D3.8 ADVANCED BENCHMARKING, REPORTING AND ANALYTICAL FEATURE DESCRIPTION

PROJECT

Acronym:	**OpenDataMonitor**
Title:	Monitoring, Analysis and Visualisation of Open Data Catalogues, Hubs and Repositories
Coordinator:	SYNYO GmbH
Reference:	**611988**
Type:	Collaborative project
Programme:	FP7-ICT
Start:	November 2013
Duration:	24 months
Website:	http://project.opendatamonitor.eu
E-Mail:	office@opendatamonitor.eu
Consortium:	**SYNYO GmbH**, Research & Development Department, Austria, (SYNYO) **Open Data Institute**, Research Department, UK, (ODI) **Athena Research and Innovation Center**, IMIS, Greece, (ATHENA) **University of Southampton**, Web and Internet Science Group, UK, (SOTON) **Potsdam eGovernment Competence Center**, Research Department, Germany, (IFG.CC) **City of Munich**, Department of Labor and Economic Development, Germany, (MUNICH) **Entidad Publica Empresarial Red.es**, Shared Service Department, Spain, (RED.ES)

DELIVERABLE

Number:	**D3.8**
Title:	**Advanced benchmarking, reporting and analytical feature description**
Lead beneficiary:	SOTON
Work package:	WP3: Concept Design and Software Development
Dissemination level:	Public (PU)
Nature:	Report (R)
Due date:	September 30, 2015
Submission date:	September 30, 2015

Authors: **Elena Simperl**, SOTON
Yunjia Li, SOTON
Qiong Bu, SOTON
Ejona Sauli, SYNYO

Contributors: **Vassilis Kaffes,** ATHENA
Amanda Smith, ODI
Stefan Ruscher, SYNYO
Johannes Burger, SYNYO
Peter Leitner, SYNYO

Reviewers: **Bernhard Jäger,** SYNYO

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

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1. INTRODUCTION

1.1. Purpose

The OpenDataMonitor (ODM) project has developed the ODM platform (http://opendatamonitor.eu) to harvest, harmonise and visualise the open data catalogues across Europe, which presents an overview of the open data ecosystem. In the second half of the project, as more feedbacks are collected from open data stakeholders in Alpha release (February 2015) and Beta release (June 2015), a series of enhancements and advanced features are added to further meet user requirements. The research has been focusing on defining benchmark metrics both quantitatively and qualitatively. Advanced benchmarking, advanced search and social plugin features are integrated into the platform.

This purpose of this document is to describe the following advanced features in detail:

- Advanced Benchmarking for both country and catalogue level
- Advanced Search for both data catalogues and datasets
- Social plugin to share page or embed visualisation chart from ODM platform

1.2. Scope

There are many enhancements incorporated into the demonstration site of ODM platform after usability evaluation in June. The feedbacks collected during D4.2 Usability Study were summarized in the usability report and the actionable items were reviewed in early September. Various changes have been implemented to improve user experience such as adding tooltips and description for the metrics and charts, enriching the content in methodology section of the demonstration site, linking the bubbles of metrics indicator on map to the corresponding country, etc. These smaller changes will not be elaborated in this document. The scope of this document is centred on the major user cases identified in previous deliverables and the main action item summarized as F-6 (comparison) and F-9 (enhanced filter/search) from D4.2 user evaluation feedback report.
We have identified ten User Cases in “D3.2 Tool specifications, user cases, mockups and functionalities report 1” and refined these user cases in “D3.5 Tool specifications, user cases, mockups and functionalities report 1” as well as adding three new scenarios allowing users to embed/share the graphical elements built in ODM platform. At the time of the ODM platform evaluation and user study in June 2015 when we hold focus group workshops and online questionnaires to collect user feedback, the following scenarios associated with advanced benchmarking, reporting and analytics features are not implemented yet. The design and implementation of these advanced features, which have incorporated the feedbacks from end users, will be described in the next chapter 2,3 and 4. The user interface is described in chapter 5. In the last chapter, we will have a summary and outlook of related advanced features.

Table 1.1: Use Case scenarios review[1]

Ref. No	User Case Name	Scope Status
UC3	Compare the open data situation between different entities - Comparison Dashboard	This use case is still in place and it is been broke down in smaller user stories (Theme 4). The user interface will look almost the same as UI 4. (Section 6)
UC5	Create a report from the data presented on the portal	This use case remains in place. The option to create a report will be given from UI4. In addition to create, the end user will be able to save, print and download the data for the displayed graphs.
UC6	Trigger automatic creation of a report	A change in the implementation of the reporting tool has been done, in order to address performance and decrease waiting time for processing and generation. The automatic generation will be possible when the registered user will place the respective settings for generating a report (frequency).
UC11	Embed graph content from country dashboard or catalogue dashboard into user's website	The user may want to embed a certain graph from country dashboard to their national open data portal, data catalogue owner may want to embed certain graph from catalogue dashboard to their website. An 'Embed' button could be provided and allow user to choose which graph they want to embed to their own site.
UC12	Share ODM dashboards link to social media (Twitter)	The user likes the content, information the visualisation presented on ODM platform and want to share with more people on social media. Provide a twitter button allowing user to share the current page (applicable for all three major dashboards- European dashboard, country dashboard and catalogue dashboard)
UC13	Embed catalogue score to user's website	If a data catalogue has better score in terms the quantitative and qualitative metrics, user may want to embed that score

1 D3.5 "Tool-specifications,-use-cases,-mockups-and-functionalities-status-report-2", OpenDataMonitor FP7-ICT 611988

	(In current design, the score is not something in any existing graph, otherwise this could combine with UC11)	to their website. Have a global button similar to Search button to allow user to choose “view my score”, then they can choose/input their catalogue name. They will be redirected to a page with their calculated score and an iframe link which they can embed to their website. Or Combine this with “Embed“ button on Country/Catalogue dashboard (as UC11), and then allow them to choose among graphs and their score. Then provide corresponding iframe link.

2. ADVANCED BENCHMARKING AND REPORTING

In the previous deliverable of D2.3[2], we have defined metrics and key figures of the Open Data Monitor, which include the measures over the aggregate, per-geography measures, per-catalogue measures and per-dataset measures. The per-geography measures are mainly reflected in the European dashboard on the home page and the country dashboard in our demonstration website. The per-catalogue metrics are visually represented in the new catalogue dashboard, which is implemented after the beta release in June 2015. As part of the new advanced benchmarking and reporting feature, some of these metrics have been enhanced and clearly defined, and corresponding comparison dashboard is developed to support comparison at country level and catalogue level. This chapter will describe the architecture, the changes introduced to the metrics, the detail design and workflow of the comparison dashboard.

2.1. Architecture Overview

The figure below integrated from deliverable D3.6[3] presents the architecture of ODM platform and the main components, which are responsible for metadata harvesting, harmonising, and how metadata was processed in analysis engine and eventually presented on the demonstration site in an intuitive visualisation form. To adapt to the needs of advanced benchmarking, the analysis engine is especially enhanced in the 2nd half year of the development to calculate the refined metrics upon the harmonized metadata. The outputs from the analysis engine are the calculated results which can be accessed via RESTful APIs by the visualisation engine to show charts and dashboards to the end users on demonstration site. The latest APIs used to calculate these metrics are shared[4] to the community. Section 2.2 illustrate the changes of these metrics.

[2] D2.3 “Best practice visualisation, dashboard and key figures report”, OpenDataMonitor FP7-ICT 611988
[3] D3.6 “Tool architecture and components plugins programing status report 2”, OpenDataMonitor FP7-ICT 611988
[4] http://odmapi.magellan.imis.athenainnovation.gr/api/v1.0/_commands

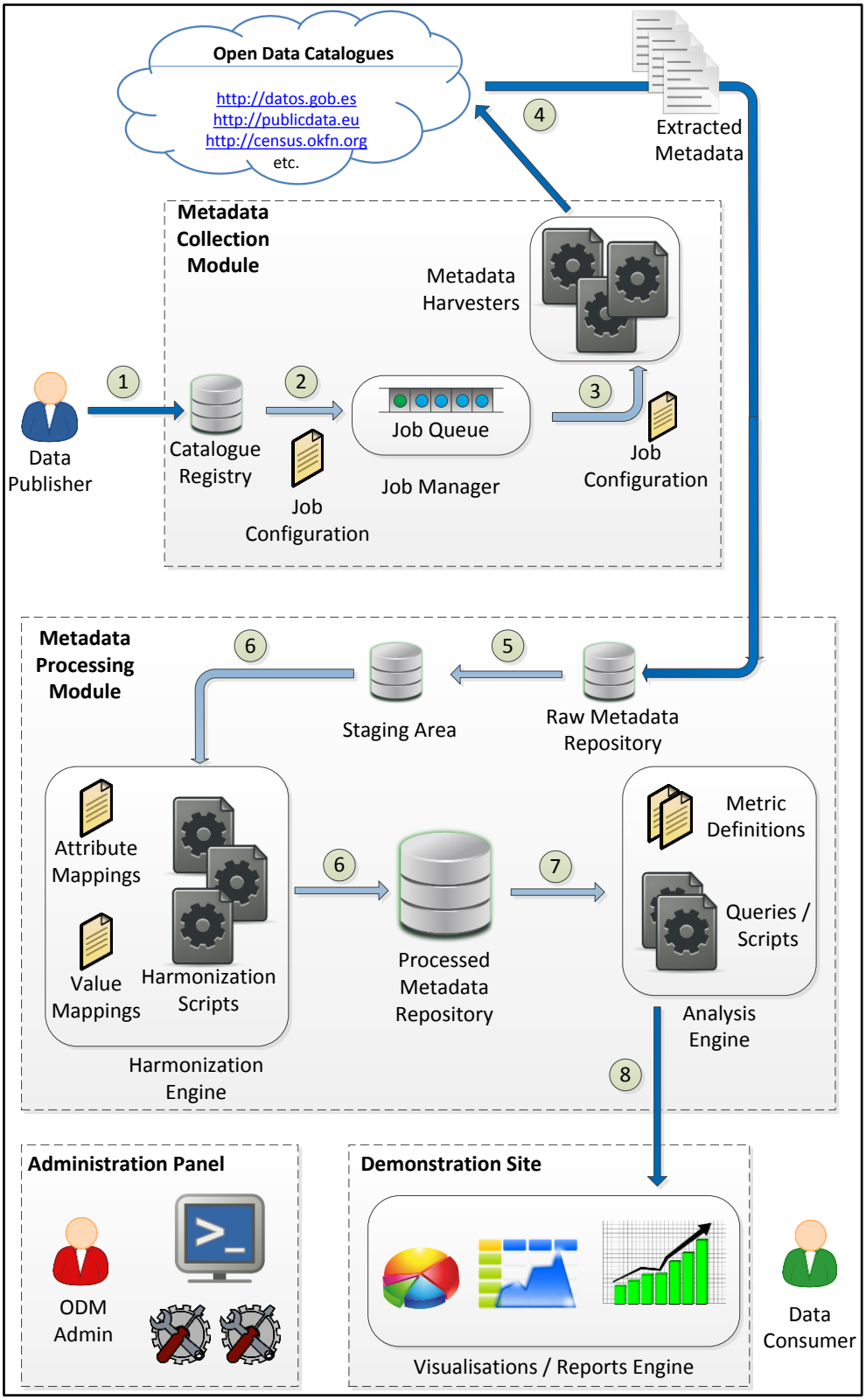

Figure 2-1 ODM Architecture

2.2. Changes in Metrics

2.2.1. Overall Changes

All the metrics we used in ODM platform are categorised into two groups, quantity metrics and quality country. Quantity metric is used to count and aggregate on selected attributes of the harvested catalogues and datasets. Quality metrics are mainly to help us understand how good a selected attribute is. For example, we count the number of datasets in a catalogue, or the total number of distributions of all datasets in a catalogue using absolute numbers, which can quantify the aggregated amount of the catalogue's attribute. On the other hand, for a specific catalogue, we also want to understand what is the percentage of datasets have open licenses, what is the proportion of datasets which provides sufficient core metadata, etc. These are quantified by relative figures calculated with the qualitative metrics to reflect how open the data catalogue is, and how useful the metadata the catalogue can provide, per se.

In total, we now have 12 metrics with 5 quantity metrics, 6 quality metrics and one aggregated overall quality score. These metrics are applicable for catalogue level and country level across the European countries respectively. We have refined the European dashboard in the home page, the Country dashboard and Catalogue dashboard to present corresponding metrics as summarized below:

- **European Dashboard (Home)**: gives an overview of the overview of European open data landscape. Quantity metric "Number of datasets" and Qualitative metrics "Open licenses", "Machine readable", "Open access", "Open metadata" are used measure the performance of member countries of Europe whose catalogues have been harvested by ODM platform. Sorting is enabled for all these five metrics so that the end-user could sort the list of country based on their choice of metric, such as Open Access.
- **Country Dashboard**: From the home page, user can navigate to Country dashboard by clicking the bubble indicators on the European map or click on each country at the bottom of the home page.
    - For each country, quality metric Open licenses, Machine readable, Open access, and Open metadata are presented at the top to reflect an overall quality of open data in that country. Total number of catalogues harvested from that country is also shown right below the top 4 charts representing qualitative performance.
    - Within each country, 8 metrics have been used to measure the performance of catalogues from that country. These metrics include 4 quality metrics (open licenses, open formats, machine readable, and open metadata) and 4 quantity metrics (number of datasets, number of distributions, distribution size, unique publishers)
- **Catalogue Dashboard**: Catalogue dashboard can be accessed in two ways. Firstly, user could choose specific catalogue they want to view from the country dashboard. Alternatively, they could search using the general search button at the top right corner or go to Advanced Search page to locate the catalogue they are interested in. Then click on the

Title of the corresponding catalogue to see its details on catalogue dashboard. Metrics used to measure the performance a catalogue are divided into three main section:

- Overall measures: figures and charts on the top right of the catalogue dashboard show the quantitative measure of Total datasets, Total distributions and Total size, as well qualitative measures in Open licenses, Machine readable, Open access and Open metadata. This gives an overall idea of the performance of the corresponding catalogue.
- Trending: Tim dimension is newly added after the beta release. Time-based charts are presented in the middle of the page to show the quality and quantity changes of that catalogue over time. Similarly as what we have used in the bottom of the country dashboard to measure catalogue performance, 4 quality indicators (open licenses, open formats, machine readable, and open metadata) and 4 quantity indicators (number of datasets, number of distributions, distribution size, unique publishers) are used in this time-based charts section.
- Detail distribution of License and Category: at the bottom of the Catalogue Dashboard, detail distribution of different license type and categories found in this catalogue are presented in a tabular format.

- **Benchmark (Comparison) Dashboard**: a new tab is added, allowing users to compare and benchmark between different countries and catalogues and generate a report as needed.
  - Benchmarking: The metrics used on the benchmark section are consistent with the metrics which have been presented on Country dashboard and Catalogue dashboard respectively to enable the comparison on country or catalogue level based on your choice. All information on metrics definition is illustrated in section 2.2.2 and the detail design of the benchmarking features is elaborated in section 2.3.
  - Reporting: On the same dashboard, the reporting feature is integrated. This is a new simplified implementation, which is mainly adjusted based on the latest user feedback and thus is different from the previous planned solution in early deliverables. More details on the reporting feature design changes are available in section 2.3.2.

- **Advanced Search**: Search feature has been there since the very beginning of the ODM platform. The original search is accessible by clicking the search icon at the top menu and then input the keyword. It is important more sophisticated advanced search is available to help user easily locate the needed catalogue or dataset especially when the amount of data to be searched from is huge. As a result, apart from existing search icon at the top right, a new tab of Advanced Search is added. On the Advanced Search page, users can switch between catalogue view and dataset view where corresponding Search Filters are provided to ease the search process. Detail description of the feature and the user interface are available in chapter 5.

2.2.2. Metric Definition and Calculation

In the second half year of development of ODM platform, especially after the beta release, more focus has been put on benchmarking, metric verification with data, metric refining and visualisation enhancement. The following Table 2.1 has the detail definition of each metric, how it is calculated, and links to github with further information on the metric related data.

It is worth mentioning that there are a few terminology changes and calculation changes upon three metrics:

- **Open Metadata**: originally named as “Complete core metadata” on ODM platform and renamed to “Open Metadata”. The completeness of each dataset is calculated based on 4 fields: licence, author, organisation and (date released or date updated). For each of these field, if it is contained in the metadata harvested, we score 0.25 for it and then sum all these number for all the datasets under a catalogue. If a dataset has all three
- **Open Access**: originally named as “Accessibility” and renamed to “Open access”. This metric is used to show how accessible the catalogue is. For each dataset, when all three conditions are satisfied -- description, at least one resource with a functional link, and email of the author, the dataset will be treated as publicly available dataset. The open access score is calculated by the number of datasets that are qualified as publicly available divided by the total number of datasets in the catalogue.
- **Open format**: originally named as “Non-proprietary format” and renamed to “open format”. This metric is a ratio between the distribution with a non-proprietary format and the total number of distribution.

Following metrics are newly added to better reflect a performance of a country/catalogue/dataset:

- **Discoverability**: This newly added metric leverages the traffic ranking systems Google and Alexa to estimate how important a catalogue is. This metric will be presented in both Catalogue Dashboard and Benchmarking (Catalogue comparison) dashboard.
- **Quantity and Quality metrics with time dimension**: This captures the performance of 8 different indicators (details in section 2.3) of selected catalogue/country and shows the changes via timeline-based charts. The metrics are the same as they were defined in Table 2.1 and used in other dashboards, but the representation of showing them over user specified time period is new in the latest implementation.

Table 2.1: Metric Definition

Metric	Definition	Category
Open licence	This metric represents total count of open licences over total count of distributions with a licence. You can access our published open licence mappings here.	Quality
Open metadata	This metric represents the average completeness of metadata across a defined set of fields: licence, author, organisation and the existence of one of the date released or date updated.	Quality
Open access	This metric represents the number of datasets qualified as publicly available over the total number of dataset in a catalogue. The availability score is calculated across a defined set of fields: a description, at least one resource with a functional link and an available email of the author.	Quality
Open formats	This compliments the machine-readable metric. It represents the total count of distributions with a non-proprietary format over total count of distributions with a format. You can access our published list of non-proprietary formats here.	Quality
Machine readable	This metric represents the readability of one dataset while calculating the machine readability of its distributions formats. The distribution formats include file types such as CSV, XLS, JSON, XML and RDF You can access the full published list of machine readable formats here.	Quality
Discoverability	This metric represents an estimation of how important a catalogue is in the web based on two traffic ranking systems: Google and Alexa.	Quality
Total distribution size	This metric represents the total size of all resources, regardless of their format, for every dataset in a catalogue. The size is in KBytes.	Quantity
Number of distribution	This metric represents the total number of distributions of a datasets or the total number of distributions of all datasets in a specific catalogue. A distribution of a certain data set refers to a specific available form of that data set.	Quantity
Number of datasets	This metric represents the total number of available datasets in a catalogue.	Quantity
Number of unique publishers	This metric represents the total number of unique publishing organisations of a specific catalogue.	Quantity

Catalogues harvested	Number of catalogues per country harvested and harmonized.	Quantity
Overall quality score	The overall quality score is calculated as the average of open licenses, machine readable, open access and open metadata. This score involves the most important metrics defined based on the current open data standards, and is used to rank and sort catalogues and countries respectively.	Quality

2.3. Design of Comparison Dashboard

In the early release of ODM platform, visualisation is provided in terms of each defined metric to reflect the corresponding performance of a single data catalogue or a country in a chart, or an overview of the different metrics across Europe on a map or in a tabular format. There is no comparison dashboard and reporting feature implemented before though this has been initially designed in D3.4[5]. This section focus on describing the advanced benchmarking in comparison dashboard and the simplified reporting features provided with the comparison dashboard.

2.3.1. Benchmarking

The primary use case of advanced benchmarking is UC3 as described in Table 1. The initial Comparison dashboard design is based on the stakeholder requirements we have collected in working package 2 (D2.4 an D2.6 in particular) and have been refined in the final development stage as we finalising the metrics for measuring catalogues and catalogues of each country. Table 2.2 has the detail user stories that have been fully considered during the comparison dashboard design. The initial design of the interface in D3.5[1] is described in Figure 2-2: Comparison, benchmarking and advance reporting user interface. User interface of the benchmarking feature has progressed in the latest release as both country dashboard and catalogue dashboard have been finalised. It is becoming obvious that it is making sense to keep similar look and feel for the comparison dashboard so that there is no additional learning curve for the end users. Meanwhile, having the predefined quantity and quality metrics for comprehensively comparing between the selected entities are necessary.

5 D3.4 “Visualisation Dashboard and Multi-lingual Interface Status report 1”, OpenDataMonitor FP7-ICT 611988

Table 2.2: User stories grouped into the benchmarking and advance reporting theme[6]

Theme: Comparison, benchmarking and advance reporting				
ID	**As a/an**	**I want to...**	**so that...**	**Notes**
THEME4-1	data publisher	know about developments in open data in my topic area (common licenses, kinds of data sets)	I can adjust the way I publish open data	
THEME4-2	advanced/sophisticated catalogue administrator	design a report with self-selected metrics	scan the development of the catalogue I administer	report design tool, all available metrics, PDF-export
THEME4-3	basic catalogue administrator	know about important performance indicators of my catalogue	scan the development of the catalogue I administer	pre-selected metrics, pre-designed report (e.g. most frequently contributing organisations, share of "truely open" data, homogeneity of licenses, prevalence of formats), PDF-export
THEME4-4	catalogue administrator	design a report with certain metrics across a limited number of catalogues	I can compare the performance of the catalogue I administer to other similar catalogues	selection of metrics, suggest comparable catalogues (date of initialisation, level of government, size, comparable kinds of data [e.g. general catalogue, topic-specific catalogue]), compare to "best" catalogue in Europe, compare to all catalogues of this level of government, compare to all catalogues from this country)
THEME4-5	local open data activist	design a report with self-selected metrics	permanently scan the development of open data in my local area/ region	
THEME4-6	OD activist group	design a report with self-selected metrics	permanently scan the development of open data in Europe and or	selection by region, country; comparison of several or all countries and/or catalogues;

6 D3.5 “Tool specifications, user cases, mockups and functionalities status report 2”, OpenDataMonitor FP7-ICT 611988

			specific countries	metrics for specific subsets of open data (e.g. health data)
THEME4-7	OD user	be alerted when specific kinds of data sets are added for my local area / region / country	I do not have to search frequently to know what is out there	alerts by search term, data categories, specific licenses
THEME4-8	OD user	get the data used to make each visualization	I can reuse those data	
THEME4-9	OD user	have predefined charts in the first load of the interface	I have an example on how to use the comparison tool	
THEME4-10	OD user	have the opportunity to add additional charts or data indicators to benchmark my country or catalogue	I can reuse the benchmarking tool at an extensive level	
THEME4-11	OD user	be able to save the report generated	so that I can refer to it at a later stage	
THEME4-12	OD user	be able to see the list of saved reports	so that I can refer to it at a later stage	
THEME4-13	OD user	be able to download the data used to create the graphs presented into the tool	I can allow myself to be creative and prove the certainty of the results	
THEME4-14	OD user	print the report I just created	I can use it in meetings or presented it to my managers or higher level of management	

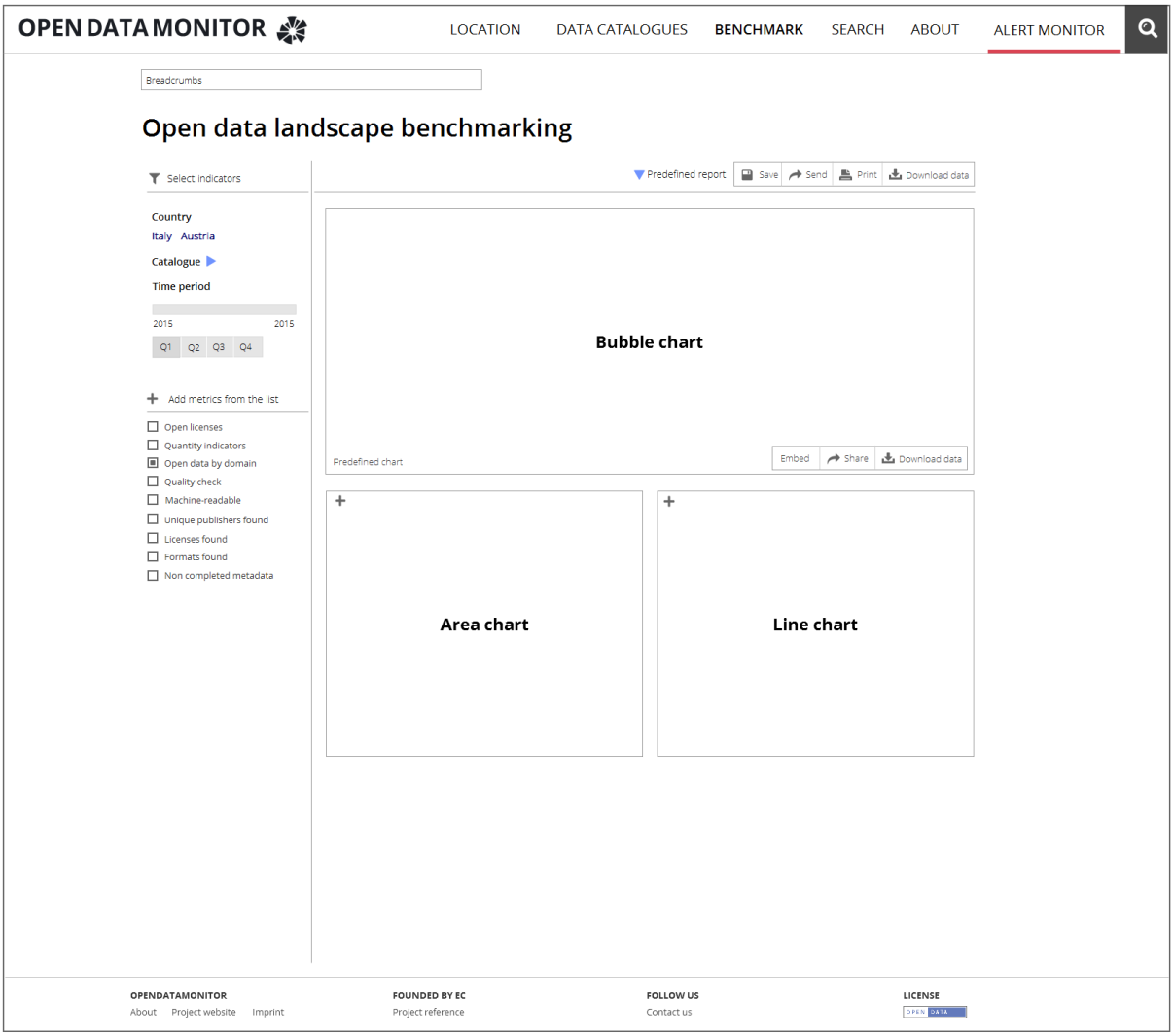

Figure 2-2: Comparison, benchmarking and advance reporting user interface[7]

The principle behind the comparison dashboard is to enable an easy selection of entities to be compared by reducing the efforts to figure out which metric to use, at the same time allow flexibility to set the preferred timeline. For instance, all the harvested catalogues are available in the drop-down list (see Figure 5-5) so that user can easily choose from it or type a keyword to narrow down the catalogues they intend to compare with. For users who only want to compare their catalogue's performance with other catalogues before Q3 of 2015, they could set the time period to customize the comparison using the scroll bar and quarter selection button provided (see Figure 5-2). Regardless if the comparison is between a country or a specific catalogue, the same layout is used to present the comparison result:

- Coverage: a map view with the country of your selected entities (country or catalogue) highlighted. The map view could be zoomed in or zoomed out as needed.

7 1

- Overall Quality Score: This is the aggregated score calculated based on four qualitative metrics: Open licenses, Machine readable, Open access, Open metadata. The overall quality score and all these four metrics have been used at the top of each country dashboard and catalogue dashboard page to highlight the quality perspective, hence it comes naturally to have comparison on these metrics highlighted.
- Quality and Quantity Indicators: In this main part, we have the four charts presenting 4 quality indicators as mentioned above. Four quantity metric based charts are presented at the bottom, including Dataset number, Distribution number, Distribution size and Unique publishers.

Tooltips are also available for all the charts on the comparison dashboard: for the coverage map, mouse over on the highlighted countries user have selected, the country name will be displayed. From the Overall Quality Score chart, user can view the detail scores of compared entities by hovering over the lines on the chart. Similarly, for each quality or quantity indicator chart, mouse over the line and point on the chart will show the name of the country/catalogue (depending on which comparison you are doing) and the score for the corresponding time point. Graphical design of the comparison dashboard related element can be found in chapter 5.

2.3.2. Reporting

Reporting was considered as an integral part of the ODM project since the beginning and has been planned in previous deliverable D3.1[8] and D3.4. The main purpose of having reporting feature is to allow user creating customized report and make it available in different layout and format, such as in print format. At that moment, a thorough study has been done and potential solutions have been proposed in section 4.5 of D3.1. Generally speaking, we could choose from three different strategies: building everything from scratch, or using a template which could be filled in beforehand, or converting existing visualisation into the desired report format. In section 4 of D3.4, we have also researched the template-based report generation tools and explored a list of report generation solutions, including BIRT(http://www.eclipse.org/birt/), Crystal Reports (http://www.crystalreports.com/), DataVision (http://datavision.sourceforge.net/), ReportLab (http://www.reportlab.com/), JasperReports (http://community.jaspersoft.com/), etc. BIRT was considered as the most appropriate tool that meets our requirement at that moment, considering it offers convenient report creation feature and allow visualisation of the report in HTML within server component.

In terms of the frequency and how the report is generated, different strategies have been compared in D3.4. Scheduled report generation and automatically triggering of report sending was one of the options have been considered. Another approach would be triggering the report generation in an ad-hoc manner. Each solution has its pros and cons. However, automatically triggering report generation described as UC6 in Table 1 will require a complete user management in backend which is not

8 D3.1 “Development of a scalable open data monitoring concept and framework design”, OpenDataMonitor FP7-ICT 611988

something critical to the open data stakeholders at the stage according to comments gathered from the workshop and online questionnaire during user evaluation. So eventually the report will be created in an ad-hoc manner.

The decision on reporting feature and the approach to implement it is largely based on the feedback collected from the usability study, as well as similarities between the existing dashboards and the desired layout/content of the report. What we have learnt from the userability study is open data stakeholders are much more interested in the data and the interpretation of various indicators/metrics at the moment, having a nice reporting comes in second place. It is quite clear that a heavy-weight reporting feature is not necessary in the ODM project context. On the other hand, open data users show special interest in understanding the performance of their country/catalogue in comparison with other countries/catalogues. To satisfy the needs of comparison as well as a easy to use reporting tool, we have integrated the simplified reporting feature into the Benchmark (Comparison) dashboard. The reporting features include:

- **Share**: This feature allows user to share the current report page on social media such as Facebook, Google+, Twitter, and Github. At the same time, it also allow user to send the link via email.
- **Create report**: This feature will create a PDF version of the current report based on user selected entities and time period (without the left side filter) and allow user to save it and store it for future references.
- **Export data**: This will allow user to download the data of the selected entries. So far we only export data in JSON format. Other formats could be considered if there is such user requirement later on.

Graphical design of the reporting element can be found in chapter 5.

3. ADVANCED SEARCH

The Search feature has been greatly improved since our beta release. We have the basic search functionality since the very beginning of the ODM platform. However, the original search is pretty simple by providing a search icon at the top menu and allow user to input the keyword they have in mind. During the workshop sessions of ODM user evaluation, we have received a number of feedbacks demanding more advanced search and filtering features. The design principle behind the advanced search is to help user easily locate the needed catalogue or dataset from the large amount of data (catalogues and datasets). As a result, a new tab of Advanced Search is added on the top menu in the latest release. On the Advanced Search page, users can switch between catalogue view and dataset view where corresponding Search Filters are provided to ease the search process. All the search results are represented in list view and available to be downloaded.

3.1. Data Catalogues

In the Data Catalogue search view (refer to Figure 5-21), a Filter panel (Figure 5-12) is provided at the left side of the page. The idea is the predefined filter field would help user to quickly narrow down to the particular catalogue(s) they are looking for. Firstly, user can choose or input which country they are interested in. Or they can directly use the drop down list for “Filter by selecting one or more catalogues“ to locate certain catalogues by either scrolling down the harvested catalogues list or typing in their keyword. For stakeholders such as catalogue administrators, data generators, or infrastructure providers, they might be interested in having a look at the catalogues by platforms. Policy makers, individual or business users, might want to locate catalogue by their status or simply want to understand the harvesting status of catalogues. Therefore, Platform filters (ckan, socrata, html) and Harvesting status (Harvested, Pending, Bad catalogues, Other sources) filter are also provided on the Filter panel.

Once users submit their filter/search criteria, result will be available in a list view that has the same style with the data catalogue list, which was available in previous alpha and beta release of ODM platform. User could use reset button to clear the current searching criteria. As to the presentation of the search result, if the search result is more than 10 entities, it will be put into multiple pages among which users can navigate by clicking on the corresponding page number at the bottom. User could switch the pagination view to “All” by having all the result in one page if they like, by using the button All/Page right above the list view. The resulting list view has Actions, Title, Description, Platform, Country and Active filed:

- The hierarchy icon in the Actions column will bring user to the detail profile of the selected catalogue.
- The title of the catalogue is a hyperlink and will forward user to the corresponding catalogue dashboard. It is worth mentioning the tick icon next to the tile of the catalogue is an indicator that catalogue is an official catalogue. In addition there is a dashboard icon at the actions column where it will take the users the same as the title hyperlink.

- The information field has brief information about the corresponding catalogue in case it's an active catalogue and the reason why it's not harvested for bad catalogues.
- Platform filed shows which platform the catalogue is published on. For instance, some catalogue is ckan-based; ckan will be shown in the platform column.
- Country column shows the country information of the selected catalogue.
- Active column is used to indicate whether the catalogue is currently active or not.

3.2. Datasets

In the Dataset search view (refer to Figure 5-22), a Filter panel (Figure 5-13) is provided at the top the page. There are five filter option groups (6 different filter fields) provided on the dataset search page:

- Filter by selecting one or more catalogues: This field allow user to choose from a list of existing catalogues or type in keyword to identify relevant catalogues.
- Filter by selecting a format: This will be useful for users who know the exact format of dataset they need so they can specifically filter that format of resource. For instance, user wants to find a dataset in RDF format, he could then choose RDF from the format drop-down list.
- Filter by selecting one or more category: This allows user to find datasets in a certain category, such as Education, Energy, Food Safety, etc.
- Filter by selecting one or more platform(s): This option allows user to choose the dataset which is deployed based on certain platform, such as ckan, socrata, opendatasoft or html.
- License related filters: We have two license related filters. "Filter by selecting a license type" allow user to choose from Open/Not Open/Not Specified and "Filter by selecting one or more license(s)" lets you to choose more specific type of licenses they want to search, like "CC BY".

Similar to data catalogue search, after users submit their filter/search criteria, result will be available in a list view at the bottom of the page. By default, if the search result is more than 10 entities, they will be put into multiple pages among which users can navigate by clicking on the corresponding page number at the bottom. User then could choose to show all result in one page if they like (not recommended if your results contain a large number of datasets). The resulting list view has Title, Country Url, License, Category, Platform, Organization Title, Country, Format, Mimetype and Size filed:

- The Title filed has brief title about the corresponding dataset. This filed is hyperlinked so that user can directly click on the Title filed to view the dataset profile details.
- The Country Url field is self-explanatory. It has the url to access the corresponding catalogue where the selected dataset is coming from.
- The License field indicates the license of the selected dataset.

- The Category filed shows the category the data in the corresponding dataset belongs to.
- The Organisation Title filed contains the brief title of the organisation who owns the dataset.
- The Platform column shows which platform the dataset is published on.
- The Country column shows the country information of the selected dataset.
- The Format column is used to indicate the format of the selected dataset.
- The Mimetype field is to indicate MIME type of the corresponding dataset resource. It depends on the format of the dataset distribution, not all dataset has this field. For instance, if a dataset provides distribution via Sparql endpoint, it might have the Mimetype as “application/sparql-results+xml”. On the other hand, if a dataset is in zip format or csv format, the Mimetype field might be empty.
- The Size filed is used to indicate the size of the corresponding dataset in kb.

4. SOCIAL PLUGIN

This section has a brief introduction of the social plugin feature to help advocate open data initiative and present the result of OpenDataMonitor (ODM) to a wider communities.

4.1. Share Current Page

The main idea of sharing the current page is from UC12 in Table 1. The purpose of having such social media plugin is to allow user to share the content they like from the ODM platform, at the same time, promoting ODM result to more users so that they can leverage the platform as needed. A Share button will be provided on ODM dashboards and it will give user the flexibility to choose which social media they want to share the current page, such as Twitter, Facebook, Google+ and Github. In particular, the share button on the benchmark page also allows user to send the current link to specified email address so the recipient could check the same comparison result and decide to save or export it into their desired format as necessary.

4.2. Embed Specific Graph

This feature of embedding specific graph is derived from UC11 and UC13 in Table 1. Embedding graph content from country dashboard, catalogue dashboard or comparison dashboard should be available to users so that they can easily put the charts they like onto their own website. For instance, a user may want to embed a certain graph from country dashboard to their national open data portal, a data catalogue owner may want to embed certain graph (e.g. Quality score charts for Open License, Open access, open metadata and machine readable) from catalogue dashboard to their website. An ‘Embed’ button could be provided and allow user to choose which graph they want to embed to their own site. Alternatively, the embed button can be provided along with each graph or graph group. In the early prototype stage, we have proposed to generate the corresponding iframe link for the corresponding chart which users can directly include into their web page, the implementation might change in the final stage as we have gone through many changes of the user interfaces of all dashboards and charts since the alpha and beta release.

5. INTERFACE

This section describes the Graphic User Interface related with advanced benchmarking, reporting and analytical features. The type of graphs and overall design of the presentation of these graphs have been greatly improved especially after we received the feedback from end users. The development team has gone through a comprehensive review and discussion to finalise the most relevant metrics and corresponding representations. Since our previous report D3.7[9], the look and feel of each graph has been basically finalized. The same type of charts from Country Dashboard and Catalogue Dashboard is correspondingly used in the Country comparison and Catalogue comparison comparison dashboard. New filters for customising report and reporting features are integrated in this comparison dashboard.

5.1. Interface Elements

5.1.1. Comparison Filters:

The graphs in the comparison dashboard are been directed by the filter on the left. Without making any previous selection the dashboard will not show any results. The respective filters for Country comparison and Catalogue Comparison are as shown in and Figure 5-2. The switch between country and catalogue comparison is enabled by clicking the tab buttons on the top of the sidebar.

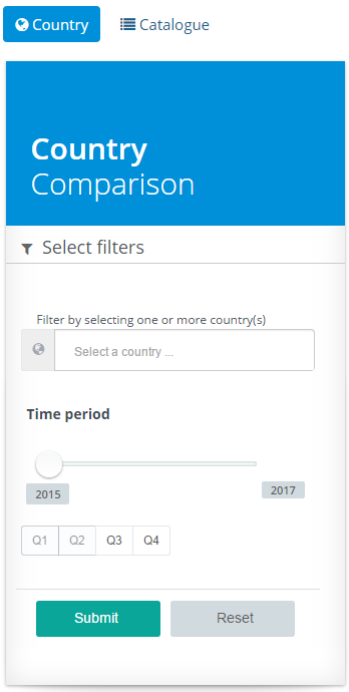

Figure 5-1: Country Comparison Filter

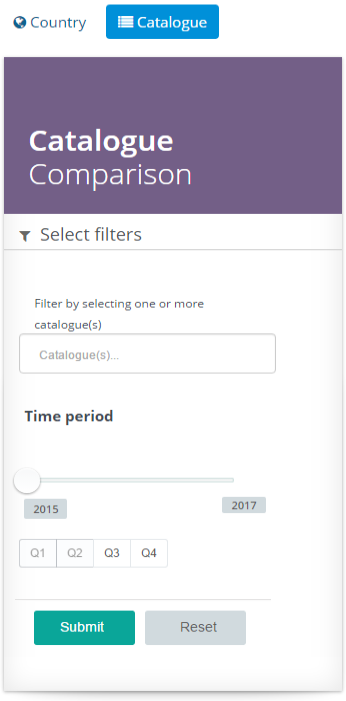

Figure 5-2: Catalogue Comparison Filter

9 D3.7 “Visualisations Dashboard and Multi-lingual Interface Status Report 2”, OpenDataMonitor FP7-ICT 611988

5.1.2. Input fields in the Comparison Filters:

The side bar filters include the main entities for comparing. During the business analysis of the project, the end users were mostly interested to compare their countries with other best practices and especially down to catalogues. The time factor is added to the filter in order to add a higher level of comparison. The input fields have enabled search as you type functionality offering a faster interaction with the select options. Below are some print screens on action of the input fields.

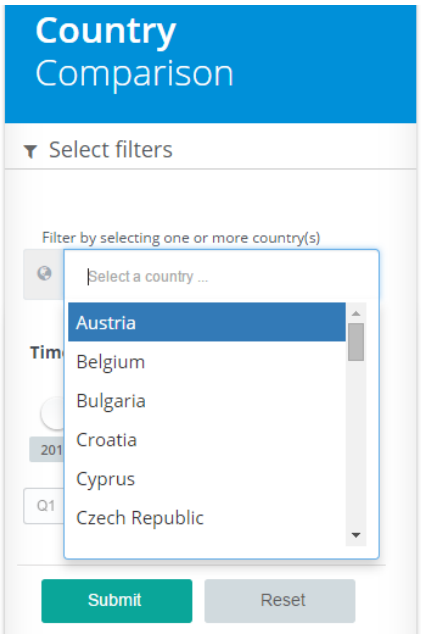

Figure 5-3 Filter by country - select

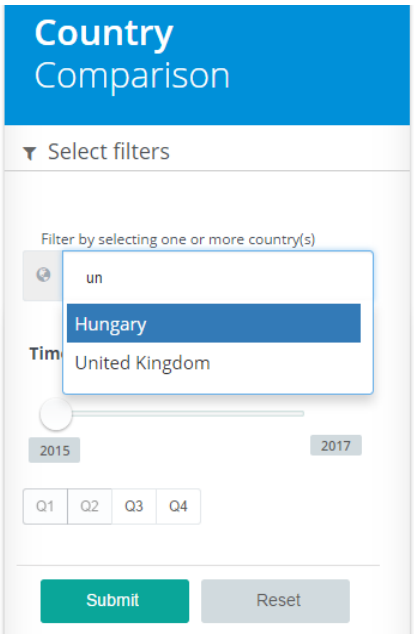

Figure 5-4 Filter by country - type in keyword

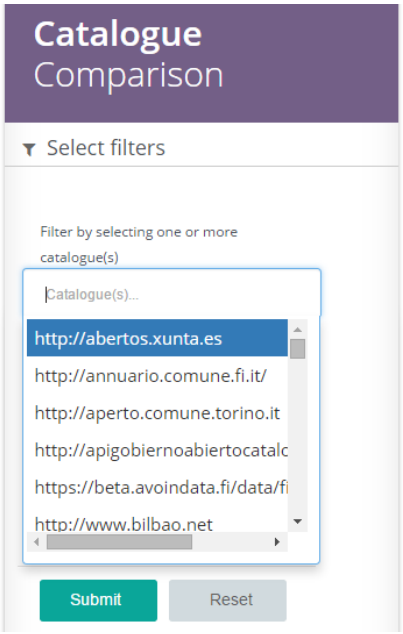

Figure 5-5: Filter by catalogue – select

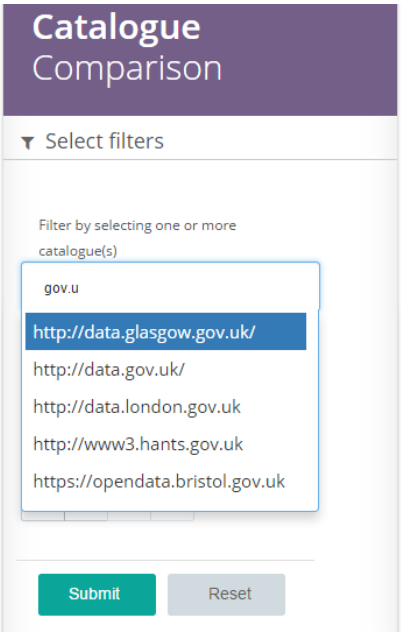

Figure 5-6: Filter by catalogue - type in keyword

5.1.3. Coverage and Overall Quality Score:

This section is at the top of the benchmark dashboard for both country comparison and catalogue comparison. It gives a quick map view with selected entities related countries highlighted and the overall quality score of the selected entities. Figure 5-7 and Figure 5-8 are example screenshots from corresponding country comparison and catalogue comparison dashboard.

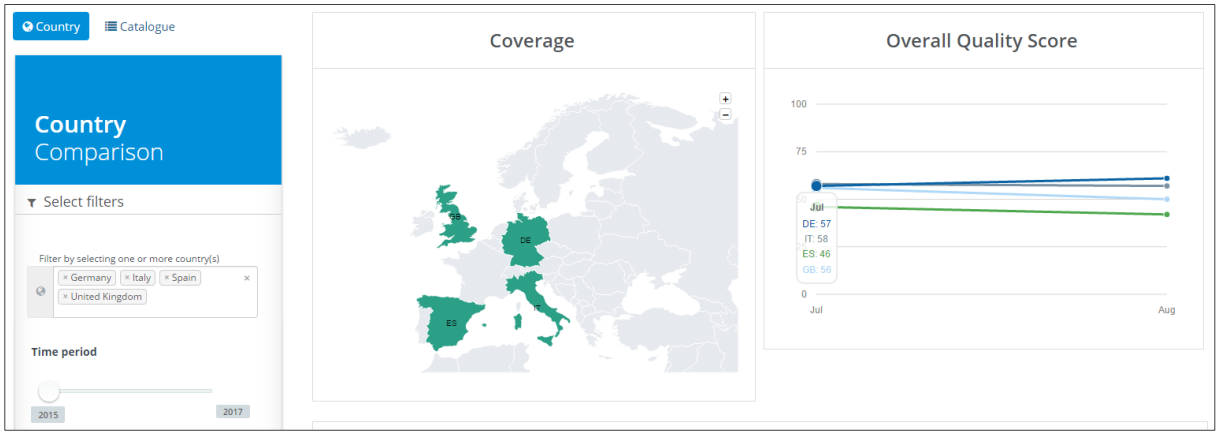

Figure 5-7: Country Comparison - Overall

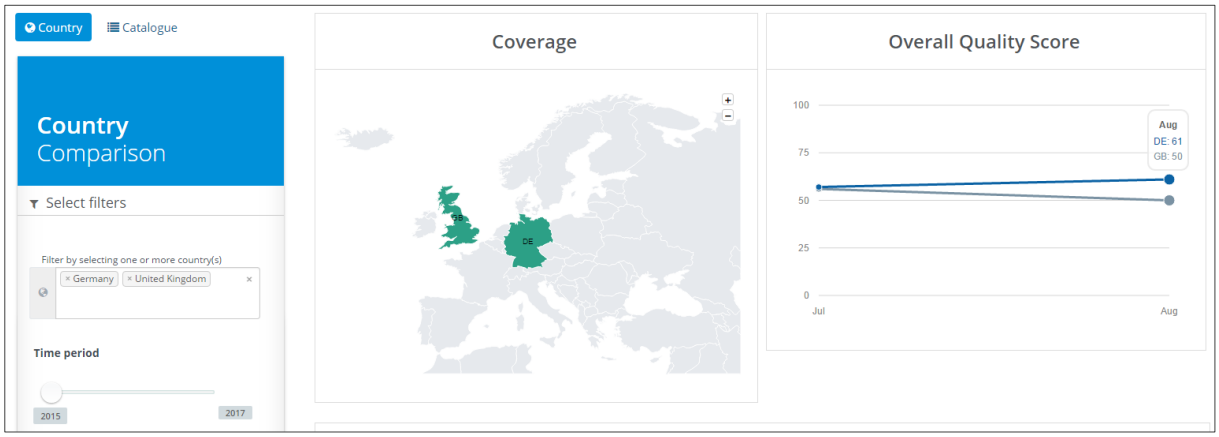

Figure 5-8: Catalogue Comparison - Overall

5.1.4. Quality and Quantity Indicators:

This section is the main body of the comparison dashboard (report) where the comparison on 8 different metrics are presented in time-based charts. In Figure 5-9, you could find the detail charts of a comparison between three selected countries. Similarly, Figure 5-9 is an example of comparison between two catalogues.

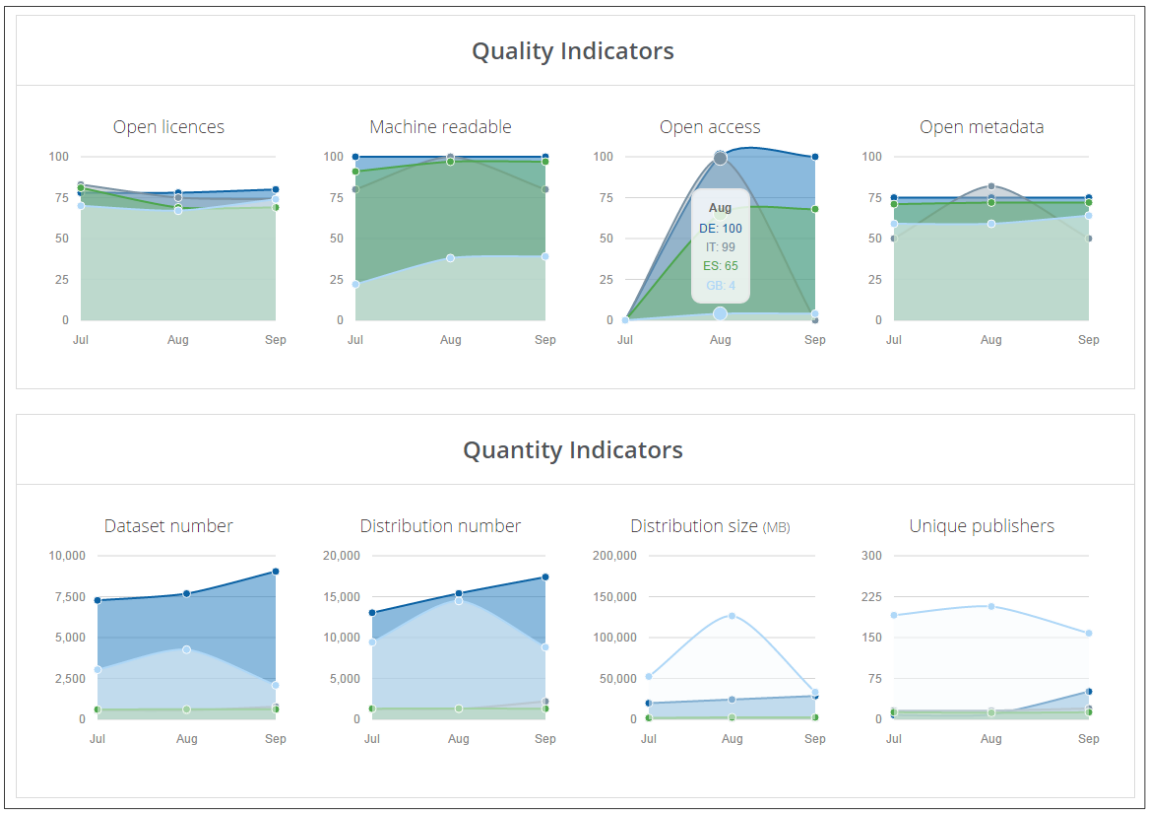

Figure 5-9: Country Comparison - Detail Indicators

Figure 5-10: Catalogue Comparison - Detail Indicators

5.1.5. Reporting:

This is the reporting feature related buttons integrated in the Benchmark page, which allows user to share, generate or export the comparison result into preferred format, such as pdf or json.

Figure 5-11: Create a report in .pdf

5.1.6. Search for catalogue:

Figure 5-12 shows the available filters user can use to search for catalogue(s). This filter panel is put on the left side of the Advanced Search page, and the results will be displayed on the right side of the Advanced Search page.

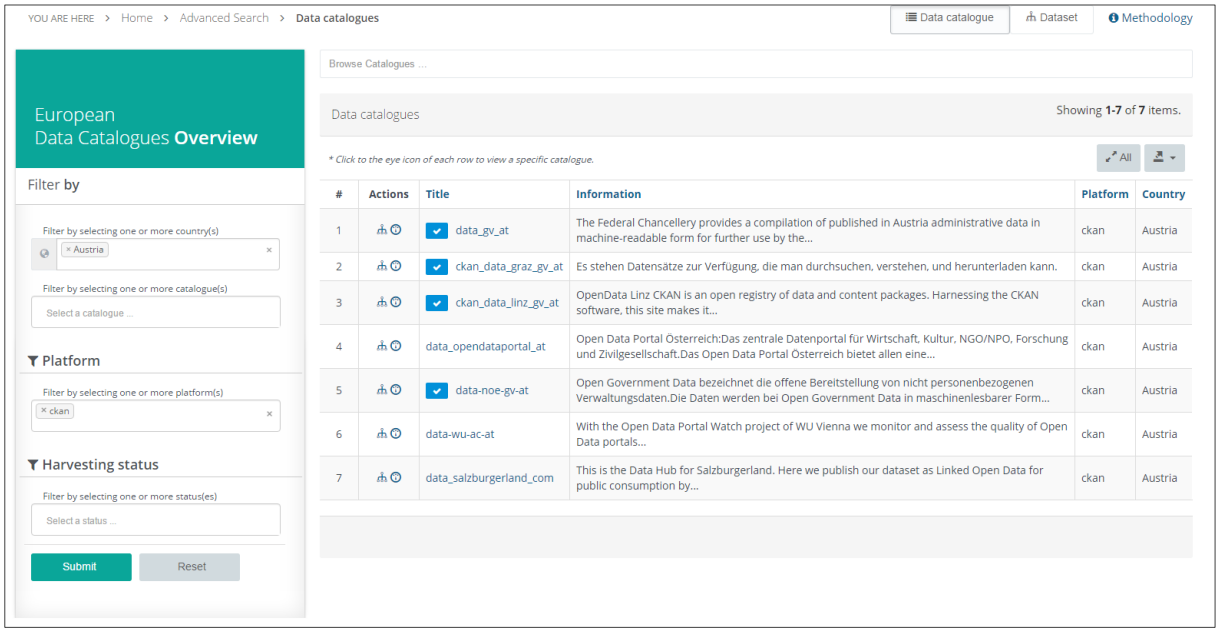

Figure 5-12: Advanced Search - Catalogue

5.1.7. Search for dataset:

Figure 5-13, Figure 5-14 and Figure 5-15 show the available filters user can use to search for dataset(s). It is notable that we have more six dataset specific filters, (a) and (b) are the zoomed version to give a better idea how these filters are placed on the user interface. This filter panel is put at the top of the Advanced Search page for dataset, and the results will be displayed at the bottom. This layout is different than the data catalogue advanced searched view and the main reason is not only because we have more filters/search criteria for dataset, but also because the results would contain more fields which need a relatively wider space.

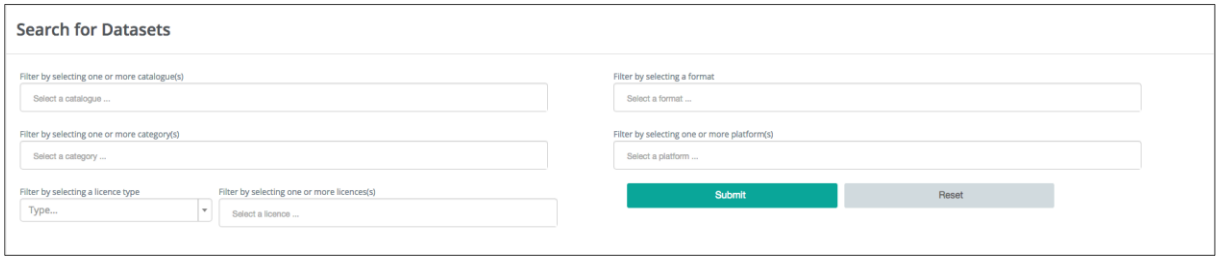

Figure 5-13: Advanced Search - Dataset

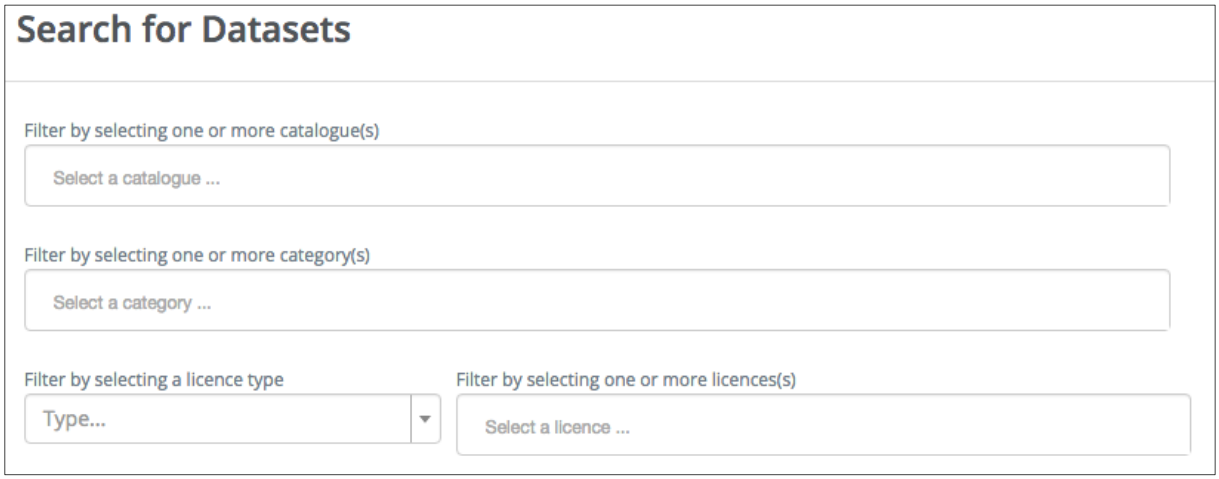

Figure 5-14: Advanced Search - Dataset (a)

Figure 5-15: Advanced Search - Dataset (b)

5.2. Screenshots

5.2.1. Advanced Benchmark

By clicking on Benchmark tab at the top menu, users will be directed to the comparison and benchmarking tool where they can choose to compare between countries or between catalogues. The switch is at the left top of the page. The corresponding comparison could also be accessed from Catalogue dashboard and Country dashboard, each have a "Compare my catalogue" (Figure 5-16) or "Compare my country" (Figure 5-17) shortcut at the top right menu above the quality charts. The default view of the benchmark page, the country comparison and catalogue comparison are respectively reflected in Figure 5-18, Figure 5-19 and Figure 5-20.

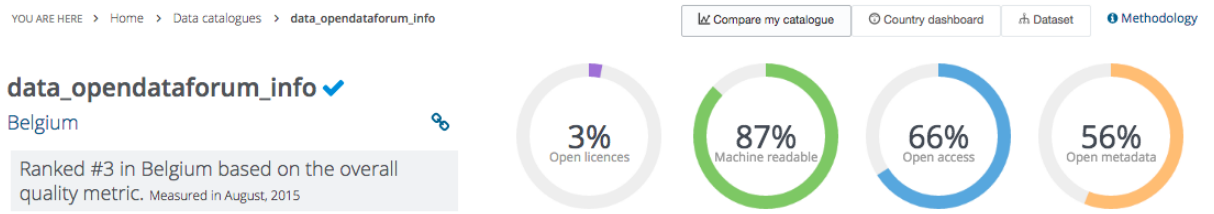

Figure 5-16: Compare my catalogue

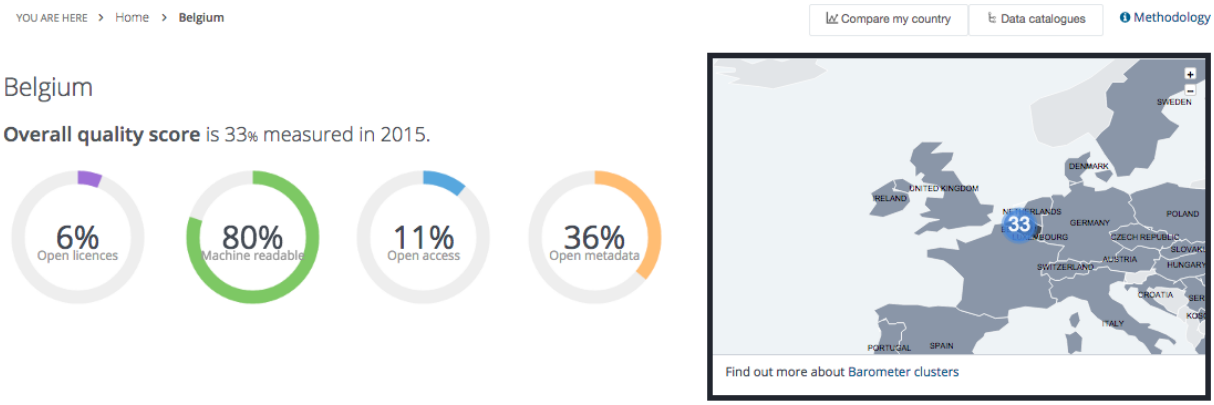

Figure 5-17: Compare my country

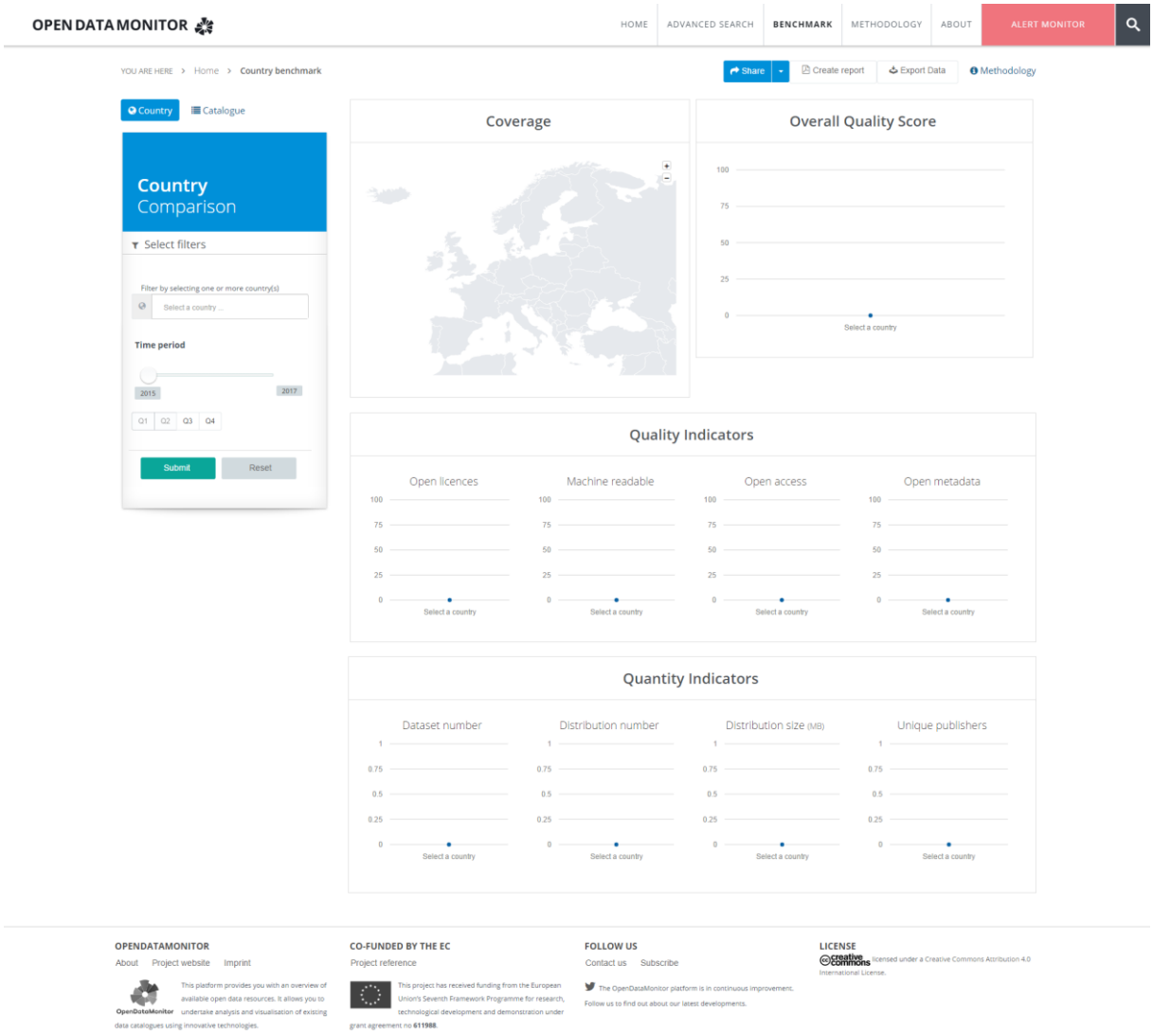

Figure 5-18: Benchmark – Comparison

On the first load of the dashboard the graphical content will be empty until the user is selecting the preferred entities for comparison.

Figure 5-19 is a print screen from a comparison between countries like Austria and France. The map is highlighted with the selection and all the graphs are showing the projection of various metrics over time. The time series graphs selected to show the various positions of the countries will change colour for each selection.

There is no limit on the selection of the countries chosen from the filter on the side bar. The user can select as many countries to compare as it is in interest for the context of data.

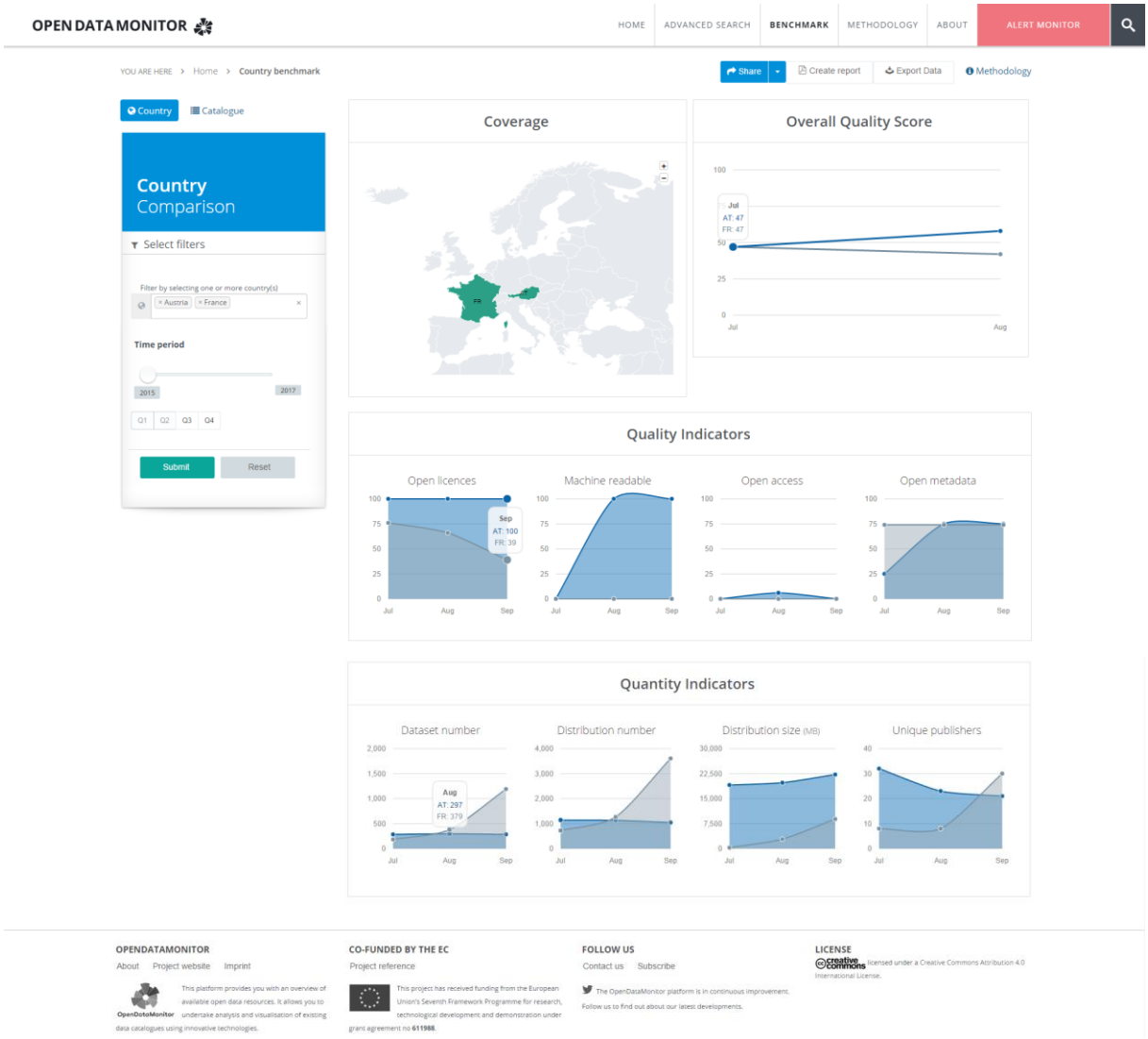

Figure 5-19: Benchmark - Country Comparison

Instead in figure 5-20, a catalogue comparison between two official catalogues (data.gov.uk and data.gov.gr) is been displayed highlighting the most performant catalogue. The countries in the map will highlight the respective countries where the catalogues are been found. The end user is able to select as many catalogues as is in its interest in order to have a full view of catalogues over time measured by key indicators.

In the example taken the official catalogue of United Kingdom is performing mostly in all the indicators displayed with exception in the last two months for machine readable and open data. The granularity of time gives a closer look in the work that these catalogues are doing over time to offer a clean and healthy open dataset.

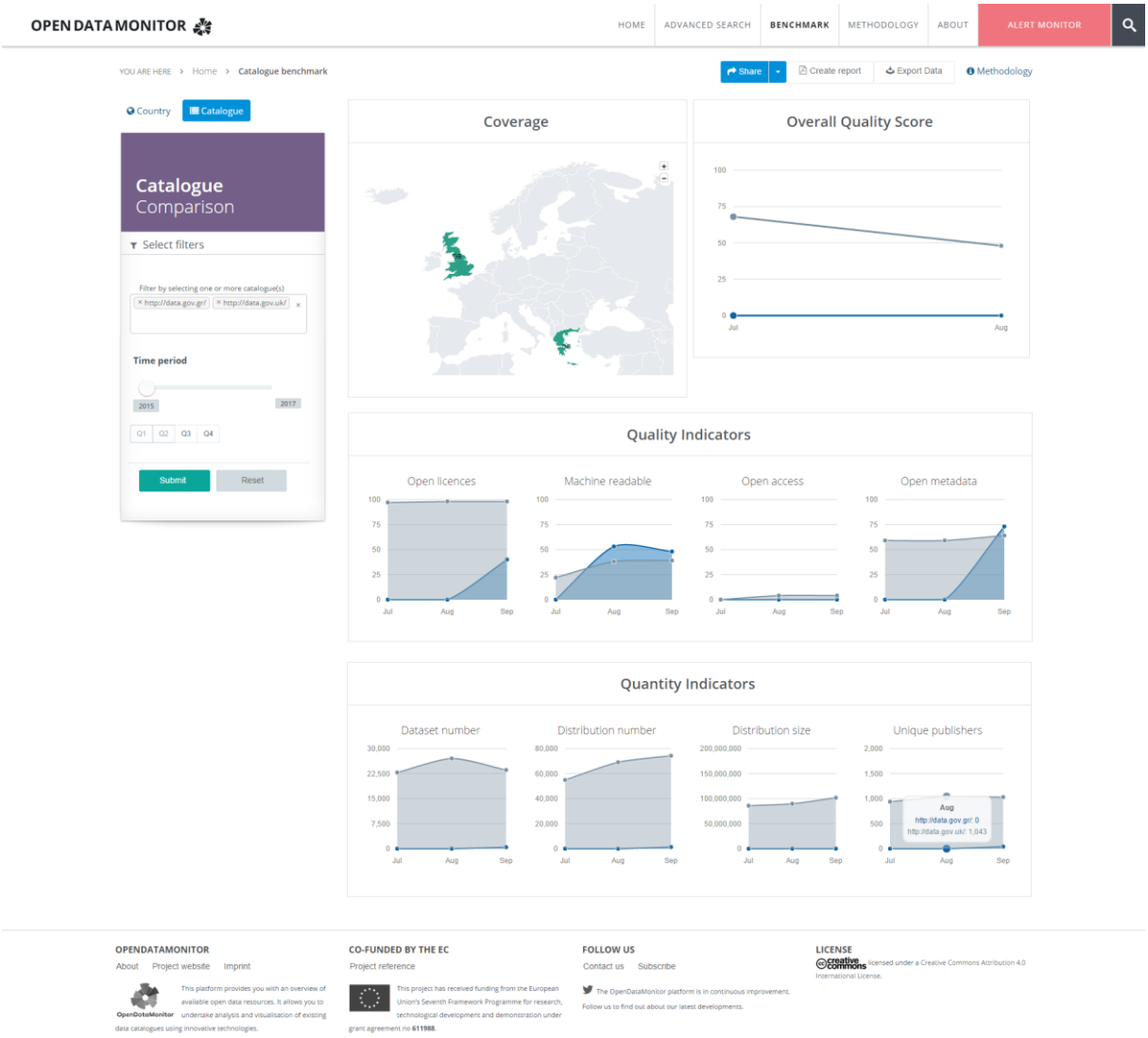

Figure 5-20: Benchmark - Catalogue Comparison

5.2.2. Advanced Search

The advanced search user interface has similar look and feel with previous data catalogue list view. The new changes added are the filter part. Considering the number of filter fields and the number of columns to be presented in the search results, the catalogue advanced search view (Figure 5-21) uses a left-right layout and the dataset advanced search view (Figure 5-22) uses a top-down layout.

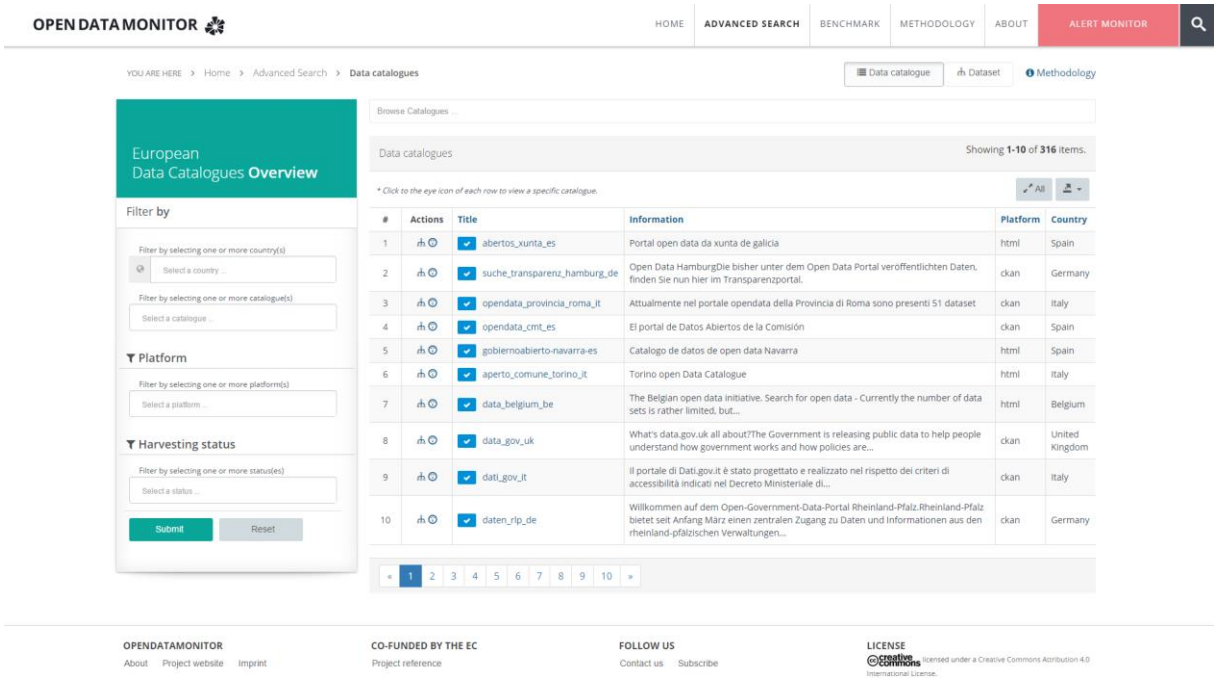

Figure 5-21: Advanced Search Data Catalogue View

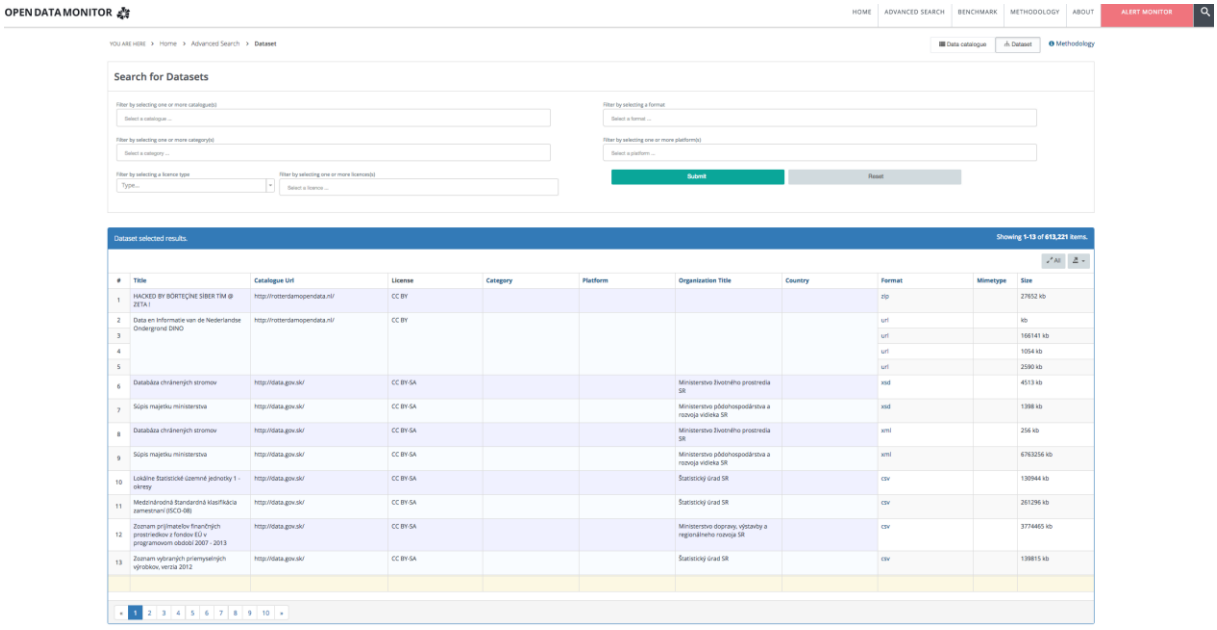

Figure 5-22: Advanced Search Dataset View

5.2.3. Social Plugin

On share to a specific page the end user will be prompted to add a custom content before sharing it on their own respective social media profile. In addition, the users are able to share the page as a link by email to the interested parties.

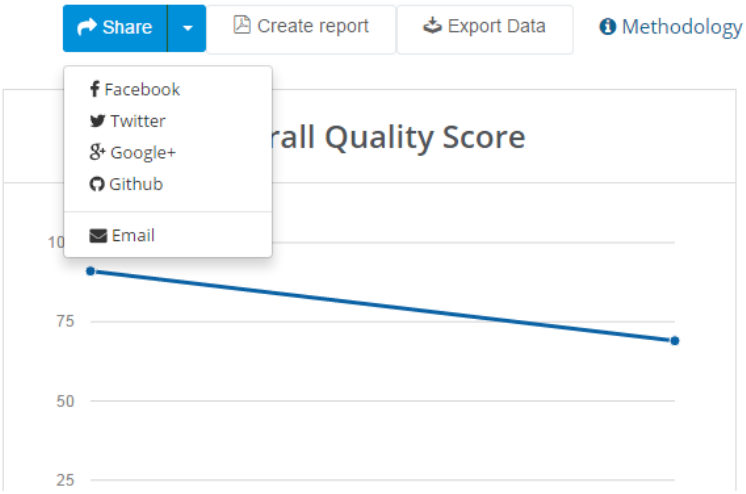

Figure 5-23: Share on social media

6 SUMMARY & OUTLOOK

The new features(Benchmarking Comparison dashboard and reporting, Social Plugin) and changes (Advanced Search, Metrics and charts) in the design are based on the user evaluation feedbacks (D4.2), potential user cases identified in previous deliverables (D3.2) and proposed mock-ups (D3.4, D3.7). In the second year of development, especially after the beta release in June 2015, the main application features of European, Country and Catalogue dashboard have gone through dramatic changes in terms of the metrics used and the charts types. Also the Benchmarking dashboard of catalogues and countries are re-designed in the August-September sprint. 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 particular, we verified the metrics and refined the definition and calculation accordingly to use the most relevant metrics to measure the performance of selected entities in both country and catalogue level. The enhanced search provides more sophisticated filters to make it easy for user to find the useful information. Social plugin adds value in promoting the results from ODM platform to more open data users who may find a central location to locate data catalogue/dataset, understand the performance of their catalogue/country, have an overview of the platform/license used, download harmonised data for their own analysis purpose, or even get more ideas on the backend technologies such as how each component of ODM works so that they can reuse or leverage part of the codes which have been shared to the community.