 doi:10.20944/preprints201801.0066.v1

Case Report

Data visualization of European regional operational programmes: unleashing the informative potential of open data for performance assessment

Emanuele Frontoni [1,*] and Roberto Palloni [2]

[1] Dipartimento di Ingegneria dell'Informazione (DII), Università Politecnica delle Marche, Via Brecce Bianche, 60131 Ancona, Italy; E-Mail: e.frontoni@unipvm.it

[2] Dipartimento di Ingegneria dell'Informazione (DII), Università Politecnica delle Marche,Via Brecce Bianche, 60131 Ancona, Italy and t33 srl, Via Calatafimi, 60121, Ancona, Italy; E-Mail: r.palloni@t33.it

* Author to whom correspondence should be addressed; E-Mail:e.frontoni@unipvm.it; Tel.: +39-071-2204458; Fax: +39-071-2204224.

Abstract: The implementation of the European Cohesion Policy aiming at fostering regions competitiveness, economic growth and creation of new jobs is documented over the period 2014-2020 in the publicly available Open Data Portal for the European Structural and Investment funds. On the base of this source, this paper aims at describing the process of data mining and visualization for information production on regional programmes performace in achieving effective expenditure of resouces.

Keywords: cohesion policy; data visualization; open data.

1. Introduction

The EU Cohesion Policy is one of the western world's largest local and regional development policy operating under broadly one overall legal and institutional framework. It targets all regions and cities in the European Union and is aimed at fostering regions competitiveness, economic growth and creation of new jobs. The planned resources over the period 2014-2020 by different funds are more than EUR 350 billion, almost one third of the total budget of the Union.

Each european regions can access the planned financing in the different funds according to specific investment strategies defined within their Operational Programmes (OP). These documents have a predefined structure for the selection and categorisation of the investment decisions and their reporting each year. In particular, projects selected, expenditure, output and result indicators have to be monitored by the Managing Authorities (MAs) responsible for the management of funds and shall be reported in the framework of the Annual Implementation Report (AIR).

In the last years the political context and the financial and economic pressure on countries budgets have imposed a change in term of accountability and justification of public expenditure, a stronger need to inform taxpayers on public investments and to broaden the debate on cohesion policy and its future orientation.

At the same time, compared to the previous programming period 2007-2013, the information management capacity and availability of quality data have largely improved with the development of many agile solutions for data usability.

Furthermore, the current legislative framework has a stronger focus on result-orientation on the performance of regions in achieving investment objectives defined during the programme design phase in terms of development needs and benefits for citizens.

 doi:10.20944/preprints201801.0066.v1

For these reasons, the reported data on financing and achievements under all the ESI Funds 2014-2020 are available under the criteria of transparency and accountability at the Open Data Portal for the ESIF[1]. The inteded users of these data is anyone interested in monitoring the development of the policy and in particular citizens of the Union, Member State administrations, EU Institutions, policy makers, researchers and practitioners in regional development studies. However, the indicators and visualizations currently provided allows only for a broaden and general view of the policy evolution and the structure, size and complexity of data represents a major challange for any user, mainly interested in the progress and performance status of her own area. At this regard, the paper describe the development of an application infrastructure[2] and the relative algorithms developed in the R language to syntetize and visualize the large sets of data in insightful indicators and statistics. Finally, the paper provides some inputs as regards the main messages and concepts to observe in the data comparison.

2. Literature review and other tools for open data analysis

The increasing and heterogeneous group of data users debating the implementation of programmes and the use of structural funds require an information dissemination strategy based on generally understandable concepts. At this regard, the use of visualizations as the easiest and fastest tools for human eye perception of patterns and trends recognition appears to be the best solution. However, the vastity of multi-dimensional information available rises the problem of successfully and easily stimulate visual reasoning synthetizing data using relatively simple tools.

Researchers are analysing increasingly large economic data sets generated in greater volumes adopting adequate tools and technologies. Big data often offers valuable information to be extracted and interpreted and time when simple bar charts or scatter plots were enough is long gone. Thus, the development of advanced data visualization techniques is becoming a necessary and challenging area of research and interest. Data visualization can help in making sense of large data sets by presenting the contents in an innovative visual format that does not require multiple tables or lots of rows and columns. Furthermore, the connection between several data sources generates newer and larger datasets leading to further discovery and information.

However, the increased complexity and volume of data collected, stored and made available by instituions and public bodies does not advance smoothly. Literature shows that open data government datasets still presents several barriers: lacks of adequate collection, classification, processing and presentation tools; non-standardized data description and formats, missingness or inchoerent data,thus poorly usable among different users and analytical approaches;[1].

IT investments and skills devoted mainly to storing systems, architecture, software, hardware, security, networks and Web technologies without an explicit purpose of data exploitation adapts poorly to the new paradigm on the use of data as an asset for business intelligence and data science. This in turn affects the benefits of open data initiatives and sharing, especially at even lower level (i.e. local administrations, municipalities, etc).

As a response to the challenges of managing vast amounts of government data and making it accessible for different purposes and informational needs, Dawes explains the concepts of stewardship and usefulness. At this regards, among "stewardship proposals" for improvement he suggests to create and improve metadata for each data source, improve the data management system, adopt standard data formats. As "usefulness proposals", he suggests to provide easy-to-use basic features and improve and enhance searching and display of data [2]. Noveck [3] adds that is it also important to own high quality standards for dissemination among different needs and usage across citizens and other social actors. Merino et al. [4] consider the delivery of public data as opportune and reliable for better

---

1 https://cohesiondata.ec.europa.eu/
2 https://cohesion-data-visualization.shinyapps.io/ms-op/

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 8 January 2018 doi:10.20944/preprints201801.0066.v1

decisions making in government as well as for government accountability of public decisions and actions. The use of different technology tool for implementing open data initiatives is recognised as a "fit-the-right-tool-for-the-job" situation, i.e. each complex economic, social, political issue and the data it generates relate to different approaches and methods for information production and use [5],[6],[4],[7].

Government open data across different end-users are available by the intensive use of technology as IT tools and Web applications [1], [3].

IT tools and web applications are currently the engine of the debate based on open data as they allow to both provide the "raw material" to different typologies of users and receive new information and data from the same users, either being decision-makers, analysts, researchers and citizens [8],[9],[10]. There is a wide range of different technological tools available for policy analysis and data visualization. The use of flexible and powerful information technologies and various analytical methods are supported by several open data initiatives. This scenario is constantly evolving, but a brief overview of some common tools and platforms used to visualize and analyse open data is as follows:

- ESIF - Viewer[3] is a tool to search planned investments in European Structural and Investment Funds (ESIF) data (ERDF, CF, ESF and YEI). The tool contains data from the ESIF Operational Programmes (OP). The amounts in this tool are presented at regional level and include data from regional OPs, but also shares of national and transnational cooperation programmes. The user can search for planned investments per country, region, OP-type and different categories of intervention;
- ICT Monitoring[4] contains data from the ESIF Operational Programmes (OP) on planned ICT related investments. The amounts in this tool are presented at regional level. Users can search within three broad dimensions, amounts, keywords and financial forms;
- Regional Benchmarking[5] is an interactive tool for Regional Benchmarking which helps identifying structurally similar regions across Europe through statistical indicators;
- EU Trade[6] is a fully interactive web-based application for the visualization and the analysis of inter-regional trade flows and the competitive position of regions in Europe. The purpose of this tool is to make possible to assess regional assets and analyse a region's economic position as a first fundamental step in the process of building place-based and evidence-based regional policies and smart specialisation strategies;
- R&I Regional Viewer[7] allows to visualize and compare Research & Innovation investments under different funding channels and EU programmes across EU Regions, i.e. economic indicators from Eurostat, planned R&I-related investments under ESIF, and Horizon 2020 funding captured by stakeholders;
- YDS - Your Data Stories[8] is a platform that helps make sense of open and social data;
- ROUTE-TO-PA[9] is a multidisciplinary innovation project, that, by combining expertise and research in the fields of e-government, computer science, learning science and economy, is aiming at improving the impact, towards citizens and within society, of ICT-based technology platforms for transparency;
- Smarticipate[10] gives citizens access to data about their city enabling them to better support the decision-making process. Residents will also play an active role in verifying and contributing to data;

---

3 http://s3platform.jrc.ec.europa.eu/esif-viewer
4 http://s3platform.jrc.ec.europa.eu/ict-monitoring
5 http://s3platform.jrc.ec.europa.eu/regional-benchmarking
6 http://s3platform.jrc.ec.europa.eu/s3-trade-tool
7 http://s3platform.jrc.ec.europa.eu/synergies-tool
8 https://yourdatastories.eu/
9 http://routetopa.eu/
10 https://www.smarticipate.eu/

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 8 January 2018 doi:10.20944/preprints201801.0066.v1

- Big Data Europe[11] undertake the foundational work for enabling European companies to build innovative multilingual products and services based on semantically interoperable, large-scale, multi-lingual data assets and knowledge, available under a variety of licenses and business models;
- OpenBudgets[12] offers a toolbox to everyone who wants to upload, visualise and analyse fiscal data. From easy to use visualisations and high-level analytics to fun games and accessible explanations of public budgeting and corruption practices along with participatory budgeting tools, it caters to the needs of journalists, researchers, policy makers and citizens alike.
- OpenCoesione[13] shares information on the italian projects financed through cohesion policy resources.

3. Data description

The data available in the Open Data Portal for the ESIF covers more than 530 Operational Programmes under the five ESI Funds: the European Agricultural Fund for Rural Development (EAFRD), the European Regional Development Fund (ERDF), the European Social Fund (ESF) with distinct data for the Youth Employment Initiative, the Cohesion Fund (CF) and the European Maritime and Fisheries fund (EMFF). Data are available in three financial datasets related to planned, implemented and payed resources and on a single achievement dataset with data on selected common indicators targets and implementation. The two most important financial variables related to the performance of regions in implementing their operational programmes are the project selection (resources allocated to investments) and the expenditure declared (resources actually disbursed to beneficiaries) as reported by the Managing Authorities of the programmes. The progress and performance of each Operational Programme is monitored against the financial planned amount decided during the planning phase at the beginning of the programming period in 2014.
These data are available disaggregated by fund, Operational Programme, Priority Axis, Thematic Objectives[14] (i.e. the macro priorities of investment of the policy) and category of regions (more developed, less developed, transition). In terms of update, while the financial planned is subject to update only in case of within OP reallocation of resouces, the financial implementation data are updated three times per year in the end of January, July and September. Likewise, disaggregation also applies to common indicators data whose update of implementation data is scheduled at the end of each year whereas targets are not subject to variations unless OPs modifications occurs. Vastity of information, complexity of data structure and timing of update suggests the adoption of agile tools for easily fetch, parse, aggregate and visualize information instantaneously.

4. Cohesion open data: API and information management

In order to allow the largest degree of accessibility and exploitation by specialists and general public, datasets are accessible and usable in different format from the ESIF portal. Among these, a web service ensures the programmatic and continuous access to programmes information through an API. Data are exposed through several endpoints in a JSON structure that allows a fast fetching and parsing of data. The availability of data as web service has driven the architecture of the visualization application namely requiring the development of specific fetching, parsing and plotting functions. The following figure shows an overview of the application architecture.

---

11 https://www.big-data-europe.eu/
12 https://openbudgets.eu/
13 http://www.opencoesione.gov.it/
14 http://ec.europa.eu/regional_policy/en/policy/how/priorities

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 8 January 2018 doi:10.20944/preprints201801.0066.v1

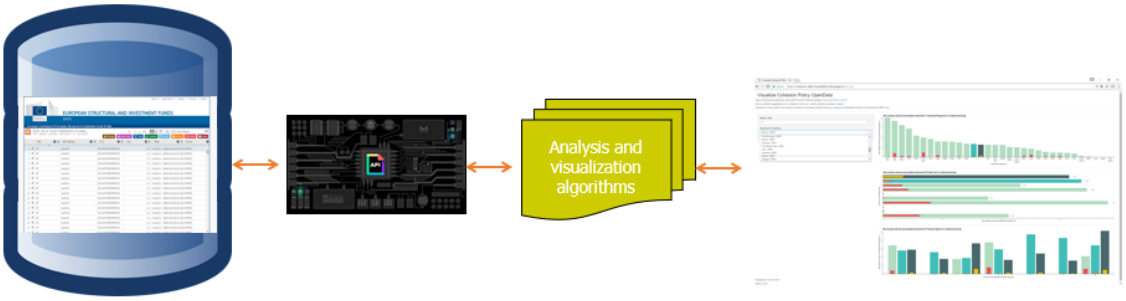

Figure 1. Web application architecture

When the application loads, data are parsed from the JSON endpoints exposing the database tables and are processed server-side to extract, calculate and plot the OP information based on the user inputs, namely the Member State and the name of the Operational Programme of interest. The simple inputs menu allows to search for specific Member State and Operational Programme but also to access the information of all the other EU programmes for an immediate comparison even beyond national borders.

5. Data visualization and interactive statistics

As the main objective of the data analysis and visualization shall be to easily compare regional programmes performance to justify EU investment and inform taxpayers on the progress of deployed resources, all visualisations use a benchmarking approach either between different geographical levels (EU, Member State, Operational Programme) or over time, observing the progress since the beginning of the programming period. As regards the design and aesthetic of the figures, four specific features have been taken into account when developing the visualization:

1. the barplot has been adopted being one of the most common and easy typology of data visualization, especially for policy makers and general public users;
2. the barplot has been improved through the logic of the progress bar and nesting the two main variables, namely the resources allocated for selection and disbured for expenditure: this visualization approach highliths their strict dependence and warns on anomalous progressing patterns (e.g. high ratio of selection and low ratio of expenditure);
3. as already discussed, benchmarks either in terms of space or time have been added within the same visualization aiming at emphasising the performance as a relative concept;
4. the adoption of dimension-specific patterns for easier concept-insight association, i.e. OPs in descending order for ranking focus, Priority Axis with coordinate flip as a sort of race line, Thematic Objectives as groups of bars to compare specific policy intervention fields.

These typology of visualization have received feedbacks and validation by a set of final users.

5.1. Developing performance indicators

A preliminary aggregation and cumulation of data by geographical level allows to calculate the main magnitudes used to develop the performance indicators. These are mainly represented by ratios of progress on planned, namely the Project Selection as share of Planned Financing (EUR) and the Expenditure Declared as share of Planned Financing (EUR).

$$\text{Rate of selection} = \frac{\sum_{op=1}^{N} Selection}{\sum_{op=1}^{N} Planned} * 100 \quad (1)$$

$$\text{Rate of expenditure} = \frac{\sum_{op=1}^{N} Expenditure}{\sum_{op=1}^{N} Planned} * 100 \quad (2)$$

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 8 January 2018 doi:10.20944/preprints201801.0066.v1

5.2. Informative data visualization

The calculation of indicators by itself does not serve the purpose of providing information to the user. For this reason, the visualization should be able to provide as many relevant information as possible without affecting the comprehension of the messsage. Thus, a trade-off between the complexity of the figure in terms of variables and dimensions considered and the informative power of the figure has to be considered when structuring the view. At this regards, depending on the specific information to transfer, each view aggregates and groups spatial or time dimensions as presented in the following figures:

- **Ranking OPs within the same MS**: ratio of selection and expenditure by OP, comparing all OPs in each MS with reference to the EU level in decreasing order;

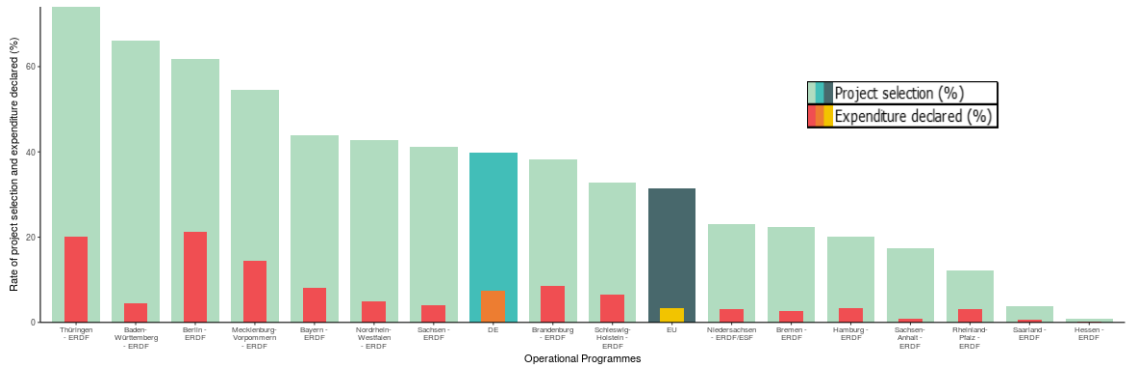

Figure 2. Rate of project selection and expenditure declared of Germany Operational Programmes (share of planned financing)

The algorithm behind the visualization ranks the OPs within a specific MS by rate of selection as well as the MS and EU values. The figure shows that, despite the highest level of selection, the first two regions have lower levels of expenditure. Despite some programmes show a very low level of progress, Germany and the majority of german OPs are above the EU average both in terms of selection and expenditusre. However, in terms of planned resources , percentages of implementation do suggest a slow advancement with only a maximum of 20% of resources disbursed at almost half of the programming period 2014-2020.

- **Comparing Priority Axes of investment within OP**: ratio of selection and expenditure by Priority Axis, comparing all the axis in each OP with reference to the overall OP level, the MS and EU progress;

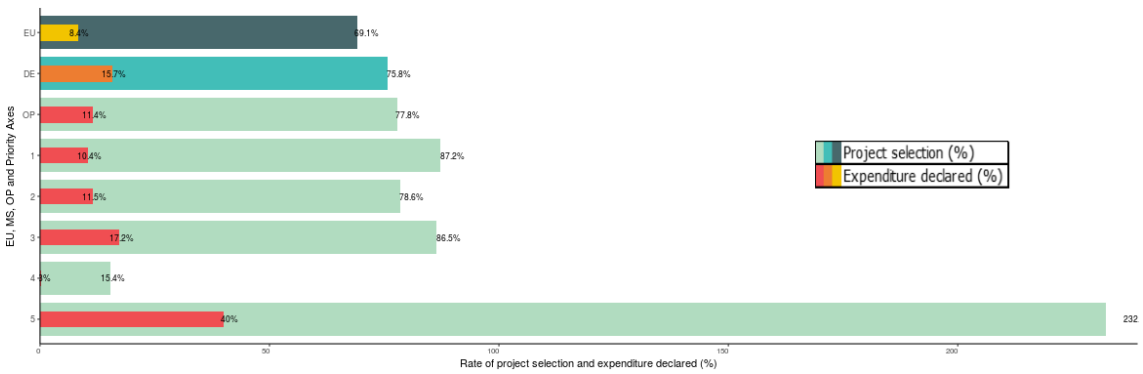

Figure 3. Rate of project selection and expenditure declared by Priority Axis (share of planned financing)

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 8 January 2018 doi:10.20944/preprints201801.0066.v1

The figure shows a progress of the overall OP under assessment almost in line with the EU and MS progress. Almost all the Priority Axis of the OP are progressing in terms of selection more than the EU and MS level. However, Axis 4 shows a very low rate of selection and absence of expenditure affecting the overall performance of the OP. This warns on possible issues and obstacles in the delivery of resources for the specific type of investments related to the priority and affects the homogeneous progress of the OP.

- **Comparing progress over time**: ratio of selection and expenditure by year between 2014 and 2017 with an overall visualization of the OP (left figure) progress and the development over time of the OP specific axes (right figure);

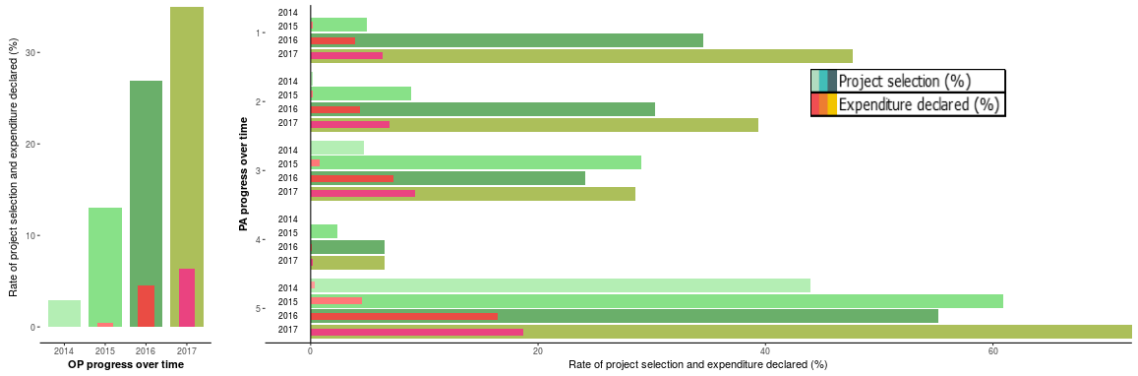

Figure 4. Rate of project selection and expenditure declared over time: OP and Axes details

Coherently with the natural progress of OP over the programming period, the left-side figure shows the evolution of selection and expenditure over time with almost absence of resources either allocated or disbursed in 2014 but an almost constant increase over the following years. The right-side view shows the progress over time of each Priority Axis with a similar pattern of development and, as already described, a critical situation in Axis 4 were in the last two years there hasn't been any change either in selection or expenditure.

- **Comparing areas of investment**: ratio of selection and expenditure by Thematic Objectives of the overall OP with benchmark to the Member State and EU levels.

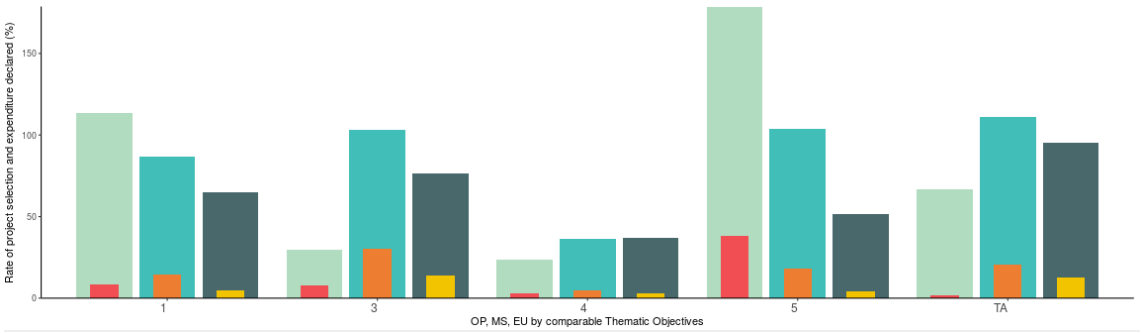

Figure 5. Rate of project selection and expenditure declared by Thematic Objective (share of planned financing)

The goal of these objectives is to focus financial resources on areas that deliver the highest benefits to citizens, fostering synergies between these fields and avoiding an excessive fragmentation of funding. This view allows for two typologies of comparison: comparing levels within TOs and

 doi:10.20944/preprints201801.0066.v1

comparing each level between TOs. The figure shows a relative better selection performance of the OP compared to the MS and the EU level in TO 1 - *Strengthening research, technological development and innovation* and TO 5 - *Promoting climate change adaptation, risk prevention and management*. On the contrary, other OPs are doing better with the resources dedicated to the other TOs as MS values are higher. As regards the comparison of the different TOs of the OP, it shows a higher degree of progress variability if compared with the MS and EU variability between TOs.

6. Conclusions

The main aim of this paper is to describe the functioning of an agile web tool for data visualization of ESI funds to anyone interested in monitoring the development of the regional programmes and in particular citizens of the Union, Member State administrations, EU Institutions, policy makers, researchers and practitioners in regional policy. The political context and the financial and economic pressure on EU countries and regions budgets of the recent years have imposed a change in term of accountability and justification of public expenditure, a stronger need to inform taxpayers on public investments and to broaden the debate on cohesion policy and its future orientation. The analysis developed through the web tool are intended as a way of improving the awareness of regions on their effective management of funds and in informing the decision-making process with real-time available data. The benckmarking approach adopted has several limitations but also possible fields of improvements as it does not refer to measures based on the whole set of regions but only to levels and time comparison. However, beyond the realm of cohesion policy and structural funds, this data-driven approach could be further used for other aspects of regional planning and decision making. Moreover, this methodology could prove to be useful in the context of data-driven evaluation and policy-learning, especially taking into account the evolution of data over time. As regards this last point, it is both useful and productive for regions to put more effort into monitoring and analysis of implementation data to maximise impact of public resources in the current period of financial and economic pressure of countries budgets.

Abbreviations

The following abbreviations are used in this manuscript:

CF	Cohesion Fund
EAFRD	European Agricultural Fund for Rural Development
EMFF	European Maritime and Fisheries Fund
ERDF	European Regional Development Fund
ESF	European Social Fund
MA	Managing Authorities
OP	Operational Programme

References

1. Dawes, S.; Helbig, N. Information strategies for open government: Challenges and prospects for deriving public value from government transparency. *Electronic government* **2010**, pp. 50–60.
2. Dawes, S.S. Stewardship and usefulness: Policy principles for information-based transparency. *Government Information Quarterly* **2010**, *27*, 377–383.
3. Noveck, B.S. *Wiki government: how technology can make government better, democracy stronger, and citizens more powerful*; Brookings Institution Press, 2009.
4. Merino, M.A.G.; Cejudo, D.; Sour, G.M.; Santibáñez, L.; Mariscal, L.; Casar, J.; Maldonado, M.A.; Gil-García, C.; Bracho, J.R.; Merino, T. Problemas, decisiones y soluciones: enfoques de política pública. Technical report, Fondo de Cultura Económica, 2010.

5. Birkland, T.A. *An introduction to the policy process: Theories, concepts and models of public policy making*; Routledge, 2014.
6. McCool, D. *Public policy theories, models, and concepts: An anthology*; Pearson College Div, 1995.
7. Sabatier, P.A. The need for better theories. *Theories of the policy process* **1999**, *2*, 3–17.
8. Hood, C.C.; Margetts, H.Z. *The tools of government in the digital age*; Palgrave Macmillan, 2007.
9. Hood, C. Intellectual obsolescence and intellectual makeovers: Reflections on the tools of government after two decades. *Governance* **2007**, *20*, 127–144.
10. Kő, A.; Gábor, A.; Szabó, Z. Policy modeling in risk-driven environment. Proceedings of the 5th International Conference on Theory and Practice of Electronic Governance. ACM, 2011, pp. 195–203.