



Performance model's development: a novel approach encompassing ontology-based data access and visual analytics

Marco Angelini¹ · Cinzia Daraio¹ · Maurizio Lenzerini¹ · Francesco Leotta¹ · Giuseppe Santucci¹

Received: 15 December 2019 / Published online: 14 September 2020
© The Author(s) 2020

Abstract

The quantitative evaluation of research is currently carried out by means of indicators calculated on data extracted and integrated by analysts who elaborate them by creating illustrative tables and plots of results. In this approach, the robustness of the metrics used and the possibility for users of the metrics to intervene in the evaluation process are completely neglected. We propose a new approach which is able to move forward, from indicators' development to an interactive performance model's development. It combines the advantages of the ontology-based data access paradigm with the flexibility and robustness of a visual analytics environment putting the consumer/stakeholder at the centre of the evaluation. A detailed description of such an approach is presented in the paper. The approach is illustrated and evaluated through a comprehensive user's study that proves the added capabilities and the benefits that a user of performance models can have by using this approach.

Keywords Education and research · Performance assessment · Performance modelling · Ontology-based data access · Visual analytics

✉ Marco Angelini
angelini@diag.uniroma1.it

Cinzia Daraio
daraio@diag.uniroma1.it

Maurizio Lenzerini
lenzerini@diag.uniroma1.it

Francesco Leotta
leotta@diag.uniroma1.it

Giuseppe Santucci
santucci@diag.uniroma1.it

¹ DIAG Department, Sapienza University of Rome, Via Ariosto, 25, 00185 Rome, Italy

Introduction

In recent decades, the rapid changes taking place in the production, communication and evaluation of research have been signs of an ongoing transformation. It has been stated that “we are living a sort of Middle-Age guided by the information and communication technologies (ICT) revolution, or the so-called *fourth revolution* as described by Floridi (2014) which emphasizes the importance of information” (Daraio 2019, p. 636). Largely, the current Middle-Age of research evaluation might be understood as the transition from a traditional evaluation model, based on bibliometric indicators of publications and citations to a modern evaluation, characterized by a multiplicity of distinct, complementary dimensions. This step is guided by the development and increasing availability of data, together with statistical and computerized techniques for their treatment, including among others the recent advancements in artificial intelligence and machine learning. Daraio and Glänzel (2016) show that the complexity of research systems requires a continuous information exchange.

These changes produce different effects (see further details and references in Daraio 2019, Table 24.2, p. 644) (i) on the *demand side* (those that ask for research assessment) including an increase of institutional and internal assessments, (ii) on the *supply side* (those that offer research assessment) including proliferation of rankings, development of Altmetrics, open access repositories, new assessment tools and desktop bibliometrics), (iii) on *scholars* (the increase of “publish or perish” pressure, impact on the incentives, behaviour and misconduct, and increasing critics against traditional bibliometric indicators), (iv) on the assessment process (increasing the complexity of the research assessment) and on the indicators’ development.

Daraio (2017) showed that the formulation of models (in this paper we will use metrics and indicators interchangeably) is necessary to assess the meaning, validity and robustness of metrics. It was observed that developing models is important for *learning* about the explicit consequences of assumptions, for testing the assumptions, for documenting and verifying the assumptions, for systematizing the problem and the choices done.

One of the main grand challenges that remains to address is the exploitation of data availability and Information Technology in a data integration framework in use for multiple purposes. Supporting the interaction of stakeholders with this framework is crucial (Daraio and Glänzel 2016).

Little attention is paid to the problem of developing performance indicators. As highlighted in Daraio (2017, 2019) it is necessary to describe the theoretical, methodological and data components that constitute the model of a metric in order to evaluate its appropriateness. Without a reference model, it is not possible to evaluate the robustness of the metric used in a performance evaluation.

Another overlooked aspect is that of the user or consumer of metrics. Generally, the interaction with the users of the metrics is not taken into consideration nor is the possibility of intervening on the metrics in the evaluation process.

To fill these existing gaps, we propose a new model of development of performance indicators, based on a visual analytics environment, that puts the user at the centre, allowing her/him to interact with the data and compare the metric models before choosing the ones that will then be used for the evaluation.

The paper is structured as follows: the next section illustrates the background of our approach and the existing related literature. “[Aim and contribution](#)” section presents the aim and the main contribution of the paper. “[Method and material](#)” section describes

the methods and the techniques proposed in our framework, in particular the OBDA system based on *Sapientia: The Ontology of Multidimensional Research Assessment* and the visual analytics environment supporting the users in the evaluation process. “[Results and discussion](#)” section reports on a usage scenario of the proposed system and on the results of a user study that tested the proposed system operationally. “[Concluding remarks](#)” section concludes the paper.

Background and related works

OBDA for indicators development

Our contribution builds on previous research carried out at Sapienza university based on an OBDA system for Research and Innovation (R&I) data integration and access. An ontology-based data access (OBDA) system is an information management system constituted by three components: an ontology, a set of data sources, and the mapping between the two. An *ontology* in Description Logic (DL) is a *knowledge base*, i.e., a couple (pair) $O = \langle \text{TBox}, \text{ABox} \rangle$, where TBox is the Terminological Box that represents the *intensional* level of the knowledge or the *conceptual* model of the portion of the reality of interest expressed in a formal way and ABox is the Assertion Box that represents the *extensional* level of the knowledge or the *concrete* model of the portion of the reality expressed by means of assertions on instances (see e.g. Calvanese et al. 1998). The *data sources* are the repositories accessible by the organization where data concerning the domain are stored. In the general case, such repositories are numerous, heterogeneous, each one managed and maintained independently from the others. The *mappings* are precise specifications of the correspondence between the data contained in the data sources and the elements of the ontology. The main purpose of an OBDA system is to allow information users to query the data using the elements in the ontology as predicates.

The OBDA system supporting *Sapientia* is based on Mastro (Calvanese et al. 2011, <http://www.obdasystems.com/mastro>). Other OBDA inference engines are anyway available. OnTop for example (Calvanese et al. 2017, <https://ontop-vkg.org/>) shares common research origins with Mastro. The two systems are, at the current development stage, different with respect to supported fragment of SPARQL queries.

The OBDA system, implemented with *Sapientia*, represents the ontology of multidimensional research assessment (Daraio et al. 2015) and permits the extraction of relevant data coming from heterogeneous sources—maintained independently, and reasoning about the performance indicators (PI) of interest.

Daraio et al. (2016a) showed the advantages of an OBDA system for R&I integration and Daraio et al. (2016b) showed that an OBDA approach allows for an unambiguous specification of indicators according to its four main dimensions: ontological, logical, functional and qualitative. See also Lenzerini and Daraio (2019) where a detailed illustration of the usefulness of an OBDA approach for reasoning over the ontology about indicators of performance is reported. Even the simplest indicator of performance, such as number of publications, has different conceptual aspects that the ontological commitment of the domain offers to the analyst (for additional details the reader is referred to Fig. 15.9 and 15.10 of Lenzerini and Daraio 2019, p. 368 and p. 369).

Other recent works on OBDA

In the last years, OBDA has been successfully applied to several case studies in challenging real, non-research, contexts. In this section, we will focus on the application of OBDA for the purpose of integrating several data sources to improve services and knowledge in public administrations.

Authors in Antonioli et al. (2013, 2014) applied OBDA to the case study the Italian Department of Treasury. In this context these works propose a Public Debt Ontology formalizing the whole domain of the Italian public debt. In particular, it describes both the public debt composition, namely the state liabilities and assets, and the financial instruments used by the Italian public administrations to manage the public debt. Importantly, it provides an historical view of the public debt, by focusing not only on the current state, but also on its evolution through past states.

Authors in Aracri et al. (2017) show the application of OBDA to the context of ISTAT, i.e., the Italian national agency in charge of producing statistics about all aspects of the Italian society, which are used as a basis for governmental decisions. As part of its activity, ISTAT produces statistics about individuals. In order to improve these statistics, to obtain new indicators and to allow citizens to query data in a guided manner, ISTAT produced an ontology covering people, families, geographical distribution and related statistical measurements.

Authors in Santarelli et al. (2019) applied the concept of OBDA to the case study of ACI—Automobile Club d'Italia. ACI is the Italian institution in charge of monitoring the circulating Italian vehicles with taxation purposes. In this application context, the authors have defined an ontology covering the domains of Public Vehicle Register (PRA) and vehicle taxation, and connected such ontology to the data source of ACI, so as to exploit semantic technology for various data government tasks.

All of the previous examples have in common with the approach presented in this paper the goal of exploiting available data to provide services and integrated information to citizens and companies. Closer to our approach, Mosca et al. (2018) propose a system based on OBDA to establish connections between the worlds of research and industry in Tuscany. Even though the final goal is similar to ours, the data sources employed are rather different, more focusing on the involvement of researchers in national and European projects with respect to scientific production. Finally, Sivertsen (2019) introduces and discusses current research information systems (CRIS) that can also be used as interoperable data sources for comparable studies across institutions and countries.

More generally, OBDA has been applied also to industrial contexts. Kharlamov et al. (2017) applied it to Statoil, an international energy company with main activities in gas and oil extraction. Here, an ontology has been employed to provide integrated data access to a number of large databases containing information about historical exploration data (e.g., layers of rocks, porosity), production logs, maps, and business information such as license areas and companies. This system also features the visual query language proposed in Giese et al. (2015) and Soylu et al. (2018).

Visual analytics

Several visual analytics solutions exist that address somehow related analytical and visual activities, e.g., comparing the performances of different complex elements, dealing with ontologies or displaying relevant pieces of information at geographical level.

Moral-Muñoz et al. (2019) offer a systematic review of science mapping software tools, showing their strengths and limitations. They analyse six software tools, namely BibExcel, CiteSpace II, CitNetExplorer, SciMAT, Sci 2 Tool, VOSviewer. They evaluate and compare the data processing, analysis options, and visualization of these tools concluding that the choice of a particular tool relies on the type of actor to be analyzed and the output expected.

Angelini et al. (2018) present the CLAIRE system that allows for comparing the performances of different Information Retrieval engines, using a visual mechanism that share the same main goal of the present paper. However, the main issue addressed by CLAIRE is the combinatorial explosion of the analysis and the large number of items that are compared at the same time, while this paper focuses on comparing few items on a larger number of characteristics. The system presented in Catarci et al. (2003) visually supports the user for query formulation but with the main goal of helping the user in selecting the right terms; the proposal described in Silva et al. (2019) supports the user in the task of exploring an ontology structure and content. Both proposals differ from the role that the ontology has in this paper, which aims at extracting data coming from heterogeneous sources. Finally, the systems presented in Angelini and Santucci (2017) and Angelini et al. (2019a) deal with performance models with respect to cybersecurity risk. The former shares with this paper the idea of visually presenting the results of the risk analysis at geographical level. Differently from this paper, however, it deals with physical elements (power network nodes) with a finer grain scale, i.e., city headquarters. Finally, while some previous works exist on the usage of visual analytics to evaluate performance indicators, like the work in Belton et al. (1993) for DEA analysis, or the work by Erhan et al. (2009) for visual sensitivity analysis of general parametric models, these works focus on analysing several performance indicators and not on supporting the performance model building and evaluation. The problem of supporting analysis of performance models with visual analytics solutions remains an active area of research.

Moving from indicators development to performance models development

In our previous research built on *Sapientia* (mentioned above) we showed the usefulness of OBDA for indicators development. In this paper we propose to move from indicators development to performance models development taking the centrality of the user into account and allowing the interaction of the user in comparing data, indicators and models, thanks to the addition of the visual analytics environment.

In the introduction, we have highlighted the importance of developing multidimensional models for the assessment of research and its impact. The modelling activity is not an easy task because defining a model requires choosing a level of analysis, identifying the main variables to describe the reality and being able to identify also the relevant dimensions that were not included in the model (e.g., for lack of data). Developing models is important for learning about the explicit consequences of assumptions, testing the assumptions, and highlighting relevant relations. It is also important for improving, documenting

and verifying the assumptions and the choices done. Some of the difficulties of modelling relate to the possibility that the targets are not quantifiable, together with the complexity, uncertainty and changeability of the environment in which the controlled system works.

The literature on performance measurement and indicators development is very rich. A state-of-the-art review on the *h*-index and its related literature can be found in Schubert and Schubert (2019) and Wildgaard (2019) offers a detailed description of the available indicators of research performance at individual level.

The main components of a performance evaluation model can be found in Daraio (2017, 2019). We have actors that are involved in processes which consist in the combination and or transformation of inputs in outputs, taking into account the main objectives of the activities. We may consider different measures of performance, ranging from efficiency (defined as the relationship between the outputs produced with respect to the resources/inputs used) to effectiveness and impacts. The constitutive elements for the development of a performance model proposed are:

1. Purpose of the assessment (objectives, stakeholder and policy) answering to the question “Why are we carrying out the assessment?”
2. Level of analysis (actors: scholars, organizations, regions or countries) answering to the question “Who are we assessing?”
3. Object of the evaluation (outputs, efficiency, results, effectiveness, impact) answering to the question “What are we assessing?”
4. Means of the evaluation (qualitative, quantitative, mixed methods; data) answering to the question “How are we assessing?”
5. Internal and external conditional factors (actors, processes, results; time, context, heterogeneity factors, rules, standards, incentives, actions, consequences) answering to the question “How, when and where are we assessing?”

The approach for developing performance models briefly outlined in this section is the basis for the users’ evaluation described in “[Users’ evaluation](#)” section.

Aim and contribution

We propose a novel approach to explore different/alternative definitions of the performance models while looking at their differences, hypothesize and test new performance models, and illustrate the results of these analyses in an interactive platform (see Battle et al. 2018).

This proposal departs from the traditional approach to indicators’ development, based on the selection of a specific set of indicators, collection of the relevant data, cleaning of the gathered data, computation of the indicators and illustration of them in a plot or table. According to this traditional approach if one wants to add a new data source or wants a different indicator, one has to restart the process from scratch. Moreover, different analyses based on different subsets of the data can be difficult to compare or to project on the used data selection. In this way both comparability and generality can be difficult to obtain as properties of the developed indicator, requiring a lot of additional work usually conducted with specific tools and/or data analysis processes.

In contrast, the proposed approach exploits ontology-based data access (OBDA) techniques to obtain data integration as a prerequisite for the performance models development, mitigating heterogeneity coming from manual data integration usually dependent on

the subset of available data. At the same time, it exploits visual analytics (VA) techniques, in the form of a proposed VA environment, to support creation, exploration with respect to original data, comparison and validation of these models.

The main contributions of this paper, that extends the work of Angelini et al. (2019b), are:

- The proposal of an integrated framework including OBDA and visual analytics techniques that:
 - using OBDA techniques for data integration in development of performance models for evaluation of research activities, instantiated in *Sapientia*, allows overcoming the heterogeneity and biases resulting by classical data-integration methods used for modelling performance indicators;
 - through the development of a novel visual analytics environment supports the creation, exploration, comparison and validation of performance models for evaluation of research activities by an analyst in an interactive way.
- A deep user evaluation conducted on real activities related to development of performance models, that demonstrates the appreciation and usefulness of the proposed approach in conducting these complex activities.

Method and material

The traditional way to define indicators relies on an *informal definition* of the indicator as the relationship between variables selected among a set of data collected and integrated “ad hoc”, specific for the user needs (*silos based* data integration approach). This means that when a new indicator has to be calculated, the process of data integration has to restart from the beginning because the dataset created “ad hoc” for an indicator is not reusable for another one.

The contribution of an OBDA approach to overcome this traditional indicator development approach is twofold. Firstly, it permits the *free* exploration of the *knowledge base* (or information platform) created to identify and specify new indicators, not planned or defined in advance by the users. This feature is particularly useful to face two recent trends in user requirements, namely *granularity* and *cross-referencing* (see Daraio and Bonaccorsi 2017 for a discussion on university-based indicators). Secondly, it allows us to specify a given indicator in a more precise way as described in Lenzerini and Daraio (2019).

In this paper we develop further this approach combining it with the main strengths of visual analytics. Visual analytics (Cook and Thomas 2005; Keim et al. 2008) is “the science of analytic reasoning facilitated by interactive visual interfaces”; through the connection of the analytical calculation with visualization and interaction by the human user, this interdisciplinary approach enhances the exploratory analysis of data, allowing to represent multidimensional data in a simple way through innovative abstract visual metaphors. Further to obtain an overview of the data, navigable by the user to the required level of detail, and the ability to apply complex analysis workflows that aim at explanation and reporting of the findings discovered during the analysis (see Fig. 1 for an overview).

The visual analytics approach developed in this paper allows us to move from performance indicators (PIs) development to performance model development, by exploring and

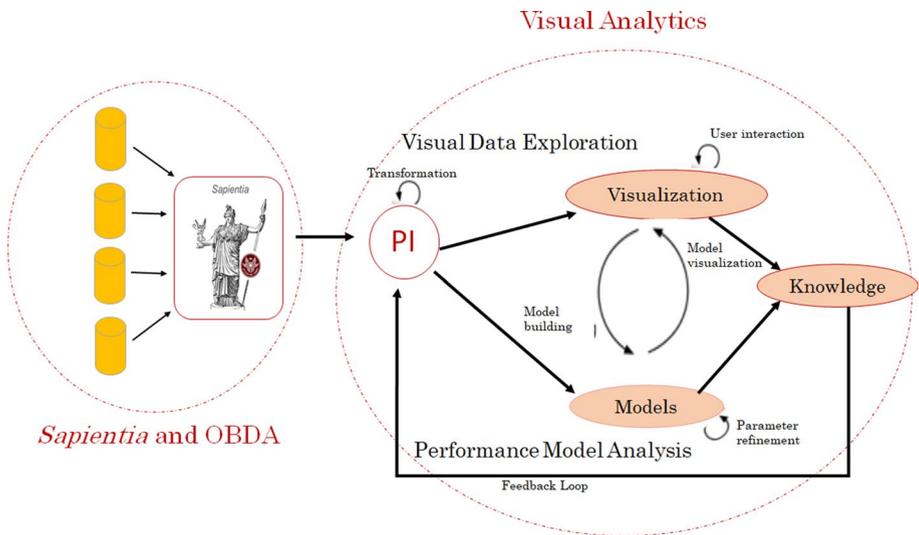


Fig. 1 An illustration of our approach that combines *Sapiaentia*, OBDA and visual analytics. *PI* performance indicator

exploiting the modelling and the data features within the flexibility of a visual analytics environment.

This allows a multi-stakeholder viewpoint on the model of *PI* and the assessment of the sensitivity and robustness of the *PI* model in a multidimensional framework.

In the next section we outline the main features of *Sapiaentia* (the Ontology of Multidimensional Research Assessment).

OBDA and *Sapiaentia*

Sapiaentia, the Ontology of Multidimensional Research Assessment (Daraio et al. 2015, 2016a, b), models all the activities relevant for the evaluation of research and for assessing its impact (see Fig. 2 for an outline of its modules). For impact, in a broad sense, we mean any effect, change or benefit, to the economy, society, culture, public policy or services, health, the environment or quality of life, beyond academia.

The *Sapiaentia* ontology has been developed using the Graphol visual language (<http://www.dis.uniroma1.it/~graphol/>, Lembo et al. 2016), that can be easily translated into standard ontology languages like Owl.

Sapiaentia acquires information from multiple sources, whose content can be overlapping. The same entity modelled in the *Sapiaentia* ontology can be represented in more than one data source, and even one data source could present (due to internal inconsistencies or design choices) the same entity multiple times in different forms.

Hence, we have the need to identify duplicated items and integrate the information obtained for each entity from any of the available sources.

In particular, at the ontology level we have created the concept of Representation. Entities modelled in the ontology of which we have different views from different data sources may have their own representation, which specializes the general Representation concept.

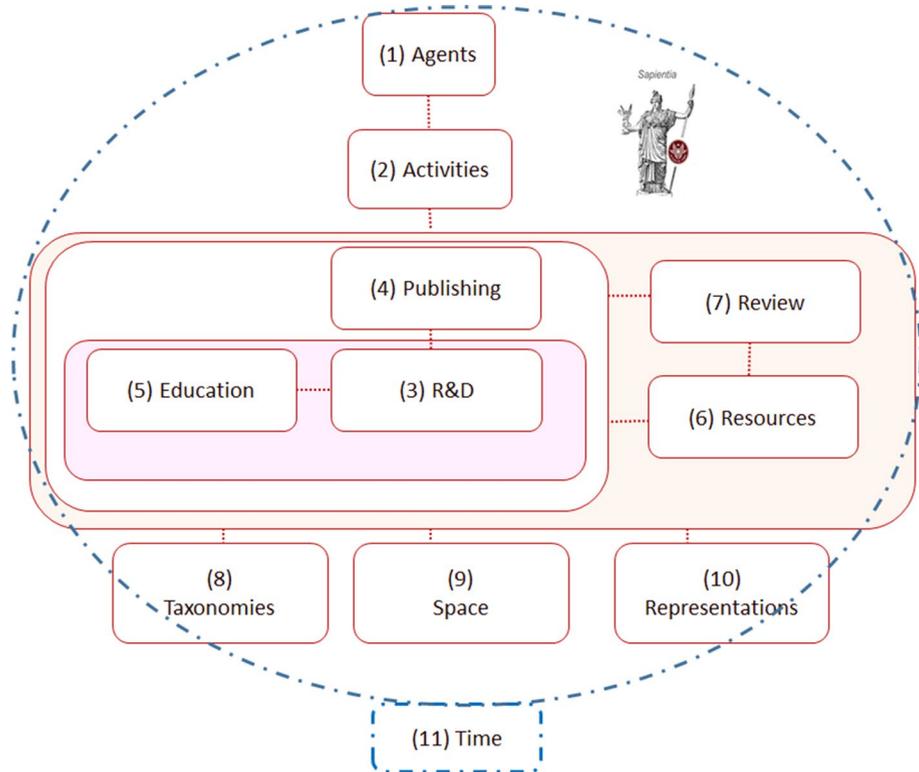


Fig. 2 Modules of *Sapientia 3.0*

This makes it possible to keep track in the ontology, through the mappings, not only of the modelled entities, but also of the way in which the information relative to the entities has been gathered from the data sources.

Data acquisition from the external sources makes use of the web service standards (REST, SOAP) when available. For less frequently updated sources and sources that do not implement an API, data acquisition leverages in some cases the open source edition of Pentaho Data Integration (<http://community.pentaho.com/projects/data-integration/>). Imported data are saved in a relational database (MySQL). Each source is modeled independently so that its peculiar structure can be fully exploited.

Figure 2 shows the modules of the last version of *Sapientia* (v3.0). They are:

1. *Agents*, that describes all human actors and institutions involved in the education, research and innovation process.
2. *Activities*, that describes the activities and projects the agents of the previous module are involved in.
3. *R&D*, that describes the different products (e.g., publications, patents) that are produced in the knowledge production process.
4. *Publishing* that describes how knowledge products are published and made available to the public.
5. *Education* that formalizes the concepts related to universities and courses.

6. *Resources* that describes all the ways an institution can be funded.
7. *Review*, that describes the process entities related to the publishing activity.
8. *Taxonomy*, that describes the elements that allows defining taxonomies applied to the different modules.
- 9 and 11. *Space and Time*, that formalizes respectively geographical entities and time instants and ranges.
10. *Representation*, that describes that the modeling mechanism by which single instances of other modules can be represented in different ways by the different sources used in *Sapientia*.

Visual analytics system

The developed solution uses visual analytics techniques to represent data from publications and education obtained by the OBDA approach. The system is implemented through Web technology.

The large quantity of indicators and basic features for the different units of analysis, including the territorial ones, and the different years of analysis increase exponentially the cardinality of data to be analyzed; in this respect, the proposed environment allows to obtain a visual overview of the data in a very simple form, and the interaction capabilities allow the analyst to navigate this overview and conduct detailed analysis up to the desired level of detail. The analyst is also supported in the discovery of new elements of interest through a process of *data exploration* that does not require a prior analysis goal.

In addition to the data exploration capabilities, there is a second area designed to analyze the performance model development. The environment is instantiated on European research and education institutions but is applicable in principle to any dataset. The analyst can, on one hand, analyze the performance of the various institutions with respect to a performance model, in order to analyze the ranking of the institutions of interest and their behavior with respect to the chosen model. Additionally, it allows to explore different performance models and to evaluate their goodness and fitness; it is also possible to evaluate the goodness of the proposed models, analyzing their variability and conducting sensitivity analysis in order to evaluate which parameters of the model (whether inputs, resources, contextual factors or outputs) contribute more to the performance of the institution with respect to the chosen model. The following subsections provide a description of the features of the visual analytics environment.

Data exploration environment

The first panel that composes the visual analytics environment is the data exploration environment. This environment consists of three main views depicted in Fig. 3.

These three views are:

- *Geographic view* (Fig. 3 top) which allows for geolocating the different institutions with respect to territorial units on a geographic layer (using Leaflet.js framework, based on OpenStreetmap). The map is navigable on 5 different levels of detail, where the first four follow the NUTS categorization from 0 (Nations) to 3 (Provinces) and the last one relates to single institutions. The user can at any time change the level of aggregation through a tab that shows the different available levels.

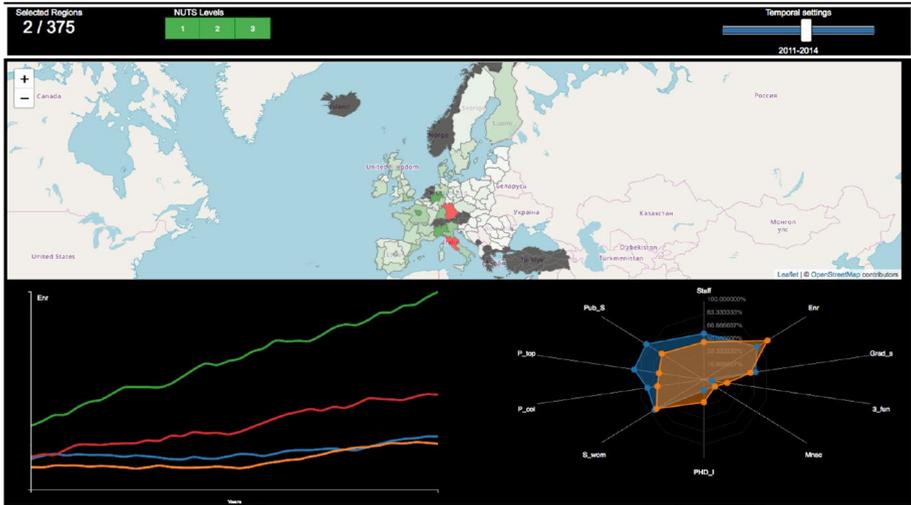


Fig. 3 Data exploration environment

The color of each element of the map reflects an indicator (basic or derived), on a green scale that identifies the values (white: low value, dark green: high value). The gray color visually encodes the absence of data for the specific territorial unit. A slider allows the analyst to scroll through the various years and conduct a temporal analysis on the available data, looking for institutions showing a high variability through a “time-lapse”.

- **Radar view** (Fig. 3 bottom-right) This view follows the visual paradigm of the radar diagrams (Von Mayr 1877), which represent the dimensions of a dataset one per axis, with the axes arranged in radial layout starting from the center. The indicators are arranged one per axis and the chart represents each data tuples (territorial unit) as a line that join the points on each axis. When the user selects one or more territorial units, the corresponding splines are highlighted, in order to allow an easy visual comparison between the selected territorial units on their dimensions. It is also possible to highlight a dynamic average trend, consisting of a line that connects the different averages on the respective axes, in order to compare the performance of a territorial unit, or generally of a given unit to the average behavior.
- **Linechart view** (Fig. 3 bottom-left) This visualization allows analyzing the time course of the indicators used for the territorial units under analysis. It is possible to analyze multiple territorial units to compare the trend of the same indicator on them, or to analyze multiple indicators on the same territorial unit, in order to have an overview of the progress of the unit itself, or a combination based on multiple territorial units and multiple indicators. In this last case the color-coding outlines all the indicators belonging to each single territorial unit.

The combined use of these views, possibly guided by the definition of specific PIs, allows more powerful dynamic exploration of the performance of the territorial units compared to the classical approaches, making the user able to obtain an overview of the general trend and specific details on the individual units, subsequently allowing to refine the analysis through the visual selection of appropriate subsets of information. The approach

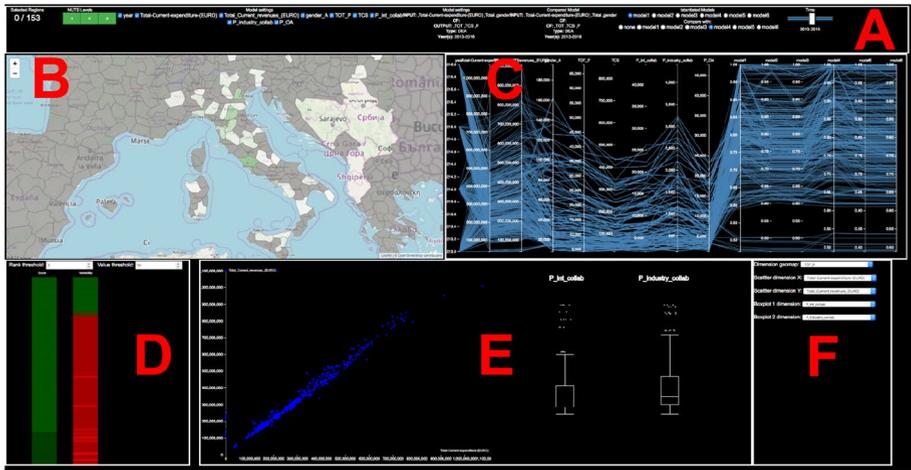


Fig. 4 New configuration of the performance model analysis environment

therefore allows the exploration of *specific scenarios* chosen by the user in *real-time*, without precomputation, which better support the formation and validation of hypotheses and the identification of areas of interest on which to conduct further analysis or to be used for reporting activities. At the same time, in this part of the environment the indicators are considered atomically, and they are used to only evaluate performances of territorial units., without any considerations on the indicator's characteristics. The next section describes instead the part of the visual analytics environment designed to create and analyze performance models starting from the available measures and indicators.

Performance model analysis environment

This environment is the core of the visual analytics system, and it is dedicated to the analysis of performance models used for analyzing the territorial units (e.g. universities for scientific evaluation). This part of the system has been evolved with respect to the work by Angelini et al. (2019b) in order to improve its functionalities and informativeness, after having been tested and used by performance models creators. An overview of the new environment is provided in Fig. 4.

The environment consists of a commands bar (A), a geographical view borrowed from the *Data Exploration* environment (B), a view based on parallel coordinates (C), a view of the rankings produced by the selected performance model(s) (D), a view based on scatter-plot and box-plot that allows to conduct sensitivity analysis on the variables of the selected model (E), and finally a configuration area where it is possible to re-map variables to the presented views (F). In addition to the functionalities presented in Angelini et al. (2019b) the presented version has a richer commands bar (A), re-mapping capabilities that allows a more versatile environment (F) and expanded functionality for the model performance evaluation (D).

The features of the individual views are described below.

Commands bar This area, revised with respect to the previous version of the environment, better identifies the main analysis commands that affect the activities in all the remaining views. Looking at Fig. 4, from left to right we have:

- the counter of the active territorial units (the territorial units contained in the current selection) with respect to their total number;
- a tab menu that allows to select the territorial units geographical aggregation level, composed by green tabular buttons;
- the dimensions (features and classic performance indicators) of the dataset, useful for creating and evaluating performance models, which can be activated using the related checkboxes. This command allows to visualize the subset of selected features, very useful in case of the presence of a big number of them for which only a subset is relevant, and eventually to re-parameterize a model (among those available) in order to conduct a different type of performance analysis;
- the information for the instantiated performance model. It is possible to instantiate a developed performance model to evaluate the territorial units; it is even possible to compare it with a second developed performance model, as shown in Fig. 4. Information reported includes the model's name, its inputs, conditioning factors and outputs, the model's type (e.g. DEA, FDH) and the time interval considered in the model. The same information is replicated in case of the presence of a second model compared to the instantiated one;
- the model selector, which allows to choose among developed performance models, ranging from custom model defined by the Analyst (e.g. Model 1, Model 2) to efficiency models,¹ Data Envelopment Analysis (DEA, Charnes et al. 1978; Banker et al. 1984), Free Disposal Hull (FDH, Deprins et al. 1984), orderM, and their conditional variants ZDEA, ZFDH, ZorderM (for an overview on these models, see Daraio and Simar 2007). The first row allows to instantiate a developed performance model to inspect, while the second row allows to choose a performance model used as reference against which the instantiated one is compared. All the performance models can be used in both modes;
- the time selector, which allows to evaluate the results of the chosen performance model with respect to a specific temporal interval, that can be controlled by means of a slider.

Geographical view This visualization follows the same operating principle illustrated for the Data Exploration Environment. In this instance, however, the color linked to each individual territorial unit is by default proportional to the unit's performance score with respect to the selected performance model. In this way the user can immediately get an overview of the different performance levels given the chosen aggregation level, performance model and time interval. The user can zoom in and out on the map in order to get more details on individual portions of the map. It is also possible to use the map as a highlighting mechanism: by clicking on one or more territorial units, these are highlighted in red on the map and in all other coordinated views, allowing to identify a subset of data of interest for the analysis starting from geographical information. Using the configuration area, the analyst can map to this view the desired feature, performance indicator or developed performance model.

Parallel coordinates view This view, based on the parallel coordinates visual paradigm (Inselberg 2009), shows all the dimensions that are part of the model (inputs, possible conditioning factors, outputs) plus the year of analysis and the ID of the units, with eventually other dimensions not considered in the model for quick comparison and

¹ An introduction to the efficiency analysis literature can be found in Cooper et al. (1999), Coelli et al. (2004) and Fare et al. (1994).

substitution. The purpose of this visualization is to explore the relationships that exist between these features, in order to decide whether to keep them in the developed performance model. From the visual point of view, each of the dimensions is represented as a vertical axis, and each unit as a line that joins the values it has on each axis. Through the commands bar it is possible to filter the features that are represented through the parallel coordinates, in order to avoid clutter effects produced by the plot of a high number of features all at the same time.

Through brushing operations on individual axes, it is possible to perform multi-filter operations on several dimensions, making possible to create very complex filtering expressions while maintaining the ease of use. In addition, by drag and drop interaction, it is possible to exchange all the axes with each other, in order to better highlight any existing correlation, anti-correlation or similarity characteristics on specific subsets of features. Any finding, as mentioned above, serves to better understand the results coming from the performance model or for its creation and eventual modification in terms of features to include/exclude.

Sensitivity analysis view This view allows to conduct sensitivity analysis. The visualization uses two different visual paradigms to relate the different features (inputs, conditioning factors, outputs) that constitute the performance model: the scatter-plot allows to a more detailed analysis of possible correlation between two features, e.g. output and a conditional factor of the performance model. Additional boxplots can be used to map additional features of the performance model (e.g. input factors), and they are coordinated with the scatterplot. The boxplots are designed to allow the interactive selection of disjoint sets of values exploiting their peculiar areas (e.g. median, upper/lower quartiles, outliers) with this filter propagated to the entire visual environment. In this way it is possible to analyze the relationship between the various elements of the performance model in a more precise and granular form, identifying from the distribution subsets of interest which will eventually correspond to the selection of a subset of units that respect the imposed constraints. The effect will therefore support the sensitivity analysis of the performance model (e.g. studying which part of an input factor has the more effect on the performance model) but also support the explorative analysis of the data through filter operations based on features of the model. By using this view, the analyst can inspect with more details the features composing a model and their effects on the model, eventually confirming them. The mapping between features and visual paradigms is free and can be defined by the analyst using the configuration area.

Model analysis and comparison view This view supports the task of exploring the performance scores of the individual units, and expand the sensitivity analysis of the model, in terms of estimating the contribution of each individual features of the model to the performance scores. The visualization is composed of two bars representing rankings, where the units are ordered according to the performance score from top (high performance score) to bottom (low performance score). Each unit is represented as a rectangle, whose color derives from the calculation of the distribution of the performance scores and from the assignment of a color to each of the 4 quartiles (the 3rd and 4th quartiles with deeper shades of green, the 1st and 2nd with deeper shades of red). An informative tooltip, activated by mouse-hover on each rectangle, allows to obtain accurate information on the performance of the unit. The second bar (comparison bar) is initially completely gray, and is activated when individual elements (inputs, conditioning factors) of the model are selected/deselected from the command bar: in this way it is possible to evaluate the displacement in the rank of each single unit with respect to addition/deletion of a feature of the model, and therefore to evaluate the stability of the model compared to the performance scores

produced. Additionally, it is possible to compare two performance models in order to evaluate their differences.

Configuration area This newly introduced area allows to apply all the described analyses on a customized set of features, by mapping them to the different views. The system executes an initial automatic analysis on all the features for identifying their characteristics (e.g. if they are numerical or categorical), and then propose for each of the available views (e.g. for the scatterplot x and y coordinates) the subset of features that are suitable for them. In this way the analyst is helped in her exploration by not having to try wrong or inefficient configurations of the proposed visual analytics environment, allowing her to focus only on the interesting combinations of features/models.

Results and discussion

In order to validate the utility of the proposed solution, we first report on a usage scenario of the proposed system to illustrate the possible execution steps and benefits it brings on performance model development and analysis. After this illustration, we report on the results of a user study that tested operationally the proposed system on a broader and task specific set of performance models.

Usage scenario of the visual analytics system

The usage scenario is based on data collected from 2011 to 2016 from ETER (European Tertiary Education Register), DBLP, Scopus, *InCites* (<https://incites.clarivate.com/>), *Geonames* (<http://www.geonames.org/>), *Eurostat* (<https://ec.europa.eu/eurostat/data/database>), already integrated through OBDA techniques and the *Sapientia* ontology.

The performance models concern the evaluation of research activities in Europe. The units of analysis are Universities and Research centres, considered singularly or aggregated by *Nomenclature of Territorial Units for Statistics* (NUTS). In this analysis units are considered at regional level of detail (NUTS2).

The analysis process begins with all the units considered by default, and the analyst can select a subset of them through the geographical view for analysis on specific sub-areas. In the Parallel coordinates view the basic features and performance indicators are reported. While they do not constitute a model in themselves, they are the basic blocks on which build and test possible performance models. The goal of this step of analysis is to select interesting features to be used for building a performance model. By dynamically defining new intervals on the various dimensions through the brushing filtering of Parallel Coordinates, and immediately verify the cardinality and the characteristics of the resulting selected subset of units, the analyst can explore several combinations and discover relations among dimensions.

Figure 5 represents the output of this analysis (after several cycles of analysis): from left to right, UID is the institution id, E_FDH is the FDH (in)efficiency score (equal to 1 means efficient; the higher it is, the more outputs the unit could proportionally produce to become efficient) STAFF is number of academic staff in FTE (Full Time Equivalent), ENR_S is number of total enrolled students per academic staff, PUB_S is number of publications in WoS (fractional count) per academic staff, P_TOP is number of publications in top 10% of highly cited journals per academic staff, P_COL is percentage of papers done with international collaborations, S_WOM is share of women professors on total academic staff,

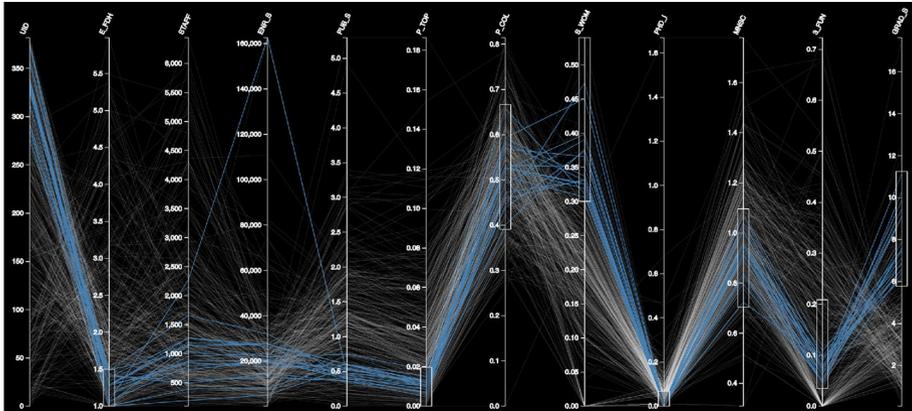


Fig. 5 Example of a complex parallel coordinates filter

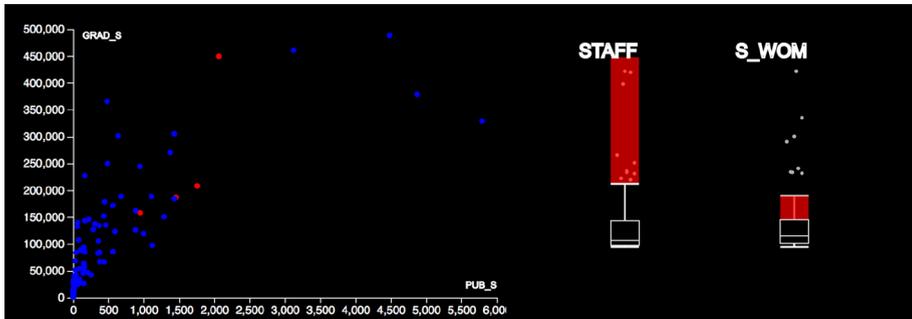


Fig. 6 Example of sensitivity analysis for features of a performance model

PHD_I is PhD intensity, MNCS is Mean Normalized Citation Score (1 corresponds to the world average, > 1 above, < 1 below world average), 3_FUN is share of third party funds in percentage, GRAD_S is total number of graduates per academic staff. Results show that among the most efficient units in teaching and research (i.e. $E_FDH=[1\ 1.5]$) there are those teaching oriented institutions (with the highest values of GRAD_S) in which the S_WOM is the highest ([0.30–0.50]): these are universities with almost zero PhD intensity that are able nevertheless to produce a small fraction of P_TOP publications with MNCS around the world average. Overall, after this phase, the analyst refines the number of initial features to 6 (STAFF, PUB_S, P_COL, S_WOM, PHD_I, GRAD_S) by looking at their trends and their correlations with other features.

Starting from those six features, the analysis proceeds in the *Sensitivity analysis view*, where the indicators can be explored jointly to help in forming performance model configurations and in analyzing contribution to the model by each feature. Figure 6 shows an example of this analysis, where GRAD_S and PUB_S are mapped to the scatterplot and STAFF and S_WOM to the boxplots. With respect to all the units, the analyst's selection is composed by high outliers for academic staff (STAFF) and the 4th quartile for percentage of women staff (S_WOM); the resulting points are highlighted in red in

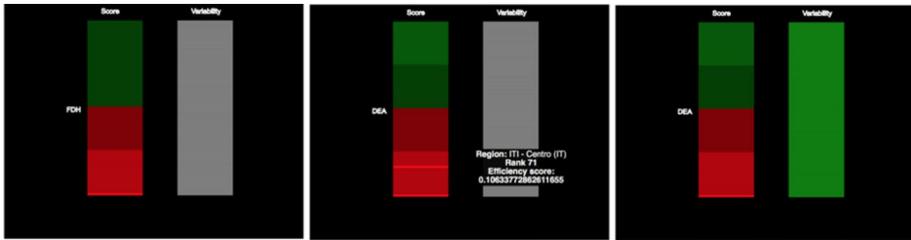


Fig. 7 rank analysis obtained using a complete FDH model (left); the same chart is instantiated through a DEA model (right), and the tooltip reports the score for the Central Italy (Italia Centro) territorial unit (center)

the scatter plot, and the unit can be identified by mouse-hover. Additionally, GRAD_S and PUB_S show good correlation for low values of both features, and an interesting number of outliers for higher values. After the analysis is completed, these 4 indicators are used to construct two models, with GRAD_S and STAFF as inputs, PUB_S as output and S_WOM as a conditional factor, using FDH (model 1) and DEA (model 4) as the base for each model. Additional variations without conditional factors (model 2 and model 3) are created but not chosen for continuing the analysis.

Having obtained the performance models, we now want to analyze them. The analysis is conducted first singularly on each model using the *Model analysis and comparison* view. Results are visible in Fig. 7. It shows a better distribution of scores for model 1 (left) with respect to model 4 (center). In order to improve model 4, an additional variation of it was tested, removing the output factor PUB_S and including P_TOP instead. As visible in Fig. 7 (right), the whole bar is green, which means that the units rank remains stable with respect to this variation, meaning that the variation works similarly to the original.

The analysis continues to the comparison of Model 1 and Model 4. The *Model analysis and comparison view* implements two separate thresholds, one for the ranking (T_{rank}) and one for the model's values (T_{value}). This new feature allows to explore the similarity between two performance models and inspect better their sensitivity. Given two performance models, M_1 , M_2 , and a unit U included in both models, the comparison bar (on the right) reports for this unit a new color-encoding:

- *Dark green*, meaning a situation in which both rank and value are below the chosen threshold such that:

$$|\text{Rank}_{M_1}(U) - \text{Rank}_{M_2}(U)| \leq T_{rank} \text{ and } |M_1(U) - M_2(U)| \leq T_{value}$$

- *Light green*, meaning that while the ranking of the unit is preserved, the associated performance scores differ significantly such that:

$$|\text{Rank}_{M_1}(U) - \text{Rank}_{M_2}(U)| \leq T_{rank} \text{ and } |M_1(U) - M_2(U)| > T_{value}$$

- *Light red*, meaning that the rank is not preserved anymore, even if the two performance scores do not differ significantly, such that:

$$|\text{Rank}_{M_1}(U) - \text{Rank}_{M_2}(U)| > T_{rank} \text{ and } |M_1(U) - M_2(U)| \leq T_{value}$$

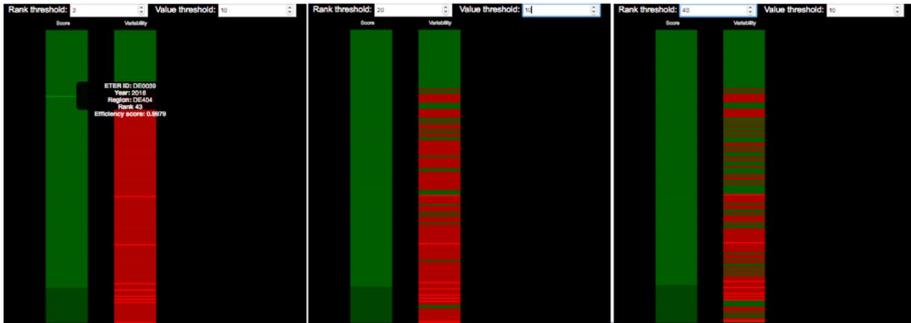


Fig. 8 Illustration of a comparison between two models

- *Dark red*, meaning that both rank and score value are not preserved, identifying strong differences between the two models on that unit, such that:

$$|\text{Rank}_{M1}(U) - \text{Rank}_{M2}(U)| > T_{\text{rank}} \text{ and } |M1(U) - M2(U)| > T_{\text{value}}$$

This analysis can be conducted at run-time for all the units, allowing to grasp these differences or similarities in visual form. Additionally, the values of the thresholds T_{rank} and T_{value} can be dynamically changed during the usage of the system by the analyst, allowing to inspect how much the similarity between two performance models is sensible to the threshold values (see Fig. 7). The analyst can explore this sensitivity by inspecting in real-time the results of different (incremental or random) values for T_{rank} and/or T_{value} and making a better idea of how much models differ or not. The system additionally visualizes state-of-the-art correlation (Pearson, Spearman) and similarity (Kendall-tau) indicators.

We used these features of the *Model analysis and comparison* view to compare more deeply the two hypothesized performance models created in the previous steps, where model 1 is instantiated and model 2 is used for comparison. The comparison is made on the overall 661 units, while Fig. 7 displays the first 200 units by ranking (the view can be scrolled down to visualize more units). It is visible that with the first parameterization of $T_{\text{rank}} = 2$ and $T_{\text{value}} = 10$ the two models produce the same ranks and values for the first 43 units (Fig. 8 left). Being less strict on the rank and imposing $T_{\text{rank}} = 20$ shows additional units presenting similar behavior yet scattered through the ranking (Fig. 8 center). Imposing $T_{\text{rank}} = 40$ produces an additional improvement, even if this can be a too strong assumption (Fig. 8 right). T_{value} does not show any significant impact for these models. We can conclude that Model 1 and Model 4 show good similarity for the first 100 positions of the ranking. Given the differences on the remaining units, overall model 1 results the better model with respect to Model 4.

The presented analysis workflow can be applied an arbitrary number of times until the analyst obtains one or more performance models of interest. In order to evaluate the quality of the obtained performance models, on top of the analyst evaluation, we have conducted a user evaluation described in the next section.

Table 1 Descriptive statistics on general and modelling questions

| | Age | Total time to answer | Avg-total-time | Total-answering-time | Q. 2 (n. of variables) | Q. 3 (n. of units) | Q. 4 (n. of performance models) |
|--------|-------|----------------------|----------------|----------------------|------------------------|--------------------|---------------------------------|
| Min | 22 | 2.65 | 0.1 | 2 | 2 | 396 | 2 |
| Max | 34 | 83.1 | 3.2 | 81 | 28 | 4020 | 12 |
| Mean | 24.17 | 15.75 | 0.61 | 14.85 | 10.53 | 1136.26 | 3.25 |
| Median | 23 | 8.76 | 0.335 | 8.5 | 9 | 1000 | 3 |

Users' evaluation

We evaluated the proposed approach developing a user's study. The study was carried out with the participation of about 70 master's degree students at their last semester, attending the Productivity and Efficiency Analysis course of the Faculty of Management Engineering of Sapienza University of Rome. Within the course the students received theoretical lessons on the development of performance models for 3 months. The students also received training on the main quantitative models for efficiency analysis and laboratory sessions to implement these models, calculating the results in terms of performance (or efficiency) related to the models they formulated on real data to which they had access for the realization of their project work. The students then attended an introductory seminar of the system, lasting 1:30 h plus 30 min for questions, which explained them how to use the functionalities of the proposed system. After this phase, the participants spent around 10 days of use of the tool with their data and for actual model building and evaluation, with frequent interactions with the authors in order to obtain detailed explanations on the functionalities of the system or proposing specific problems of analysis/bugs of the system. Finally, after finishing their works, they were asked to fill in the Questionnaire "Performance models through visual analytics" that is reported in "Appendix". This methodology effectively challenged the system in being used for real scenarios of analysis, where heterogeneity of performance models building and evaluation were captured and characteristics of workflow of analysis with the system were observed. The choice of having master students at their final stage before graduation as participants is motivated by the following reasons:

- they can be gathered in higher number with respect to experts (70 experts would have been difficult to gather) and engaged for more time in the analysis process (Considering all the activities the students were engaged for approximately 1 month), effectively supporting a quantitative evaluation;
- They presented a real task of analysis to be conducted using the proposed system, avoiding the creation of synthetic scenarios of usage that would be needed by experts (given that different experts could have different problems and data) and would have influenced the evaluation task;
- The expertise of the students was high enough in order to be proficient in using both the theoretical concepts of performance modelling and the proposed visual analytics environment.

We obtained 46 filled questionnaires. The respondents were all with a bachelor's degree as last earned title, 43% of which were females, and 37% of the respondents worked most on the visual analytics environment for their project work.

Table 2 Distribution of the answers to Questions 5, 6, 7, 8, 9, 12, 13, 14 and 15

| Q. 5 | % | Answer | Q. 6 (%) | Q. 7 (%) | Q. 8 (%) | Q. 9 (%) | Q. 12 (%) | Q. 13 (%) | Q. 14 (%) | Q. 15 (%) |
|------------------------------|-------|---------------------------------|----------|----------|----------|----------|-----------|-----------|-----------|-----------|
| Between 1 and 3 h | 26.09 | Very useful | 17.39 | 32.61 | 10.87 | 15.22 | 13.04 | 2.17 | 10.87 | 17.39 |
| Between 3 and 8 h | 41.30 | Useful | 26.09 | 26.09 | 39.13 | 43.48 | 43.48 | 17.39 | 36.96 | 41.30 |
| Between 8 and 15 h | 23.91 | Medium useful | 41.30 | 28.26 | 34.78 | 30.43 | 39.13 | 39.13 | 28.26 | 30.43 |
| Less than 1 h/more than 14 h | 8.70 | Not useful/Compl. not useful | 15.22 | 13.04 | 15.22 | 10.87 | 4.35 | 41.30 | 23.91 | 10.87 |
| Total | 100 | | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |

Table 1 reports a descriptive analysis on the general questions and on the questions related to the features and performance models developed and analysed with the system. The average age of the respondents is 24 and their average total answering time has been of around 15 min. The number of features analysed in the visual analytics environment ranges from 2 to 28, with an average of 11. The number of analysed observations ranges from 396 to 4020 and the performance models evaluated with the tool go from 2 to 12. More than 40% of users, spent between 3 and 8 h practicing with the visual analytics environment, while around 24% spent between 8 and 15 h (see the first column of Table 2). Overall, the users express a very high appreciation for the usefulness of the visual analytics environment for executing their tasks. Around 96% of users evaluate as useful the tool (see Table 2, Q. 12 column). The functionality better evaluated from the users has been the *Model analysis and comparison view* in the environment to develop and finalize their performance models (Question 9, Table 2), appreciated by around 90% of users. This appreciation is confirmed by Question 15 (see Table 2) where around 90% of users declare they appreciated the most the task of performance model evaluation included in the visual analytics environment.

Questions 6 to 8 asked an evaluation of the main components of the environment. The Geographic visualization (Question 6) and the Boxplot/Scatterplot visualization (Question 8) were appreciated by around 85% of users. The Parallel Coordinates visualization (Question 7) received appreciation by 87% of users.

The users identify some areas of the environment requiring extensions and improvements. Questions 13 and 14 show that more than 40% of users was not satisfied by the features selection functionality and around 24% of users found not useful the correlation analysis implemented in the visual environment (see Table 2).

Answers to Questions 10 and 11 provided several detailed comments about the usefulness of the environment. The most significant were: “The most useful visual environment tool has been the Parallel coordinates view, it made possible for us to think on some of the results we got and try to understand the reasons behind the relationships among the variables.” “It is useful specially to have a clear vision of the ranking of the variables”. “It is a very interesting way to facilitate the interpretation of results”. “The most useful visual environment for me is Model analysis and comparison visualization since it helps us to choose between DEA and FDH model and between CRS and VRS for our analysis.” These comments show the good appreciation of visually enabled analysis, in particular for performance model exploration, explanation and configuration tasks.

Finally, the answers to Question 16 provided many suggestions for further extensions and improvements of the visual analytics environment that can be summarized in:

- *Instruction manual*, including information about the main components and functionalities of the visual analytics environment;
- possibility to *maximize* one component of the environment and have it full-screen;
- inclusion of *additional visual paradigms*;
- including the possibility to *export* the plots with the units selected during the visual environment exploration for reporting activities;
- improving the integration of the different components of the environment.

Results about the usability of the system

We tested even the usability of the system, using the well-known System Usability Scale (SUS) (Brooke 1996). The results are shown in Fig. 9, where on the left we have the SUS

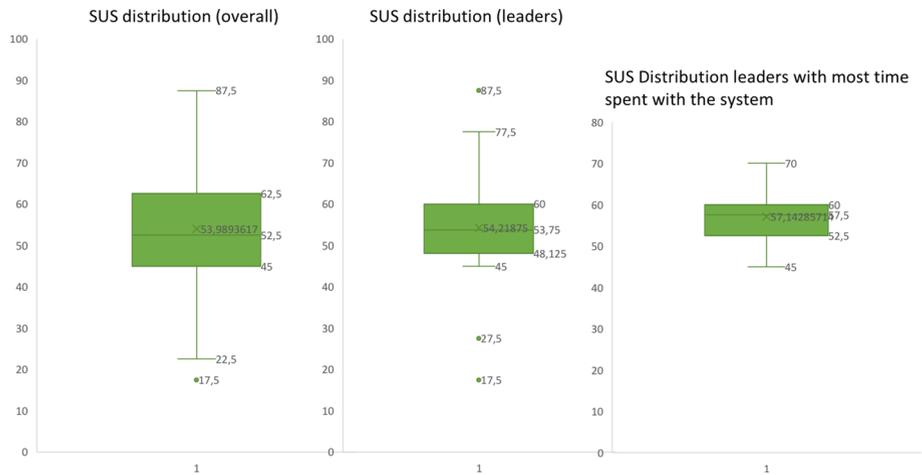


Fig. 9 Boxplots of the usability of the system for three categories of users: all users (left), leaders (centre) and leaders with most time spent with the system (right)

scores computed for the overall population, in the centre the computed score only for persons that declared to be heavily involved in using the system (labelled as leaders), and on the right the computed score for leaders with the maximum interval of system usage (8–15 h).

The results show that usability is a characteristic that should be improved in future development of the system; a possible cause for these results is the heterogeneity of systems from participants that some time did not respect the minimum requirements (e.g. a screen resolution of at least $1920 * 1080$ pixels), that could have prevented some of the users to obtain the desired user experience. Even the request for instruction manual can identify the need for more training. The results, however, are quite good, with an average medium to good scores for each of the 10 questions, resulting in a final score of 53.98 for the overall population, that raises to 54.21 for the leaders and 57.14 for leaders that spent most time with the system (8–15 h) showing a sufficient usability for the system (answers ranges from average 3 to 3.5 score for each of the questions). Nonetheless, more efforts must be produced in order to improve this characteristic and bring the score near the 68 threshold level in order to fully enable the capabilities of the system.

Concluding remarks

In this paper we consolidated the research based on *Sapientia* and OBDA combining it with a visual analytics approach. The new approach proposed allows us to move from Performance Indicators (PI) development to performance model's development, by exploring and exploiting the modelling and the data features within the flexibility of a visual analytics environment. This allows a multi-stakeholder viewpoint on the model of PI, the assessment of the sensitivity and robustness of the performance model in a multidimensional framework. The extensions of the visual analytics environment, described in the previous sections, have been assessed through a user evaluation based on 46 respondents. Overall, 96% of the users express a very high appreciation for the usefulness of the visual analytics

environment for executing their tasks. The functionality better evaluated from the users has been the *Model analysis and comparison view* in the environment to develop and finalize their performance models. Further extensions and improvements of the visual analytics environment suggested by the users include the preparation of an Instruction Manual, improvement of the possibility to maximize one component of the environment and have it full screen; inclusion of additional plots; the possibility to export the plots with the units selected during the visual environment exploration; improving the integration of the different components of the environments. All these suggestions will be taken into account in future works.

Acknowledgements The financial support of the Italian Ministry of Education and Research (through the PRIN Project N. 2015RJARX7), of Sapienza University of Rome (through the Sapienza Awards No. PH11715C8239C105, KIMAR Project and Sapienza Awards No. RM11916B8853C925), and of Clarivate Analytics through the KOL Project, is gratefully acknowledged. This work has been partially supported by MIUR, under the PRIN 2017 project “HOPE” (prot. 2017MMJJRE).

Funding Open access funding provided by Università degli Studi di Roma La Sapienza within the CRUI-CARE Agreement.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

Appendix: Performance models through visual analytics—evaluation questionnaire

Positioning questions:

Question A. Please provide your age

Answer: Number

Question B. Please provide your gender

Answer: F or M

Question C. Please provide your last degree.

Answer: High School degree, Machelor's degree, Master's degree, Doctorate

Technical questions:

Question 1. The test is individual: each participant should answer individually. Before starting the questionnaire please indicate if in your group you are the person that worked most with the tool. Choose one of the following answers:

- I worked most on the visual analytics tool
- I focused on different parts of the project

Question 2. How many variables do you have in your dataset?

Answer: Number

Question 3. How many units of analysis (number of observations) do you have in your dataset?

Answer: Number

Question 4. How many performance models did you analysed?

Answer: Number

Question 5. Please provide us an estimate of the amount of time you spent in using the visual environment:

Answer:

- A. Less than one hour
- B. Between one and three hours
- C. Between three and eight hours
- D. Between eight and 15 hours
- E. More than 15 hours

Question 6. On a scale from 1 to 5 (1 = Completely not useful, 5 = Completely useful) Could you please rate how much useful you found the Geographic visualization (visualization A) in the tool to develop and finalize your performance model/s?

Answer: Scale ranging from 1 to 5

Question 7. On a scale from 1 to 5 (1 = Completely not useful, 5 = Completely useful) Could you please rate how much useful you found the Parallel Coordinates visualization (visualization B) in the tool to develop and finalize your performance model/s?

Answer: Scale ranging from 1 to 5

Question 8. On a scale from 1 to 5 (1 = Completely not useful, 5 = Completely useful) Could you please rate how much useful you found the Boxplot/Scatterplot visualization (visualization C) in the tool to develop and finalize your performance model/s?

Answer: Scale ranging from 1 to 5

Question 9. On a scale from 1 to 5 (1 = Completely not useful, 5 = Completely useful) Could you please rate how much useful you found the Model Analysis and comparison visualization (visualization D) in the tool to develop and finalize your performance model/s?

Answer: Scale ranging from 1 to 5

Question 10. With respect to the more useful visual environment, could you briefly explain the reason for your choice?

Answer: Free text

Question 11. With respect to the least useful visual environment, could you briefly explain the reason for your choice?

Answer: Free text

Question 12. On a scale from 1 to 5 (1 = Not helpful at all, 5 = Completely helpful) How much the visual environment has been helpful for the development of your task compared to the scenario in which you did not use it?

Answer: Scale ranging from 1 to 5

Question 13. On a scale from 1 to 5 (1 = Not useful at all, 5 = Completely useful) With respect to the task of *variable selection* for your model/s. How useful has been the tool?

Answer: Scale ranging from 1 to 5

Question 14. On a scale from 1 to 5 (1 = Not useful at all, 5 = Completely useful) With respect to the task of *variable correlation* for your model/s. How useful has been the tool?

Answer: Scale ranging from 1 to 5

Question 15. On a scale from 1 to 5 (1 = Not useful at all, 5 = Completely useful) With respect to the task of *model evaluation*. How useful has been the tool?

Answer: Scale ranging from 1 to 5

Question 16. Please suggest us a functionality that may be helpful for your model development, you would like to see implemented in the system that is not yet present.

Answer: Free text

Usability questions (SUS)

Question 17. On a scale from 1 to 5 (1 = Strongly Disagree, 5 = Strongly Agree) please answer the following question: ‘I think that I would like to use this system frequently’.

Answer: Scale ranging from 1 to 5

Question 18. On a scale from 1 to 5 (1 = Strongly Disagree, 5 = Strongly Agree) please answer the following question: ‘I found the system unnecessarily complex.’

Answer: Scale ranging from 1 to 5

Question 19. On a scale from 1 to 5 (1 = Strongly Disagree, 5 = Strongly Agree) please answer the following question: ‘I thought the system was easy to use.’

Answer: Scale ranging from 1 to 5

Question 20. On a scale from 1 to 5 (1 = Strongly Disagree, 5 = Strongly Agree) please answer the following question: ‘I think that I would need the support of a technical person to be able to use this system.’

Answer: Scale ranging from 1 to 5

Question 21. On a scale from 1 to 5 (1 = Strongly Disagree, 5 = Strongly Agree) please answer the following question: ‘I found the various functions in this system were well integrated.’

Answer: Scale ranging from 1 to 5

Question 22. On a scale from 1 to 5 (1 = Strongly Disagree, 5 = Strongly Agree) please answer the following question: ‘I thought there was too much inconsistency in this system.’

Answer: Scale ranging from 1 to 5

Question 23. On a scale from 1 to 5 (1 = Strongly Disagree, 5 = Strongly Agree) please answer the following question: ‘I would imagine that most people would learn to use this system very quickly.’

Answer: Scale ranging from 1 to 5

Question 24. On a scale from 1 to 5 (1 = Strongly Disagree, 5 = Strongly Agree) please answer the following question: ‘I found the system very cumbersome to use.’

Answer: Scale ranging from 1 to 5

Question 25. On a scale from 1 to 5 (1 = Strongly Disagree, 5 = Strongly Agree) please answer the following question: ‘I felt very confident using the system.’

Answer: Scale ranging from 1 to 5

Question 26. On a scale from 1 to 5 (1 = Strongly Disagree, 5 = Strongly Agree) please answer the following question: ‘I needed to learn a lot of things before I could get going with this system.’

Answer: Scale ranging from 1 to 5

Final questions

Question 27. Please leave additional comments (if any).

Answer: Free text

References

- Angelini, M., Blasilli, G., Catarci, T., Lenti, S., & Santucci, G. (2019a). Vulnus: Visual vulnerability analysis for network security. *IEEE Transactions on Visualization and Computer Graphics*, 25(1), 183–192. <https://doi.org/10.1109/TVCG.2018.2865028>.
- Angelini, M., Daraio, C., Lenzerini, M., Leotta, F., & Santucci, G. (2019b). Performance model's development: A novel approach encompassing ontology-based data access and visual analytics. In *2019 17th International conference on scientometrics and infometrics, ISSI 2019—Proceedings* (Vol. 2, pp. 1912–1923).
- Angelini, M., Fazzini, V., Ferro, N., Santucci, G., & Silvello, G. (2018). CLAIRE: A combinatorial visual analytics system for information retrieval evaluation. *Information Processing & Management*, 54(6), 1077–1100.
- Angelini, M., & Santucci, G. (2017). Cyber situational awareness: from geographical alerts to high-level management. *Journal of Visualization*, 20(3), 453–459.
- Antonoli, N., Castanò, F., Civili, C., Coletta, S., Grossi, S., Lembo, D., et al. (2013). Ontology-based data access: The Experience at the Italian Department of Treasury. In *Conference on advanced information systems engineering, CAiSE 2013-Co-located with 25th international conference on advanced information systems engineering, CAiSE 2013* (Vol. 1017, pp. 9–16).
- Antonoli, N., Castanò, F., Coletta, S., Grossi, S., Lembo, D., Lenzerini, M., et al. (2014). Ontology-based data management for the Italian Public Debt. In *Formal ontology in information systems: Proceedings of the eighth international conference (FOIS 2014)*. IOS Press.
- Aracri, R. M., Radini, R., Scannapieco, M., & Tosco, L. (2017, June). Using ontologies for official statistics: The ISTAT experience. In *International conference on web engineering* (pp. 166–172). Cham: Springer.
- Banker, R. D., Charnes, A., & Cooper, W. W. (1984). Some models for estimating technical and scale inefficiencies in data envelopment analysis. *Management Science*, 30(9), 1078–1092.
- Battle, L., Angelini, M., Binnig, C., Catarci, T., Eichmann, P., Fekete, J. D., Santucci, G., Sedlmair, M., & Willett, W. (2018). Evaluating visual data analysis systems: A discussion report. In *Proceedings of the workshop on human-in-the-loop data analytics (HILDA'18)* (Vol. 4, pp. 1–6). New York, NY, USA: Association for Computing Machinery. <https://doi.org/10.1145/3209900.3209901>.
- Belton, V., Stephen, P., & Vickers, S. P. (1993). Demystifying DEA—A visual interactive approach based on multiple criteria analysis. *Journal of the Operational Research Society*, 44(9), 883–896.
- Brooke, J. (1996). SUS-A quick and dirty usability scale. *Usability Evaluation in Industry*, 189(194), 4–7.
- Calvanese, D., Cogrel, B., Komla-Ebri, S., Kontchakov, R., Lanti, D., Rezk, M., et al. (2017). Ontop: Answering SPARQL queries over relational databases. *Semantic Web*, 8(3), 471–487.
- Calvanese, D., De Giacomo, G., Lembo, D., Lenzerini, M., Poggi, A., Rodriguez-Muro, M., et al. (2011). The MASTRO system for ontology-based data access. *Semantic Web*, 2(1), 43–53.
- Calvanese, D., De Giacomo, G., Lenzerini, M., Nardi, D., & Rosati, R. (1998). Description logic framework for information integration. In *Proceedings of KR* (pp. 2–13).
- Catarci, T., Di Mascio, T., Franconi, E., Santucci, G., & Tessaris, S. (2003). An ontology based visual tool for query formulation support. In *European conference on Artificial Intelligence* (Vol. 2889, pp. 32–33). Lecture notes in computer science.
- Charnes, A., Cooper, W. W., & Rhodes, E. (1978). Measuring the inefficiency of decision making units. *European Journal of Operational Research*, 2, 429–444.
- Coelli, T., Rao, D. S. P., & Battese, G. E. (2004). *An introduction to efficiency analysis* (2nd ed.). Dordrecht: Kluwer Academic Publishers.
- Cook, K. A., & Thomas, J. J. (2005). Illuminating the path: The research and development agenda for visual analytics. *Tech. rep.*, Pacific Northwest National Lab. (PNNL), Richland, WA, USA.
- Cooper, W. W., Seiford, L. M., & Tone, K. (1999). *Data envelopment analysis: A comprehensive text with models, applications, references and DEA-Solver software*. Boston: Kluwer Academic Publishers.

- Daraio, C. (2017). A framework for the assessment of research and its impacts. *Journal of Data and Information Science*, 2(4), 7–42.
- Daraio, C. (2019). Econometric approaches to the measurement of research productivity. In W. Glänzel, H. F. Moed, H. Schmoch, & M. Thelwall (Eds.), *Springer handbook of science and technology indicators*. Cham: Springer.
- Daraio, C., & Bonaccorsi, A. (2017). Beyond university rankings? Generating new indicators on universities by linking data in open platforms. *Journal of the Association for Information Science and Technology*. <https://doi.org/10.1002/asi.23679>.
- Daraio, C., & Glänzel, W. (2016). Grand challenges in data integration. State of the art and future perspectives: An introduction. *Scientometrics*, 108(1), 391–400.
- Daraio, C., Lenzerini, M., Leporelli, C., Moed, H. F., Naggari, P., Bonaccorsi, A., et al. (2015). Sapientia: The ontology of multi-dimensional research assessment. In A. A. Salah, Y. Tonta, A. A. Akdag Salah, C. Sugimoto, & U. Al (Eds.), *Proceedings of ISSI 2015 Istanbul: 15th international society of scientometrics and informetrics conference, Istanbul, Turkey, 29 June to 3 July, 2015, Bogaziçi University Printhouse* (pp. 965–977).
- Daraio, C., Lenzerini, M., Leporelli, C., Moed, H. F., Naggari, P., Bonaccorsi, A., et al. (2016a). Data integration for research and innovation policy: An ontology-based data management approach. *Scientometrics*, 106(2), 857–871.
- Daraio, C., Lenzerini, M., Leporelli, C., Naggari, P., Bonaccorsi, A., & Bartolucci, A. (2016b). The advantages of an ontology-based data management approach: Openness, interoperability and data quality. *Scientometrics*, 108(1), 441–455.
- Daraio, C., & Simar, L. (2007). *Advanced robust and nonparametric methods in efficiency analysis. Methodology and applications*. New York: Springer.
- Deprins, D., Simar, L., & Tulkens, H. (1984). Measuring labor inefficiency in post offices. In M. Marchand, P. Pestieau, & H. Tulkens (Eds.), *The performance of public enterprises: Concepts and measurements* (pp. 243–267). Amsterdam: North-Holland.
- Erhan, H., Woodbury, R., & Salmasi, N. H. (2009). Visual sensitivity analysis of parametric design models: Improving agility in design. In T. Tidafi & T. Dorta (Eds.), *Proceedings of joining languages, cultures and visions: CAAD futures 2009, PUM, 2009* (pp. 815–829).
- Fare, R., Grosskopf, S., & Lovell, C. A. K. (1994). *Production frontiers*. Cambridge: Cambridge University Press.
- Floridi, L. (2014). *The fourth revolution: How the infosphere is reshaping human reality*. OUP Oxford.
- Giese, M., Soylu, A., Vega-Gorgojo, G., Waaler, A., Haase, P., Jiménez-Ruiz, E., et al. (2015). Optique: Zooming in on big data. *Computer*, 48(3), 60–67.
- Inselberg, A. (2009). *Parallel coordinates: Visual multidimensional geometry and its applications* (Vol. 20). Berlin: Springer.
- Keim, D., Andrienko, G., Fekete, J. D., Gorg, C., Kohlhammer, J., & Melancon, G. (2008). *Visual analytics: Definition, process, and challenges* (pp. 154–175). Berlin: Springer.
- Kharlamov, E., Hovland, D., Skjæveland, M. G., Bilidas, D., Jiménez-Ruiz, E., Xiao, G., et al. (2017). Ontology based data access in Statoil. *Journal of Web Semantics*, 44, 3–36.
- Lembo, D., Pantaleone, D., Santarelli, V., & Savo, D. F. (2016). Easy OWL drawing with the Graphol visual ontology language. In *KR* (pp. 573–576).
- Lenzerini, M., & Daraio, C. (2019). Challenges, approaches and solutions in data integration for research and innovation. In W. Glänzel, H. F. Moed, H. Schmoch, & M. Thelwall (Eds.), *Springer handbook of science and technology indicators*. Cham: Springer.
- Moral-Muñoz, J. A., López-Herrera, A. G., Herrera-Viedma, E., & Cobo, M. J. (2019). Science mapping analysis software tools: A review. In W. Glänzel, H. F. Moed, H. Schmoch, & M. Thelwall (Eds.), *Springer handbook of science and technology indicators* (pp. 159–185). Cham: Springer.
- Mosca, A., Roda, F., & Rull, G. (2018). The ontology for research and innovation policy making. In *Formal ontology in information systems: Proceedings of the 10th international conference (FOIS 2018)* (Vol. 306, p. 200). IOS Press.
- Santarelli, V., Lembo, D., Ruzzi, M., Ronconi, G., Bouquet, P., Molinari, A., et al. (2019). Semantic technologies for the production and publication of open data in ACI—Automobile Club d’Italia. In *Proceedings of the ISWC 2019 satellite tracks (posters & demonstrations, industry, and outrageous ideas) co-located with 18th international semantic web conference (ISWC)* (Vol. 2456, pp. 307–308).
- Schubert, A., & Schubert, G. (2019). All along the *h*-index-related literature: A guided tour. In W. Glänzel, H. F. Moed, H. Schmoch, & M. Thelwall (Eds.), *Springer handbook of science and technology indicators* (pp. 301–334). Cham: Springer.

- Silva, I. C. S., Santucci, G., & Freitas, C. M. D. S. (2019). Visualization and analysis of schema and instances of ontologies for improving user tasks and knowledge discovery. *Journal of Computer Languages*, 51, 28–47.
- Sivertsen, G. (2019). Developing current research information systems (CRIS) as data sources for studies of research. In W. Glänzel, H. F. Moed, H. Schmoch, & M. Thelwall (Eds.), *Springer handbook of science and technology indicators* (pp. 667–683). Cham: Springer.
- Soylu, A., Kharlamov, E., Zheleznyakov, D., Jimenez-Ruiz, E., Giese, M., Skjæveland, M. G., et al. (2018). OptiqueVQS: A visual query system over ontologies for industry. *Semantic Web*, 9(5), 627–660.
- von Mayr, G. (1877). *Die Gesetzmäßigkeit im Gesellschaftsleben: Statistische Studien*. München: Oldenbourg.
- Wildgaard, L. (2019). An overview of author-level indicators of research performance. In W. Glänzel, H. F. Moed, H. Schmoch, & M. Thelwall (Eds.), *Springer handbook of science and technology indicators* (pp. 361–396). Cham: Springer.