Benchmarking Open Data Availability across Europe: The Case of EU Structural Funds

Although Open Government, Government 2.0 and Open Government Data have been at the centre of the debate on e-Government policy over the last two years, the European Union still lacks comparable data on transparency. European Regional Policy is the ideal context to test a *Benchmarking 2.0*: it involves all Member States and EU regions, influences national and regional policies and can push the transparency agenda in those areas of Europe where administrative culture and capacity is traditionally low. In this paper the datasets on beneficiaries of European Structural Funds provided by the Managing Authorities of the Operational Programmes of EU27 are evaluated through a scheme based on the 8 principles of Open Government Data. The paper compares the performance of the European Countries for the first time and sets new targets that could be considered as possible requirements for the next programming period.

Disclaimer:

The views expressed in this article are those of the author and, in particular, do not necessarily reflect those of the Ministry of Economic Development.

Luigi Reggi

Department for Development and Economic Cohesion, Italian Ministry of Economic Development

Keywords

European Structural Funds, regional policy, benchmarking, government transparency, open government data, methodology, method, evaluation

> The results of this web-based survey show that the European Cohesion Policy is only halfway to accomplishing a paradigm shift to open data, with differences in performance both between and - in some cases - within European Countries.

1. Introduction

1.1 Benchmarking Government 2.0: from e-Government interactivity to open data availability

More than two years ago, in this journal, David Osimo raised the question of how to benchmark e-Government in the web 2.0 era. After having analysed the traditional e-Government benchmarking method adopted at European level, he identified transparency as a flagship initiative for e-Government policy and Open Government Data (OGD) availability as the way to measure it (Osimo, 2008). Since then, the debate on the so-called 'Government 2.0' has rapidly evolved and the "2.0" meme has risen from obscurity to mainstream in e-Government policy (Osimo, 2010b).

The new European e-Government Action Plan 2011-2015 (European Commission, 2010c) highlights the two main sources of potential benefits of Government 2.0: transparency and data re-use. The former is a direct effect of opening up data and information on government decision making, such as laws and regulations, policies and finance. The latter, which implies a call for "raw data" (Robinson et al., 2010; Berners-Lee, 2006), focuses on the creation of value by combining data from different sources and making "mash-ups" and new applications, both for commercial and non-commercial purposes. The economic value generated by the exploitation of Public Sector Information (PSI) by the public and private sector is estimated as being significant by several studies (see among others Pollock, 2009; Dekkers et al., 2006).

Although these topics are at the centre of the current debate on e-Government policy, the need to develop and test new methodologies to benchmark open data availability among the Member States of the European Union remains largely unsatisfied. The new edition of the well-established supply side benchmarking exercise confirms the importance of updating and expanding the scope of the analysis by including new metrics on *"Transparent and Open Government"* (Capgemini, 2010, 134). Since traditional rankings on online services availability and sophistication can play a role in giving an impulse to countries still engaged in developing digital infrastructures and delivering their basic public services online (Reggi, 2009), new indicators on transparency should be introduced in a seamless way, while maintaining core indicators and comparability of datasets over time.

Osimo (2008) proposes a list of "basic public data" to be measured at EU level. The first in the list is "beneficiaries of public funding", which includes the beneficiaries of European Structural Funds. Regional policy is in fact the second biggest item in the EU budget after agriculture - it absorbs approximately one third of the total budget - and, moreover, it forces each Member State to share the same rules and regulations, which can improve the comparability of data. Beneficiaries of Regional Policy could indeed represent an interesting testing domain for a new open data benchmarking.

As shown in more detail in Section 3, though, the surveys commissioned in the last two years by the European Parliament and the European Commission were either too early (carried out when the 2007-13 period implementation was in its early phase) or focused mainly on qualitative information rather than on quantitative and comparable data.

The aim of this paper is therefore to provide an objective, web-based benchmarking of the publicly available data on projects and beneficiaries of the 2007-13 Structural Funds across Europe. The methodology and the results of this test (see Sections 4 and 5 respectively) could represent a methodological basis for an extended benchmarking exercise that should include other kinds of data from European Policies such as the beneficiaries of the Common Agricultural Policy.

2. The Regional Policy and the European Transparency Initiative

European Regional Policy (otherwise named European Cohesion Policy) *"aims to promote harmonious development of the Union and its regions by reducing regional disparities"* (Article 174 of the Treaty).

The policy *"underpins the growth model of the Europe 2020 strategy including the need to respond to societal and employment challenges all Member States and regions face. [...] The multilevel governance system for the policy helps to make the EU more visible to its citizens"* (European Commission, 2010a). The role of Structural Funds in financing the Europe 2020 strategy is in fact more and more significant (European Commission, 2010b).

Regional policy is implemented mostly thanks to two Structural funds, namely the European Regional Development Fund (ERDF) and the European Social Fund (ESF). ERDF is aimed at levelling economic differences among regions and it finances, for example, initiatives for research and innovation, local development and employment, infrastructure, and protection and improvement of the environment. ESF was established to improve the quality and accessibility of jobs and employment opportunities within the European Union.

The amount of Community resources dedicated to Regional Policy in 2007-13 is EUR 347 billion (European Commission, 2008), 3.75% of which is dedicated to ICT and Information Society (European Commission, 2007). It constitutes the second largest item in the Community budget after agriculture.

Financial resources are concentrated on the lagging regions that fall under the Convergence objective, with 81.5% of the investment available. The declared rationale of the Convergence objective is to promote growth-enhancing conditions and factors. Outside the Convergence regions, the Regional Competitiveness and Employment objective (Competitiveness) aims at strengthening competitiveness and attractiveness, as well as employment, especially through innovation and the promotion of the knowledge society. The European Regional Cooperation objective (Cooperation) strengthens cross-border co-operation through joint local and regional initiatives, trans-national co-operation and interregional co-operation and exchange of experience.

In addition to the Community financing, substantial national and regional budgets are mobilised, which must conform to EU rules and regulations. One of the goals of the Regional Policy is in fact to improve the quality of national and regional policies and to strengthen administrative capacity in the disadvantaged regions (see for example Bache, 2008; Baum & Marek, 2008; Barca, 2009).

Such a positive influence on national and regional policies could be leveraged also to foster transparency across Europe.

Structural Funds regulations for the 2007-13 programming period require the Managing Authorities (Member States and Regions managing an Operational Programme financed by Structural Funds) to publish the names of the beneficiaries, the name of the project co-financed with Structural Funds and the corresponding amount of public funding received. In fact, according to Article 69 of the Council Regulation No 1083/2006 of 11 July 2006 and repealing Regulation (No 1260/1999), *"the Member States and the Managing Authority for the operational programme shall provide information on and publicise operations and co-financed programmes. The information shall be addressed to European Union citizens and beneficiaries with the aim of highlighting the role of the Community and ensuring that assistance from the Funds is transparent"*. In particular, Commission Regulation No 1828/2006 of 8 December 2006 (art. 7) states that *"the managing authority shall be responsible for organising the publication, electronically or otherwise, of the list of beneficiaries, the names of the operations and the amount of public funding allocated to the operations"*.

Moreover, in November 2005 the European Commission launched a 'European Transparency Initiative', which is promoted and implemented through different regulatory texts and documents aiming at increasing financial accountability and strengthening personal integrity and institutional independence.

A Green Paper presented by the Commission on 3 May 2006 identifies the four main components of the ETI in (1) the public access to documents, (2) the rules and standards on professional ethics of public office holders in the European institutions, (3) the lobbying transparency and (4) the information on beneficiaries of EU funds.

In 2008 the Commission provided guidance to Member States on the practical implications of implementing the Transparency Initiative with a detailed *Guidance Note* that the European Commission and Member States agreed on in the COCOF of 23 April 2008 (European Commission, 2008b). The note commits to the Commission the coordinating role of facilitating access to the data available on the websites of the managing authorities and proposes a common standard for the publication of data, so as to enable interested parties to carry out consistent analyses across the EU. Although the set of minimum information is still relatively small and should be extended, the European Transparency Initiative certainly represents a breakthrough innovation in the way most European Countries implement public policy.

An *"indicative table for setting the list of beneficiaries of EU funding"* is annexed, focusing on 6 main standard designators of the database fields that should be included:

1. The name of beneficiaries (defined by Article 2 of the Council Regulation No 1083/06 as *"operator, body or firm, whether public or private, responsible for initiating and implementing operations. In the context of aid schemes under Article 87 of the Treaty, beneficiaries are public or private firms carrying out an individual project and receiving public aid"*).
2. The name of the operation (defined as *"a project or group of projects selected by the managing authority of the operational programme concerned or under its responsibility according to criteria laid down by the monitoring committee and implemented by one or more beneficiaries allowing achievement of the goals of the priority axis to which it relates"*).
3. The amount of public funding committed to the operation.
4. The amount of public funding paid to the beneficiary at the end of the operation.
5. The year of final payment.
6. The date of the last update.

3. Previous studies evaluating the lists of beneficiaries of Structural Funds

Two reports recently commissioned by European Institutions have dealt with the evaluation of existing data on projects and beneficiaries of European Structural Funds in the 2007-13 programming period.

The first study was funded by the European Parliament's Committee on Regional Development and presented in July 2010, though it was completed in June 2008, only one year after the beginning of the 2007-13 period (CSIL, 2008). The report, entitled "The Data Transparency Initiative and its Impact on Cohesion Policy", evaluates the implementation of the European Transparency Initiative

by providing some quantitative data and four case studies about Finland, Italy, the Netherlands and Poland.

The situation reported, mainly due to the very early phase of policy implementation, *"results in incomparable, often not machine readable and in some countries almost unusable data in different EU languages and different currencies"*. Only 78% of the European regions managing an ERDF operational programme provide the minimum information required. 19% provide a description of the operations, 41% a location of the projects, 27% the amount of national co-funding. Moreover, while 44% of EU regions publish data on the total amount of funding, only 32% of available datasets specify the amount of public money actually paid out. PDF is the prevailing format in which data are released (52%), followed by XLS (27%) and HTML (21%); a situation that had not changed almost two years later (Reggi, 2010).

The report draws some final recommendations:

- to provide additional essential information, such as contact details, localisation, project summaries, description of project partners, etc.
- to make databases fully searchable and compatible, so as to make possible an EU-wide outlook of the data.
- to describe the data in English and not only in the local language.

The second report - "Study on the quality of websites containing lists of beneficiaries of EU Structural Funds" by Technopolis Group - was funded by the DG Regional Policy of the European Commission (Technopolis Group, 2010). The study adopted mainly qualitative methods in answering its evaluation questions, and data are collected through a series of interviews and an online questionnaire.

The interviews addressed the questions of comparability and compatibility of data across the Member States and regions and concludes with a review of the technical approaches to the presentation of information.

Table 1 - Member state approaches in publishing data on beneficiaries of Structural Funds

Centralised / national approach	Regional / decentralised approach
Cyprus, Czech Republic, Denmark, Estonia, Finland, France, Greece, Hungary, Latvia, Lithuania, Luxembourg, Malta, Romania, Slovakia, Slovenia, Spain, Sweden	Australia, Belgium, Bulgaria, Germany, Ireland, Italy, Netherlands, Poland, Portugal, United Kingdom

Source: Technopolis Group, 2010.

The study also classifies the Member States in two groups: countries following a centralised approach, which have developed centralised information systems resulting in one shared publicly available dataset, and countries following a regional approach, which implies decentralised and fragmented databases.

4. Methodology and data collection

A web-based survey was conducted in October 2010 in order to explore the availability and quality of the lists of projects and beneficiaries of the European Regional Development Fund (ERDF) and the European Social Fund (ESF) published by the managing authorities.

Table 2: Operational programmes analysed by objective, fund, regional scope, type of information system used and Member State

	Regional objective*			Fund		Nat / reg		Info system		All Operational Programmes
	Convergence	Comp.	Cooperation	ERDF	ESF	National or multireg.	Regional	Centralized	Not centralized	
BG	7	-	-	5	2	7	-	-	7	**7**
BE	2	8	-	4	6	1	9	-	10	**10**
CZ	15	2	-	14	3	8	9	17	-	**17**
DK	-	2	-	1	1	2	-	2	-	**2**
DE	14	22	-	18	18	1	35	-	36	**36**
EE	3	-	-	2	1	3	-	3	-	**3**
GR	14	-	-	10	4	5	9	14	-	**14**
ES	23	22	-	23	22	7	38	45	-	**45**
FR	9	27	-	31	5	5	30	36	-	**36**
IE	-	3	-	2	1	1	2	-	3	**3**
IT	19	33	-	28	24	9	43	-	52	**52**
CY	1	1	-	1	1	2	-	2	-	**2**
LV	3	-	-	2	1	3	-	-	3	**3**
LT	4	-	-	2	2	4	-	4	-	**4**
LU	-	2	-	1	1	2	-	2	-	**2**
HU	14	1	-	13	2	8	7	15	-	**15**
MT	2	-	-	1	1	2	-	2	-	**2**
NL	-	5	-	4	1	5	-	-	5	**5**
AT	2	9	-	9	2	1	10	-	11	**11**
PL	21	-	-	20	1	5	16	21	-	**21**
PT	11	3	-	10	4	7	7	-	14	**14**
RO	7	-	-	5	2	7	-	7	-	**7**
SI	3	-	-	2	1	3	-	3	-	**3**
SK	10	1	-	9	2	9	2	11	-	**11**
FI	-	7	-	5	2	-	7	7	-	**7**
SE	-	9	-	8	1	1	8	9	-	**9**
UK	6	16	-	16	6	-	22	-	22	**22**
Cross-border cooperation	-	-	54	54	-	-	-	-	54	**54**
Interreg cooperation	-	-	3	3	-	-	-	-	3	**3**
Trans-national cooperation	-	-	14	14	-	-	-	-	14	**14**
Total	**190**	**173**	**71**	**317**	**117**	**108**	**254**	**200**	**234**	**434**

* Programmes belonging to both Convergence and Competitiveness objectives are classified into Convergence objective

Source: Author's elaboration based on European Commission - DG Regional Policy data (June 2009)

The survey is based on the official database on the approved Operational Programmes provided by the DG Regional Policy in June 2009. All the 434 Programmes approved at that time were taken into

account[1]. As showed in Table 2, the programmes are classified into various categories depending on the objective (Convergence, Regional Competitiveness and Employment, European Regional Cooperation), the fund (European Regional Development Fund, European Social Fund) and the regional scale (National or Multi-regional, Regional). The studies on the lists of beneficiaries of structural funds mentioned before helped classify the programmes also by type of information system used (centralised, decentralised). Programmes with regional cooperation objectives, by definition, involve more than one Member State, and therefore could not be connected to any particular country.

Datasets published on the web were identified through a visit to the URIs indicated by the managing authorities and reported in the *Inforegio*[2] web site (managed by the DG Regional Policy of the European Commission) or in the web site of the European Social Fund[3] (managed by the DG Employment). When the link was broken or unavailable, a search in the websites of regional operational programmes and of regional managing authorities was performed.

Transparency of every single operational programme is assessed against a 4-stage model reflecting the quality of data provided. Stage models have been widely used in the literature to define a sort of evolutionary path toward excellence in service delivery, within both an "e-Government" (Capgemini, 2009; United Nations, 2010, Baum & Di Maio, 2000; Layne & Lee, 2001; Andersen & Henriksen, 2006) and "Goverment 2.0" (Osimo, 2008; Johnson, 2010) paradigm.

In order to analyse the datasets from an *open data* perspective, the eight principles of Open Government Data are considered as evaluation variables (Open Government Working Group, 2007). These principles were developed by 30 Open Government advocates[4] during a meeting in Sebastopol, California on 7-8 December, 2007, coordinated by Tim O'Reilly of O'Reilly Media and Carl Malamud of Public.Resource.Org. The group suggested eight desirable properties for government data that, if implemented, *"would empower the public's use of government-held data"*. These eight principles are now considered as a worldwide *de facto* standard for open data evaluation and are cited as a key reference by practitioners and academics in policy discussions as well as in the top academic journals in the field (Davies, 2010; Osimo, 2010a; Bertot et al, 2009; Johnson, 2010). They are identified as follows[5]:

1. *Complete* - All public data are made available. Public data are data that are not subject to valid privacy, security or privilege limitations
2. *Primary* - Data are as collected at the source, with the highest possible level of granularity, not in aggregate or modified forms
3. *Timely* - Data are made available as quickly as necessary to preserve the value of the data
4. *Accessible* - Data are available to the widest range of users for the widest range of purposes
5. *Machine processable* - Data are reasonably structured to allow automated processing.
6. *Non-discriminatory* - Data are available to anyone, with no requirement of registration.
7. *Non-proprietary* - Data are available in a format over which no entity has exclusive control
8. *License-free* - Data are not subject to any copyright, patent, trademark or trade secret regulation. Reasonable privacy, security and privilege restrictions may be allowed.

1 A list of operational programmes can be found at http://ec.europa.eu/regional_policy/country/prordn/index_en.cfm
2 http://ec.europa.eu/regional_policy/country/commu/beneficiaries/index_en.htm
3 http://ec.europa.eu/employment_social/esf/discover/article_7093_en.htm
4 A list of the 30 OGD advocates can be found at http://public.resource.org/open_government_meeting.html
5 See also the additional notes: http://www.opengovdata.org/home/8principles/annotations

The definition of the 4 stages for each principle (see Table 3) is mainly based on the W3C and UK Central Office of Information guidelines. For each principle a score is attributed as follows:

Stage 0 = 0%

Stage 1 = 33%

Stage 2 = 66%

Stage 3 = 100%

A composite indicator measuring the overall quality of each Operational Programme is obtained as a simple mean of the scores attributed to the 8 principles.

Table 3: Evaluation scheme and description of the stages

	Principle	Description	Stage 0 (0%)	Stage 1 (33%)	Stage 2 (66%)	Stage 3 (100%)
1	Complete	All public data are made available. Public data are data that are not subject to valid privacy, security or privilege limitations.	Data not available	Low completeness: presence of project, beneficiary and total cost of the project	Good completeness: all European Transparency Initiative (ETI) recommendations are met	High completeness: ETI recommendations met + detail for EU or other kind of co-financing funds and status of the project provided.
2	Primary	Data are as collected at the source, with the highest possible level of granularity, not in aggregate or modified forms.	Data not available	Low granularity: aggregated data	-	High granularity: information available for each beneficiary (raw data)
3	Timely	Data are made available as quickly as necessary to preserve the value of the data.	No info on update	-	Day or month of the last update is provided	Information on the frequency of update is provided
4	Accessible	Data are available to the widest range of users for the widest range of purposes.	No accessibility: broken link to the DB from Inforegio or ESF websites, no description provided	Low accessibility: the link from Inforegio or ESF websites is correct	Good accessibility: DB available by 3 clicks from the HP.	High accessibility: good description of the data or metadata is provided. The DB is located by 3 clicks from the HP. Columns are translated into English.
5	Machine processable	Data are reasonably structured to allow automated processing.	Not machine-processabile format: PDF, DOC, results displayed in “HTML reports”	Machine-processable format: CSV, HTML, XLS, ODT	Data interchange format: XML, JSON	Linked data: RDF
6	Non-discriminatory	Data are available to anyone, with no requirement of registration.	Requirement of registration and approval	Requirement of registration and download for everyone	-	Non-discriminatory: no requirement of registration

	Principle	Description	Stage 0 (0%)	Stage 1 (33%)	Stage 2 (66%)	Stage 3 (100%)
7	Non-proprietary	Data are available in a format over which no entity has exclusive control.	Proprietary formats: XLS, DOC	-	-	Open formats and standard formats: CSV, RDF, XML
8	License-free	Data are not subject to any copyright, patent, trademark or trade secret regulation. Reasonable privacy, security and privilege restrictions may be allowed.	Not license-free: data are subject to copyright, patent, trademark or trade secret regulation	No license specified. Terms of use as given in law	-	Licence is compatible with Creative commons (by and reuse for commercial) or Open data commons

5. Results

Table 4: Average scores of the Operational Programmes of EU27, by objective, fund, regional scope and type of information system used

	Overall performance	Complete	Primary	Timely	Accessible	Machine processable	Non-discriminatory	Non-proprietary	License-free
Regional objective									
I - Convergence (n=190)	**49.6%**	48.7%	92.1%	54.3%	53.2%	15.1%	100.0%	0.5%	33.0%
II - Competitiveness (n=173)	**48.6%**	40.5%	91.3%	54.1%	54.3%	11.4%	100.0%	4.0%	33.0%
III - Cooperation (n=71)	**46.3%**	28.8%	76.5%	62.3%	65.5%	4.2%	100.0%	0.0%	33.0%
One way ANOVA test (F value)	2.34*	18.86***	9.04***	2.08	4.37**	13.07***	-	3.94**	-
Bonferroni test (F value)	*I > II: 0.01* *I > III: 0.03** *II > III: 0.42*	*I > II: 0.08**** *I > III: 0.19**** *II > III: 0.11****	*I > II: 0.007* *I > III: 0.15**** *II > III: 0.14****	*I > II: 0.002* *I > III: -0.07* *II > III: -0.08*	*I > II: -0.011* *I > III: -0.12*** *II > III: -0.11***	*I > II: 0.03** *I > III: 0.1**** *II > III: 0.07****		I > II: -0.03** I > III: 0.005 II > III: 0.04**	
Fund									
ERDF (n=317)	49.3%	42.5%	88.0%	59.8%	56.6%	13.0%	100.0%	1.9%	33.0%
ESF (n=117)	46.8%	41.2%	92.6%	44.0%	53.1%	8.7%	100.0%	1.7%	33.0%
t-test (t value)	-2.130***	-0.473	1.511*	-4.881***	-1.018	-2.504***	-	-0.125	-
Regional scope									
National or multireg. (n=108)	48.3%	47.3%	87.6%	53.3%	50.3%	14.7%	100.0%	0.0%	33.0%

	Overall performance	Complete	Primary	Timely	Accessible	Machine processable	Non-discriminatory	Non-proprietary	License-free
Regional (n=254)	49.5%	43.9%	93.8%	54.6%	55.1%	12.9%	100.0%	3.1%	33.0%
t-test (t value)	1.046	-1.193	2.294***	0.336	1.456*	-0.967	-	1.868**	-
Information system									
Centralised (n=200)	52.1%	47.6%	98.8%	52.3%	61.4%	19.8%	100.0%	3.5%	33.0%
Not centralised (n=234)	45.8%	37.5%	81.0%	58.3%	50.7%	5.1%	100.0%	0.4%	33.0%
t-test (t value)	*-6.160****	*-4.311****	*-6.931****	*2.011***	*-3.589****	*-10.873****	-	*-2.382****	-
All Operational Programmes	**48.7%**	42.2%	89.2%	55.5%	55.6%	11.9%	100.0%	1.8%	33.0%

* Significant at 10% level. ** Significant at 5% level. *** Significant at 1% level

Bonferroni, Scheffe and Sidak tests provided the same results in terms of significance

First of all, findings suggest that European Cohesion Policy is only halfway to accomplishing a paradigm shift to open data, which is ideally correspondent to the 100% score. The overall performance of all Operational Programmes (48.7%) is mainly driven by *Non-discriminatory* (100%) and *Primary* (89.2%) principles, which make a major contribution to the average score. This could be considered a direct effect of the current regulations of Structural Funds. In fact, the provision of the highest possible level of granularity (project and beneficiary) is one of the requirements of regulations, while the publication without restrictions could also be interpreted as mandatory. On the contrary, the aspects not covered either by regulations or the European Transparency Initiative show very low results; this is the case of the format in which data are published (*Machine-processable* and *Non-proprietary* are the principles with the lowest values, 11.9% and 1.8% respectively). These findings seem to imply that managing authorities of the programmes are more interested in formally meeting the requirements of the regulation than pursuing real transparency.

Considering the variation among the different categories, we first notice that, on average, programmes belonging to the *Convergence* and *Competitiveness* objectives show higher scores on overall quality (49.2% and 48.6% respectively) than those belonging to the *Cooperation* objective (46.3%)[6]. A statistically significant difference emerges between the programmes funded by European Regional Development Fund (49.3%) and by European Social Fund (46.8%), while no significant variation is found between Regional and National or Multiregional Programmes.

A considerable difference (*t value* is significant at 1% level) in performance is shown when comparing datasets that are shared and centralised at national level (52.1%) with those which are managed by a single regional authority (45.8%). This variation is also statistically significant with regard to all the indicators examined, and is probably due to the fact that a centrally managed programme has the advantage that information flows are easier to manage and local actions are more easily coordinated.

6 Bonferroni, Scheffe and Sidak tests (Hochberg & Tamhane, 1987), which all provide the same results, indicate that the only statistically significant difference in overall performance is between the average scores of the Convergence and the Cooperation objectives.

With regard to the *Non-discriminatory* and *Licence-free* principles, all Operational Programmes obtain the same scores. Every dataset in EU27 is publicly available with no requirement for registration, and a licence is never specified. On the contrary, the principles *Complete* and *Machine processable* show the highest variance and, as already said, are quite important in explaining the variation of the overall performance.

On average, the completeness of information provided is 42.2%, far below the 66% level (stage 2), which means that the ETI requirements are still far from being met.

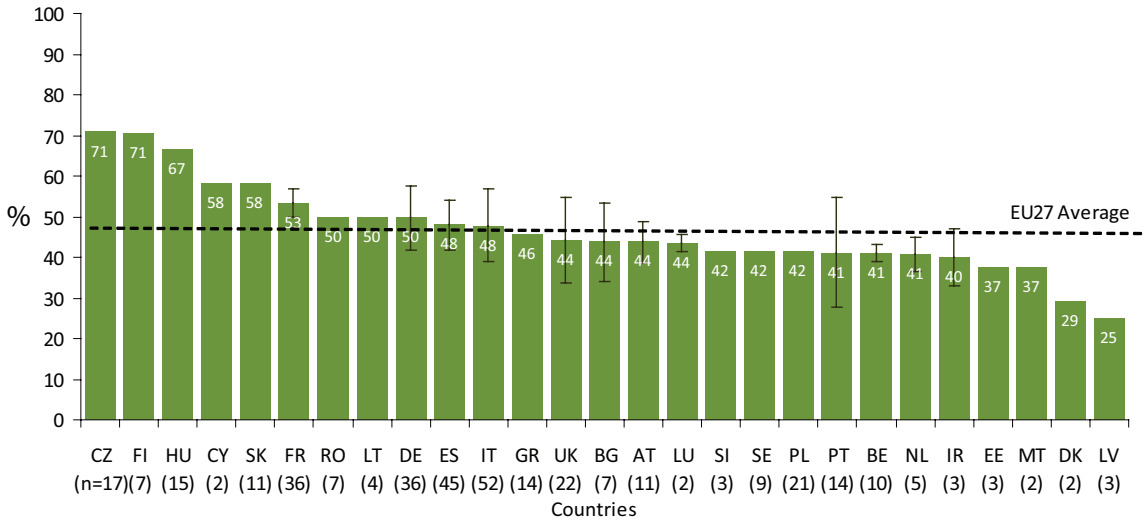
The two principles *Machine processable* and *Non proprietary* seem even more urgent to be applied. The way data are formatted and delivered makes a big difference. Government-produced reports, charts, and analyses can be very valuable, but *"it is essential to also publish the underlying data itself in a computer-friendly format that makes it easy for the vibrant community of civic technologists to make and share a broad range of tools for public engagement"* (Robinson, Yu and Felten, 2010).

For now, the situation is rather discouraging, with results showing very low scores for both principles (11.9% for *Machine processable* and 1.8% for *Non proprietary*). PDF is by far the most common format in which data is published, followed by XLS and DOC, while none of the programmes use data interchange formats such as XML or *linked data* formats such as RDF (Berners-Lee, 2006). The only open format actually used is CSV. This implies that data publicly available on Structural Funds are not really "open data" as commonly defined (see for example the Open Knowledge Definition[7] and Chernoff, 2010).

Better results are shown regarding the *Primary* principle. Most of the information on beneficiaries is presented at the highest level of disaggregation (that is, the beneficiary of funding), as required by the regulation. Otherwise, the information is provided in aggregated tables, which, by the way, can hardly be defined "lists of beneficiaries".

As for the *Timely* principle, it is quite common to find information on the day or month of last update as recommended by the European Transparency Initiative (stage 2). However, the stage 3 (declaration of the update frequency) is almost never reached.

Graph 1 - Average scores and standard deviation of the Operational Programmes of EU27 (Cooperation objective not included), by Member State



7 http://www.opendefinition.org/okd/

Finally, the results show a considerable variation in the overall quality of data among Member States. Best performing countries such as the Czech Republic and Finland obtain a score of 71%, while the worst performing Member State is Latvia with 25%.

It is worth noting that countries from the eastern Europe often appear in the first half of the chart. A possible explanation for this may rest in the specific actions and positive influence that the DG Regional Policy of the European Commission has exerted on these countries during the last few years in the official Monitoring Committees.

The chart also illustrates the disparities within the Member States by showing the standard deviation of the scores obtained by each country. As explained before, the variation equals zero in presence of an integrated information system at the national level. Portugal shows the highest dispersion from the average (13.5%), followed by the United Kingdom (10.5%), Bulgaria (9.6%) and Italy (8.9%).

7. Conclusions and policy recommendations

This empirical study, based on indications of previous studies and on existing evaluation schemes, is the first attempt to compare transparency of the Cohesion Policy across Europe and could represent a first step in developing a new European benchmarking framework aimed at comparing European countries in terms of the availability and quality of Open Government Data provided.

Structural Funds are the ideal context for testing a new methodology for Open Government Data evaluation because they involve all Member States and regions with common rules and regulations and influence national and regional policies and strategies, and so they play a pivotal role in spreading the administrative culture of transparency and openness across Europe.

The results of this web-based survey show that the European Cohesion Policy is only halfway to accomplishing a paradigm shift to open data, with differences in performance both between and – in some cases - within European countries.

Low scores are attributed to the formats the authorities are choosing when publishing their data on the web, while other indicators such as the level of granularity are positively influenced by the requirements of current regulations.

The use of open, machine-processable and *linked-data* formats have unexpected advantages in terms of transparency and re-use of the data by the public and private sector. These aspects are already highlighted in the current e-Government Action Plan, in the European Directive on the re-use of Public Sector Information and in the Open Government policies that are being implemented in many OECD Countries around the world.

The application of these technical principles does not need extra budget or major changes in government organisation and information management; nor does it require the update of existing software and infrastructures. What is needed today is the promotion among national and local authorities of the culture of transparency and the raising of awareness of the benefits that could derive from opening up existing data and information in a re-usable way.

As demonstrated, in Cohesion Policy implementation a key role is played by the regulations and the consequent actions that the European Commission or the national authorities should put in place. In this regard, the evaluation scheme proposed sets of specific targets in terms of quality, openness

and completeness that could be considered as possible requirements for the next funding period. For example, Managing Authorities of Structural Funds should:

- use only open and machine-processable formats. In particular, the linked data paradigm should be adopted
- provide a comprehensive description of the data including information on the frequency of update
- improve the data accessibility by sharing it with other administrations so as to develop new and larger data-sets enabling direct comparisons between countries.

Acknowledgments

I would like to thank Chiara Assunta Ricci for her contribution to the data collection. I also thank Gianfranco Andriola, Sergio Scicchitano and David Osimo for their useful comments on a previous version of this paper.

Furthermore, I am grateful to the referee for the careful review and the valuable suggestions that helped improve the paper.

References

Andersen, K. V., Henriksen, H. Z. (2006). E-government maturity models: Extension of the Layne and Lee model. Government Information Quarterly, 23(2), 236.

Bache, I. (2008). Europeanization and multilevel governance, Lanham, MD: Roman and Littlefield Publishers.

Barca, F. (2009). An agenda for a reformed cohesion policy A place-based approach to meeting European Union challenges and expectations. Independent Report prepared at the request of Danuta Hübner, Commissioner for Regional Policy, Brussels, retrieved December 30, 2010 from http://ec.europa.eu/regional_policy/policy/future/barca_en.htm.

Baum, C., Di Maio, A. (2000). Gartner's four phases of e-government model. Stamford, Ct., Gartner Group, 21, 12-6113.

Baum, M. & Marek, D. (2008). EU Cohesion Policy and Sub-National Authorities in the New Member States. Contemporary European Studies, 2, 5-20.

Bertot, J.C., Jaeger P.T., Shuler, J.A., Simmons, S.N. & Grimes, J.M. (2009). Reconciling government documents and e-government: Government information in policy, librarianship, and education. Government Information Quarterly, 26, 433–436.

Berners-Lee, T. (2006). Linked Data. International Journal on Semantic Web and Information Systems. w3c.org. Retrieved December 30, 2010 from http://www.w3.org/DesignIssues/LinkedData.html.

Capgemini (2010). Digitizing Public Services in Europe: Putting ambition into action - 9th Benchmark Measurement, retrieved February 23, 2011 from http://ec.europa.eu/information_society/newsroom/cf/item-detail-dae.cfm?item_id=6537.

Chernoff, M. (2010). What "open data" means – and what it doesn't, retrieved December 30, 2010 from http://opensource.com/government/10/12/what-"open-data"-means---and-what-it-

doesn't?sc_cid=70160000000Sz26AA.

Central Office of Information (2010). Underlying data publication: guidance for public sector communicators, website managers and policy teams, retrieved December 30, 2010 from http://coi.gov.uk/blogs/digigov/wp-content/uploads/2010/07/TG135-Underlying-Data-Publication-v0-4.pdf.

CSIL - Centre for Industrial Studies (2008). The Data Transparency Initiative and its Impact on Cohesion Policy, report for the European Parliament's Committee on Regional Development, retrieved December 30, 2010 from http://www.europarl.europa.eu/document/activities/cont/201006/20100611ATT75925/20100611ATT75925EN.pdf.

Davies, T. (2010). Open data, democracy and public sector reform. A look at open government data use from data.gov.uk, MSc Dissertation, University of Oxford.

Dekkers, M., Polman F., te Velde R. & de Vries, M. (2006). MEPSIR - Measuring European Public Sector Information Resources, retrieved December 30, 2010 from http://ec.europa.eu/information_society/policy/psi/actions_eu/policy_actions/ mepsir/index_en.htm.

European Commission (2003). The Role of eGovernment for Europe's Future, COM(2003) 567 final, Brussels.

European Commission (2006). European Transparency Initiative, COM (2006) 194, Brussels.

European Commission (2007). Regions delivering innovation through cohesion policy, Commission staff working document, SEC(2007) 1547, Brussels.

European Commission (2008a). Communication on the results of the negotiations concerning cohesion policy strategies and programmes for the programming period 2007-2013, COM(2008) 301 final, Brussels.

European Commission (2008b). European Transparency Initiative: Implementation of the Financial Regulation regarding the publication of data on beneficiaries of Community Funds under the shared management mode, Note to the COCOF 07/0071/03 of 23/04/2008, Brussels.

European Commission (2010a). Conclusions of the fifth report on economic, social and regional cohesion: the future of cohesion policy, COM(2010) 642 final, Brussels.

European Commission (2010b). Regional Policy contributing to smart growth in Europe 2020, COM(2010) 553 final, Brussels.

European Commission. (2010c). The European eGovernment Action Plan 2011-2015 - Harnessing ICT to promote smart, sustainable & innovative Government, COM(2010) 743, Brussels.

Hochberg Y. & Tamhane A. C. (1987). Multiple Comparison Procedures. New York: Wiley.

Johnson, C. (2010). Evaluating Open Pages, Sunlight LABS, Sunlight Foundation, retrieved December 30, 2010 from http://sunlightlabs.com/blog/2010/evaluating-open-pages/.

Layne, K. & Lee, J. (2001). Developing fully functional E-government: A four stage model. Government Information Quarterly, 18(2), 122-136.

Open Government Working Group (2007). 8 Principles of Open Government Data, retrieved December 30, 2010 from http://www.opengovdata.org/.

Osimo, D. (2008). Benchmarking eGovernment in the Web 2.0 era: what to measure, and how.

European Journal of ePractice, n.4. Retrieved December 30, 2010 from http://www.epractice.eu/en/document/287915.

Osimo, D. (2010a). A short history of Goverment 2.0: from cool projects to policy impact. In Gotze, J. and Pedersen, C.B. (Eds.) State of eUnion, 21Gov.net.

Osimo, D. (2010b). Government 2.0 - Hype, Hope, or Reality? European Journal of ePractice, n.9. Retrieved December 30, 2010 from http://www.epractice.eu/en/editorial/345317.

Pollock, R. (2009). The Economics of Public Sector Information. Retrieved December 30, 2010 from http://econpapers.repec.org/paper/camcamdae/0920.htm.

Reggi, L. (2009). Don't forget 'traditional' e-government. Retrieved December 30, 2010 from http://www.luigireggi.eu/Innovation-policies/Home/Entries/2009/12/12_DONT_FORGET_TRADITIONAL_E-GOVERNMENT.html.

Reggi, L. (2010). The open data paradigm and the European structural funds. Retrieved December 30, 2010 from http://www.scribd.com/doc/33597904/The-Open-Data-Paradigm-and-the-European-Structural-Funds.

Robinson, D. & Harlan, Y., Zeller, W., & Felten, E. (2009). Government Data and the Invisible Hand. Yale Journal of Law and Technology, 11, 160.

Technopolis Group (2010). Study on the quality of websites containing lists of beneficiaries of EU Structural Funds, retrieved December 30, 2010 from http://ec.europa.eu/regional_policy/country/commu/pdf/final_report_erdf_en.pdf.

United Nations (2010). E-government survey 2010. Leveraging e-government at a time of financial and economic crisis, retrieved December 30, 2010 from http://www2.unpan.org/egovkb/global_reports/10report.htm.

W3C (2009). Improving Access to Government through Better Use of the Web, retrieved December 30, 2010 from http://www.w3.org/TR/2009/WD-egov-improving-20090310/.

Author

Luigi Reggi
Department for Development and Economic Cohesion,
Italian Ministry of Economic Development
luigi.reggi@gmail.com
http://www.epractice.eu/en/people/157670