

Deliverable 6.3

Quality index of EU structural funds data

Dissemination Level	Public
Due Date of Deliverable	Month 21 31.01.2017
Actual Submission Date	17.03.2017
Work Package	WP6, Test Beds and Evaluations - Transparency
Task	T6.2
Type	Report
Approval Status	Approved
Version	0.2
Number of Pages	15
Filename	D6.3_ Quality index of EU structural funds data

Abstract: This simple benchmark focuses on the quality of the EU member states' beneficiary data released for the European Structural and Investment Funds for the funding periods of 2007-2013 and 2014-2020. This index serves as the basis of the analysis and finding contained in deliverable 6.6. Data quality report.

The information in this document reflects only the author's views and the European Community is not liable for any use that may be made of the information contained therein. The information in this document is provided "as is" without guarantee or warranty of any kind, express or implied, including but not limited to the fitness of the information for a particular purpose. The user thereof uses the information at his/ her sole risk and liability.



History

Version	Date	Reason	Revised by
0.1	09.02.2017	Internal revision	Anna Alberts
0.2	16.02.2017	External revision	Nick Aiossa

Author List

Organisation	Name	Contact Information
OKFDE	Michael Peters	michael.peters@okfn.de
OKFDE	Anna Alberts	anna.alberts@okfn.de

Executive Summary

This data index serves as basis for the report on the quality of the EU member states' beneficiary data released for the European Structural and Investment Funds for the funding periods of 2007-2013 and 2014-2020. EU Regulation No 1303/2013 from December 2013 requires the member states to create a single website providing all viable information on their operational programmes and publishing their beneficiary data in a machine-readable format. The Data index focuses on three main components. First, it considers the adherence to the aforementioned regulatory provisions. Second, it considers the accessibility and usability of the data via the managing authorities' websites and, third, the quality and format of this data.

For their previous project subsidystories.eu Open Knowledge International and Open Knowledge Germany collected all data for the 2007-2013 and 2014-2020 funding periods, which set the foundation for the quantitative analysis in this report. All EU member states' ESIF websites were analyzed and evaluated against the governing EU regulation with special attention towards usability, data access and their availability in English.

Abbreviations and Acronyms

CF	Cohesion Fund
CSV	Comma Separated Value
DG	Directorate General
EAFRD	European Agricultural Fund for Rural Development
EFF	European Fisheries Fund
ERDF	European Regional and Development Fund
ESF	European Social Fund
EU	European Union
HTML	HyperText Markup Language
JSON	JavaScript Object Notation
NUTS	Nomenclatura of territorial units for statistics
OP	Operational Programme
XLSX	Microsoft Excel File

Table of Contents

1 Introduction	7
2 EU Policy Background.....	7
3 Quality Index	9
3.1 Data Portals	10
3.2 Index: Practical Usability	11
3.3 Index: Regulatory Publication Requirements.....	12
3.4 Index: Data Formats	12
4 Conclusion	13
5 Appendix	13

List of Tables

Table 1: Overview Data Portals	10
Table 3: Data Formats ERDF 2014-2020	13

1 Introduction

This index was written for the EU financed project OpenBudgets.eu. It corresponds to the deliverable 6.3 Data Quality report of EU structural funds data. The data that this report relies on is based on earlier work by the Open Knowledge Foundation Germany and Open Knowledge International in their “Subsidystories.eu” project. In this project the ERDF, ESF and CF data for both the 2007-2013 and 2014-2020 period were collected for all EU member states. Data was mapped and visualized with the Open Fiscal Data package and is open and available at www.subsidystories.eu. The data index and benchmarks will relate to three main areas: practical usability, adherence to EU regulation, and the data formats available from the Member States.

2 EU Policy Background

To give some context to what the European Structural Investment Funds are and how they work, the EU’s investment policy will be discussed. The EU Commission laid out their Horizon 2020 strategy for generating smart, sustainable and inclusive growth in the EU. In order to achieve these goals, the EU manages the European Structural Investment Funds, which are the EU’s main investment policy tools. To assure that the funds are used to achieve the EU’s goals, detailed investment priorities and thematic objectives are defined, which function as guidelines for the use of the funds. The European framework constitutes funding periods of seven years with the last period ranging from 2007-2013 and the current period lasting from 2014 until 2020.

Institutionally, the member states and the European Commission (through its directorates general) negotiate a Partnership Agreement within the benchmarks that are set by the regulations for the structural and cohesion funds. Partnership agreements are contracts governing the funding process between the European Commission and the member states. Thereafter, the operational programme (OP) have to be submitted based on how applicants are planning to achieve the Commission’s goals by funding local projects. The applicants for these operational programmes are the member states’ regions as defined by the NUTS classification (Nomenclatura of territorial units for statistics). Within the regions a management authority has to be declared such as ministries of finance or regional administrations. While application is always handled by the region, countries with a strong central state often administer the funds on a national level. This leads to spending data being released on a national level. For countries with a federal structure such as Germany, Spain and Austria, data is usually published on the regional level.

The management authorities have to give detailed descriptions on their goals and how they plan to achieve these with the respective ESIF funds. Goals have to be in line with the thematic objectives and investment priorities published by the European Commission. After submitting the OP, they are reviewed by the responsible directorate general (DG). If



accepted, the management authorities receive the funds from the DG and use their own websites to advocate funding. Thereafter, individual project application starts.

The European Structural Investment Funds (ESIF) cover five different instruments:

- European Regional and Development Fund (ERDF)
- European Social Fund (ESF)
- Cohesion Fund (CF)
- European Agricultural Fund for Rural Development (EAFRD)
- European Fisheries Fund (EFF)

With subsidystories.eu, we focused on three of these ESIF funds: The ERDF and Cohesion Fund managed by the Directorate General for Regional and Urban Policy and the ESF overseen by the Directorate General for Employment, Social Affairs & Inclusion. While the ERDF aims to strengthen economic and social cohesion in the European Union by correcting imbalances between its regions ([here](#)), the ESF is Europe's main instrument for supporting jobs, helping people get better jobs and ensuring fairer job opportunities for all EU citizens ([link](#)).

While all member states can apply for ERDF/ESF funding, the Cohesion Fund only applies to member states whose Gross National Income (GNI) per inhabitant is less than 90 % of the EU average. For the current period this concerns: Bulgaria, Croatia, Cyprus, the Czech Republic, Estonia, Greece, Hungary, Latvia, Lithuania, Malta, Poland, Portugal, Romania, Slovakia and Slovenia.

While the EU provides spending data on the aggregate (member state or regional) level, this project gathered all available data on which beneficiaries receive European funding and which projects are implemented. Our aim is to improve fiscal transparency in the European Union by fostering the access to its spending data and allowing for cross country comparison for the first time.

3 Quality Index

EU member states have been required to publish the data online since the 2007-2013 period. However, the 2007 regulation was still vague and led to some member states publishing detailed datasets, while the majority only published basic information on beneficiary names, amounts and dates. The management authorities usually create a website regarding the European Structural Investment Funds (ESIF), where they offer information on funding opportunities for possible beneficiaries and list previous projects etc. In some cases, this means there is one website / online portal, where information on all funds (ERDF, ESF and CF if applicable) is provided such as France, Cyprus or Denmark. In countries with a decentralized state - like Germany, Austria and Belgium - regions function as management authorities and hence, publish the data on a regional website. For Germany's 16 regions this leads to 16 different websites, however, the websites are often separately distinguished by funds, meaning the actual number of websites for Germany is 27. You can find an overview on the country specific portals in table 1.

Country	EU Data Portal
Austria	http://www.esf.at/esf/service/dokumente-2007-2013/
Belgium Flanders	http://www.vlaio.be/
Bulgaria	http://umispublic.government.bg/
Croatia	http://www.strukturifondovi.hr/
Cyprus	http://www.structuralfunds.org.cy/
Czech Republic	http://www.dotaceeu.cz/cs/Informace-o-cerpani/Seznamy-prijemcu
Denmark	https://regionalt.erhvervsstyrelsen.dk/projekter-0
Estonia	http://www.strukturifondid.ee/programming-2014-2020/
Finland	https://www.eura2014.fi/rrtiepa/?lang=en
France	http://www.europe-en-france.gouv.fr/
Germany	http://www.esf.de/portal/DE/Startseite/inhalt.html
Greece	https://www.espa.gr/en/pages/default.aspx
Hungary	http://eupalyazatiportal.hu/
Ireland	http://eustructuralfunds.gov.ie/
Italy	http://www.opencoesione.gov.it/

Latvia	http://www.esfondi.lv/es-fondu-projektu-mekletajs
Lithuania	http://www.esinvesticijos.lt/
Luxembourg	http://www.fonds-europeens.public.lu/
Malta	https://investinginyourfuture.gov.mt/projects?lang=mt
Netherlands	https://www.europaomdehoek.nl/ ¹
Poland	http://www.mapadotacji.gov.pl/en
Portugal	https://www.portugal2020.pt/Portal2020/
Romania ²	http://www.fonduri-ue.ro/
Slovakia	https://www.itms2014.sk/
Slovenia	http://www.eu-skladi.si/
Spain	http://www.dgfc.sepg.minhafp.gob.es/sitios/dgfc/en-GB/Paginas/inicio.aspx
Sweden	http://projektbank.tillvaxtverket.se/projektbanken2020#page=eruf
UK - England	https://www.gov.uk/government/publications/european-structural-and-investment-funds-useful-resources

Table 1: [Overview Data Portals]

The EU provides an overview on some of the websites in their own portal [here](#). It is a good starting point, but not necessarily up to date. Online searches of “ERDF/ESF + beneficiary + respective country/region” usually lead to the required portals. While some websites are available in English, others are not and require using website translation. Obtaining the data can therefore be quite troublesome.

3.1 Data Portals

The following section focuses upon the ranking that we came up with concerning all 28 European data portals. The ranking is based on criteria such as availability of the website in English, ease of use, functionality and how easily beneficiary data can be found. The

¹ For the Netherlands and Romania, the data was found on different portals than officially indicated by the EU or on other portals. For Romania, the data for 2007-2013 is available on the Open Data Portal, and was partially send to us directly. For The Netherlands different files are available for the European Social Funds on national level, and for the ERDF on regional level in different formats and from different quality. For Chapter 3, we decided to only evaluate the data portals as indicated. However, in chapter 4 – 6, the data as eventually located was used in our evaluation.

² See Footnote 1.

[regulation](#) reads: “[...] giving examples of operations, by operational programme, on the single website or on the operational programme's website that is accessible through the single website portal; the examples should be in a widely spoken official language of the Union other than the official language or languages of the Member State concerned.”

We used the EU’s own data portal as a starting point for our search, and if a specific website was not included, we searched for it. The first obstacle when confronted with a foreign countries data portal is usually the language, even though a “widely spoken official language” of the Union is required, 12 out of the 28 countries do not provide any English assistance. This is problematic, because the websites have to be translated first, in order to allow for any further research. We used <http://itools.com/tool/google-translate-web-page-translator> for this task. It remains to be said, that even if websites offer translations, this does not guarantee their helpfulness. Often the translated pages just cover a small part of the original website and in some cases do not allow for finding the beneficiary data while in the English mode, such as the German and French portals.

Finding a coherent way of evaluating the country portals and the beneficiary data is difficult, due to their differences in conception. As discussed, countries with a strong federal state tend to distribute the ESIF funds on a regional level, leading to multiple and different portals. Some even have distinguished platforms for the ERDF and ESF. For the 2014-2020 period we looked closely at the ERDF data and respective portals, and noted if they included all or fund specific information. In case of countries that published the data regionally, we considered one regional dataset such as the Belgian region Flanders or the German region Berlin. However, it should be noted, that not all Belgian or German regions have published their data yet. In case there was no data available for the 2014-2020 period (Austria, Spain, Romania, and Cyprus), we still evaluated the webpages based on the 2007-2013 period.

3.2 Index: Practical Usability

The index and the corresponding scoring are a combination of a few simple questions that we wanted to be answered by the portal:

- Was the website available in English?
- How easily could the portal be located by using Google search?
- How long did it take to find the beneficiary data?
- Could the data be downloaded directly or did it require scraping?

These questions do have subjective nuances, e.g. finding the beneficiary data on the website can to an extent be fostered by luck of clicking on the correct subpage. However, this is influenced by the fact that the pages are available in English or follow a clear and intuitive structure. The subjectiveness of “ease of use” should be considered when viewing this ranking. Factors such as design or “look” of the website were neglected unless they specifically aided the access to beneficiary data. Furthermore, we are only considering the data format here and not the data quality, which will be evaluated by itself later on. Scores were awarded on a scale from 1-5 with one being the lowest and five the highest possible score. Countries that fulfilled all our criteria received a five, while minor issues led to a four, if

no data could be found, websites could not be located or other major issues existed they received a one.

3.3 Index: Regulatory Publication Requirements

The benchmark used to evaluate Member State data will also align to the regulatory requirements. The new EU legislation “Regulation (EU) No 1303/2013 of the European Parliament and of the Council of 17 December 2013” mandated the form the data should be presented in. The data shall be uploaded in the aforementioned online portals in a machine readable format and at least include the variables: beneficiary name, project name, operation summary, start & end date, total eligible expenditure, union co-financing rate, operation postcode, name of category of intervention and date of last update. 2014 – 2020 data is not yet available for every member state, because some have simply not released it yet.

3.4 Index: Data Formats

Our research confronted us with many different formats in which the data was presented. This is despite the fact that the regulation for the 2014-2020 period clearly states that machine-readable formats shall be used (such as CSV). The index will use this regulatory baseline and look at the machine-readability of the ERDF data. For this case we are counting XLS, XLSX and CSV as machine-readable, although only CSV truly is. However, XLS and XLSX can usually be converted to CSV rather easily.

Format (before scraping)	#	Format (after scraping)	#
JSON	0	JSON	6
CSV	4	CSV	26
XLSX	5	XLSX	0
XLS	13	XLS	0
WEB	6	WEB	0
PDF	37	PDF	0
Total	22	Total	32

Table 2: [Data Formats ERDF 2014-2020]

However, getting the data out of the PDF format is a lot more tedious, since the data cannot be accessed directly. In order to extract data from a PDF the file has to be “scraped” – that is an automated way to obtain the information from the original file has to be found. This can be done by coding, if you are an experienced developer or with automated tools such as Tabula.

To get an impression of the overall progress in data formats and a possible effect the newly introduced EU regulation might have had, table 4 is presented. It shows the distribution of data formats for the 2007-2013 and the 2014-2020 period. While more data for the 2007-2013 period was collected (75) vs. (47) for 2014-2020, the number of datasets in machine readable formats have improved. This is most visible in the number of datasets presented in PDF (49 in 2007 vs. 2 in 2014) and XLSX (4 in 2007 vs. 30 in 2014). This is a positive development that we want to highlight, although many of the datasets do not comply with the self-prescribed EU standards.

4 Conclusion

In the EU funds data quality index we established benchmarks regarding the portals, practical usability, regulatory requirements and data format will serve as a basis for evaluating the overall data quality of European Structural and Investment Funds for the funding periods of 2007-2013 and 2014-2020. The accompanying analysis and findings are contained in deliverable 6.6. Data Quality Report. Therein, we find that availability of the member states' ESIF websites in English language is still low. Furthermore, the use of open data formats such as CSV or XLSX requires further improvement. While not all regulatory requirements for publication are met in the 2014-2020 period, overall quality has improved in comparison to 2007-2013.

5 Appendix

Appendix I: List of variables used in fiscal data model

Name	Description	Variable Type
beneficiary_name	name of the beneficiary (person, company, organisation)	string
project_name	name of project	string
project_description	description of the project	string
project_id	unique code of the project (generated by authority itself)	numeric
beneficiary_person	name of person responsible	string
project_status	status of the project	string
starting_date	starting date of the project	numeric
completion_date	completion date of the project	numeric

approval_date	approval date of the project	numeric
final_payment_date	date on which the final payment was made	numeric
theme_name	name of the thematic objective	string
theme_code	code of the thematic objective	numeric
cci_program_code	CCI codes identifying operational programs	numeric
priority_label	description of the priority number of the grant agreement	string
priority_number	priority number of the grant agreement	numeric
management_authority	management authority	string
operational_programme	information which operational program the project is governed	string
total_amount	total cost of project	numeric
total_amount_eligible	total eligible expenditure	numeric
member_state_amount	amount that is awarded from national funds	numeric
eu_cofinancing_amount	amount of co-financing from the EU	numeric
eu_cofinancing_amount_eligible	amount of co-financing a project is eligible for	numeric
eu_cofinancing_rate	rate (percent) of co-financing from the EU	numeric
third_party_amount	total amount additional to the action over third party funding	numeric
fund_acronym	acronym of the fund (ERDF, ESF, CF)	string
beneficiary_address	full address of the beneficiary	string
beneficiary_city	city of beneficiary	string
beneficiary_postal_code	postal code of beneficiary	string
beneficiary_nuts_region	region matching the NUTS code	string
beneficiary_nuts_code	NUTS code of beneficiary region	numeric
beneficiary_county	county of beneficiary	string
beneficiary_country	country of beneficiary	string
beneficiary_country_code	two digit NUTS country code of beneficiary	numeric
beneficiary_url	URL of the project	string
source	a source url of the original data	string