

ISSN 2281-4299



DIPARTIMENTO DI INGEGNERIA INFORMATICA  
AUTOMATICA E GESTIONALE ANTONIO RUBERTI



**SAPIENZA**  
UNIVERSITÀ DI ROMA

**Performance Model's development: A  
Novel Approach encompassing Ontology-  
Based Data Access and Visual Analytics**

Marco Angelini  
Cinzia Daraio  
Maurizio Lenzerini  
Francesco Leotta  
Giuseppe Santucci

Technical Report n. 11, 2019

# Performance Model's development: A Novel Approach encompassing Ontology-Based Data Access and Visual Analytics

Marco Angelini<sup>1</sup>, Cinzia Daraio<sup>1</sup>, Maurizio Lenzerini<sup>1</sup>, Francesco Leotta<sup>1</sup>, Giuseppe Santucci<sup>1</sup>

<sup>1</sup>{angelini, daraio, lenzerini, leotta, santucci}@diag.uniroma1.it

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

## 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 paper we propose a new approach which is able to move forward, from indicators' development to performance model's development. It combines the advantages of the Ontology-based data Access (OBDA) integration with the flexibility and robustness of a Visual Analytics (VA) environment. A detailed description of such an approach is presented in the paper. The approach is evaluated through a comprehensive user's study that proves the added capabilities and the benefits that an analyst of performance models can have by using this approach.

**Keywords:** Education and research, performance assessment, performance modelling, ontology-based data access, visual analytics

## Introduction: An advanced models' development approach

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 and 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 (2017a) showed that the formulation of models of metrics (in this paper we will use metrics and indicators as synonyms) 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, test the assumptions, highlight relevant relations; and for *improving*, document/verify the assumptions, systematize the problem and the evaluation/choice done, explicit the dependence of the choice to the scenario. Moreover, there are several *drawbacks* in modelling, which have to be taken into account. The main pitfalls relate to the targets that are not quantifiable; the complexity, uncertainty and changeability of

the environment in which the system works, to the limits in the decision context, and, last but not least, to the intrinsic complexity of calculation of the objective of the analysis.

In this paper we depart 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 you want to add a new data source or you want a different indicator you have to restart the process from the scratch.

We support an alternative approach based on an OBDA system for Research & Innovation (R&I in the following) 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*. It is a couple (pair)  $O = \langle TBox, ABox \rangle$ , 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 (instances). An ontology populated by instances and completed by rules of inference is defined as *knowledge base* (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, implemented with *Sapientia*, represents the ontology of multidimensional research assessment (Daraio, Lenzerini 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, Lenzerini et al. (2016a) showed the advantages of an OBDA system for R&I integration and Daraio, Lenzerini 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, pag. 368 and pag. 369).

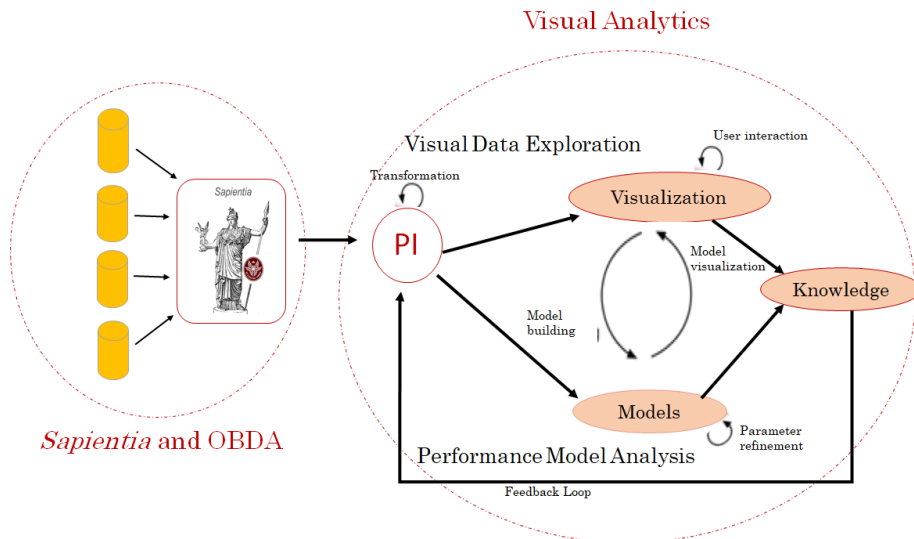
The main contribution of this paper, that extends the work of Angelini et al. (Angelini, 2019), is making a step further, on our previous researches and to propose a new approach for the multidimensional assessment of research and its impact based on the combination of OBDA and Visual Analytics. This novel approach allows for the development and evaluation of performance models instead of the traditional indicators' building system.

## Combining OBDA and Visual Analytics

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. First of all, 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 would be 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 & 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 it allows navigation in the data space, in order to obtain an overview of the data, eventually tunable to the required level of detail, and the ability to apply complex analysis workflows that aim at explainability and the ability to obtain summary reports of the findings discovered during the analysis phase. See Figure 1 for an overview.



**Figure 1. An illustration of our approach that combines *Sapientia*, OBDA and Visual Analytics. PI states for Performance Indicator.**

While some previous work exists on the subject, like the work in (Belton & Vickers, 1993) for DEA analysis, or the work by Erhan et al. (Erhan, 2009) for visual sensitivity analysis of general parametric models, most of the works focuses 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.

The Visual Analytics approach developed in this paper allows us to move from performance indicators (PIs) development to performance model 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 and the assessment of the sensitivity and robustness of the PI model in a multidimensional framework.

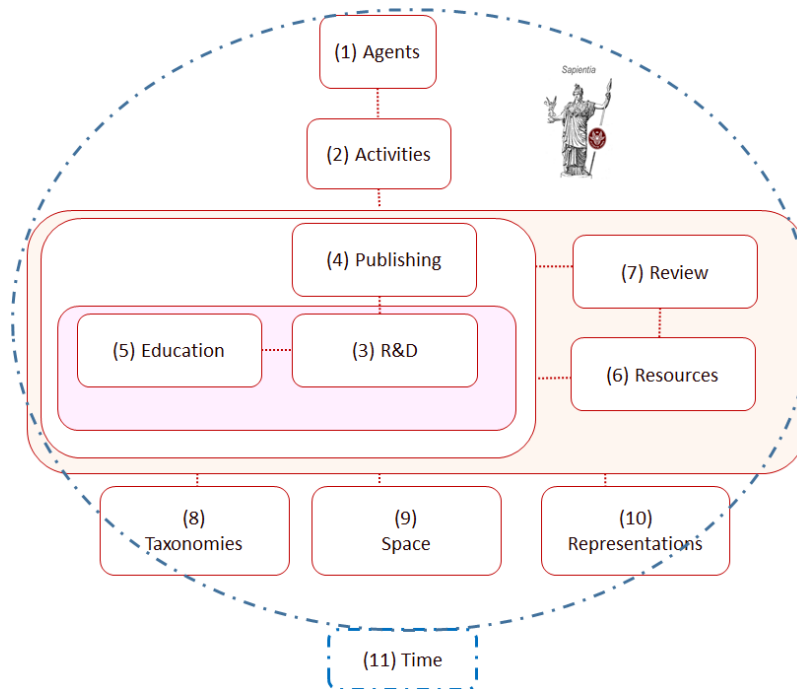
In the next section we outline the main features of *Sapientia* (the Ontology of Multidimensional Research Assessment). After that we present our Visual Analytics environment for the performance model's development together with an illustration of its potentialities. After this

section we present the methodology and results of the users study conducted for evaluating the approach. The final section concludes the paper.

### **OBDA at work through *Sapientia*: The Ontology of multidimensional research assessment**

*Sapientia*, the Ontology of Multidimensional Research Assessment (Daraio et al. 2015, 2016a, 2016b), models all the activities relevant for the evaluation of research and for assessing its impact (see Figure 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 *Sapientia* 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.



**Figure 2. Modules of *Sapientia* 3.0.** 1. **Agents:** describes all human actors and institutions involved in the education, research and innovation process. 2. **Activities:** describes the activities and projects the agents of the previous module are involved in. 3. **R&D:** describes the different products (e.g., publications, patents) that are produced in the knowledge production process. 4. **Publishing:** describes how knowledge products are published and made available to the public. 5. **Education:** introduces concept related to universities and courses. 6. **Resources:** describes all the ways and institution can be funded. 7. **Review:** describes the process entities related to the publishing activity. 8. **Taxonomy:** describes the elements that allows to define taxonomies applied to the different modules. 9 and 11: **Space and Time:** allow to describe respectively geographical entities and time instants and ranges. 10. **Representation:** allows to describe the fact that single instances of other modules can be represented in different ways by the different sources used in *Sapientia*.

*Sapientia* acquires information from multiple sources, whose content can be overlapping. The same entity modelled in the *Sapientia* 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. 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.

*Sapientia* extract information, among others, from the following datasets; for each dataset a table is presented with the number of records imported for each category (and optionally subcategory) of data:

- *Scopus*. A very large abstract and citation database of peer-reviewed literature, containing information about scientific journals, books and conference proceedings. Scopus provides information about authors' affiliations as well. The available REST interface allows to retrieve: document information, document citations data, percentiles data and journal percentiles data. Here the data cannot be massively downloaded, requiring instead a per author search.
- *ETER*. The ETER (European Tertiary Education Register) consortium acquired extensive information pertinent to tertiary educational institution of many European countries. Data have been acquired by the consortium for the years 2011-2016 and are publicly available (<https://eter-project.com>).

Category	Subcategories	Cardinality
Institutions		2892
Institution Variables		520

- *DBLP*. A service that provides open bibliographic information on major computer science journals and proceedings. Data is available through massive XML files.

Category	Subcategories	Cardinality
Document	Article	1594942
	Book	13086
	In Collection Papers	40382
	In Proceedings Papers	1948419
	PhD Thesis	44351
	Proceedings	33136
	Web Links	1864331
Citations		41478638

- The *InCites* (<https://incites.thomsonreuters.com>) dataset contains research indicators organized on a geographical base. Data can be downloaded in the form of CSV files that are then imported using an ad-hoc procedure.

Category	Subcategories	Cardinality
Geographic Entities	Country	214

	State	104
	Country Cluster	11
Geographic Entity Variables		28
Institutions		7616
Institution Variable		48

- *Geonames* (<http://www.geonames.org/>) is a dataset that contains information about geographical areas at any level. The dataset can be freely download, and has been employed to match geographical entities from the different data sources.

Category	Subcategories	Cardinality
Geographic Entities	City	27358
	Province	20971
	Region	3618
	Country	192

- *Eurostat* (<https://ec.europa.eu/eurostat/data/database>) is a dataset containing valuable information about regions and countries composing European Union and their population.

Category	Subcategories	Cardinality
Geographic Entities	Country	195
	Province	1818
	Region	448
	Region Cluster	190
Geographic Entity Variables		360

- *Web of Science* database is going to be included as well.

From above tables, we can see that imported sources contain data about:

- Objects including geographic entities (e.g. cities, regions and countries), institutions (e.g., universities, research centres) and documents (e.g., article books)
- Variables on objects. To each variable a value is assigned for each year and for each period of the year (e.g., quarters, months). For example, a variable may report number of students for each quarter of a university in several different years.
- Relationships between objects (e.g., citations among articles, containment relationships between geographic entities)

Each data source may talk about the same objects using a different terminology, and this is why the data integration of the *Sapientia* is in place.

The data manipulation layer of the *Sapientia*, which allows to populate the ontology from the data sources, is composed of an indexing module, an entity-resolution module and a normalization module. In general, operations included in data manipulation are expensive in terms of processing power and time. Throughout this section we will report computation time required by specific operation on a Linux machine with 16GB or RAM memory with two quad core processors Intel i7-6700.

In general terms, the *indexing module* creates and maintains up to date the indices that are used by the *entity resolution module* to implement the blocking functionalities that allow to keep the



time complexity of the entity-resolution algorithms under control. This module has the dual purpose of easing the definition of the mappings toward the *Sapientia* Ontology, and creating the basis for a common interface of the entity-resolution algorithms. Indices inside the *Sapientia* application are implemented using the Hibernate search (<http://hibernate.org/search/>) library and the Lucene indexer and searcher (<http://lucene.apache.org/>). The indexing activity is performed while the database is populated, thus it adds a very low additional computation time. Entity resolution is the task of connecting matching entities between different data sources. As this kind of process is exponential in complexity with the number of data sources and entities per data sources, it is split in two phases:

- *Blocking*, by employing indices to create groups of potential matching entities. Blocking is an expensive task, taking three full days of computation.
- *Entity matching*, which finds matching entities inside clusters identified in blocking phase. Entity matching task took one day of computation.

After matching entities have been recognized by entity resolution, the *normalization step* is employed in order to provide a uniform representation for the information contained in different and heterogeneous data sources. These uniform representations are called mappable entities. These mappable entities are mapped to ontology entities through an operation called mapping. *Sapientia* uses the Mastro Ontology-Based Data Access (OBDA) management system (<http://www.dis.uniroma1.it/~mastro/?q=node/1>). The *Sapientia* ontology, however, is defined over a richer language than the one supported by Mastro. Hence, we used the OWL2DL tool in order to obtain a simplified version of the *Sapientia* ontology that conforms to the DL-light language supported by Mastro.

The definition of the mappings in Mastro is XML based. There are three types of ontology predicate mappings: concept, role and attribute.

As suggested by the names, the concept predicate mapping refers to entities, the role predicate mapping puts entities in relation, populating a role, while the attribute mapping relates an entity with a constant, which is the value of its attribute.

#### *Some examples of extraction and mapping of relevant data*

In order to show the potential of the proposed approach, we will show how indicators can be extracted from the ontology and grouped according to a specific level of analysis. In the illustration units are identified with European denoted Nomenclature of Territorial Units for Statistics (NUTS) code. The modules of the ontology interested in this query are:

- The *Agents module*, which contains the concept of University as a specialization of the concept of Organization. An Organization has an Organization State, which represents the evolution of the Organization in time, and that refers to the Residence.
- The *Space module* that contains the concept of Residence as a specialization of a Position. A Position has an Entrance, which is localized in an Address inside a City. The City is a Territory, and European Cities are European Territories that can be aggregated by NUTS codes.
- The *Taxonomy module* where an Organization is contained in a Taxonomic Unit. Each Taxonomic Unit has a State that has indicators as attributes.

For a specific university denoted by its Eter ID, we can for example compute the cardinality of academic staff with the following SPARQL query:

```
select ?academic_staff {
  ?org sapientia:has_place_in ?taxon_unit .
  ?org a sapientia:University .
  ?taxon_unit sapientia:has_state_of_taxonomic_unit ?state_tax .
```



```

        ?state_tax a sapientia:Present_state .
        ?state_tax sapientia:teacher_population ?academic_staff .
    }

```

In order to group by a specific NUTS codes, it is possible to extend the previous query as follows:

```

select SUM(?academic_staff), ?nuts2 {
    ?org sapientia:has_place_in ?taxon_unit .
    ?org a sapientia:University .
    ?taxon_unit sapientia:has_state_of_taxonomic_unit ?state_tax .
    ?state_tax sapientia:teacher_population ?academic_staff .
    ?state_tax a sapientia:Present_state .
    ?org sapientia:has_state_of_organization ?org_state .
    ?org_state a sapientia:Present_state .
    ?org_state sapientia:has_residence ?resid .
    ?resid a sapientia:Legal_residence .
    ?resid sapientia:has_entrance ?entr .
    ?entr a sapientia:Address .
    ?entr sapientia:is_in_the_city ?city .
    ?city a sapientia:European_territory .
    ?city sapientia:is_territory_part_of ?region .
    ?region a sapientia:Small_european_region .
    ?region sapientia:NUTS2ref ?nuts2 .
    ?region sapientia:NUTS2ref ?nuts1 .
    ?region sapientia:NUTS2ref ?nuts3
}
GROUP BY ?nuts2

```

where the results have been grouped by NUTS2. It is possible to easily modify the query in order to group by other levels of NUTS. In a similar way, *mutatis mutandis*, it is possible to extract the data and indicators that will be used for the Performance Indicator and model development that is described in the next sections.

The memory required by MASTRO and the query computation time strongly depends on the complexity of the issued query. For the abovementioned query, a memory occupation of 1GB was required and the computation time was quite low (less than a second). Anyway, the memory requirements represent a considerable bottleneck of the system, as a lot of data require to be materialized.

## The Visual Analytics environment

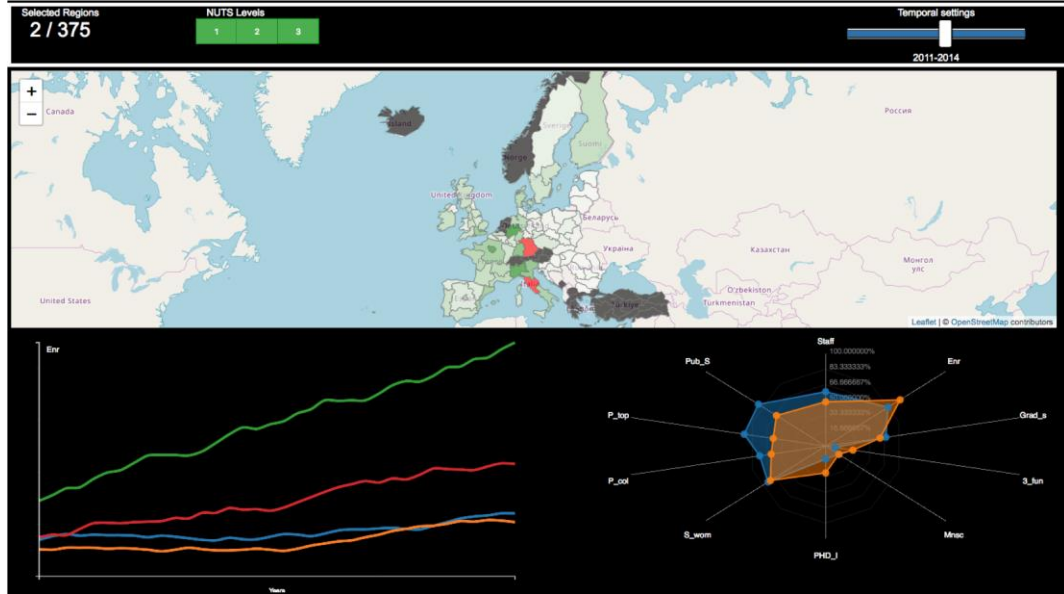
This section describes the Visual Analytics environment and its main features. The developed solution uses Visual Analytics techniques to represent data from publications and education obtained from the OBDA approach described in the previous section and complete the workflow. The system is implemented through Web technology. Clearly the large quantity of indicators and basic features for the different units of analysis, including the territorial ones, and the different years of analysis increases 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 in 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 explicitly aimed at analyzing the model development and performance computed on these indicators, based on the definition and exploration of performance models. The environment is instantiated on European research and education institutions as a case study 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 positioning of the institutions of

interest; additionally, it allows to explore different performance models and to evaluate their goodness and fitness. Further, 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 or resources, contextual factors or outputs) contribute more to the performance of the institution with respect to the chosen model. The following subsections will 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 Figure 3.



**Figure 3. Data Exploration Environment**

These three views are:

- *Geographic view*: which allows for geolocating of 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.

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 particular 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*: 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 form starting from the center. The indicators are arranged one per axis and the chart presents several lines that join the points on each axis in the number of one per institution or territorial unit. When the user selects one or more territorial units, the corresponding splines are highlighted, in order to allow an easy visual comparison between the different territorial units selected on their different 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

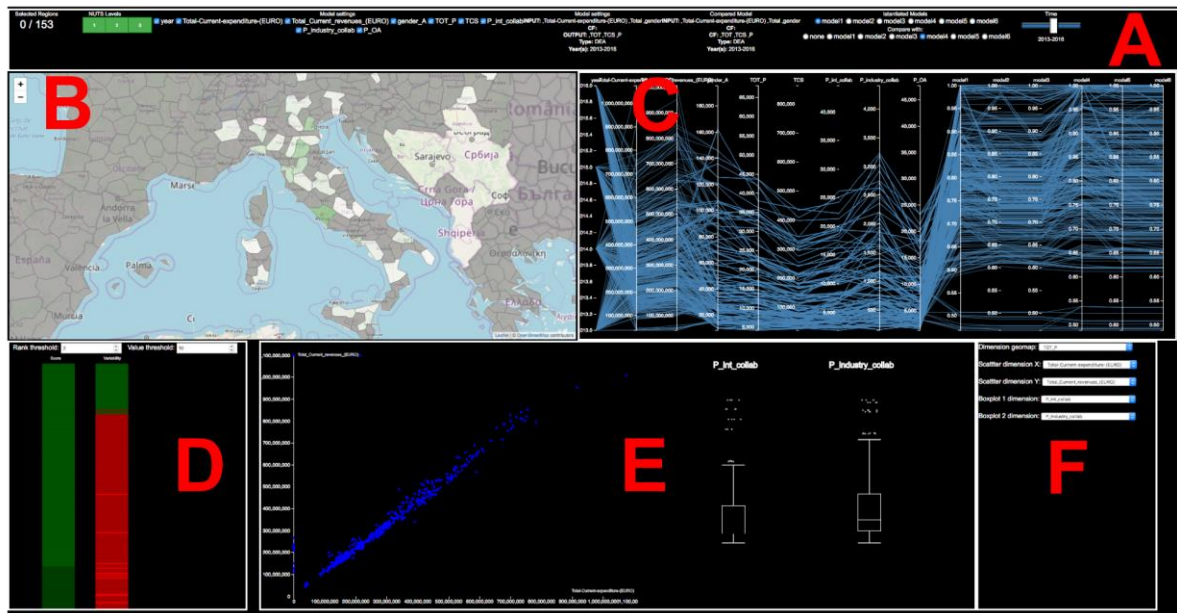
performance of a territorial unit, or generally of a given unit of analysis, not only to other units but also to the aggregate behavior between the territorial units.

*-Linechart view:* This visualization allows analyzing the time course of the evaluation measures used for the units under analysis. It is possible to analyze both multiple territorial units to compare the trend of the same measure on them, and to analyze multiple measures on the same territorial unit, in order to have an overview of the progress of the unit itself, and a combination based on multiple territorial units and multiple measures. In this case the color-coding outlines all the measures 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 model data 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 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.

### Performance Model Analysis Environment

This environment is the core of the system, and it is dedicated to the analysis of performances of the model used for analyzing the units. This part of the system has been evolved with respect to the work by Angelini et al. (Angelini, 2019) in order to improve its functionalities and informativeness, after having been tested and used by performance model creators. An overview of the new environment is visible in Figure 4.



**Figure 4. New configuration of the Performance Model Analysis Environment.** In addition to the functionalities presented in Angelini, et al (2019) it has a more general commands bar (A), additional selectors for the whole environment presented in the configuration area (F) and a redesigned functionality for the model performance evaluation (D).

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 parameters of the selected model (E), and

finally a configuration area where is possible to select the different features to map on the other views (F). 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 will affect the selections in all remaining views. From left to right we have:

- the counter of the territorial units active with respect to the total (the territorial units contained in the current selection);
- a tab that allows to select the geographical aggregation level on which to conduct the analysis, represented as green tabular buttons;
- the parameters and features of the dataset, useful for creating and evaluating performance models, which can be activated using the appropriate checkboxes. This command allows to visualize the subset of selected features, very useful in case of the presence of a big numbers of features 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 analysis of performance;
- the information for an instantiated performance model. It is possible to instantiate it for a selected model and eventually, as in Figure 3, for a second model used for comparison. Information reported includes the model's name, the inputs, conditioning factors and outputs, the model's type (e.g. DEA, FDH) and the time interval considered by the model. The same information is replicated in case of the presence of a second model compared to the selected one;
- the model selector, which allows to choose between families of performance models, ranging from custom model defined by the Analyst (e.g. Model 1, Model 2) to efficiency models<sup>1</sup>, 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 select a performance model to inspect, while the second row allows to choose a performance model used as reference against which the selected model is compared. All the models can be used in both modes;
- the time selector, which allows to evaluate the result of the chosen model with respect to a 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 model. In this way the user can immediately get an overview of the different performance levels given the chosen hierarchical level, model and time interval. The user can zoom in 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 units, these are highlighted in red on the map and in all other coordinated views, allowing to identify a subset of data of interests starting from geographical coordinates of the unit. Using the configuration area the analyst is able to map to this view a different feature or model from the dataset.

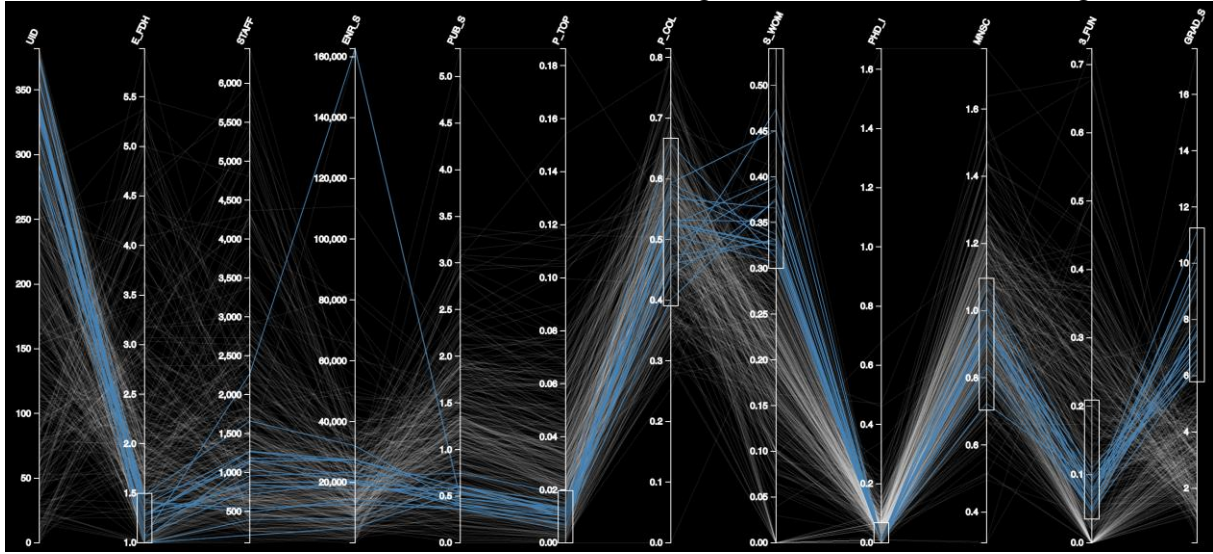
*Parallel coordinates*: this view, based on the parallel coordinates visual paradigm (Inselberg, 2009), shows all the dimensions that are part of the model (inputs, possible conditioning factors,

---

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

outputs) plus the year of analysis and the ID of the units, with eventually other dimensions not considered in the model. The purpose of this visualization is to explore the relationships that exist between these quantities, in order to decide whether or not to keep them in the selected 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 even 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 select very complex filtering expressions while maintaining the ease of creating these filters.

By dynamically define new intervals on the various dimensions, and immediately verify the cardinality and the characteristics of the resulting subset, the analyst can explore several combinations and discover relations among dimensions (see Figure 5).



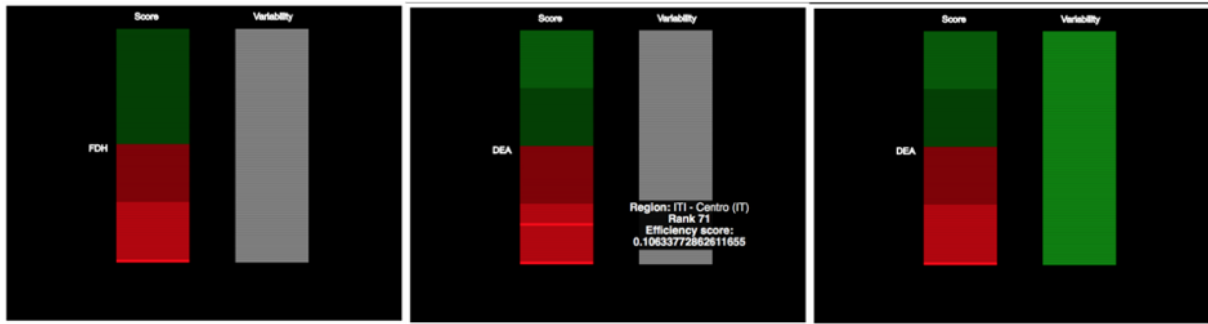
**Figure 5. Example of parallel coordinates filter.** Axes, from the left: UID is the institution id number, E\_FDH is the FDH (in)efficiency score (equal to 1 means efficient; the higher it is, more outputs the unit could proportionally produce to become efficient) STAFF is number of academic staff 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, 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 %, GRAD\_S is total number of graduates per academic staff. The filter shows 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.

In addition, by drag and drop interaction, it is possible to exchange all the axes with each other, in order to better highlight any correlations, anti-correlations or similarity characteristics on specific subsets of data among dimensions. Any findings, as mentioned above, serve to better understand the results coming from the performance model used and its eventual modification in terms of features to include/exclude.

*Model Analysis and comparison:* This view supports the tasks of exploring the performance scores of the individual units, and the sensitivity analysis on the model, in terms of estimating



the contribution of each individual parameter 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 3<sup>rd</sup> and 4<sup>th</sup> quartiles with deeper shades of green, the 1<sup>st</sup> and 2<sup>nd</sup> 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 parameter of the model, and therefore be able to evaluate the stability of the model compared to the performance scores produced, and the sensitivity of the performance model in terms of contribution that any parameter produces in the ranking (see Figure 6).



**Figure 6.** rank analysis obtained using a complete FDH model (left); the same chart is instantiated through a DEA model, and the tooltip reports the score for the “Italia Centro” territorial unit (center); finally, the result on the variability obtained by removing the output factor PUB\_S and including P\_TOP (right). As you can see, the whole bar is green, which means that the units rank remains stable with respect to this input, which could be replaced by another more significant input.

In order to make this analysis more explorative, the current Visual Analytics Environment expanded the capabilities of this view. It is now possible even to compare two different performance models (and not anymore only a variation of a single model) and the view even implements two separate thresholds, one for the ranking ( $T_{rank}$ ) and one for the model's values ( $T_{value}$ ). This new feature allows us 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:  

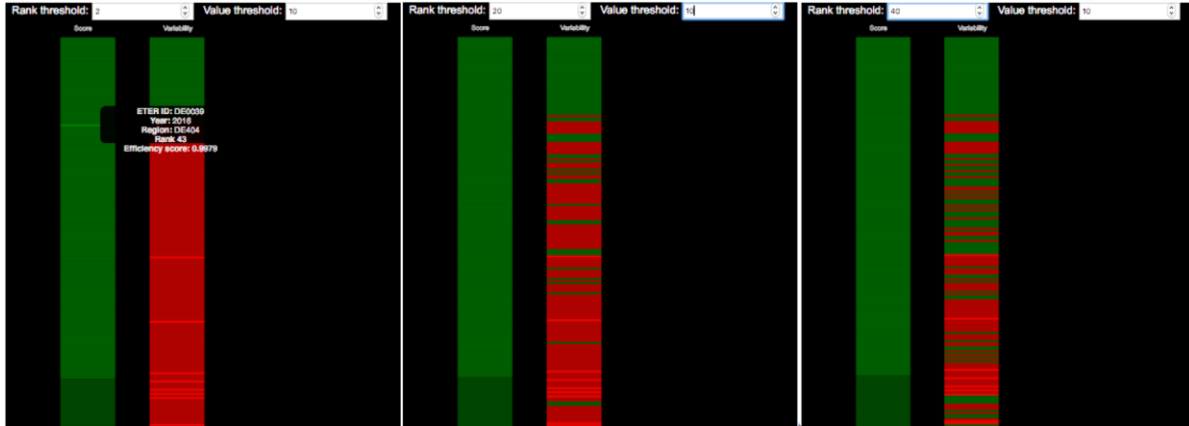
$$|\text{Rank}_{M_1}(U) - \text{Rank}_{M_2}(U)| \leq T_{rank} \text{ and } |M_1(U) - M_2(U)| \leq T_{value}$$
- *Light green*, meaning that while the ranking of the unit is preserved, the associated performance scores differ significantly such that:  

$$|\text{Rank}_{M_1}(U) - \text{Rank}_{M_2}(U)| \leq T_{rank} \text{ and } |M_1(U) - M_2(U)| > T_{value}$$
- *Light red*, meaning that the rank is not preserved anymore, even if the two performance scores do not differ significantly, such that:  

$$|\text{Rank}_{M_1}(U) - \text{Rank}_{M_2}(U)| > T_{rank} \text{ and } |M_1(U) - M_2(U)| \leq T_{value}$$
- *Dark red*, meaning that both rank and score value are not preserved, identifying strong differences between the two models on that unit, such that:  

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

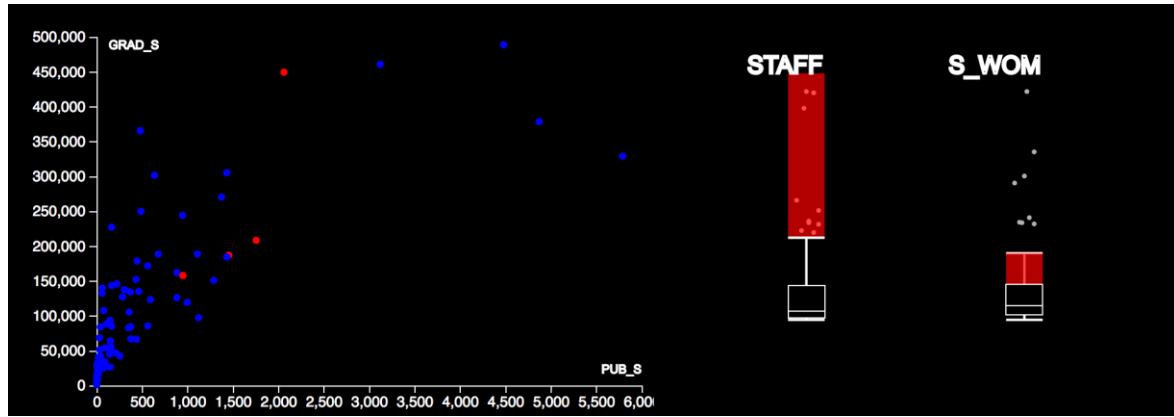
This analysis can be conducted at run-time for all the units, allowing to grasp these differences or similarities for all the units visually. Additionally, the values of the thresholds  $T_{rank}$  and  $T_{value}$  can be dynamically changed during the usage of the system, allowing to inspect how much the similarity between two models is sensible to the thresholds values (see Figure 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 (Paerson, Spearman) and similarity (Kendall-tau) indicators.



**Figure 7. Illustration of a comparison between two models.** The figure illustrates two models, Model1 and Model4, where the first is selected and the second used for comparison. The models are instantiated on 661 units. The illustration reports the first 200 units by efficiency ranking (the view can be scrolled down to visualize more units). It is visible as with the first parameterization of  $T_{rank}=2$  and  $T_{value}=10$ . The two models produce the same rank and values for the first 43 units (graph on the left). Being less strict on the rank and imposing  $T_{rank}=20$  shows additional units presenting similar behavior yet scattered through the ranking (graph in the center). Imposing  $T_{rank}=40$  produces and additional improvement, even if this can be a too strong condition (graph on the right).  $T_{value}$  does not show any significant impact in this example. We can conclude that Model1 and Model4 show good similarity for the first 100 positions of the ranking.

*Sensitivity analysis:* This view expands the sensitivity analysis capabilities, already introduced in the *Model Analysis and comparison* view. The visualization uses two different visual paradigms to relate the different parameters (inputs, conditioning factors, outputs) that constitute the performance model: as an example, in the first one, a scatter plot, the relation between the conditioning factors (if present) and the outputs is reported. Input factors are instead reported as a distribution in the form of a boxplot for each input factor. The interactivity of this chart allows to select disjoint sets of values from each boxplot and inspect the propagated filter on the entire visual environment. It will be 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 model but also support the explorative analysis of the data through filter operations based on factors of the model (see Figure 8). The mapping between features and visual paradigms is free and can be defined in the configuration area.





**Figure 8. Example of data filtering: with respect to all the units, the 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 the scatter plot, and the unit can be identified by mouse-hover.**

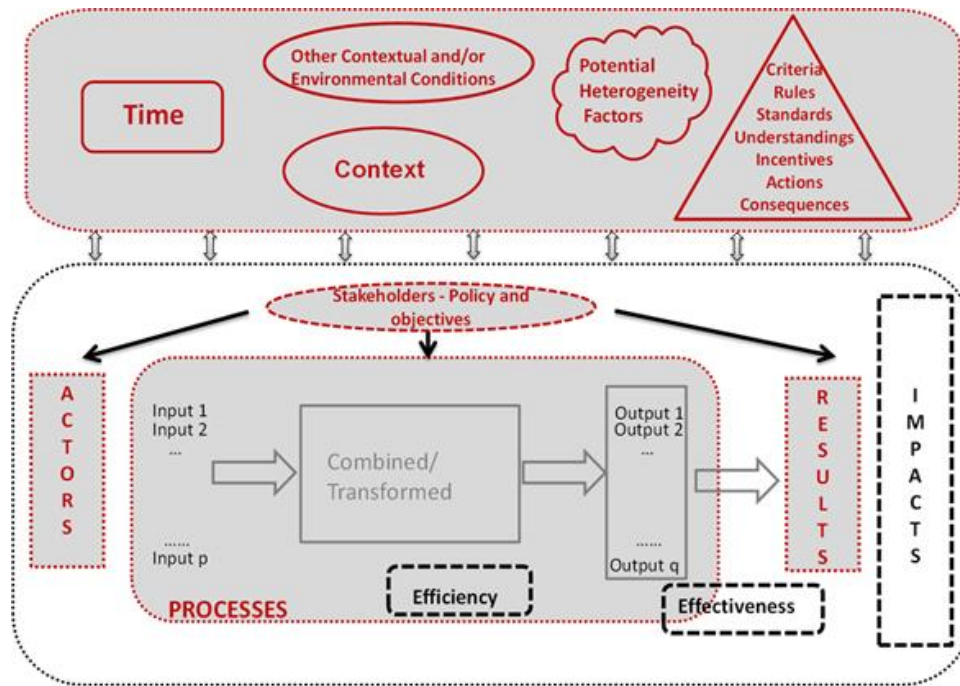
*Configuration area:* this newly introduced areas allows to apply all the described analyses on a general set of features, of which the visual mapping is set in this area. The system does an automatic analysis on all the features for identifying their characteristics (e.g. if they are numerical or categorical features), and then propose for each of the visual coordinates available (e.g. for the scatterplot the x and y coordinates) the subset of the features that are suitable for that visualization. In this way the analyst is helped in her exploration in not having to try wrong or not efficient configuration of the proposed visualizations, allowing her to focus only on the interesting combinations of features/models.

## Models building and their assessment

In the introduction, we described the importance of developing multidimensional models for the assessment of research and its impact. The modelling activity is not an easy task because as discussed in Daraio (2017) 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, test the assumptions, and highlight relevant relations. It is also important for improving, for 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; the complexity, uncertainty and changeability of the environment in which the controlled system works and the limits in the decision context; the intrinsic complexity of calculation.

Figure 9 shows the main components of a performance evaluation model. 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. It considers different measures of performance, ranging from efficiency (as the relationship between the outputs produced with respect to the resources/inputs used) to effectiveness and impacts. Figure 9 illustrates in the top also different “conditioning” dimensions, which relate to contextual/environmental factors that may affect the performance, but are not under the control of the analysed units.

Table 1 details the constitutive elements of the performance model illustrated in Figure 9.



**Figure 9. Performance evaluation model components (source: Daraio, 2017b).**

Performance evaluation model component	Constitutive elements (see Figure 4)	Main question
Purpose of the assessment	Objectives, stakeholder and policy	<i>Why</i>
Level of analysis	Actors (micro level: scholars, organization; meso-level: regional system; macro-level: country)	<i>Who</i>
Object of the evaluation	outputs, efficiency, results, effectiveness, impact	<i>What</i>
Means of the evaluation	(1. qualitative, quantitative, mixed methods; 2 data*)	<i>How</i>
Internal conditional factors	Actors, processes, results	<i>How, When and Where</i>
External conditional factors	Time, context, other contextual factors, potential heterogeneity factors, criteria, rules, standard, understandings, incentives, actions, consequences	<i>How, When and Where</i>

**Table 1 Constitutive elements for the development of a performance model (Source: Daraio, 2019).**

The approach for developing performance models briefly outlined in this section can be considered a base for the model development of analysts that performed the users' evaluation described in the next section.

### Users' evaluation

We tested the proposed approach described in this paper developing a user's evaluation study. The user study was carried out with the participation of about 70 master's degree students, attending the Productivity and Efficiency Analysis course of Sapienza University of Rome. Within the course the students received theoretical lessons on the development of performance models summarized in Figures 8 and Table 1. 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 hours plus 30 minutes 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 A. 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 results of this activity are described in the next section.

## Results

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

Table 2 reports some descriptive analysis on the general questions and on the questions related to the variables and models of performance analysed in the system. The average age of the respondents is 24 and their average total answering time has been of around 15 minutes. The number of variables analysed in the visual analytics tool ranges from 2 to 28, with an average of 11. The number of observations visualized 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 hours practicing with the visual analytics tool, while around 24% spent 8 and 15 hours (see the first column of Table 3). Overall, the users express a very high appreciation for the usefulness of the Visual Analytics tool for executing their tasks. Around 96% of users evaluate as useful the tool (see Table 3, Q. 12 column). The functionality better evaluated from the users has been the Model Analysis and comparison visualization in the tool to develop and finalize their performance models (Question 9, Table 3), appreciated by around 90% of users. This appreciation is confirmed by Question 15 (see Table 3) where around 90% of users declare appreciated most the task of model evaluation included in the Visual Analytics tool.

Questions 6 to 8 asked an evaluation on 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 variable selection functionality and around 24% of users found not useful the correlation analysis implemented in the visual environment (see Table 3).

	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)
<i>Min</i>	22	2.65	0.1	2	2	396	2
<i>Max</i>	34	83.1	3.2	81	28	4020	12
<i>Mean</i>	24.17	15.75	0.61	14.85	10.53	1136.26	3.25
<i>Median</i>	23	8.76	0.335	8.5	9	1000	3

**Table 2. Descriptive statistics on general and modelling questions**

Q. 5	%	Answer	Q. 6 (%)	Q. 7 (%)	Q. 8 (%)	Q. 9 (%)	Q. 12 (%)	Q. 13 (%)	Q. 14 (%)	Q. 15 (%)
Between 1 and 3 hours	26.09	Very useful	17.39	32.61	10.87	15.22	13.04	2.17	10.87	17.39
Between 3 and 8 hours	41.30	Useful	26.09	26.09	39.13	43.48	43.48	17.39	36.96	41.30
Between 8 and 15 hours	23.91	Medium useful	41.30	28.26	34.78	30.43	39.13	39.13	28.26	30.43
Less than 1 hour/more than 14 hours	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 3. Distribution of the answers to Questions 5, 6, 7, 8, 9, 12, 13, 14 and 15**

Answers to Questions 10 and 11 provided several detailed comments about the usefulness of the tool. 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 model exploration, explanation and configuration tasks.

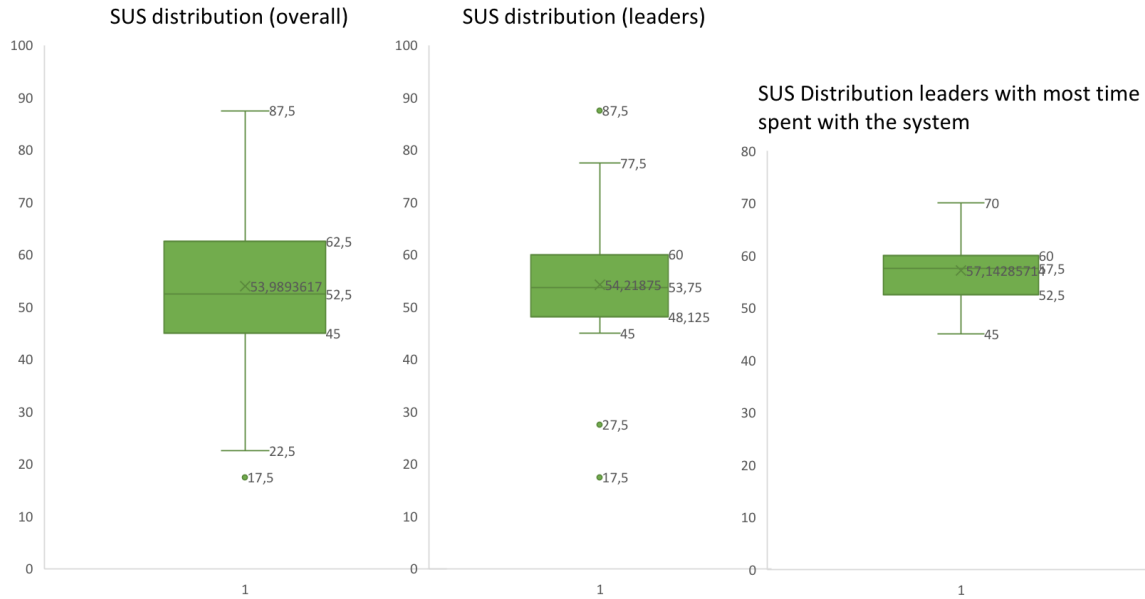
Finally, the answers to Question 16 provided many suggestions for further extensions and improvements of the visual environment that can be summarized in:

- *Instruction manual*, including information about the main components and functionalities of the visual environment;
- possibility to *zoom* on one component of the environment and have it full screen;
- inclusion of *additional plots*;
- including 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.

In the next section, we show the results of the Usability Questions 17 to 27.

### *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 Figure 10, where on the left we have the SUS scores computed for the overall population, in the centre the score is computed only for persons that declared to be heavily involved in using the system (labelled as leaders), and on the right the score is computed for leaders with the maximum interval of system usage (8-15 hours).



**Figure 10: Boxplots of the usability of the system for three categories of users: all users (graph on the left), leaders (graph at the centre) and leaders with most time spent with the system (graph on the right).**

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 hours) 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.

## Conclusions

In this paper we consolidated the research based on *Sapientia* and OBDA combining it with a Visual Analytics approach (Angelini et al. 2019). 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 tool for executing their tasks. The functionality better evaluated from the users has been the Model Analysis and comparison visualization in the tool to develop and finalize their performance models. Further extensions and improvements of the visual environment, suggested by the users include the preparation of an Instruction Manual, improvement of the possibility to zoom on 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.

## Acknowledgments

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 and KIMAR Project), and of Clarivate Analytics through the KOL Project, is gratefully acknowledged.

## References

- Angelini, M., Daraio, C., Lenzerini, M., Leotta, F., Santucci, G. (2019) Performance Model's development: A novel approach encompassing ontology-based data access and visual analytics (2019) 17th International Conference on Scientometrics and Informetrics, ISSI 2019 - Proceedings, 2, pp. 1912-1923.
- 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.
- 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., G. De Giacomo, D. Lembo, M. Lenzerini, and R. Rosati (2007). Tractable reasoning and efficient query answering in description logics: The DL-Lite family. *JAR*, 39(3): 385–429.
- Calvanese D., G. De Giacomo, M. Lenzerini, D. Nardi, and R. Rosati (1998). Description logic framework for information integration. In Proc. of KR, pages 2–13.
- Coelli, T., Rao, D.S.P. & Battese, G.E. (2004), *An Introduction to Efficiency Analysis*, Kluwer Academic Publishers, Second Edition.
- Charnes, A., Cooper, W.W. & E. Rhodes (1978), Measuring the inefficiency of decision making units. *European Journal of Operational Research*, 2, 429-444.
- Cooper, W.W., Seiford L.M., & Tone K. (1999), *Data Envelopment Analysis: A Comprehensive Text with Models, Applications, References and DEA-Solver Software*, Kluwer Academic Publishers, Boston.
- Daraio C. (2017), A framework for the assessment of Research and its Impacts, *Journal of Data and Information Science*, Vol. 2 No. 4, 2017 pp 7–42.
- Daraio C. (2019), Econometric approaches to the measurement of research productivity, in *Springer Handbook of Science and Technology Indicators* edited by Glänzel W., Moed H.F., Schmoch H. and Thelwall M.
- 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*, DOI: 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., Naggar P., Bonaccorsi A., Bartolucci A. (2015). Sapienia: The Ontology of Multi-Dimensional Research Assessment, in Salah, A.A., 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., Naggar P., Bonaccorsi A., Bartolucci A. (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., Naggar 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*, Springer, New York.
- Deprins, D., Simar, L. & H. Tulkens (1984), Measuring labor inefficiency in post offices. In *The Performance of Public Enterprises: Concepts and measurements*. M. Marchand, P. Pestieau and H. Tulkens (eds.), Amsterdam, North-Holland, 243-267.
- Erhan, H., Woodbury, R., & Salmasi, N. H. (2009). Visual sensitivity analysis of parametric design models: improving agility in design.
- Fare, R., Grosskopf, S. and Lovell, C.A.K. (1994), *Production Frontiers*, Cambridge University Press, Cambridge.
- Inselberg, A. (2009). *Parallel Coordinates: Visual Multidimensional Geometry and Its Applications* (Vol. 20). Springer Science & Business Media.
- Lenzerini M. & Daraio C. (2019), Challenges, Approaches and Solutions in Data Integration for Research and Innovation, in *Springer Handbook of Science and Technology Indicators* edited by Glänzel W., Moed H.F., Schmoch H. and Thelwall M.
- Lembo, D., Pantaleone, D., Santarelli, V., & Savo, D. F. (2016). Easy OWL Drawing with the Graphol Visual Ontology Language. In KR (pp. 573-576).
- 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 (United States).
- Keim, D., Andrienko, G., Fekete, J.D., Gorg, C., Kohlhammer, J., Melancon, G. (2008). *Visual Analytics: Definition, Process, and Challenges*. Springer Berlin Heidelberg, Berlin, Heidelberg, 154–175.
- Von Mayr, G. (1877) *Die gesetzmässigkeit im gesellschaftsleben*. Didemburg.



# Appendix A

## 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, Bachelor'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