenodo.org/record/2551455</p>
Interoperable data	Interoperability and metadata	The RDF data are fully interoperable according to the W3C Tabular Data and Metadata on the Web as outlined in the guidance for the tool table2qb used along this project.
Re-usable data	Data licensing	Publicly available data
	Quality assurance	<p>First, data is an excerpt from the Lithuanian Official Statistics Portal sources which is subject to their quality assurance processes.</p> <p>During the data cleansing process, well documented processes were run to preserve data quality.</p>
	Preservation time	Unlimited.
Allocation of resources	Cost estimation	OpenLink Virtuoso data storage engine is free subjected to the use of Insight Centre's server infrastructure. The last one is provided during the OGI project lifecycle. No future plans for maintenance are defined.
Secure Data	Data security	<p>Information about data security on the OpenLink Virtuoso platform can be found here: http://vos.openlinksw.com/owiki/wiki/VOS/VirtRDFGraphsSecurity</p> <p>RDF data has been backed up in an internal GitLab repository and stored as open statistical data in the mentioned engine. The persistence of the data is ensured by NUIG backup mechanisms.</p>

Resident population at the beginning of the year (by age groups)

Dat a	Purpose of data collection	Understanding the demographic distribution at the municipality
----------	----------------------------	--

		level.
	Data type(s) and format	Data is presented in CSV format and then converted to RDF using specific scripts and tools developed for this purpose.
	Expected size	Raw CSV file size is 31 KB
	Origin of data	Data set was downloaded from the Official Statistics Portal https://osp.stat.gov.lt/
	Data utility	Data will be useful for citizens or investors aiming at opening new businesses in Lithuania as it will provide a clear understanding of the potential target market distribution.
Findable data	Data set reference and name	1.1.LT.EnterpriseLithuania.ResidentPopulationAgeGroups
	Metadata	Last update, Title.
Openly accessible data	Open access	<p>Data set will be openly available through two different mechanisms:</p> <ul style="list-style-type: none"> a) A central storage RDF endpoint. Knowledge in SPARQL query language will be required by the interested parties to access this data. URL: http://vmogi01.deri.ie:8000/spargl b) A GraphQL based API implementation. Data accessibility will be enhanced by providing access with a widely known standard such as JSON. <p>Data are also be available for download in the following Zenodo repository: https://zenodo.org/record/2551455</p>
Interoperable data	Interoperability and metadata	The RDF data are fully interoperable according to the W3C Tabular Data and Metadata on the Web as outlined in the guidance for the tool table2qb used along this project.
Re- use	Data licensing	Publicly available data

	Quality assurance	First, data is an excerpt from the Lithuanian Official Statistics Portal sources which is subject to their quality assurance processes. During the data cleansing process, well documented processes were run to preserve data quality.
	Preservation time	Unlimited.
Allocation of resources	Cost estimation	OpenLink Virtuoso data storage engine is free subjected to the use of Insight Centre's server infrastructure. The last one is provided during the OGI project lifecycle. No future plans for maintenance are defined.
Secure Data	Data security	Information about data security on the OpenLink Virtuoso platform can be found here: http://vos.openlinksw.com/owiki/wiki/VOS/VirtRDFGraphsSecurity RDF data has been backed up in an internal GitLab repository and stored as open statistical data in the mentioned engine. The persistence of the data is ensured by NUIG backup mechanisms.

Resident population at the beginning of the year (by gender)

Data Summary	Purpose of data collection	Understanding the demographic distribution at the municipality level.
	Data type(s) and format	Data is presented in CSV format and then converted to RDF using specific scripts and tools developed for this purpose.
	Expected size	Raw CSV file size is 31 KB
	Origin of data	Data set was downloaded from the Official Statistics Portal https://osp.stat.gov.lt/ .
	Data utility	Data will be useful for citizens or investors aiming at opening new businesses in Lithuania as it will provide a clear understanding of the potential target market distribution.
Findable data	Data set reference and name	1.1.LT.EnterpriseLithuania.ResidentPopulationGender
	Metadata	Last update, Title.

Openly accessible data	Open access	<p>Data set will be openly available through two different mechanisms:</p> <ul style="list-style-type: none"> a) A central storage RDF endpoint. Knowledge in SPARQL query language will be required by the interested parties to access this data. URL: http://vmogi01.deri.ie:8000/sparql b) A GraphQL based API implementation. Data accessibility will be enhanced by providing access with a widely known standard such as JSON. <p>Data are also be available for download in the following Zenodo repository: https://zenodo.org/record/2551455</p>
Interoperable data	Interoperability and metadata	The RDF data are fully interoperable according to the W3C Tabular Data and Metadata on the Web as outlined in the guidance for the tool table2qb used along this project.
Re-usable data	Data licensing	Publicly available data
	Quality assurance	<p>First, data is an excerpt from the Lithuanian Official Statistics Portal sources which is subject to their quality assurance processes.</p> <p>During the data cleansing process, well documented processes were run to preserve data quality.</p>
	Preservation time	Unlimited.
Allocation of resources	Cost estimation	OpenLink Virtuoso data storage engine is free subjected to the use of Insight Centre's server infrastructure. The last one is provided during the OGI project lifecycle. No future plans for maintenance are defined.
Secure Data	Data security	<p>Information about data security on the OpenLink Virtuoso platform can be found here: http://vos.openlinksw.com/owiki/wiki/VOS/VirtRDFGraphsSecurity</p> <p>RDF data has been backed up in an internal GitLab repository and stored as open statistical data in the mentioned engine. The persistence of the data is ensured by NUIG backup mechanisms.</p>

Resident population at the beginning of the year (persons by municipality)

Data	Purpose of data collection	Understanding the demographic distribution at the municipality
------	----------------------------	--

		level.
	Data type(s) and format	Data is presented in CSV format and then converted to RDF using specific scripts and tools developed for this purpose.
	Expected size	Raw CSV file size is 31 KB
	Origin of data	Data set was downloaded from the Official Statistics Portal https://osp.stat.gov.lt/
	Data utility	Data will be useful for citizens or investors aiming at opening new businesses in Lithuania as it will provide a clear understanding of the potential target market distribution.
Findable data	Data set reference and name	1.1.LT.EnterpriseLithuania.ResidentPopulation
	Metadata	Last update, Title.
Openly accessible data	Open access	<p>Data set will be openly available through two different mechanisms:</p> <ul style="list-style-type: none"> a) A central storage RDF endpoint. Knowledge in SPARQL query language will be required by the interested parties to access this data. URL: http://vmogi01.deri.ie:8000/spargl b) A GraphQL based API implementation. Data accessibility will be enhanced by providing access with a widely known standard such as JSON. <p>Data are also be available for download in the following Zenodo repository: https://zenodo.org/record/2551455</p>
Interoperable data	Interoperability and metadata	The RDF data are fully interoperable according to the W3C Tabular Data and Metadata on the Web as outlined in the guidance for the tool table2qb used along this project.
Re- use	Data licensing	Publicly available data

	Quality assurance	First, data is an excerpt from the Lithuanian Official Statistics Portal sources which is subject to their quality assurance processes. During the data cleansing process, well documented processes were run to preserve data quality.
	Preservation time	Unlimited.
Allocation of resources	Cost estimation	OpenLink Virtuoso data storage engine is free subjected to the use of Insight Centre's server infrastructure. The last one is provided during the OGI project lifecycle. No future plans for maintenance are defined.
Secure Data	Data security	Information about data security on the OpenLink Virtuoso platform can be found here: http://vos.openlinksw.com/owiki/wiki/VOS/VirtRDFGraphsSecurity RDF data has been backed up in an internal GitLab repository and stored as open statistical data in the mentioned engine. The persistence of the data is ensured by NUIG backup mechanisms.

TUT data sets

Interviews' and online survey's data

Data Summary	Purpose of data collection	The data set was generated by TUT in the framework of Tasks 1.2 and 1.4 of WP1. The data was collected through e-mail interviews and online survey involving key actors in the field of open data and public sector innovation in the six pilot countries: Belgium, Greece, Estonia, Ireland, Lithuania and the UK. The respondents were chosen to represent public administrations as well as relevant business and civil society actors, based on the suggestions of OGI pilot partners in each country. The data will be used in an anonymized and generalized form in a summary report of the survey results in deliverable D1.1. The results may potentially also be used in this form in related scientific publications.
	Data type(s) and format	Data is stored in XLSX and CSV formats
	Expected size	82kb

	Origin of data	<p>The questionnaire consisted of 11 questions of varying length and depth. The questionnaire was disseminated to 120 respondents and returned 63 responses.</p> <p>The collected data is qualitative and involves the respondents' experience with open data-driven co-creation, their opinions on issues such as drivers and barriers to open data-driven co-creation, the capacities and needs of the organisations involved, good examples of policies that have been used to encourage open data innovation, etc.</p>
	Data utility	In addition to the OGI project, the generalized results could be potentially interesting for researchers, policy-makers, different types of users and providers of open data. There is also potential value in comparing the data with previous studies on open data innovation.
Findable data	Data set reference and name	4.3.TUT.InterviewsSurveyDataOGI
	Metadata	N/A
Openly accessible data	Open access	The original, identifiable data will be stored at the Tallinn University of Technology and will not be shared with any other parties outside the research team due to confidentiality reasons and GDPR requirements. Aggregated data may potentially be released, but a written request must be sent to the research team which will then be analysed by the Estonian ethics team.
Interoperable data	Interoperability and metadata	N/A
Re-usable data	Data licensing	The data set is not for public release.
	Quality assurance	N/A
	Preservation time	The data should be usable for at least five years after the end of the project date.

Allocation of resources	Cost estimation	N/A
Secure Data	Data security	<p>The original data will be stored by the Tallinn University of Technology in a secure server in Estonia, which only the research team has access to. The key allowing re-identification will be stored in an encrypted form only until necessary for the analysis, and will be deleted after the project's end.</p> <p>The anonymized data will be stored on a secure server of the Tallinn University of Technology up to 5 years after the project's end date.</p>