



## D3.7 VISUALISATIONS, DASHBOARDS AND MULTILINGUAL INTERFACE STATUS REPORT 2

### PROJECT

Acronym: **OpenDataMonitor**  
Title: Monitoring, Analysis and Visualisation of Open Data Catalogues, Hubs and Repositories  
Coordinator: SYNYO GmbH

Reference: **611988**  
Type: Collaborative project  
Programme: FP7-ICT

Start: November 2013  
Duration: 24 months

Website: <http://project.opendatamonitor.eu>  
E-Mail: [office@opendatamonitor.eu](mailto:office@opendatamonitor.eu)

Consortium: **SYNYO GmbH**, Research & Development Department, Austria, (SYNYO)  
**Open Data Institute**, Research Department, UK, (ODI)  
**Athena Research and Innovation Center**, IMIS, Greece, (ATHENA)  
**University of Southampton**, Web and Internet Science Group, UK, (SOTON)  
**Potsdam eGovernment Competence Center**, Research Department, Germany, (IFG.CC)  
**City of Munich**, Department of Labor and Economic Development, Germany, (MUNICH)  
**Entidad Publica Empresarial Red.es**, Shared Service Department, Spain, (RED.ES)

## DELIVERABLE

Number:	<b>D3.7</b>
Title:	<b>Visualisations, dashboards and multilingual interface status report 2</b>
Lead beneficiary:	SYNYO
Work package:	WP3: Concept Design and Software Development
Dissemination level:	Public (PU)
Nature:	Report (R)
Due date:	August 31, 2015
Submission date:	August 31, 2015
Authors:	<b>Ejona Sauli, SYNYO</b> <b>Peter Leitner, SYNYO</b>
Contributors:	<b>Vassilis Kaffes, ATHENA</b> <b>Dimitris Skoutas, ATHENA</b> <b>Johannes Burger, SYNYO</b> <b>Stefan Ruscher, SYNYO</b>
Reviewers:	<b>Bernhard Jäger, SYNYO</b>

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

**Disclaimer:** The content of this publication is the sole responsibility of the authors, and in no way represents the view of the European Commission or its services.

## TABLE OF CONTENTS

<b>1. Introduction .....</b>	<b>6</b>
<b>1.1. Purpose .....</b>	<b>6</b>
<b>1.2. Scope .....</b>	<b>6</b>
<b>1.3. References .....</b>	<b>7</b>
<b>2. Architecture overview of the ODM platform .....</b>	<b>9</b>
<b>3. Development Environment .....</b>	<b>12</b>
<b>4. Implementation Status .....</b>	<b>13</b>
<b>4.1. Data analysis.....</b>	<b>13</b>
<b>4.2. Methodology .....</b>	<b>13</b>
<b>4.3. Download Center .....</b>	<b>14</b>
<b>4.4. Developer API .....</b>	<b>14</b>
<b>4.5. Next Steps.....</b>	<b>16</b>
<b>5. Graphic User Interface Status Report .....</b>	<b>17</b>
<b>5.1. European Dashboard .....</b>	<b>17</b>
<b>5.2. Catalogues harvested.....</b>	<b>19</b>
<b>5.3. Distribution Size.....</b>	<b>20</b>
<b>5.4. Open licenses.....</b>	<b>20</b>
<b>5.5. Non-proprietary formats.....</b>	<b>21</b>
<b>5.6. Machine readable .....</b>	<b>22</b>
<b>5.7. Complete core metadata.....</b>	<b>22</b>
<b>5.8. Overall quality measure .....</b>	<b>23</b>
<b>5.9. Country Dashboard .....</b>	<b>24</b>
<b>5.10. Catalogue Dashboard .....</b>	<b>26</b>

<b>5.11.</b>	<b>Catalogue Listview .....</b>	<b>28</b>
<b>5.12.</b>	<b>Advanced Search .....</b>	<b>28</b>
<b>5.13.</b>	<b>Alert monitor .....</b>	<b>29</b>
<b>5.14.</b>	<b>Dataset Profile .....</b>	<b>30</b>
<b>5.15.</b>	<b>Internationalization (i18n).....</b>	<b>31</b>
<b>6.</b>	<b>Conclusion .....</b>	<b>32</b>
	<b>Appendix I. Glossary of acronyms and abbreviations .....</b>	<b>33</b>
	<b>Appendix II. Metrics explained .....</b>	<b>34</b>

## LIST OF FIGURES

Figure 1: Overview of architecture and processing workflow .....	9
Figure 2: OpenDataMonitor methodology explained .....	13
Figure 3: Download Center interface providing the end users with CSV data format.....	14
Figure 4: Developer API interface proving the end users with JSON data .....	15
Figure 5: Example of an API.....	15
Figure 6: European Dashboard User Interface.....	18
Figure 7: European vectorial map displaying catalogues harvested (catfreq).....	19
Figure 8: Discover and contribute into the Resources collection list.....	19
Figure 9: European vectorial map displaying distribution size (catdatasizetotal) .....	20
Figure 10: European vectorial map displaying open licenses datasets (catopenlicfreq) .....	20
Figure 11: European vectorial map displaying non-proprietary distribution formats (catnonproprformatfreq) .....	21
Figure 12: European vectorial map displaying machine readable distributions (catmachinereadformatfreq) .....	22
Figure 13: European vectorial map displaying complete core metadata (catcoremetadafreq).....	23
Figure 14: Overall quality score through countries.....	23
Figure 15: Navigation to the Country Dashboard .....	24
Figure 16: Country dashboard interface .....	25
Figure 17: Catalogue performance in a glance.....	26
Figure 18: Catalogue dashboard user interface .....	27
Figure 19: Catalogue listview user interface .....	28
Figure 20: Advanced search for filtering data catalogues and datasets .....	29
Figure 21: Top menu generic search .....	29
Figure 22: Alert monitor user interface.....	30
Figure 23: Dataset Profile .....	30
Figure 24: An example of a translated user interface in German .....	31

## LIST OF TABLES

Table 1: References to other resources .....	7
Table 2: Development environment specifications for the demonstration site.....	12
Table 3: List of implemented metrics.....	34

## 1. Introduction

### 1.1. Purpose

The purpose of this document is to provide a second status report on the implementation of OpenDataMonitor end-user interfaces including planned dashboards, visualisation elements and other features. This report will follow up the recent implementation achievements. The target audience of this report include project partners, the European Commission and relevant audiences with technical knowledge of open data and software development concepts.

The work being undertaken in Work Package 3, consist of the actual implementation of the platform and related supporting services and products. The technical implementation of the platform and supporting services will be extended throughout the two project years, while producing the first prototype after the first project year. This report will provide insights on the implementation status of the Presentation layer of the platform in the last development phase. The given report will take for granted the implementation of Metadata Harvester, Metadata Repository, Harmonisation Engine, Catalogue Registry and Job Manager. These components are defined in detail in D3.3 and D3.6.

### 1.2. Scope

The main objective of this project is to create sophisticated methods to monitor data catalogues, harvest and analyse the metadata of the datasets published, and provide comprehensive visualisations to compare existing open data resources. Using standardised APIs (e.g. CKAN2), it will be possible to analyse data usage, file formats, updates, licenses and further metadata to statistically describe and visualise it. This information will be used to identify trends, gaps and potentials of open data resources, and also to build a scalable open data monitoring concept using metadata, parameters and key-indicators.

The end product of this project mainly consists of an open data platform that will collect and harmonize the metadata of existing open data. The unique values of this project in comparison to other existing open data platforms include:

- Aggregate metadata of open datasets from all European countries, mainly member states, and build dashboards upon the metadata harvested.
- Creation of graphical and user friendly dashboard based on main level of data abstraction.
- Automatically harvest metadata of open data published by existing data catalogues either from published APIs or via scraping techniques.
- Use state of the art visualization user interface elements to better represent the metadata and their specific context in order to address the target audiences accordingly.

### 1.3. References

This report is based on the previous business logic and system designs documented in the previous deliverables. Additionally, references to official online resources are being used to support the concept and terms used in this document. The main references to the other project’s deliverables are listed in the following table.

**Table 1: References to other resources**

Deliverable	Description
D2.2 Monitoring methods, architectures and standards analysis report <i>July 2014</i>	Deliverable 3.2 makes use of the analysis results of D2.2 to proper define the use cases, related business requirements and non-functional requirements related to performance and software quality. The work was conducted in parallel, in order to obtain consistency of technical concepts and decisions made.
D2.3 Best practice visualisation, dashboard and key figures report <i>July 2014</i>	This report was used as a base input for conceptualizing and creating functionalities, user interfaces and respective use cases. The full report was taken into account to make proper decisions on the business requirements, specifically defining the software interfaces between different architecture layers. The work undertaken for D2.3 was conducted in parallel with the work for the present deliverable, in order to obtain consistency of project business logic and best cases to represent data.
D2.4 Open data stakeholder requirement report 1 <i>September 2014</i>	The output of the research conducted for this deliverable has been synchronised with the business requirements of D3.2. The synchronisation required is very essential in order to keep the business logic consistent.
D2.5 Open data resources, platforms and APIs collection 1 <i>September 2014</i>	This deliverable provides a list of, and detailed data on, the most significant open data resources we consider important during the ODM tool development process.
D2.6 Open Data Stakeholder Requirement Report 2 <i>March 2015</i>	This report explores and analyses the roles, interests and understandings of the various stakeholders in the open data ecosystem.
D2.7 Open data resources, platforms and APIs collection 2 <i>June 2015</i>	This document is the second iteration for task 2.5 and summarizes the further efforts built upon D2.5. It holds all relevant information on open data resources, platforms and APIs that have to be implemented in WP3 for a full functional tool prototype. This deliverable provides a list of, and detailed data on, the most

	significant open data resources we consider during the ODM platform development process.
D3.1 Scalable open data monitoring concept and framework design <i>May 2014</i>	In this deliverable the platform architecture has been defined, and this design and technical planning is used in D3.2 to further analyse and specify requirements and interfaces.
D3.2 Tool specifications, use cases, mockups and functionalities status report 1 <i>July 2014</i>	D3.2 is a complementary document to the current one, focusing rather at the system level and the product perspective, describing overall requirements and functionalities, user roles and interfaces, while the current document details the implementation status of the visualisation of dashboards and interfaces.
D3.3 Tool architecture and components/plugins programming status report 1 <i>October 2014</i>	The purpose of this document is to present the overall architecture of the ODM platform, and to report the status of the implementation for each of the main components involved
D3.4 Visualisations, dashboards and multilingual interface status report 1 <i>October 2014</i>	This report provides an initial status report on the implementation of Open Data Monitor end-user interfaces including planned dashboards, visualisation elements, reporting capabilities and other basic features.
D3.5 Tool specifications, use cases, mockups and functionalities status report 2 <i>April 2015</i>	This deliverable presents a status report of the second iteration of business analysis and system design phases. In this report are been described the approach used for system analysis, challenges and lessons learned from previous mockups and designs.



## 2. Architecture overview of the ODM platform

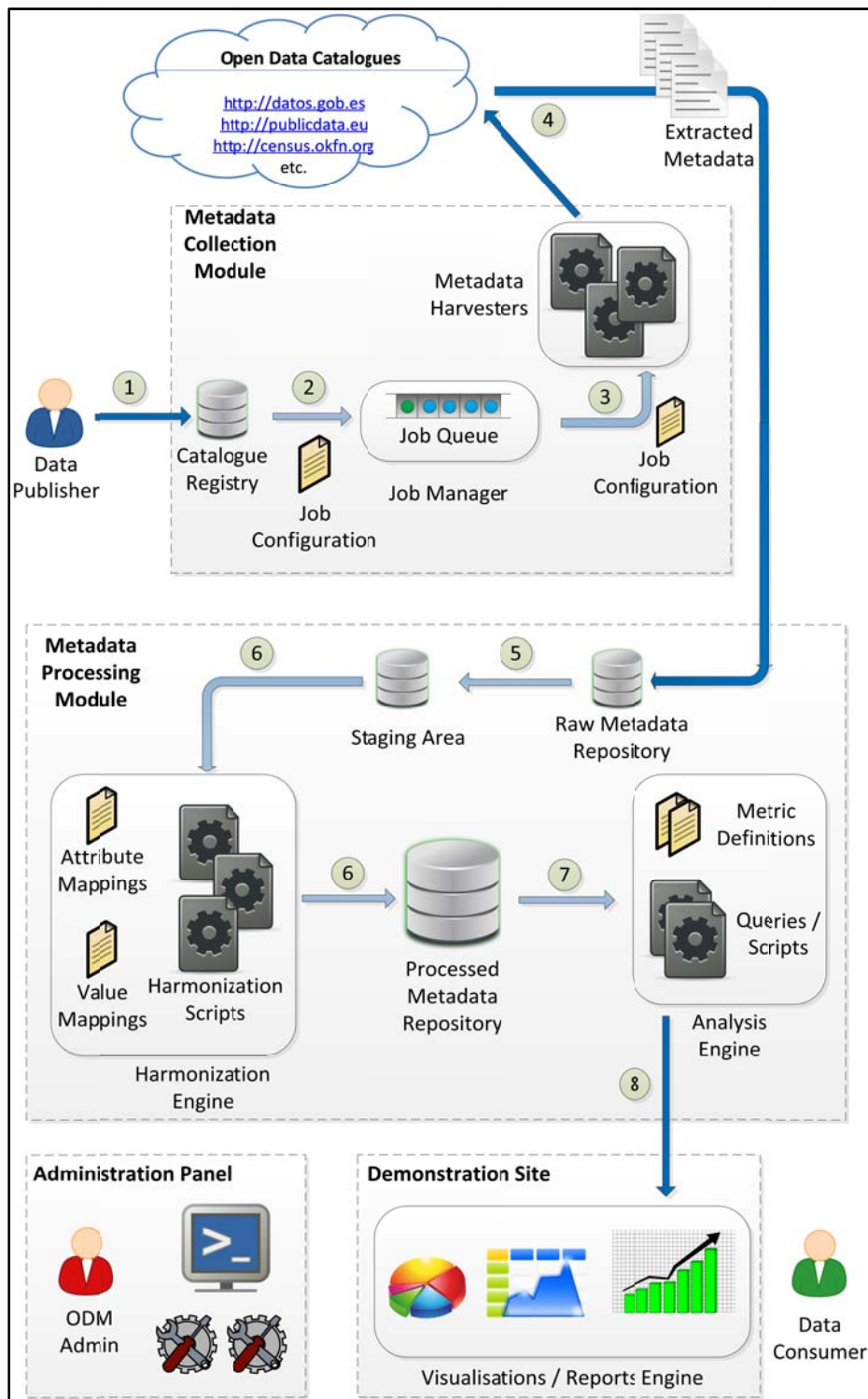


Figure 1: Overview of architecture and processing workflow.

Figure 1 shows an overview of the ODM platform architecture, presenting the main components and illustrating the metadata processing workflow, as was initially designed in the beginning of the

project and presented in Deliverable D3.3. The progress of the work throughout the second year of the project has adhered to this architecture, without any major changes or deviations.

Below, we briefly outline the main modules and components comprising the ODM platform:

- **Metadata collection module.** This module is responsible for collecting metadata from a list of registered open data catalogues. It consists of the following main components:
  - **Catalogue registry.** It allows the registration of catalogues for harvesting and monitoring. Registration is done via a Web-based User Interface (UI), where a form is completed with basic information about the registered catalogue, as well as some additional information that is needed in order to setup and configure the respective harvesting process for this catalogue (See section 2 for detailed information). This information provided during the registration step forms the catalogue profile and is stored in the catalogue registry.
  - **Job manager.** The Job Manager schedules the execution of harvesting jobs, periodically or on demand, and is responsible for monitoring their process and reporting the status of execution. A harvesting job is a task that collects metadata from a registered open data catalogue. It provides the required configuration that drives the harvesting process (e.g., which harvester to use and a set of metadata extraction rules to be applied). Harvesting jobs are maintained in a queue and are scheduled for processing.
  - **Metadata harvesters.** These are scripts executed by harvesting jobs in order to perform the actual extraction of metadata from the respective catalogue. Different harvesters are implemented and used to address the different open data platforms and APIs that exist. The configuration included in the harvesting job specifies which harvester should be used and how.
- **Metadata processing module.** This module performs the cleaning, integration and analysis of the metadata that are extracted from the various catalogues that are being monitored. It consists of the following main components:
  - **Harmonisation engine.** It processes the raw, original metadata that were retrieved by the harvesters and performs cleaning and integration tasks required to obtain a homogenized dataset in terms of both attribute names and attribute values.
  - **Analysis engine.** Once the collected metadata have been mapped to a consistent internal schema and representation, the analysis engine performs the required operations (e.g. aggregations) in order to compute the metrics that have been defined for monitoring (Refer to D3.7 for detailed information). It also makes these results available to the demonstration site for visualisation and presentation to the end users.
- **Demonstration site.** This module comprises several components for generating intuitive visualisations and reports that are presented to the end users, allowing them to obtain a comprehensive overview of trends in the evolving open data landscape, based on the

monitored open data catalogues. This module intends to add a missing layer of the open data in Europe. Through four layers of aggregation, the module will allow all kind of stakeholders to find their purposes and aims of open data. After the data is analysed and processed, the demonstration site will interpret these findings through intuitive visualisation like maps, dynamic grids, multi bar table, parallel coordinates, donuts charts, bubble charts, series and bar charts and so on,

- **Administration panel.** This module comprises a set of dashboards that allows the ODM system administrator to monitor, control and configure various aspects of the system's operation (e.g., configure options for metadata collection, monitor the status of harvesting jobs, define rules for metadata harmonisation, specify templates for visualisations).

The second iteration of visualisations, dashboards and multilingual interfaces will present to the reader the last impressions of the platform. The report will go through all the main sections and themes of the site while also giving a status update for the implementation. Based on the core architecture, this report will focus mainly on the demonstration site module and the analysis engine.

### 3. Development Environment

The technical team is distributed in Greece and Austria. Based on the core architecture (see figure 1) and the competences of each partner involved the work was initially planned and documented using internal development plans. The communication between the demonstration site with the rest of the processing modules is been done through RESTful API. Hence, except for the communication protocol for API business logic and modelling it wasn't necessary to deploy any software integration. The processing modules and the demonstration site were kept separate in order to offer flexibility for future open data initiatives.

The development environment for the demonstration site is divided in three instances (LIVE: opendatamonitor.eu, DEV: demo.opendatamonitor.eu and TEST: at SYNYO's intranet infrastructure). The database received constant updates through scheduled jobs from the processing units. Except for data synchronisation, there is also another job built for time snapshotting and report generation.

**Table 2: Development environment specifications for the demonstration site**

Resource Type	Name	Version	License
<b>HTTP server</b>	Apache	2.4	Apache 2.0 License
<b>Programing Language</b>	PHP	5	PHP License v3.01,
<b>Database</b>	MYSQL	5.5.36	MySQL Community Edition
<b>PHP Framework</b>	Yii Framwork	2.0.0	GNU Free Documentation License (GFDL)
<b>HTML, CSS &amp; JS Framework</b>	Bootstrap	3.3.0	MIT License
<b>JavaScript Library</b>	D3.js	3.4.13	Open Source
<b>JavaScript Library</b>	jQuery	2.1.1	MIT License
<b>Translator Library for raster and vector geospatial data formats</b>	gdal	1.11	X/MIT
<b>GeoJson extension</b>	TopoJson	1.1	Open Source
<b>Vectorial JS Map</b>	jQuery Mapael	1.0.1	MIT
<b>Basic charts based on JS</b>	Morris Js	0.5.1	All rights reserved
<b>Donut graph type</b>	jQuery Plugin Circliful	0.1.5	MIT
<b>Programing Language</b>	python	2.7	PSF LICENSE
<b>Operating System</b>	Linux Fedora	14	LICENSE AGREEMENT FEDORA(TM) 14
<b>Version Control Software</b>	Bitbucket	101	Atlassian Customer Agreement
<b>Browser Extension</b>	Firebug	1.4	BSD License

## 4. Implementation Status

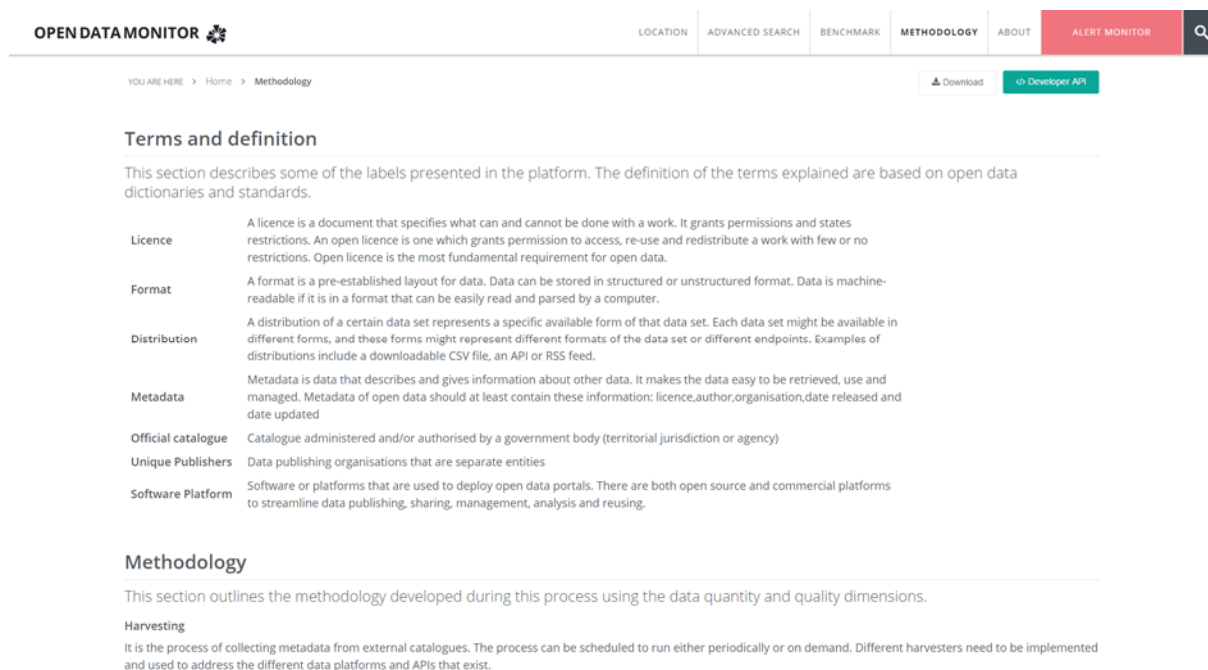
### 4.1. Data analysis

As a result of a collaborative work, **263** data catalogues have been identified by the consortium members and the community which has followed our work. From **263** data catalogues, only **153** were able to be harvested. From the catalogues harvested and harmonized, there are **228,244** datasets and **515,406** distributions found with the size of **1,173GB**. In the moment of the capture, the active (harmonised) catalogues represent **24** European Countries.

### 4.2. Methodology

The project from the aspect of data analysis has considered previous scientific and practical work conducted in the field and developed genuine metrics and open data indicators to set up another baseline. The collection of terms and definitions, in addition with metrics mentioned in the platform is explained at the “Methodology” page accessed from the top menu.

One of the most required functionality from the various stakeholders of open data and on-going projects in the field was the ability to download or use the data harmonised and processed from OpenDataMonitor. The application is providing the data to developers or other users through the Download Center and The Developer API in the CSV and JSON formats. The navigation to the downloading and APIs can be done through the sub menu displayed on the right of the interface.



The screenshot shows the 'Methodology' page on the OpenDataMonitor website. The page has a navigation bar at the top with 'OPEN DATAMONITOR' and a search icon on the left, and 'LOCATION', 'ADVANCED SEARCH', 'BENCHMARK', 'METHODOLOGY', 'ABOUT', 'ALERT MONITOR', and a search icon on the right. Below the navigation bar, there is a breadcrumb trail: 'YOU ARE HERE > Home > Methodology'. There are two buttons: 'Download' and 'Developer API'. The main content area is titled 'Terms and definition' and contains a paragraph: 'This section describes some of the labels presented in the platform. The definition of the terms explained are based on open data dictionaries and standards.' Below this is a table with the following content:

Licence	A licence is a document that specifies what can and cannot be done with a work. It grants permissions and states restrictions. An open licence is one which grants permission to access, re-use and redistribute a work with few or no restrictions. Open licence is the most fundamental requirement for open data.
Format	A format is a pre-established layout for data. Data can be stored in structured or unstructured format. Data is machine-readable if it is in a format that can be easily read and parsed by a computer.
Distribution	A distribution of a certain data set represents a specific available form of that data set. Each data set might be available in different forms, and these forms might represent different formats of the data set or different endpoints. Examples of distributions include a downloadable CSV file, an API or RSS feed.
Metadata	Metadata is data that describes and gives information about other data. It makes the data easy to be retrieved, use and managed. Metadata of open data should at least contain these information: licence,author,organisation,date released and date updated
Official catalogue	Catalogue administered and/or authorised by a government body (territorial jurisdiction or agency)
Unique Publishers	Data publishing organisations that are separate entities
Software Platform	Software or platforms that are used to deploy open data portals. There are both open source and commercial platforms to streamline data publishing, sharing, management, analysis and reusing.

Below the table is another section titled 'Methodology' with a paragraph: 'This section outlines the methodology developed during this process using the data quantity and quality dimensions.' Underneath is a sub-section 'Harvesting' with a paragraph: 'It is the process of collecting metadata from external catalogues. The process can be scheduled to run either periodically or on demand. Different harvesters need to be implemented and used to address the different data platforms and APIs that exist.'

Figure 2: OpenDataMonitor methodology explained

### 4.3. Download Center

The “Download” page is found at the methodology page. The data is been distributed under the Creative Commons Attribution 4.0 International License<sup>1</sup>. The end user has the option to select the time period to download the preferred data. The data is provided so far in the CSV format. The end user can download the files individually for each country or as a compressed file.

The data made available for download include the list of all European countries with the respective values: total dataset number, open licenses, machine readable, accessibility, complete metadata and overall quality metric. At the Country level, the data made available include all the countries covered from the project. For each country, the list of all catalogues harvested and harmonised is produced with the defined fields like the catalogue url, catalogue name, total number of datasets per catalogue, total number of distributions, total distribution size(MB), unique publishers, open licences (%), non-proprietary formats (%), machine readable (%), and complete core metadata (%).

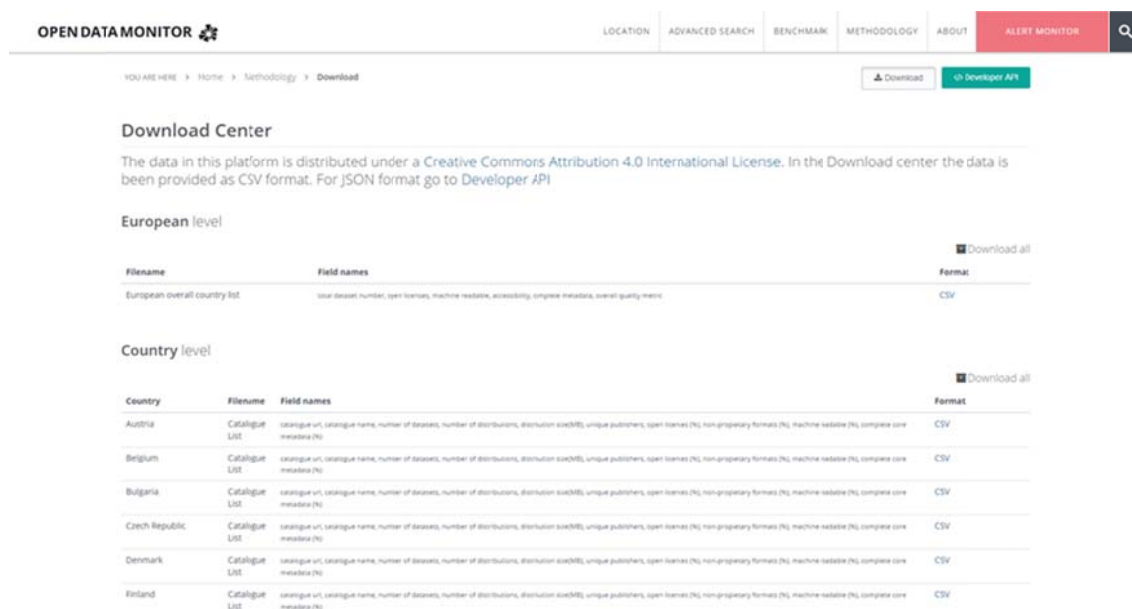


Figure 3: Download Center interface providing the end users with CSV data format.

### 4.4. Developer API

To provide an efficient way to all the application users and developers, OpenDataMonitor is publishing all the relevant APIs used by the demonstration platform and double checked by the data analysts and testers. Each API is identified with a tag name which is then added at the end of the API URL. The API URL is the same for each application interface

`http://83.212.122.164:27081/api/v1.0/_`

The table which lists all the APIs is providing the end user with additional information for each metric like label, description, metadata outputted, metric category and the accessible URL. See Appendix II for the full list of implanted metrics.

<sup>1</sup> <http://creativecommons.org/licenses/by/4.0/>

OPEN DATAMONITOR  LOCATION ADVANCED SEARCH BENCHMARK METHODOLOGY ABOUT ALERT MONITOR 

YOU ARE HERE > Home > Methodology > API Download [Developer API](#)

### Developer APIs

The data in this platform is distributed under a Creative Commons Attribution 4.0 International License. In the Developer API page the data is been provided as JSON format. For CSV format go to Download Center

*For a structured view of the APIs it is required the installation of a JSON viewer plugin compatible with your client browser.*

Tag	Label	Description	Metadata	Category	Data format
catdbylicensfreq	Frequency of distributions by license type	Count distinct licenses per catalogue	licenses_id	Per catalogue per selector	JSON
cataccessabilityfreq	Accessibility	Defined set of fields as requirements - description (notes in datasets): 1 if contains some text, 0 null/empty - not a broken link: 1 if not broken - author email/maintainer_email (only in CKAN): 1 if present sum(description && not_broken && author	count_is, tot_count	Quality	JSON
catcoremetadafreq	Complete core metadata per catalogue	Average across a defined set of fields, eg License, Author/Maintainer (maintainer field exists only in CKAN), Organisation, Date released or Date updated	count_is, tot_count	Quality	JSON
catmachinereadformatfreq	Machine-readable in % per catalogue	total count of MR datasets/total count of datasets	count_is, tot_count	Quality	JSON
catnonpropformatfreq	Non-proprietary formats per catalogue	total count of distributions with a non-proprietary format/total count of distributions with a format	count_is, tot_count	Quality	JSON
catopenlicfreq	Open licences in % per catalogue	total count of open licences / total count of distributions with a licence	licenses[], count_is, tot_count	Quality	JSON
catsitepagerank	Discoverability	page ranking, based on google ranking / alexa ranking	pagerank_id	Quality	JSON
edcoremetadafreq	Complete core metadata	Average across a defined set of fields, eg License, Author/Maintainer (maintainer field exists only in CKAN), Organisation, Date released or Date updated	freq	Quality	JSON
edmachinereadformatfreq	Machine-readable in %	total count of MR datasets / total count of datasets	freq	Quality	JSON
ednonpropformatfreq	Non-proprietary formats	total count of distributions with a non-proprietary format/total count of distributions with a format	freq	Quality	JSON
edopenlicfreq	Open licences in %	total count of open licences / total count of distributions with a licence	licenses[]	Quality	JSON

Figure 4: Developer API interface proving the end users with JSON data

```
83.212.122.164:27081/api/v1.0/_edopenlicfreq

{
  ok: 1,
  result: [
    {
      licenses: {
        Data licence Germany - attribution - version 2.0: 0.19484439572573323,
        CC BY: 0.1777217245890359,
        OGL: 0.13011429783046155,
        CC0: 0.042116570184948854,
        ODC-ODbL: 0.025704009825532784,
        CC BY-SA: 0.010577650113417693,
        ODC-BY: 0.004272666535979613,
        PDDL: 0.0032205023983741463,
        Against DRM: 0.0006160961109933149,
        GNU FDL: 0.00021603370125739614,
        DL-DE-BY-2.0: 0.00015802465184568793,
        ODC-PDDL: 0.00007401154580114498,
        W3C license: 0.00002000312048679594,
        SPL: 0.000002000312048679594
      },
      freq: 0.5896399838374786
    }
  ]
}
```

Figure 5: Example of an API

## 4.5. Next Steps

During the last project months the project will continue to maintain the synchronisation process and support the mismatching data mappings or application errors. A special attention will be placed into the following matters:

- The final implementation of the Benchmarking of the Country and Catalogue dashboards.
- The further social media integration and the embed functionalities.
- The final implementation of Multi-lingual interfaces.
- The optimization of some specific visualisations.
- Optimization of the backend and code.
- Format and comment crucial parts in the code in order to make it more comprehensive for the community.



## 5. Graphic User Interface Status Report

The following section shows and describes the current status of the OpenDataMonitor demonstration site's user interfaces. The design and type of graphs have progressed and changed since the previous report (D3.4). The technical team after a thorough prioritisation of the major and the most relevant metrics came to the decision of optimising the user experience. The new interfaces presented at the usability workshops received positive feedback and created a baseline for adding more functionalities to the final application.

### 5.1. European Dashboard

The European Dashboard is the highest aggregation of open data presented to the end user at the Home page or through the "Location" menu item. The main visualisation chosen to display the quantity and quality of open data in Europe is a vectorial map. The quantity and quality metrics are been defined and rated in between the project team as the most relevant, interesting to the end user (scientific, researchers, experts, beginners and others) and missing in the open data landscape.

On the left sidebar of the map container a top level aggregation is displayed giving the current status of open data in the dimensions; total number of data catalogues harvested and harmonised, total distribution size in Europe, the percentage of open licenses in Europe, the percentage of non-proprietary distributions formats, the percentage of machine readable distributions, and the percentage of complete core metadata in each dataset harvested.

Rather than the top level aggregation of the open data, in the European level a country distribution of metrics is been presented in order to show the geographic coverage of this project and an overview of the current open data status in the country level. The metrics are been visualised through colourful bubbles in the map. The legend of the map places horizontally the list of combined quantity and quality metrics using a radio button effect. On the selection of one circle or one metric label in the legend the bubbles distribution will change while the grey area (the coverage area) will remain the same. The list of metrics in the legend will give the end user the choice between number of catalogues harvested, distribution size, open licenses (default on load of the map), non-proprietary formats, machine readable or complete core metadata.

A need for the development of an overall quality score to rate and rank the countries and catalogues harvested was presented into the project. OpenDataMonitor calculates this score as an average of four fundamentals metrics including Open licenses, Machine readable, Accessibility and Complete core metadata. A brief description of each metric involved into the overall quality score is shown on the European Dashboard. Except for the map, a list of all countries included into the project through their open data catalogues is displayed while offering sorting functionalities. In order to raise the comprehension of the quality score, the size of Country's data is added giving the total number of datasets per country. While comparing the size and quality of each country, a fair judgement of country ranking can be done. More importantly, the end user can download the data presented in the country list as shown in the figure 6.

From the European Dashboard, one can navigate to the Country Dashboard through clicking on a country area from the map or clicking on one country name from the list below.

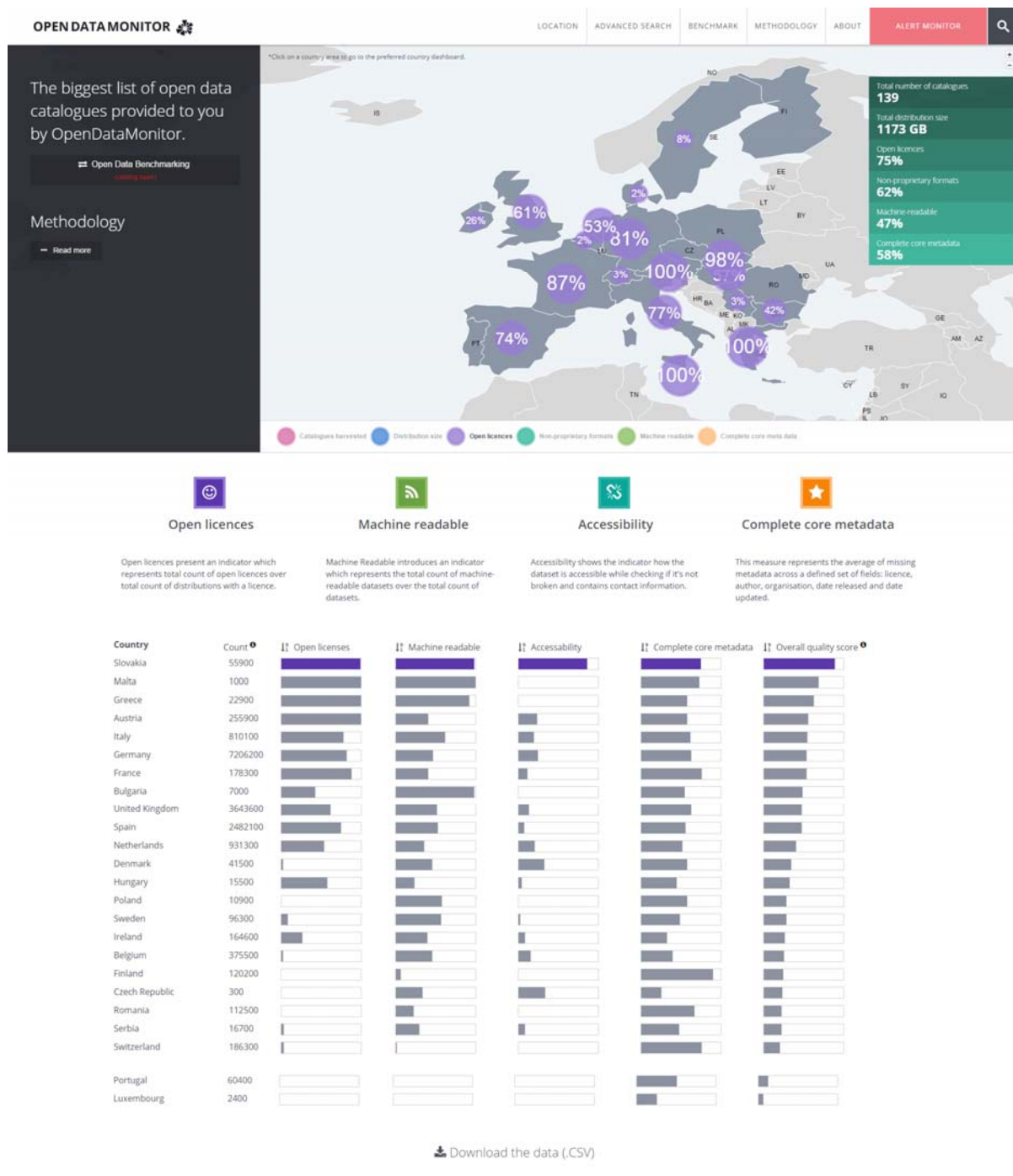


Figure 6: European Dashboard User Interface

## 5.2. Catalogues harvested

The project team dedicated resources and efforts to collect existing open data catalogues throughout Europe. From the catalogue collection only some of them were able to be harvested correctly. For the catalogues not able to be harvested for various reasons like deep hierarchy to the dataset level, access to dataset denied, cannot locate dataset and so on, a list is also presented into the “Advanced search” menu item. From the catalogues able to be harvested the platform stores two statuses: harvested and in process of harmonization, and harmonised with inclusion in all the API calculations.

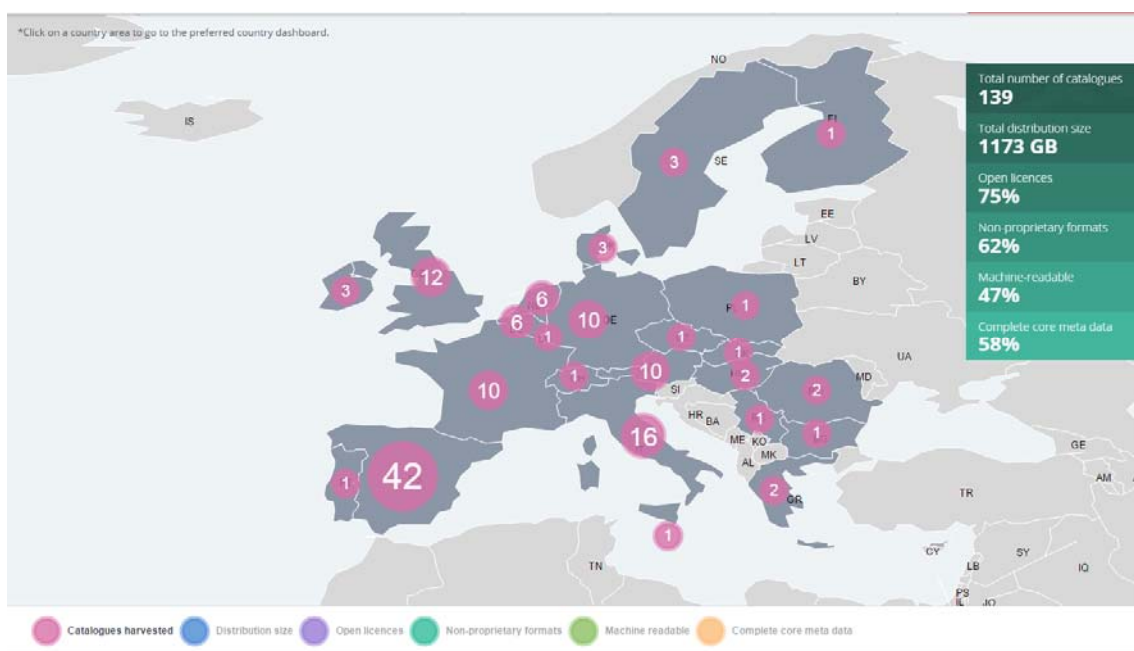


Figure 7: European vectorial map displaying catalogues harvested (catfreq)

In the vectorial map at the European Dashboard, once clicking the “Catalogue harvested” legend item the end user can have an overview of the catalogues harvested or harmonised. A closer look of the status for each catalogue can be received at the “Advanced search” menu item. A bigger number of catalogues harvested does not mean the bigger the size of a country or better quality of that country. The collection of all resources found was published on the platform and project website (project.opendatamonitor.eu) and promoted through the twitter channel while giving all the interested parties the chance to send their catalogue for harvesting and monitoring it through the platform. The project has considered harvesting catalogues from open data platforms like CKAN, Socratas and using a HTML scrapping method (further described in D3.6).

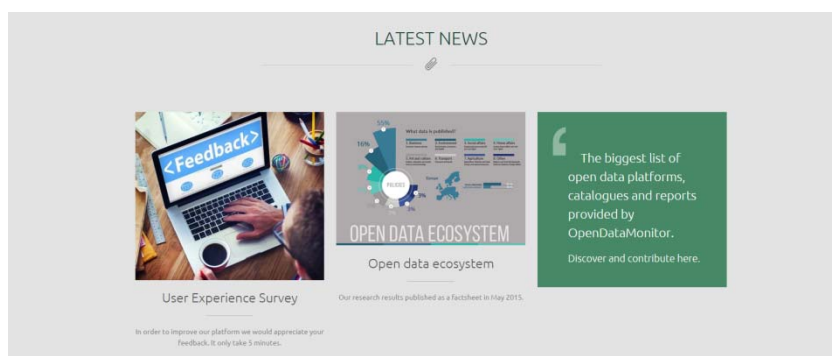


Figure 8: Discover and contribute into the Resources collection list

### 5.3. Distribution Size

The second quantity metric displayed into the vectorial map is the distribution size of each country. This metric measures the sum of distributions for each country in kb and for visualisation purposes the bubbles are been formatted into GB or MB.

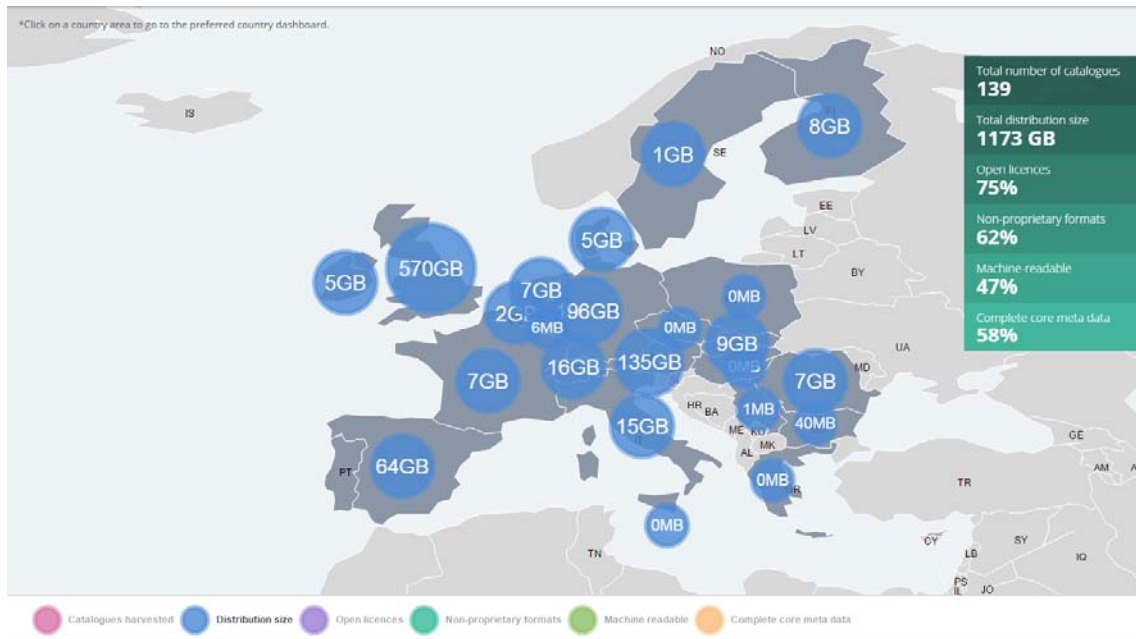


Figure 9: European vectorial map displaying distribution size (catdatasizetotal)

### 5.4. Open licenses

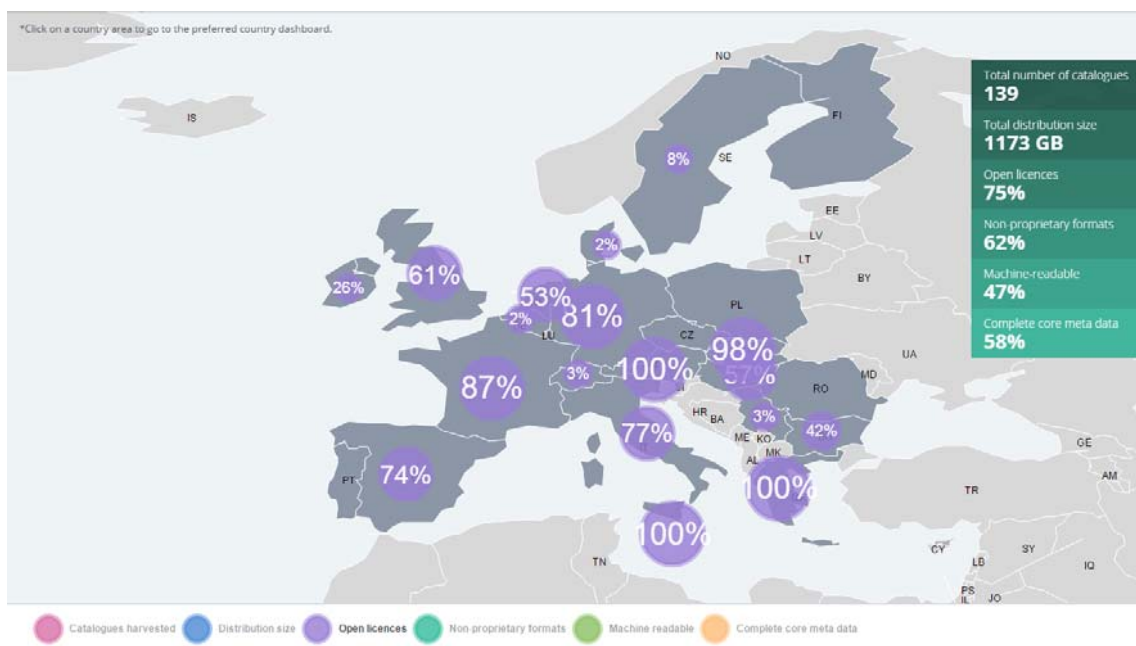


Figure 10: European vectorial map displaying open licenses datasets (catopenlicfreq)

The license information is the key to identifying the openness of a catalogue or country. The project team developed a license map which was used during the harmonisation process to filter and refine all the raw data. The license mapping is been distributed through the project’s github<sup>2</sup> account and the process is described in details at the D3.6<sup>3</sup>. The mapping assigns a value to each dataset of opened or not opened accordingly to the schema defined.

The definition of open licenses is based on the division of total count of open licenses with the total count of distributions having a license grouped by countries.

## 5.5. Non-proprietary formats

The non-proprietary format is based on the 5 star deployment schema<sup>4</sup>. These formats include any structured data made available. Based on the 5 star schema any distribution aligned with 2 stars is a grouped under non-proprietary formats. The percentage of non-proprietary formats is based on the division of the total count of distributions with the total count of distributions having a format grouped by countries.

It is noticed that some of the countries covered from the project do not provide data for this metric. It might happen that the catalogues harvested have no or very few distributions, or additionally that the formats read are non-proprietary.

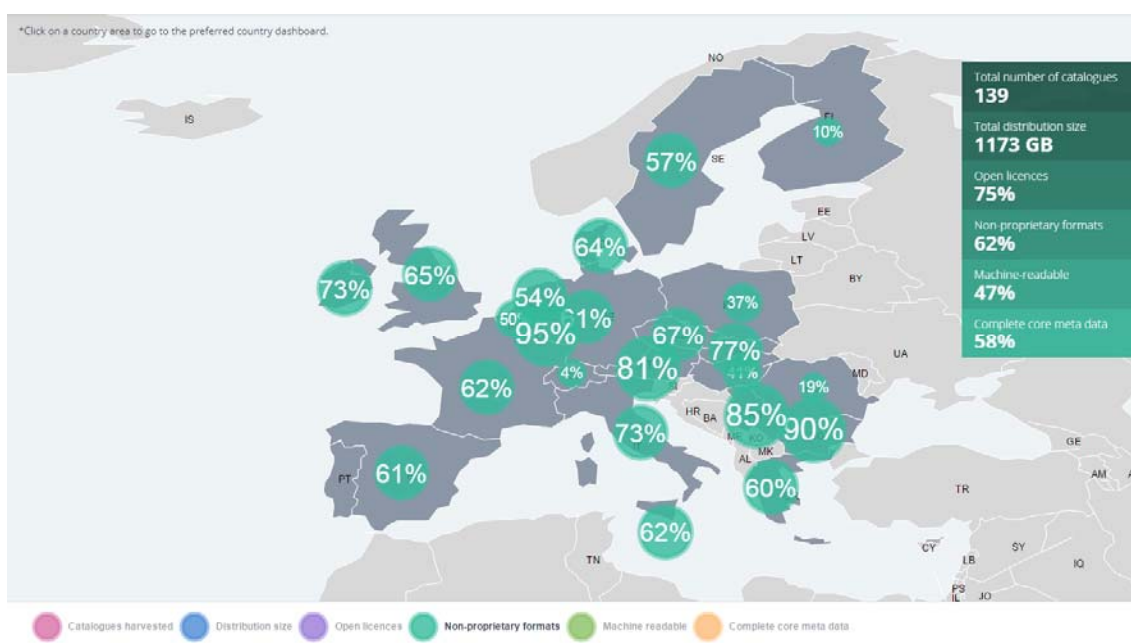


Figure 11: European vectorial map displaying non-proprietary distribution formats (catnonpropformatfreq)

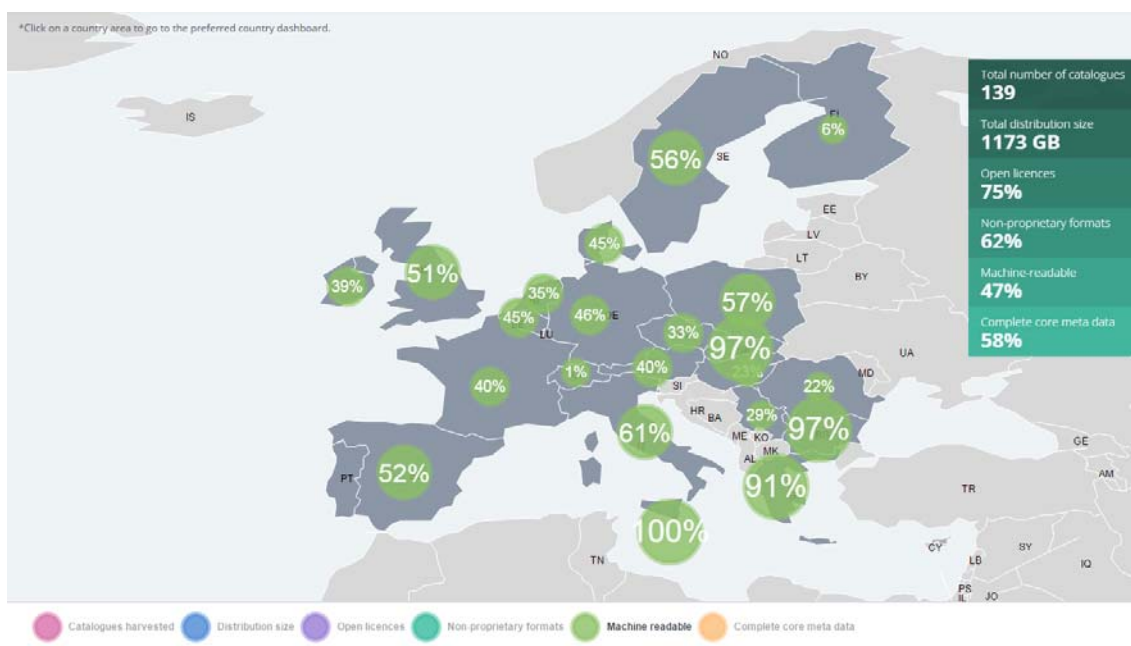
<sup>2</sup> <https://github.com/opendatamonitor>

<sup>3</sup> Deliverable 3.6: Tool architecture and components/plugins programming status report 2

<sup>4</sup> <http://5stardata.info/>

## 5.6. Machine readable

In the basic context non-proprietary formats and machine readable might look the same, since a distribution which is aligned to two stars is also readable. In this project’s context we measured it based on different key elements. While non-proprietary formats presents the results based on the distribution metadata, the machine readable does it upon datasets. OpenDataMonitor measures “Machine readable” by dividing the sum of datasets number that have at least one “Machine readable” dataset with the total count of datasets per country.



**Figure 12: European vectorial map displaying machine readable distributions (catmachinereadformatfreq)**

This metric is very relevant for the mission of open data and data in general. Improving the machine readability will lead into the increase of data reusability and overall awareness of various policy fields of open data.

## 5.7. Complete core metadata

Diving into open datasets the project noticed that the metadata given were not accurate or some relevant metadata were missing. In order to measure the completeness of one dataset the OpenDataMonitor developed a metric called “Complete Core Metadata” which checks if a dataset has a license, an author, an organisation and a date released or modified.

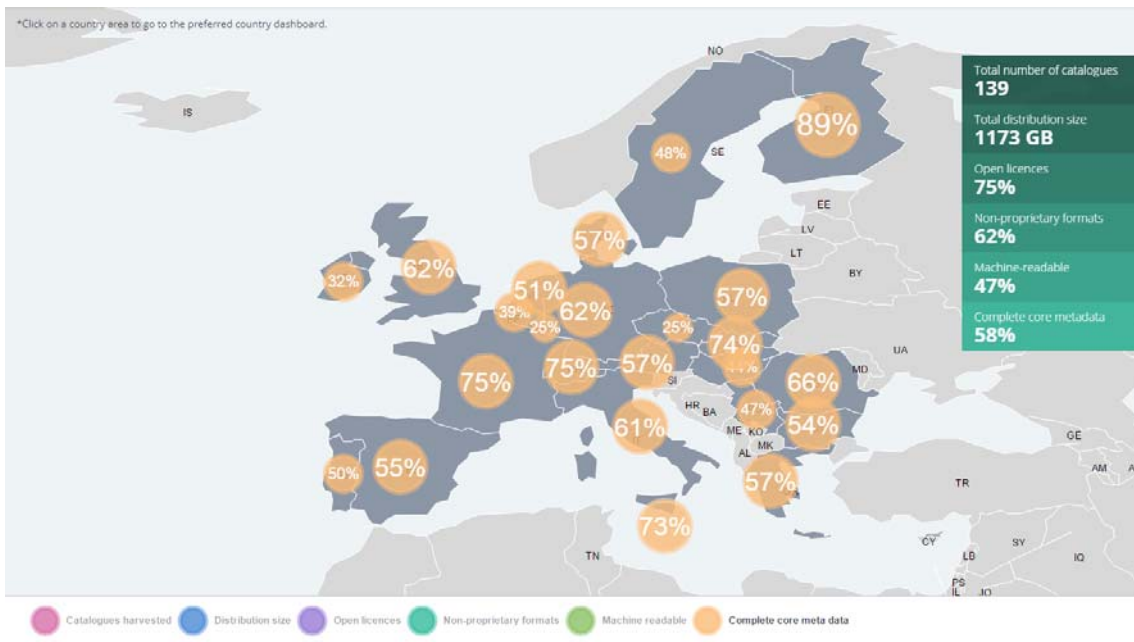


Figure 13: European vectorial map displaying complete core metadata (catcoremetadatarfq)

Considering the size of the bubble and the percentage give, the end user can easily identify which countries have a higher accuracy in providing the open data correctly.

## 5.8. Overall quality measure

The overall quality score is visualised differently in comparison to the other metrics. A simpler look and feel is been presented through a sortable multiple bar chart. This score is developed by the OpenDataMonitor project and it is integrated also in other interfaces like “Country dashboard” and “Catalogue dashboard”. The overall quality score is measured as an average of four fundamentals metrics including Open licenses, Machine readable, Accessibility and Complete core metadata.

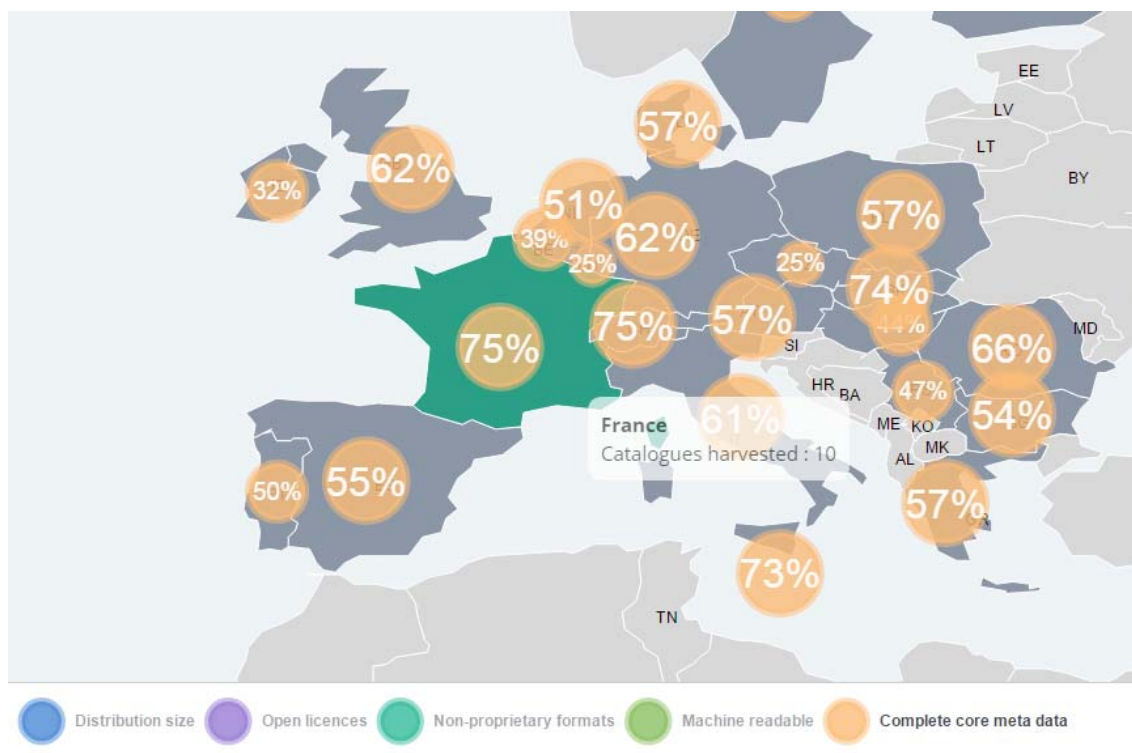


Figure 14: Overall quality score through countries

The developed score will be useful when it comes to ranking or rating the quality of an entity like a country or catalogue. The project has projected this score in four dimensions that cover the core existence of open data.

## 5.9. Country Dashboard

The country dashboard has two ways of diving from the start page. The first and the most visual is from the vectorial map. And the second way is from the overall quality score country list. In addition to this, at the catalogue dashboard or listview interface the end user can navigate straight to country dashboard by clicking on the respective country name.



**Figure 15: Navigation to the Country Dashboard**

Once the end user has selected to see the overview of a specific country, the chosen country will be projected into various dimensions in regard to open data. As noticed from figure 16, the most fundamental metrics are been highlighted through animated donut charts. While the dimensions of the overall quality measures are displayed separately through a visualisation (open licenses, machine readable, accessibility and core metadata), the score itself it's been shown on a timely basis in a textual format and as a bubble element at the map in the right.

In figure 16 France is taken as an example. France at the moment of the capture has recorded a 53% overall quality including the individual measures of open licenses with 87%, machine readable 40%, accessibility 11% and core metadata of 75%. To complete the overview for a country, the list of all catalogues harvested and harmonised is been included into a visualisation. The list of catalogues is integrated into a fixed width container in order to manage the long lists of catalogues and keep the focus in the parallel coordinate's visualisation. In the case of France, only 10 catalogues were harvested and harmonised at the moment of the capture.

In order to further promote the country results, two integration functionalities are been developed. The share button tents to share the selected page into the main social media channels like Twitter



and Facebook. While the embed one generates a code to integrate the main visualisations of this interface into the desired application.

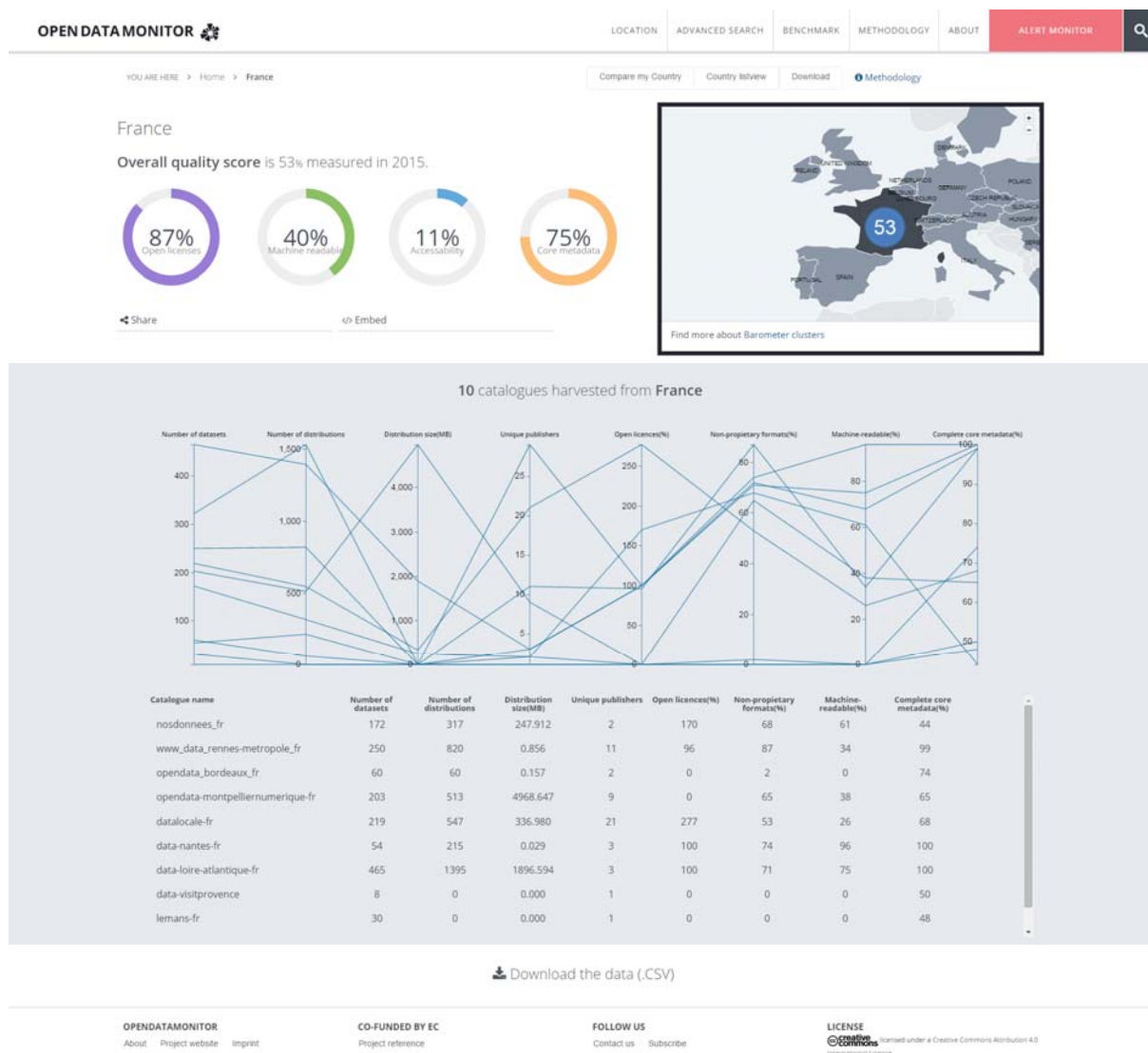


Figure 16: Country dashboard interface

Before diving into the catalogue dashboard a list of the catalogues harmonised is been extended to some relevant quantity and quality metrics. Quantity metrics include total number of datasets, total number of distributions, distribution size per catalogue, and unique publishers. The quality metrics include fields like open licenses, non-proprietary formats, machine readable and complete core metadata per data catalogue.

A multidimensional graph using parallel coordinates is used to project all the catalogues for one country in a variety of fields. The connection between the parallel points is drawn using polylines to form a sort of time series line graph without the time projection. On hover to a specific catalogue or graph line, the user will see highlighted the series of quantity and quality values presented into a parallel coordinate’s graph (see figure 17).

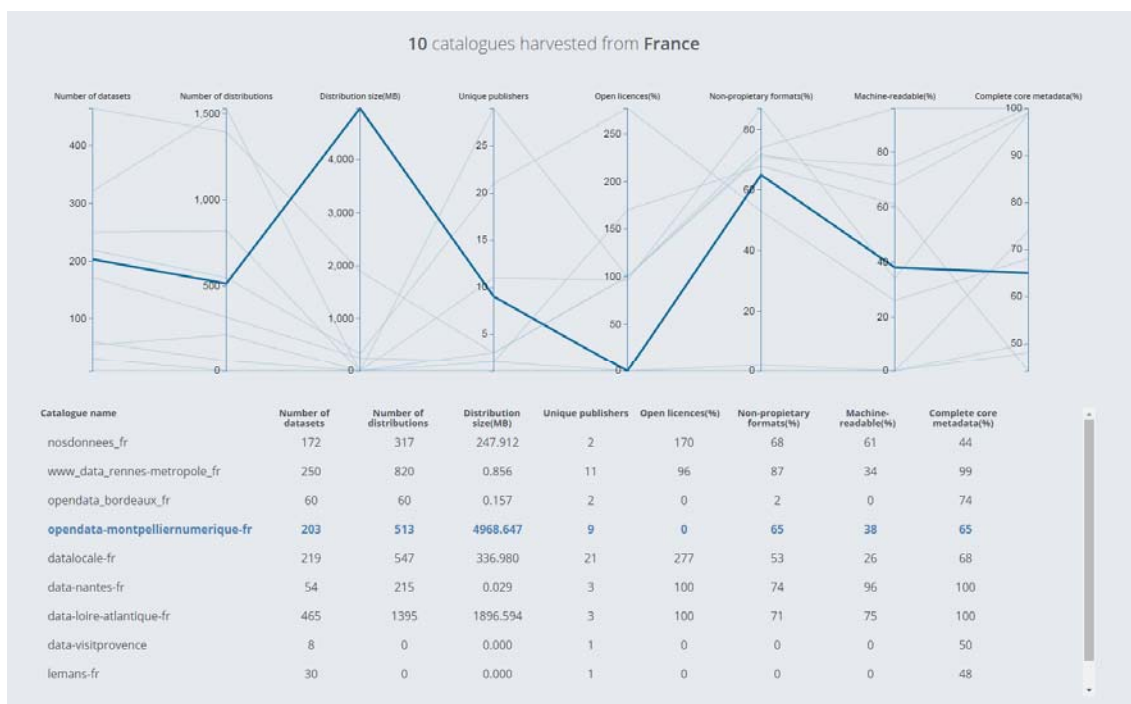


Figure 17: Catalogue performance in a glance

## 5.10. Catalogue Dashboard

The OpenDataMonitor is able to also monitor open data on the catalogue level. The Catalogue Dashboard will display the current status of the catalogue harvested and harmonized. For most of the open data catalogues this layer of data analytics doesn't exist, at least for the public view. During the stakeholder workshops, this dashboard raised many interests and also motivated them to add their catalogues into the monitor so that they could measure their work and compare it with other "competitors".

Also at the Catalogue Dashboard the consistency for the four most relevant metrics is saved. At the example capture for demonstration purpose **data-loire-atlantique-fr** catalogue, which is an official catalogue in France has 100% open licenses, 75% machine readability, 0% Accessibility and 100% core metadata completed. These figures result into a catalogue with 100% openness and considerably readable and reusable from the community but not able at all to be accessible for further maintenance. This catalogue is ranked the 3<sup>rd</sup> in France based on the overall quality measure. In addition to the quality metrics, a clear indicator is also the quantity metrics which completes the evaluation of the selected catalogue. In the example taken (see figure 18) there are 465 catalogues found, 1395 distributions collected with the size of 1897MB.

A new requirement was added later in the project which gave the platform a more interesting point of view from the data contextual perspective. The demonstration site is able to timely capture the data while demonstrating through monthly / quarterly the development of various quality and quantity indicators. The demonstration site finished the time snapshotting at the end of June and for demonstration purposes we were not able to produce more eloquent data. The late implementation of this module was also decided considering the harvesting plan.

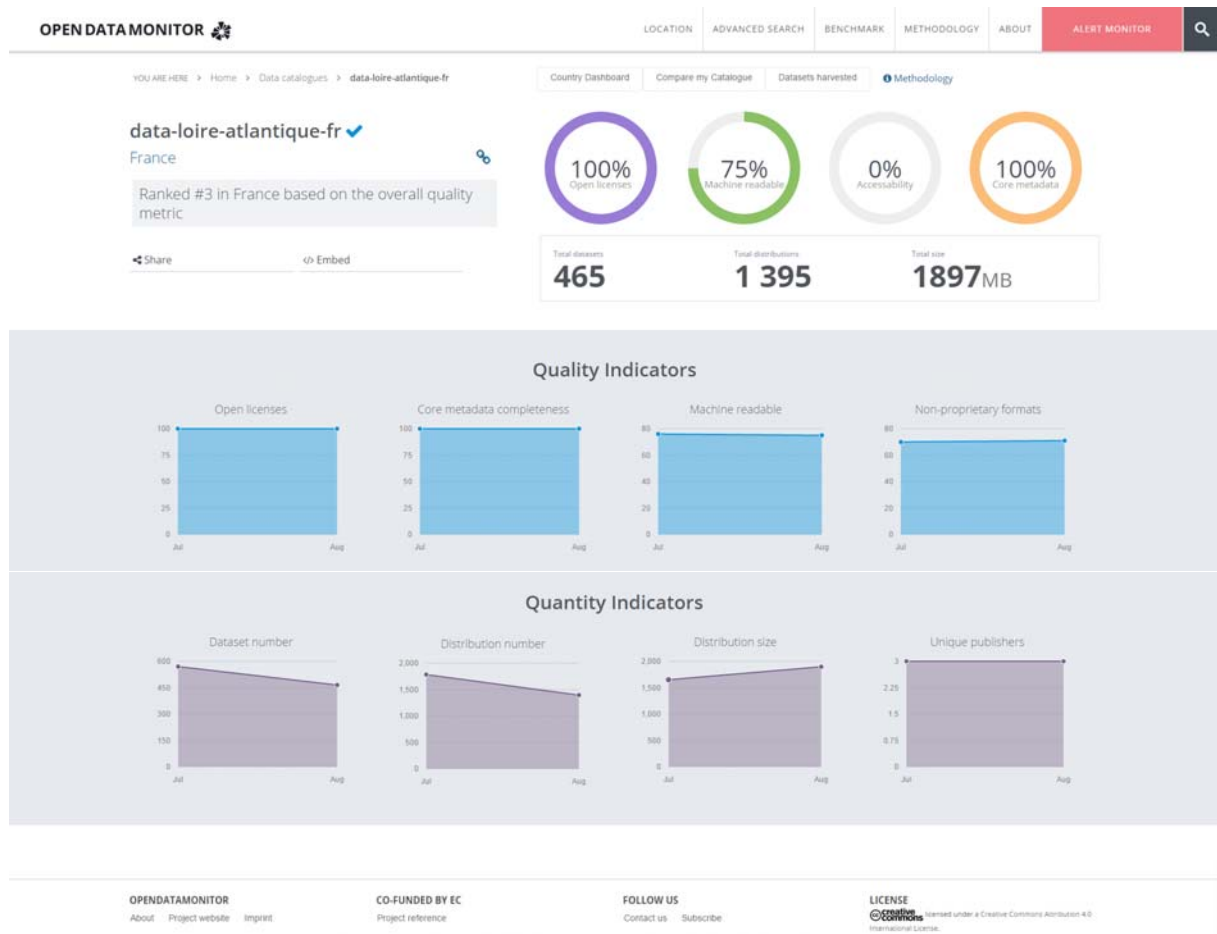


Figure 18: Catalogue dashboard user interface

## 5.11. Catalogue Listview

Once the end user has selected to view a specific Catalogue Dashboard, on the top right sub menu it displays the option to view the grid of all datasets found in that catalogue. The list of datasets is provided to the end users as a dynamic grid with easy searchable and filterable functionalities and additional export functions. In order to export the whole list of the datasets found, the user should click the “All” button and then choose the file format preferred to be exported.

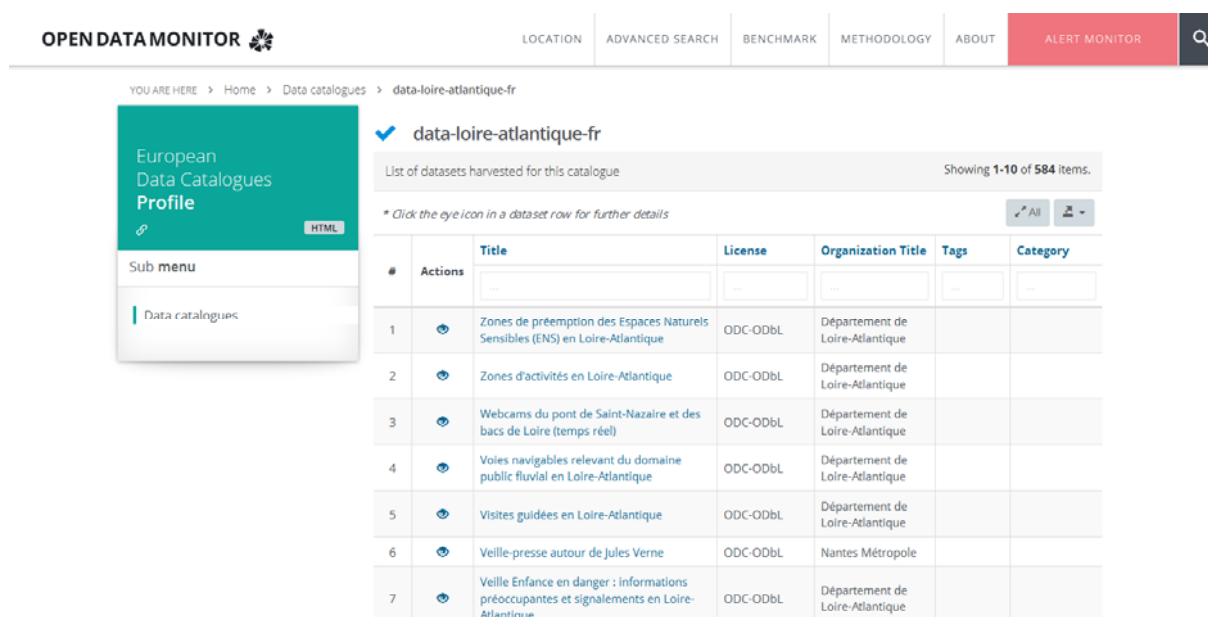
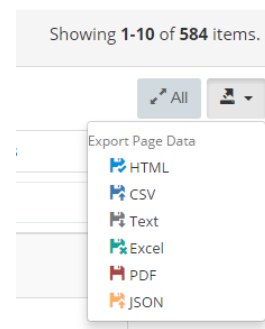


Figure 19: Catalogue listview user interface

## 5.12. Advanced Search

The search functionalities are extremely necessary in case of an application with a considerable amount of data. The search is provided at all times at the top menu. Once the icon in the shape of a loupe is pressed, an input box will be extended for the user to type the requested keywords (see figure 21).

In addition to the portable input box, there is also an “Advanced Search” made available to easily navigate through the platform. The left sidebar is offering the most important filters in order to get the preferred results. The results can be switched at the top right sub menu between catalogues and datasets while changing also the filters on the left.

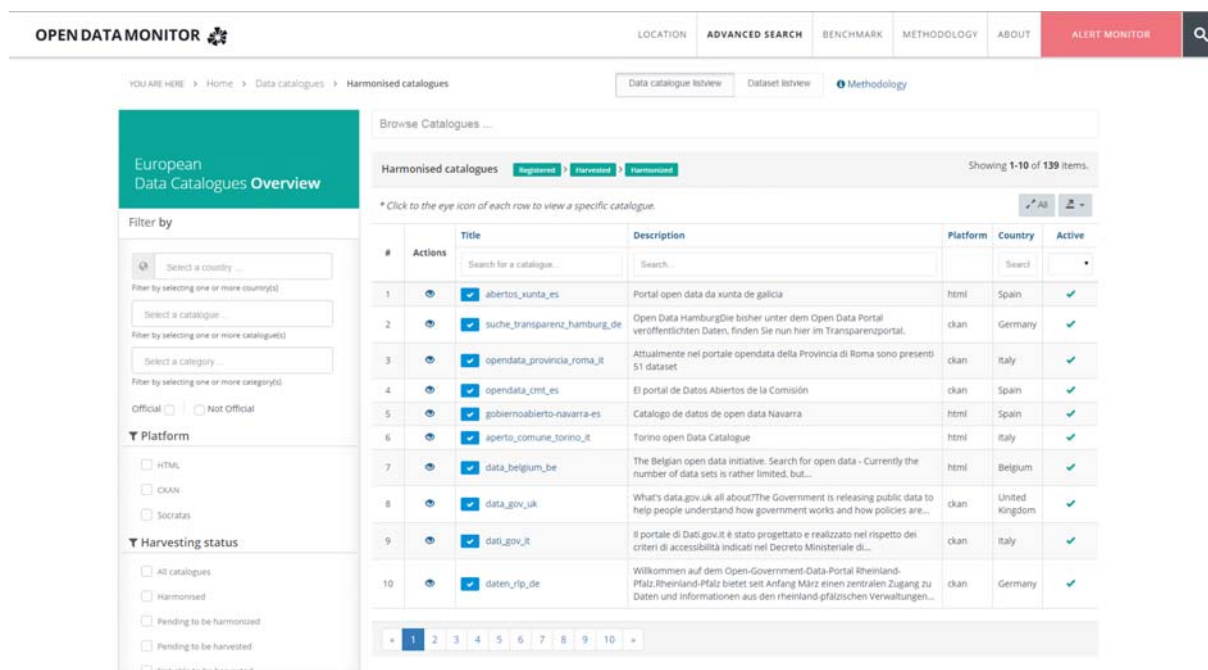


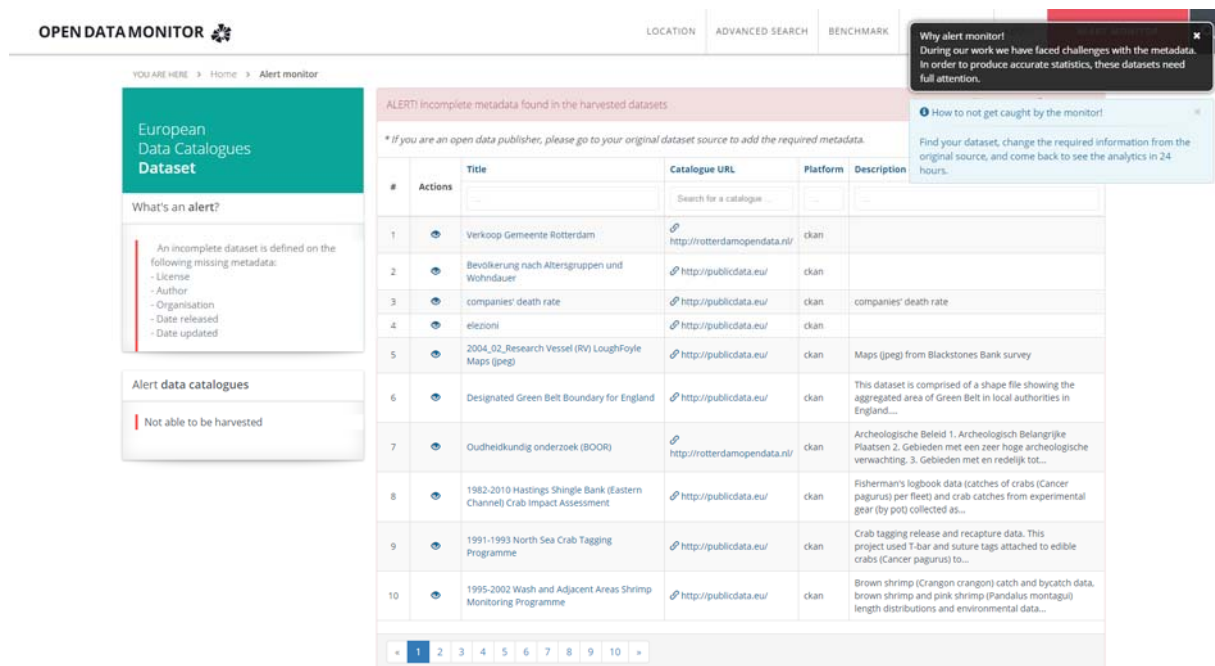
Figure 20: Advanced search for filtering data catalogues and datasets



Figure 21: Top menu generic search

### 5.13. Alert monitor

During the harmonisation process and the stakeholder interviews it was noticed that not all the publishers of open data were aware of a complete and correct dataset. The purpose of the “Alert monitor” is only to display the bad practice datasets while also offering a remedy method for deleting it from the monitor. Each dataset or catalogue alerted will take the user to the originated URL. The data publisher will be notified through a message with the problems found. That way, the data publisher has the chance to change the misinformation or the lack of information and wait for the monitor to grab the last updates from the catalogue revisit.



**Alert Monitor User Interface**

**Alert:** Incomplete metadata found in the harvested datasets

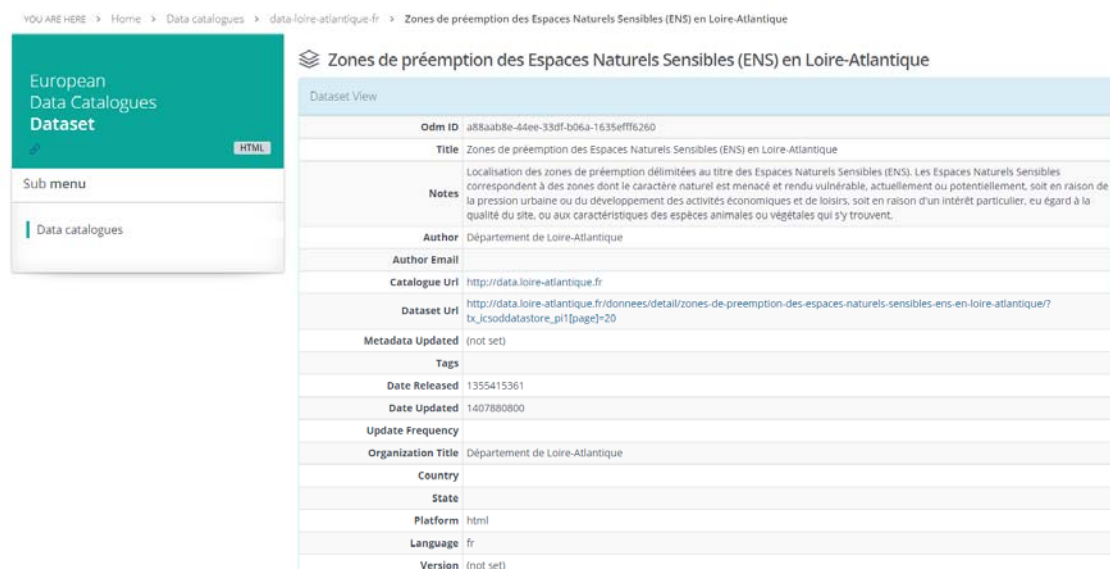
*\* If you are an open data publisher, please go to your original dataset source to add the required metadata.*

#	Actions	Title	Catalogue URL	Platform	Description
1		Verkoop Gemeente Rotterdam	<a href="http://rotterdamopendata.nl/">http://rotterdamopendata.nl/</a>	ckan	
2		Bevolkerung nach Altersgruppen und Wohndauer	<a href="http://publicdata.eu/">http://publicdata.eu/</a>	ckan	
3		companies' death rate	<a href="http://publicdata.eu/">http://publicdata.eu/</a>	ckan	companies' death rate
4		elezioni	<a href="http://publicdata.eu/">http://publicdata.eu/</a>	ckan	
5		2004_02_Research Vessel (RV) LoughFoyle Maps (jpeg)	<a href="http://publicdata.eu/">http://publicdata.eu/</a>	ckan	Maps (jpeg) from Blackstones Bank survey
6		Designated Green Belt boundary for England	<a href="http://publicdata.eu/">http://publicdata.eu/</a>	ckan	This dataset is comprised of a shape file showing the aggregated area of Green Belt in local authorities in England...
7		Oudheidkundig onderzoek (BOOR)	<a href="http://rotterdamopendata.nl/">http://rotterdamopendata.nl/</a>	ckan	Archeologische Beleid 1. Archeologisch Belangrijke Plaatsen 2. Gebieden met een zeer hoge archeologische verwachting. 3. Gebieden met en redelijk tot...
8		1982-2010 Hastings Shingle Bank (Eastern Channel) Crab Impact Assessment	<a href="http://publicdata.eu/">http://publicdata.eu/</a>	ckan	Fisherman's logbook data (catches of crabs (Cancer pagurus) per fleet) and crab catches from experimental gear (by pot) collected as...
9		1991-1993 North Sea Crab Tagging Programme	<a href="http://publicdata.eu/">http://publicdata.eu/</a>	ckan	Crab tagging release and recapture data. This project used T-bar and suture tags attached to edible crabs (Cancer pagurus) to...
10		1995-2002 Wash and Adjacent Areas Shrimp Monitoring Programme	<a href="http://publicdata.eu/">http://publicdata.eu/</a>	ckan	Brown shrimp (Crangon crangon) catch and bycatch data, brown shrimp and pink shrimp (Pandalus montagui) length distributions and environmental data...

Figure 22: Alert monitor user interface

## 5.14. Dataset Profile

Even though it was not the primary focus of this project to offer a navigation interface for the datasets harvested and harmonised, during the stakeholder workshops conducted we resulted into a change of decision. Several stakeholders were interested to search and display some best cases of datasets around Europe from the same policy field. In these terms, the project team decided to add the dataset listview and a single dataset profile.



**Dataset Profile**

**Dataset View**

Odm ID	a88aab8e-44ee-33df-b06a-1635efff6260
Title	Zones de préemption des Espaces Naturels Sensibles (ENS) en Loire-Atlantique
Notes	Localisation des zones de préemption délimitées au titre des Espaces Naturels Sensibles (ENS). Les Espaces Naturels Sensibles correspondent à des zones dont le caractère naturel est menacé et rendu vulnérable, actuellement ou potentiellement, soit en raison de la pression urbaine ou du développement des activités économiques et de loisirs, soit en raison d'un intérêt particulier, eu égard à la qualité du site, ou aux caractéristiques des espèces animales ou végétales qui s'y trouvent.
Author	Département de Loire-Atlantique
Author Email	
Catalogue Uri	<a href="http://data.loire-atlantique.fr/">http://data.loire-atlantique.fr/</a>
Dataset Uri	<a href="http://data.loire-atlantique.fr/donnees/detail/zones-de-preemption-des-espaces-naturels-sensibles-ens-en-loire-atlantique?tx_jcsoddastore_pi1[page]=20">http://data.loire-atlantique.fr/donnees/detail/zones-de-preemption-des-espaces-naturels-sensibles-ens-en-loire-atlantique?tx_jcsoddastore_pi1[page]=20</a>
Metadata Updated	(not set)
Tags	
Date Released	1355415361
Date Updated	1407880800
Update Frequency	
Organization Title	Département de Loire-Atlantique
Country	
State	
Platform	html
Language	fr
Version	(not set)

Figure 23: Dataset Profile

In the single dataset profile the metadata harvested are been displayed in a detailed view with relation to the distributions found in this dataset. The url shown in this view is the original harvesting url from where the end user can follow up with further investigation of the selected case.

## 5.15. Internationalization (i18n)

Another functionality of OpenDataMonitor is the internationalization in order to offer comprehensive interfaces to the end users with different background in Europe. The languages chosen to be translated are English, German, Greek, French and Spanish. A script is installed in the backend of the application which detects the language of the browser. If the browser language is not supported by the application, English will be the default language displayed.

Except for the automatic detection, the end user has also the option of choosing the preferred language. The select box placed to choose the language desired is positioned at the footer of each page as seen in the figure below. This solution, as already mentioned in the D3.4, is handling only the static labels and messaging of the application and not the dataset content itself.

The current status of Internationalization is in quality assurance of the labels already translated. Once the textual quality assurance is finished, some further testing will be performed in order to check the full operative internationalization module.

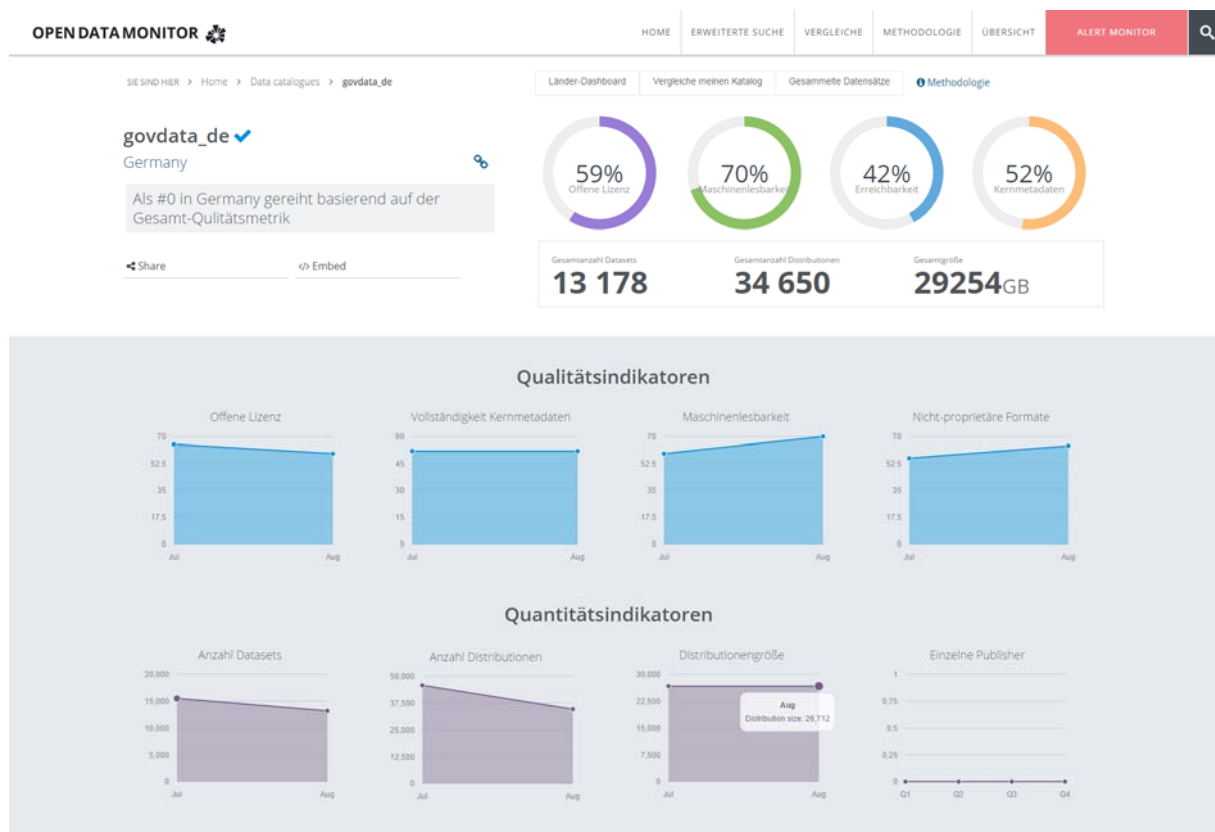


Figure 24: An example of a translated user interface in German

## 6. Conclusion

The changes in the design and UX of the platform are based on the mockups developed at D3.5. Based on the stakeholder usability workshops it was noticed that the usability of the platform was increased in comparison to the alpha version. In the second year of development the main application features were implemented including the three level dashboards (European, Country and Catalogue). The Benchmarking dashboard of catalogues and countries is designed and in process of implementation and will be finished during the next project month. The second year of design and implementation took the platform to another level of understanding and elaboration, while using the first year's prototype to lead the research and innovation for the missing values in open data. In addition to the finalisation of the three level dashboards, a missing functionality in the open data landscape is developed – snapshotting open data in 18 variables over time (See Appendix II). The disadvantage of the method used is that the snapshot over time is starting from July 2015, at the point of implementation from the platform and where the platform was robust enough regarding the data harvested. However, from this date on the snapshotting may produce new insights over time.



## Appendix I. Glossary of acronyms and abbreviations

API	Application Program Interface
BSD	Berkeley Software Distribution
CC	Creative Commons
CKAN	Comprehensive Knowledge Archive Network
CPU	Central Processing Unit
CRUD	Create Read Update Delete
CSS	Cascading Style Sheet
DB	DataBase
GB	Giga Bytes
GPL	General Public License
HTML	HyperText Markup Language
HTTP	Hypertext Transfer Protocol
ISO	International Standards Organization
JSON	JavaScript Object Notation
LESS	Leaner CSS
MIT	Massachusetts Institute of Technology license
MVC	Model Viewer Controller
ODM	Open Data Monitor
RAM	Random-Access Memory
SASS	Syntactically Awesome Stylesheets
SOAP	Simple Object Access Protocol
URI	Uniform Resource Identifier
URL	Uniform Resource Locator
UC	Use Case
WSDL	Web Service Definition Language
XD	Experience Design
XML	Extensible Markup Language

## Appendix II. Metrics explained

Table 3: List of implemented metrics

Tag	Label	Description	Metadata	Category
catdsbylicensefreq	Frequency of distributions by license type	Count distinct licenses per catalogue	licenses, _id	Per catalogue per selector
cataccessabilityfreq	Accessibility	Defined set of fields as requirement: - description (notes in database): 1 if contains some text, 0 null/empty - not a broken link: 1 if not broken - author email/maintainer_email (only in CKAN): 1 if present sum(descrtiption && not_broken && author	count, _id, tot_count	Quality
catcoremetadafreq	Complete core metadata per catalogue	Average across a defined set of fields, eg License, Author/Maintainer (maintainer field exists only in CKAN), Organisation, Date released or Date updated	count, _id, tot_count	Quality
catmachinereadformatfreq	Machine-readable in % per catalogue	total count of Machine readable / total count of datasets	count, _id, tot_count	Quality
catnonproprformatfreq	Non-proprietary formats per catalogue	total count of distributions with a non-proprietary format / total count of distributions with a format	count, _id, tot_count	Quality
catopenlicfreq	Open licences in % per catalogue	total count of open licences / total count of distributions with a licence	licenses[], count, _id, tot_count	Quality
catsitepagerank	Discoverability	page ranking, based on google ranking / alexa ranking	pagerank[], _id	Quality
edcoremetadafreq	Complete core metadata	Average across a defined set of fields, eg License, Author / Maintainer (maintainer field exists only in CKAN),	freq	Quality

		Organisation, Date released or Date updated		
edmachinereadformatfreq	Machine-readable in %	total count of machine readable datasets / total count of datasets	freq	Quality
ednonpropformatfreq	Non-proprietary formats	total count of distributions with a non-proprietary format / total count of distributions with a format	freq	Quality
edopenlicfreq	Open licences in %	total count of open licences / total count of distributions with a licence	licenses[]	Quality
catdatasetstotfreq	Total number of datasets	Count number of metadata objects per catalogue	count, _id	Quantity
catdatasizetotal	Total distribution size per catalogue	Sum resources' size per catalogue (KB)	count, _id	Quantity
catdistribstotfreq	Total number of distributions	Count number of resources per catalogue	count, _id	Quantity
categories	Categories per catalogue	Count categories found	count, _id, category[]	Quantity
catfreq	Total number of catalogues	Count number of different catalogues	count	Quantity
catpublishersfreq	Unique publishers	Organisations per catalogue	count, _id	Quantity
eddistribsize	Total distribution frequency and size	Total Sum of resources' size in Europe (KB)	freq, size	Quantity