

Visual Correlation Analysis to Explain Open Government Data based on Linked-Open Data for Decision Making

Dirk Burkhardt^{1,2}, Kawa Nazemi^{1,2}, Mohsen Parisay¹, Jörn Kohlhammer¹

¹Fraunhofer Institute for Computer Graphics Research, Darmstadt, Germany

²TU Darmstadt, Department of Computer Science, Darmstadt, Germany

Abstract

Public authorities normally consider statistical data about indicators in their decision makings. Such valid kind of data allows an objective observation about indicator developments over time. In case of a significant deviation from the normal indicator level, it is difficult to understand the reasons for upcoming problems. In this article we present an approach that allows an enhanced information gathering through an improved information overview about the depending aspects to such an indicator by considering governmental data-sources that provide also other types of data than just statistics. Even more, our approach integrates a system that allows generating explanations for Open Government Data, especially to specific indicators, based on Linked-Open Data and shows it in graphical form to enable a fast overview gathering. This allows decision-makers to get hints for unexpected reasons of concrete problems that may influence an indicator.

1. Introduction

Decision-making in public authorities is often based on statistical governmental data. Here, statistical data about certain indicator were used and analyzed for deviation from normal ranges, and therewith indicate a possible problem. However for some deviations it is difficult to find reasons that have affected this deviation. The default way is to search for reasons in additional data-sources, contact directly an expert of the field, or gather and analyze citizens' opinions, e.g. through surveys.

Another aspect is how experts deal with information. In particular on municipality level a variety of tools for different kind of data and task is used and even more, there are just a few tools that are able to provide information overview of different data sources simultaneously. But having exactly such an overview of available information enables a more efficient decision making.

This paper describes on the one hand an approach to handle multi-data through multi-visualization for a personalized information overview that enables experts to investigate more information sources and generate a visual cockpit [15] (dashboard) with

information visualizations that allows a more appropriate decision making. On the other hand we describe a new approach for visually interlinking Open Government Data with Linked-Open Data to generate and visualize explanations for certain indicator data. This is beneficial for problem finding in policy making, especially if the reason finding is complicate because of the problem complexity. Through the analysis and comprehension of statistical data against entries and properties from Linked-Open Data, correlation were extracted that may include possible reasons that can enlighten indicator deviations from the normal range.

2. The Policy Modeling Process and the Role of Statistics

The term policy is defined as “a theoretical or technical instrument that is formulated to solve specific problems affecting, directly or indirectly, societies across different periods of times and geographical spaces” [1]. In a public and political context it can result in a law. In perspective of the creation of policies, policy modeling can be defined as “an academic or empirical research work, that is supported by the use of different theories as well as quantitative or qualitative models and techniques, to analytically evaluate the past (causes) and future (effects) of any policy on society, anywhere and anytime” [1]. So the major focus lays on the policy and the causes and effects on the society. The creation of policies consists of a number of tasks and involves many stakeholders. To bring them in an efficient and effective order, policy modeling can be understood as a process, where is defined what actor has what task at which certain time. This arrangement of tasks and stakeholders, regarding the development of policies, is commonly named as policy modeling process.

2.1. Process Definitions in Public Authorities

The work in public authorities is majorly defined by concrete (administrative) processes. This should ensure an efficient and effective work. Conventional processes as they are defined in existing works ([2],[3],[25]), are using ICT and new technologies in a more traditional manner. Modern processes with

focus on the general alignment of ICT using majorly an abstract policy-making process, e.g. ([4],[5],[25]). Most of these process definitions include a gap in bringing the detailed policy definition and inclusion of ICT within the policy making process together [25]. This gap, and thereby the solution, is faced in a couple of research projects, among others, in the European research project ePolicy (<http://epolicy-project.eu>) and FUPOL (<http://fupol.eu>). In FUPOL there already exists a detailed policy process definition (see [6],[22]). To each process step a number of ICT features are aligned. The process orients on conventional process definitions and were just enhanced for the inclusion of ICT, which should enable an easier installment in public authorities.

2.2. Objective vs. Subjective Data in Policy Modeling

In spite of these advanced ICT-focusing policy processes, the most authorities do not use such a process and in fact they also use modern ICT approaches only in a limited way. Features, as for instance Social Media Analysis, are rarely used in public authorities (in contrast to a number of enterprises, which already use it for e.g. brand sentiment observations). Currently, the most work is done based on objective data, majorly in form of statistics.

The reason for this can be seen in two facts. The first is that most of the modern ICT features, e.g. Opinion Mining and Social Media, are new and also today only a small number of software products are available on the market that deals with these data in perspective of policy making. So there are currently no coherent solutions for public authorities that focus on processing and analyzing others than statistical data for governmental intentions. The second fact is that applications for public authorities need to be very formal, to ensure the serious or valid work of such public institutions. This seriousness is also ubiquitous for the considered data that is analyzed for the policy creation. Used data are commonly valid data, mostly in form of statistical data. Overall statistical data are good to observe indicators, which are reflecting the economy or other political issues. However, this kind of data has disadvantages, if reasons or solutions should be identified based on this data. It allows seeing that for instance the GDP is decreasing, but it cannot be seen why, and how this trend can be stopped.

In contrast to these statistical data e.g. social media and data as result of opinion mining are less representative and valid. They only represent the subjective opinion of a group of discussing citizens or professionals. This should not mean that these kinds of data is useless for policy making since it gives decision makers an insight of citizens' opinions and thoughts, but it can just be used as orientation, because

of its limited validity. However, this kind of data has a significant advantage in contrast to statistical data. Since the data are coming from discussions, it can also contain possible reasons and solutions, which face the existing problem. It should also be mentioned that the quality of the named reasons and solutions can vary, but overall it can contain such explanations, which statistical data never can contain.

3. Open Government Data and Linked-Open Data

Recently, the provision of data is mainly done in two ways. The most important political change for a better provision of transparency [5] is the Open Government Data (OGD) or Open Data initiative. Here, the idea is to provide statistical information about the countries, states, or municipalities open and free accessible over the internet.

In parallel there is the Linked-Open Data (LOD) approach, which was a major outcome of Semantic Web initiatives. In contrast to ontologies, LOD data-bases formalize data as semantics in a more accessible way without the use of higher description logics [21]. However, major aspects of categorizing and conceptualizing resources and interlinking them semantically are part of the LOD annotations [21].

3.1. Open Government Data

Von Lucke defines Open Government Data (sometimes also named as Open Data) as open available data sets, which are of interest for public authorities and citizens without any limitation in its free use, redistribution and further use, and which are made available for free (Open-Access) [7]. The Open Knowledge Foundation mentions as a general requirement the provision of such data for free – also for commercial usage, and provided in an open format, e.g. CSV [8]. In many countries Open Government Data is understood as the provision of statistical government data to the public to provide an advanced transparency. For this purpose the data needs to be openly accessible through a portal where the indicators are available in a simple statistical format, such as CSV, or in another open standardized format, such as SDMX (see [20]).

To visualize Open Government Data in a more advanced way than just through basic visualizations, some state-of-the art tools and services in the web exist. Examples for established free accessible tools are, e.g. Gapminder (<http://www.gapminder.org>), OECD eXplorer [9] and EZB Inflation Dashboard (<http://www.ecb.europa.eu/stats/prices/hicp/html/inflation.en.html>). Usually such Open Government Data visualizations focus only on statistical information. Just a small number of systems aim to enhance the visualizations by showing also other

information to provide a better understanding about the data as it can be useful for political decision makers. One of such systems is developed by Hienert et al. [10], which shows historical event information parallel to the statistical indicator data. This allows analyst and decision makers to identify correlation of statistics against historical events.

3.2. Linked-Open Data

Linked-Open Data (LOD) has experienced a great growth in the semantic web community. Linked Data provides conceptualization of entities and a description of semantic relations between these items that are identified by a Uniform Resource Identifier (URI) [11]. LOD is a reflection of the knowledge interpretation by communities that model domain knowledge for structuring and disseminating it to a diversified audience. A single linked database gains millions of knowledge entities per day. A key aspect of LOD is using the web to create typed links between resources from different data-bases [11].

An aspect that is even though less considered towards policy modeling is the contained knowledge of many hundreds of volunteer workers for such LOD databases (e.g. dbpedia, which mainly bases on the data of Wikipedia) to maintain, enhance, correct and enrich these data-sources with new information so that they can be seen as some of the largest open data-sources in the web (this counts in particular for dbpedia).

A major benefit of LOD lays in its readability, which allows technical systems to perform analysis and mining on the structured data. However, there are also benefits for users, if LOD will be shown to user in a graphical manner. A benefit of such structured data is the easier ability to get an overview about a certain topic, as far graphical visualizations are used to show the information structure and in intuitive manner, e.g. SemaPlorer [12]. Even more, the structured data allows also finding of resources that are less important, e.g. a search for "Merkel" would normally only show information about the German chancellor Angela Merkel, but there is a small town in Texas, USA too.

3.3. Bridging of Open Government Data to Linked-Open Data

Both data types have their special character and usage scenarios. The idea of linking OGD resources on LOD entities is the normal way, but the advantage for decision makers would be low. A different approach for any kind of statistics is described by Paulheim et al. ([13],[17]). Based on a data table, the system generates correlation (named as rules), which result set contains possible reasons for a certain circumstance. In the presented version it was applied on general statistical data.

The mentioned approach is limited on a textual representation in form of sentences of the calculated rules. In the original version no graphical metaphors and visualizations are considered. Certainly, from a more general point of view, for rules and correlations in general, some options to show them in a visual manner still exist.

Visualization of complex data can enable an easier access to the contained information and hence it can result in a more effective and efficient decision making, in particular in scope of politics [24]. To develop visualizations that enable an improved decision makings, they should cover the aspects (1) the user and his individual behaviors, (2) the visualized data and (3) the task the user aims to solve/do with the visualizations [22]. Especially in regards of the calculated rules the visualization of them can enable a faster detection of important reasons and solutions. The reason is that for a single analysis a very high number of correlations could be calculated and needs to be analyzed by the user, which of them might be important and count as sense-making reasons or solutions.

4. Design for a Visual Semantics Explanation System

Decision makers and analysts are always interested in getting an overview about exiting and relevant problems. However, in scope of problem understanding and solution finding the traditional use of only Open Government Data is often limited for this purpose. Open Government Data can just indicate a possible problem through deviations from the normal level, e.g. if the unemployment rate increased significantly within a short time. In fact, the indicator does not explain why there is such a deviation.

In the following section we explain our integration approach that allows generating explanations for certain deviations. This is realized by a merge of Open Government Data with Linked-Open Data and a graphical mapping for an easy and intuitive usage.

4.1. Embedding in the Policy Making Lifecycle

For a better analysis and decision-making, we follow the idea of providing information to stakeholders from various data-sources and from different types. This should allow stakeholders to get the big picture about a specific issue. For this, we investigate objective (valid statistical data) as well as non-objective (user-generated) data-sources, even if non-objective data-sources are not an adequate basement for decisions, but they can help to get ideas and arguments for a certain problem. It is always the responsibility of the stakeholders and experts to decide what kind of data-source and information

should be considered in the analysis phase. Based on this idea, we can sketch the problem and solution finding in a theoretical data flow diagram (see Fig. 1) and follows the idea of process-driven information visualization [14]. This data-flow diagram orients on the major tasks agenda-setting and analysis (see [4],[5],[6]). For this purpose, different kinds of data from different data sources are considered to allow stakeholders getting an overview about a certain issue. This includes data about indicators as well as user-generated data, e.g. citizens' opinions coming from social media portals. All of these data can be used to allow stakeholders to gather an encompassing overview. Through the interlinking of different data sources certain issues can be better analyzed.

All of the data should be displayed to the stakeholders in a visual form. It is often easier for humans to understand data in graphical form than in tables or text.

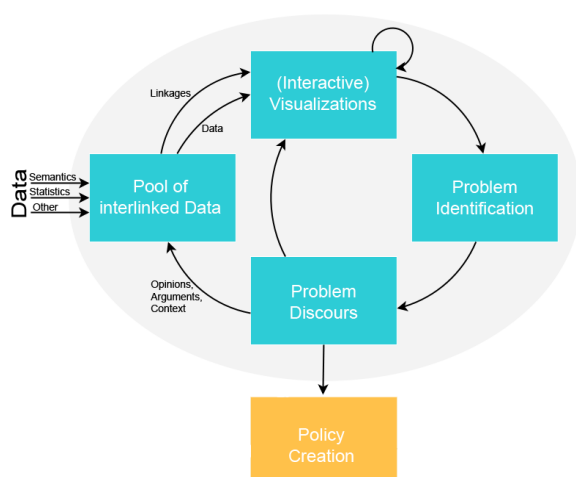


Figure 1. Interaction and data flow diagram to process the data on the server, beginning with the statistic-data input and generation of the explanations.

4.2. Semantics Visualization of Linked-Open Data

The semantics visualizations are designed to show networks of linked information and also structures (see for instance [18]). As data sources the common LOD data-sources, e.g. dbpedia, can be used as well as own LOD sources that some cities or regions do carry. Such LOD sources provide often basic information about the city, as well as references to important topics. Such data can help to understand specific behaviors of a region, e.g. traditional problems.

The visualization plays an important role in visualizing LOD for users since the gathering of an

overview about an unknown topic and problem depends on how easy complex dependencies can be shown to users. LOD sources can also provide additional (meta-) information about indicator data, e.g. dependencies and correlation between different indicators or an indicator which built upon another.

From the visual point of view the data must be shown in an explorative manner. This can be realized by cockpit integration (see therefore [15]), which allows users to orchestrate visualizations in a personalized form. As visualizations a set of different graphical layout algorithm is provided, which ensures that different aspects can be shown by the same data.

4.3. Statistics Visualizations of Open Government Data

The benefit of statistic visualizations of OGD is obvious. Since statistical data are majorly used in public authorities for decision making, this kind of data visualizations are omnipresent. Based on such data stakeholders are able to observe if a problem may occur, e.g. if an indicator has a significant deviation from the normal level.

For a more intuitive exploration of statistical data, it is beneficial to visualize the data in an interactive and explorative manner [19]. Therefore, the user-interface needs to provide the indicator data for exploration. This has to count for the indicator list itself and its categorical hierarchy, as well as the provision of the concrete statistical indicator data. To achieve such an interactive and explorative user-interface, we also designed a statistic cockpit approach (see therefore [16]), which allows users to orchestrate visualizations regarding the indicator hierarchy and list, as well as concrete statistical visualizations e.g. with LineCharts and ParallelCoordinates by their personal behavior and expectation.

4.4. FeGeLOD for Explanations to bridge Open Government Data to Linked-Open Data

FeGeLOD¹ is a program that allows generating correlation (so called rules) for a given data table with statistical contents. Explanations will be generated as a result. It runs as standalone desktop application in the default form.

For our purposes, we needed to adapt it. The major change is that we need a backend system that generated the information for our own frontend visualization technology. Another change was necessary to allow an enhanced configuration. In the desktop application the system takes as input small csv data-table were the names were automatically

¹ More information and software download on: <http://ke.tu-darmstadt.de/resources/fege lod> (accessed: 30/05/2014)

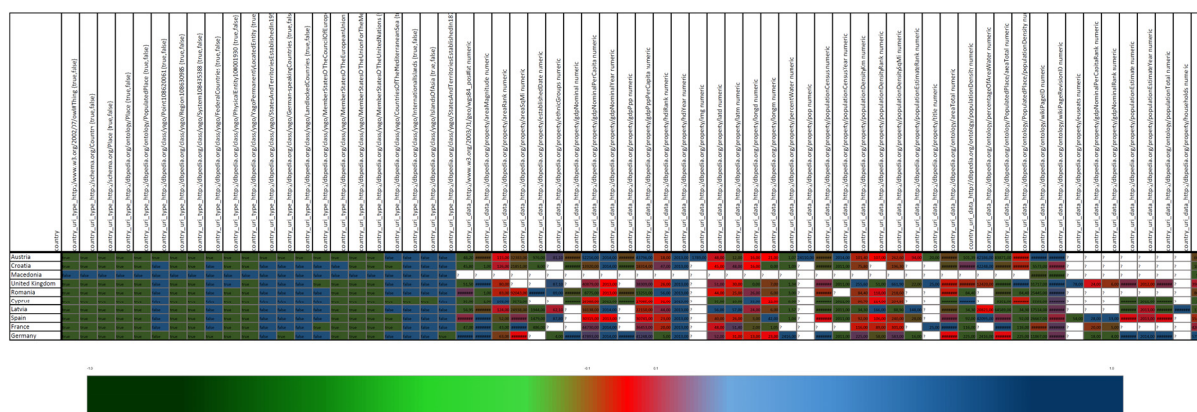


Figure 2. A matrix visualization of the geographical regions and the comparison properties. The colors indicating the range between the maximum and minimum of the available values.

tried to resolve to dbpedia resources. In perspective to the planned direct use of Open Government Data. The resolving of the geographical names is critical, because of their different spellings in the different languages. We extended the approach by a broad editing mode, where the user gets the ability to correct the aligned entities through a search for an alternative resources or the option to delete the data entry. Another change was required to decrease the generation time through a caching functionality. Especially in peak times the rule generation takes very long. We integrated a database that stores generated results so that not always a new generation of rules is required, which leads also to a real-time response on requests.

A special challenge is the parameterization for the explanations' generation. The length of the result-set can be controlled based on a number of parameters. For some request sometimes more than one passage is necessary to find the optimal parameters. Unfortunately, the identification of optimal parameters requires knowledge about how the system works. Commonly stakeholders do not have knowledge about this issue. For an easier way of use we designed an automatic optimization routine. Based on an initial passage the parameters were adjusted. If also the second run was not optimal, a third and last passage will run that changes the parameters in dependency of the results of passage one and two.

To achieve an interactive and explorative final visualization we also planned to change the result form. To allow a graphical exploration based on the generated rules, we extracted all kinds of links and parameters in separate form so that the user can select such a link and can further explore it and can retrieve further information. This includes also the trivial mapping of the geographical names from the Open Government Data base to concrete resources from dbpedia. After running explanation generations, we have an encompassing linking from OGD to LOD, which provides a significant advantage in the followed graphical analysis and exploration phase.

4.5. Correlation and Rule Visualization for an Intuitive and Fast Data Understanding

After the correlation calculation, the analyst aims to explore the rules for identifying possible problems and solutions. In a basic format the rules will be displayed in a simple text phrase like 'A country with a high value of pop has high Employment'. In fact that for larger data sets the number of rules can become large too, such sentences to not help to find the most relevant information. For this purpose a graphical representation of the rules can be helpful, which allows identification of the most significant rules.

To choose sufficient visualizations for such rules, it is essential to understand how the result set is constructed. The result set can be understand as a two dimensional data set. The first dimension is the geographical location, e.g. the countries or cities, the second dimension are the criteria, e.g. the HDI rank or a class inheritance. Based on this format a number of well-known visualization approaches seeming appropriate.

A basic visualization approach that is close to data table format is a matrix visualization (as for instance described in [23]). See also our example with real data in Fig 2.

4.6. Multi-Data and Multi-Visualization User-Interface

For analyst and decision makers it can be useful to see all relevant information parallel at the same time. This allows a better considering of most of the relevant information, which are needed for an optimal decision making. For this purpose, we extend the concept of Knowledge Cockpit (for LOD [15]) and Statistic Cockpit (for OGD [16]) to a Decision-Making Cockpit, which shows all of the previous mentioned data and visualization at the same time. This cockpit is completely dynamical and allows to compose it individually by the stakeholders. Therefore, the expert just chooses the data-source and

select or deselect the preferred visualizations and arrange it on the screen as he likes to.

During the policy making, the expert can try to find a problem in OGD. Often such a problem can be found, if there is a significant negative indicator development, e.g. an increase of the unemployment rate. To understand some basic information about such an indicator, e.g. what an indicator represents and how it is measured, he can try to understand it through LOD. If based on LOD and OGD the problem cannot be ascertained identified, the FeGeLOD engine can be used to find some possible reasons. Based on the FeGeLOD resulting explanations, the expert can find possible reasons for a certain deviation in the Open Government Data and of course, he can validate it through all the other available data sources. After that problem identification and characterization, the problem discourse can be started with the major goal to find a solution. Even though the decision-making cockpit can be used for the problem discourse, as well as for the policy creation phase.

5. Implementation of the Visual and Interactive Semantics Explanation System

The implementation consists of two technical components. The first is the FeGeLOD engine, which we transformed into a web-service so that it acts as backend and overtakes all data processing. The second technical component is the visualization frontend that allows visualizing the semantically as well statistical data, and -of course- even though the explanations coming from the FeGeLOD web-service. Both parts are web-resources and can be accessed directly with a browser.

5.1. FeGeLOD as Web-Service

To allow a flexible use of the FeGeLOD system through the web, a couple of changes were required. The major change was the general deploying of the technology as a web-service, which includes separating the internal processing into several stages that can also be accessed separately. We named the web-service Explain-a-LOD service. Based on the fact that the separation provides a number of interfaces that can lead to wrong processing and in consequence to wrong results, if it is used e.g. in a wrong order, we needed to define a clear process-driven approach. Such a process-driven implementation makes the API also better usable for users since it follows a logical better understandable order of activities [14]. We summarize the final implementation into seven stages (see also Fig. 3):

- (1) Over the first interface the data-table is shown to the user and he can select which data from an Open Government Data source should be involved for the explanation generation.

- (2) After this, a first mapping to concrete dbpedia resource is performed, e.g. the country name Germany is resolved to <http://dbpedia.org/resource/Germany>. The results are stored in a database.

- (3) In the third step, the user can check the results and if all data could be successfully mapped to a dbpedia resource. If an element could not be automatically resolved to a dbpedia resource, the user has the ability to search for a fitting resource or he can delete it, so that this entity is ignored in the further proceeding of the explanation generation.

- (4) With the beginning of the fourth step, the explanation generation starts. Therefore in sub-step (5), features to each resource are extracted from dbpedia and based on them in sub-step (6) the explanations are generated. Rules and explanations have in this context the same meaning.

- (7) The results—the explanations—are shown in the last step. These explanations are also stored in the database and can be downloaded in an export format.

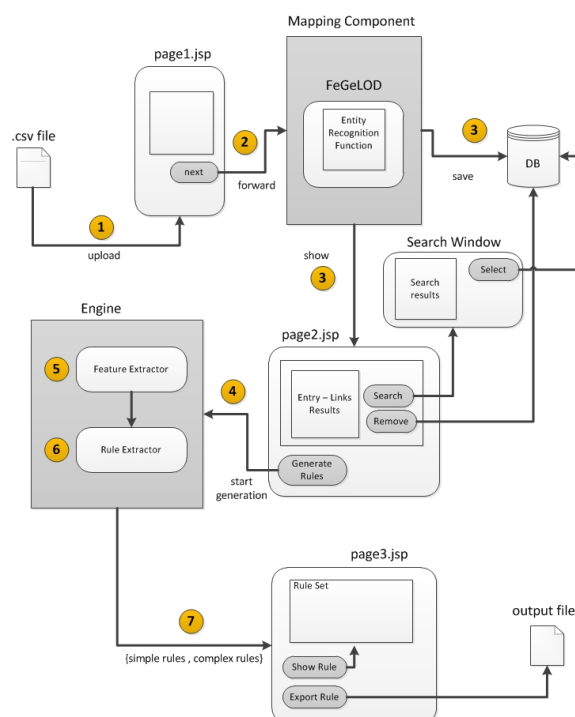


Figure 3. Interaction and data flow diagram to process the data on the server, beginning with the statistic-data input and generation of the explanations.

5.2. The Decision-Making Cockpit

We used our own web-based visualization system for the visualization of the data. It allows the visualization of heterogeneous data types (see Fig. 4). On the top the user can choose a data-source and can enter a query. On the right side a couple of visualizations are available that the user can select and orchestrate on his Decision-Making Cockpit.

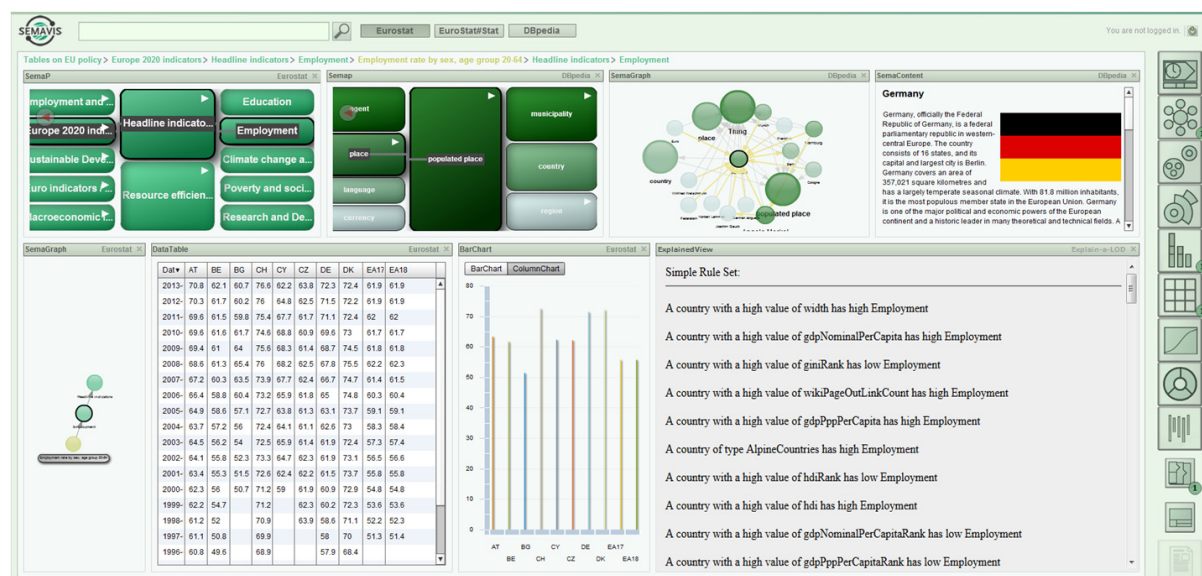


Figure 4. Screenshot of the User-Interface with which an expert is able to work with a composition of multiple data and multiple visualization to gather an overview about a certain problem.

In regards of the policy making lifecycle, the experts usually do observe the indicator data about their region. These indicator data are available through Open Government Databases. In this demo we included EuroStat. If the expert needs to clarify some issue, he can use the dbpedia. Therefore, he can perform a search by typing a query in the top. Afterwards he can navigate through the visualized result set.

To perform an explanation analysis by the Explain-a-LOD service, the user has to choose an indicator. The user can start the calculation process by marking a column at the indicator data table and followed by choosing the menu-item "Explain by LOD". In a first view, the user sees a result table where the entities from EuroStat were assigned to dbpedia resources. If an entity could not be resolved, the user can try to find an alternative resource manually, or he can delete this entry from the table so that it will not be considered in the further processing. After that, the internal data and analysis is processed and explanations will be generated. The generated simple explanations can be like this:

- *A country with a high value of pop has high Employment*
- *A country with a high value of width has high Employment*

- *A country with a high value of wikiPageOutLinkCount has high Employment*
- *A country with a high value of gdpNominalPerCapita has high Employment*
- *A country with a high value of gdpPppPerCapita has high Employment*
- *A country with a high value of hdiRank has low Employment*

The origin FeGeLOD supports also so called complex rules, but for them a higher number of entities is required that for our used EuroStat data does often not work well. These complex rules providing also complex information, e.g. only countries within a certain data range or in regards of different aspects have a high employment.

All of the mentioned information can be composed into one user-interface. The control of what visualization shows what data has always the user.

5.3. Visualizations for the generated Correlation and Rules

For the visualization of the rules we primary focused on the textual presentation in sentence form. However, because of its less intuitiveness and fast

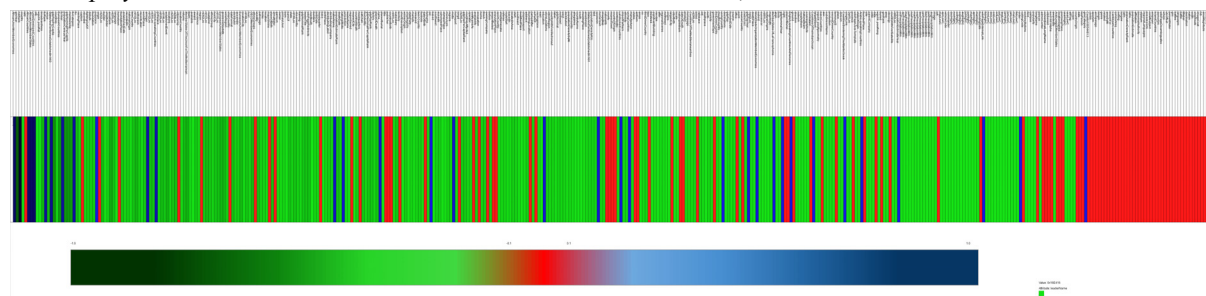


Figure 5. Overall rule ranking visualization to e.g. identify the most relevant comparison properties.

readability, we also started to develop visualizations, which can show the rules in graphical form that could be faster recognized by the human brain. Hence, we worked on two kinds of visual representation, first of all the basic matrix visualization (Fig. 2) and second a rule-overview visualization that shows the overall rule ranking (see Fig. 5).

6. Use-Case Future Policy Modeling in European Cities

The described system is presented in his first version to a couple of experts, often analyst for certain tasks, e.g. urban planners, from different municipalities of European cities. Here we could show them how especially the Explain-a-LOD feature can help to explain some indicator deviations from normal. Because of the completely technical processing the generated rules can be very canonical and not helpful, but overall it can mention unexpected reasons for some problems too. So far the system was only tested with indicator data on country level, because the data quantity and quality about cities varies in dbpedia, but only if for most cities an accurate number and quality of data is available, the system can generate useful explanations.

We are currently working on an evaluation design to evaluate the quality and usefulness for the generated explanations and the system. But it needs to be considered that an evaluation is very difficult, because it is hard to measure the quality of the generated explanations against reality. An indirect method would be an evaluation in form of questionnaire, but by this we just get more or less a subjective feedback.

7. Conclusion and Future Work

In this paper described on the hand an approach to handle multi-data through multi-visualization for a personalized information overview. This allows experts to retrieve all information sources and generate a decision-making cockpit for an optimal decision making, because of an encompassing overview about the data. On the other hand, we described a new approach to visually link Open Government Data with Linked-Open Data to generate and visualize explanations for certain indicator data. This is beneficial for problem finding in policy making, if reason finding is difficult. Through analyzing and comparing the statistical data with entries and properties of Linked-Open Data, correlations were extracted that may contain possible reasons for an indicator deviation from the normal range. Both solutions aim to support analysts and decision makers for a more effective policy modeling. This paper explained also a first implementation and integration of the Explain-a-LOD web-service, which works overall good. Even more introduced into

some first visualization implementation that can show the calculated rules in a graphical manner. However, there are some potential for improving the system. In general there is some potential in regards of usability aspects, e.g. a grouping of the available visualization metaphors (on the right selection bar in Fig. 4) in perspective of the supported data. In the future, we plan also to improve the interlinking of extracted entities and properties in the final visualizations so that the system gets more explorative and allows a more effective overview of the underlying problem. Additionally we will focus on the visualization of the generated explanations too. The textual representation is good, but from the user's point of view difficult to recognize intuitively and efficiently. Through further graphical representation of the rules, the system can enable a better and faster comprehension of generated explanations, in particular if the user can choose from a variety of available visualization layouts. This future work is an interesting research questions, because there are only a few ideas published how rules can be graphically displayed.

10. References

- [1] M. A. R. Estrada, "Policy modeling: Definition, classification and evaluation," In *Journal of Policy Modeling*, Volume 33, Issue 4, pp. 523-536, Elsevir, New York, USA, 2011.
- [2] M. Howlett, M. Ramesh and A. Perl, "Studying Public Policy: Policy Cycles and Policy Sub-systems," Oxford University Press, 2009.
- [3] H. D. Lasswell, "The Decision Process: Seven Categories of Functional Analysis," University of Maryland, College Park, USA, 1956.
- [4] A. Macintosh, "Characterizing E-Participation in Policy-Making," In *Proceedings of HICSS'04*, Vol. 5. IEEE Computer Society, Washington, DC, USA, 2004.
- [5] OECD, "Promise and Problems of e-Democracy – Challenges of Online Citizen Engagement," OECD Publication Service, Paris, France, 2003.
- [6] S. Sonntagbauer, K. Nazemi, S. Sonntagbauer, G. Prister & D. Burkhardt, "The FUPOL Policy Lifecycle," In *Handbook of Research on Advanced ICT Integration for Governance and Policy Modeling*, IGI Global, 2014.
- [7] J. von Lucke & C. P. Geiger, "Open government data - frei verfügbare daten des öffentlichen sektors," *Technical Report*, Zeppelin University Friedrichshafen, 2010.
- [8] Open Knowledge Foundation, "Defining the open in open data, open content and open services," <http://opendefinition.org/okd/>, accessed: 10/06/2014, 2009.
- [9] M. Jern, "Collaborative web-enabled geoanalytics applied to OECD regional data," In *Proceedings of the 6th international conference on Cooperative Design, Visualization, and Engineering*, Springer-Verlag, Berlin, Heidelberg, pp. 32-43, 2009.

- [10] D. Hienert, D. Wegener & S. Schomisch, "Making Sense of Open Data Statistics with Information from Wikipedia," In *Availability, Reliability, and Security in Information Systems and HCI*, LNCS 8127, Springer-Verlag, Berlin, Heidelberg, pp. 329-344, 2013.
- [11] C. Bizer, T. Heath & T. Berners-Lee, "Linked Data - The Story So Far," In *Int. Journal on Semantic Web and Information Systems*, 2009.
- [12] S. Schenk, C. Saathoff, S. Staab & A. Scherp, "SemaPlorer-Interactive semantic exploration of data and media based on a federated cloud infrastructure," In *Web Semant.* 7, 4, pp. 298-304, 2009.
- [13] H. Paulheim & J. Fürnkranz, "Unsupervised Generation of Data Mining Features from Linked Open Data," In *International Conference on Web Intelligence and Semantics (WIMS'12)*, 2012.
- [14] D. Burkhardt, T. Ruppert & K. Nazemi, "Towards process-oriented Information Visualization for supporting Users," In *Proceedings of ICL '12*, 8p., 2012.
- [15] K. Nazemi, D. Burkhardt, M. Breyer et al., "Semantic Visualization Cockpit: Adaptable Composition of Semantics-Visualization Techniques for Knowledge-Exploration," In *Proceedings of ICL '10*, University Press, Kassel, pp.163-173, 2012.
- [16] D. Burkhardt, K. Nazemi, C. Stab et al. "Visual Statistics Cockpits for Information Gathering in the Policy-Making Process," In *Advances in Visual Computing*, Springer-Verlag, pp. 86-97, 2013.
- [17] H. Paulheim, "Explain-a-LOD: using linked open data for interpreting statistics," In *Proceedings of IUI'12*. ACM, New York, USA, pp.313-314, 2012.
- [18] K. Nazemi, M. Breyer, J. Forster, D. Burkhardt & Arjan Kuijper, „Interacting with Semantics: a User-Centered Visualization Adaptation based on Semantics Data,” In *Proceedings of HI'11*, Springer-Verlag, Berlin, Heidelberg, pp. 239-248, 2011.
- [19] E. Duval, "Attention please! learning analytics for visualization and recommendation," In *Proceedings LAK '11*. pp. 9-17. ACM, New York, NY, USA, 2011.
- [20] P. Bunyakiati & P. Voravittayathorn, "Dissemination formats and major statistic data sets of the AEC countries: A survey," In *Proceedings of ISSDM'12*, pp.313-316, 2012.
- [21] T. Heath & C. Bizer: Linked Data – Evolving the Web into a Global Data Space. *SYNTHESIS LECTURES ON THE SEMANTIC WEB: THEORY AND TECHNOLOGY*. Morgan & Claypool Publishers, 2011.
- [22] D. Burkhardt, K. Nazemi, P. Sonntagbauer, S. Sonntagbauer & J. Kohlhammer, „Interactive Visualizations in the Process of Policy Modeling,” In *Proceedings of IFIP eGov 2013*, IFIP, 2013.
- [23] R. Gove; N. Gramsky; R. Kirby; E. Sefer; A. Sopan; C. Dunne; B. Shneiderman & M. Taieb-Maimon, "NetVisia:

Heat Map & Matrix Visualization of Dynamic Social Network Statistics & Content," In *Proceedings of Privacy, Security, Risk and Trust and IEEE Third International Conference on Social Computing*, pp.19,26, 2011.

[24] K. Nazemi; M. Steiger; D. Burkhardt & J. Kohlhammer, "Information Visualization and Policy Modeling." *Handbook of Research on Advanced ICT Integration for Governance and Policy Modeling*. IGI Global, pp. 175-215, 2014.

[25] D. Burkhardt; K. Nazemi & J. Kohlhammer, "Policy Modeling Methodologies." *Handbook of Research on Advanced ICT Integration for Governance and Policy Modeling*. IGI Global, pp. 48-60, 2014.

11. Acknowledgements

Part of this work has been carried out within the FUPOL project, funded by the European Union under the grant agreement no. 287119 of the 7th Framework Programme. This work is based on the SemaVis technology developed by Fraunhofer IGD (<http://www.semavis.net>). SemaVis provides a comprehensive and modular approach for visualizing heterogeneous data for various users.