Create Dashboards and Data Story with the Data & Analytics Frameworks

Chapter · December 2019

DOI: 10.1007/978-3-030-36599-8_24

3 authors:

Michele Petito
Agency for Digital Italy (AgID)

Francesca Fallucchi
Università Telematica Guglielmo Marconi

Ernesto William De Luca
Otto-von-Guericke-Universität Magdeburg

Create dashboards and data story with the Data & Analytics frameworks

Michele Petito[3[0000-0002-5463-5322]], Fallucchi Francesca[1,2[0000-0002-3288-044X]], and De Luca Ernesto William[1,2[0000-0003-3621-4118]]

[1]DIII, Guglielmo Marconi University, Rome, Italy
[2]DIFI, Georg Eckert Institute Braunschweig, Germany
[3] Università di Pisa, Italy
[f.fallucchi,ew.deluca]@unimarconi.it
[fallucchi,deluca]@gei.de
michele.petito@unipi.it

Abstract. In recent years, many data visualization tools have appeared on the market that can potentially guarantee citizens and users of the Public Administration (PA) the ability to create dashboards and data stories with just a few clicks, using open and unopened data from the PA. The Data Analytics Framework (DAF), a project of the Italian government launched at the end of 2017 and currently being tested, has the goal to improve and simplify the interoperability and exchange of data between Public Administrations, thanks to its big data platform and the integrated use of data visualization tools and semantic technologies. The DAF also has the objective of facilitating data analysis, improving the management of Open Data and facilitating the spread of linked open data (LOD) thanks to the integration of OntoPiA, a network of controlled vocabularies and ontologies, such as "IoT Events", an ontology for representing and modelling the knowledge within the domain of the Internet of Things. This paper contributes to the enhancement of the project by introducing a case study created by the author, concerns tourism of Sardinia (a region of Italy). The case study follows a process in the DAF in 5 steps, starting from selection of the dataset to the creation phase of the real dashboard through Apache Superset (a business intelligence tool) and the related data story. This case study is one of the few demonstrations of use on a real case of DAF and highlights the ability of this national platform to transform the analysis of a large amount of data into simple visual representations with clear and effective language.

Keywords: Data & Analytics Framework, Data Visualization, Dashboard, Business intelligence, Open Data.

1 Introduction

Public sector is rich of data. In recent years, the open data dataset in Italy has increased considerably: in 2018 there were about 15,000 and in 2019 the number is increased to over 25,000 [1]. Unfortunately this dataset growth has not been accompanied by an improvement in the quality of open data portals that represent only simple catalogs. The analysis and implementation phase of the dashboards can only take place using third-party tools, thus not favoring the sharing of knowledge and the birth of new ideas. But the biggest problem concerns the fragmentation of data that limits the analysis and the interpretation of national social and economic phenomena [2] [3]. Therefore, in order to make the most of the data potential, it is necessary to abandon the silo approach and adopt a systemic vision that favors access and data sharing.

In this scenario, the big data platform DAF [4] designed by the *Digital Transformation Team* [5], has the challenge to provide a single point of access for government data and support increased public participation, collaboration and cooperation. The DAF is an infrastructure of the Italian Government established in September 2016, represents Italy's latest effort to valorize public information assets. The objective of the DAF is to overcome these difficulties by using big data platforms to store in a unique repository the data of the PAs, implementing ingestion procedures, promoting standardization and interoperability. So, thanks to a framework for distributed applications such as Apache Hadoop [6], the DAF allows the exploitation of enormous public sector data that describes the realities of citizens and businesses to generate insights and information hidden in it [7] [8].

Furthermore, the DAF promotes semantic interoperability, according to the new European Interoperability Framework (EIF) [9]. To enhance interoperability DAF make use of an ecosystem of ontologies and controlled vocabularies (OntoPiA) [10]). Every dataset in DAF is accompanied by metadata that describes the dataset and its internal structure. It will be the user's responsibility to define the ontological information and controlled vocabularies associated with the data structure, through the meaning of the semantic tags. A tagging system will allow to drive the user to the correct use of controlled vocabularies and to ensure that all datasets can be effectively connected together.

In this paper, we will focus in particular on DAF data visualization technologies, providing a comparative analysis with other tools and platforms on the market (see Section 2). In Section 3 we will present the general architecture of the Data & Analytics Framework and OntoPiA, the network of ontologies and controlled vocabularies. Section 4 focuses on the process of building a dashboard through Apache Superset. Finally, a case of use of the DAF will be presented for the construction of a dashboard starting from a dataset on tourism in the Sardinia Region (see Section 5). We conclude the paper with some future developments.

2 Related works

Research supports an increasing focus on visual imagery, in all its forms, as a way of communicating with and engaging with online audiences. *Data visualization*[1] allows PA to obtain useful trends and information with maximum simplicity and speed. The

[1] The science of visual representation of 'data', which has been abstracted in some schematic form, including attributes or variables for the units of information.

DAF project embraces this approach by integrating Superset [11], a business intelligence tool for data representation. There are many other tools in the same category such as Microsoft Power BI [12], Tableau [13], Google Data Studio [14] e Ploty.ly [15] that offer the possibility of use in the cloud via API and that could allow integration with the DAF, even if with important limitations.

Public Tableau [13] allows the creation of complex dashboards with great flexibility, without requiring specific technical skills. But the use of the free version is allowed only through a desktop application.

Google Data Studio [14] is very intuitive but does not allow the use of more than one dataset within the same dashboard. It has few data connectors (only for MySql [16], and for PostgreSQL [17]) and the quality of the dashboards is not comparable to that of the competitors in the sector.

Plot.ly [15] allows the creation and sharing of quick interactive dashboards. But it only accepts datasets with a maximum size of 5 MB and the published graphics must be public (for them to be private it is necessary to pay a subscription). Among the open source categories we have instead distinguished three data visualization tools: Superset [11], Metabase [18] and Redash [19]. All these projects meet the requirements [20] that an OGD[2] visualization tool should possess. Superset and Redash are very similar. Both are powerful and give the possibility to connect to a large number of data sources. In addition, they have a powerful interface for writing and executing SQL queries.

Once saved, queries can also be used as a basis for the creation of dashboards. Superset supports a larger number of authentication systems than Redash. For example, it includes LDAP[3], the system used for the unique authentication of all modules in the DAF. Although Redash is an excellent project in rapid evolution, Superset [11] has been chosen for three main reasons: the presence of LDAP authentication, the high number of views, and the support of the Python language (used throughout the DAF project).

Superset [11] is an open source product hosted on the Apache Foundation Github[4] platform and is developed using Flask [21] a very lean Python framework [22] for web development. The part that generates the interactive graphs instead makes use of NVD3 [23] a javascript library built on D3.js [24]. Any dashboard created with Superset consists of a series of graphs (called slices). Each of these can be resized, moved relative to the others, or shown in full screen. In addition, each dataset represented in a graph can also be exported in CSV or JSON format or through SQL queries. The "slices" are created starting from a table available in the many data sources that Superset is able to manage. Superset provides two main interfaces: the first is the Rich SQL IDE (Interactive Development Environment) called *Sql Lab*[5] with which the user can have immediate and flexible access to data or write specific SQL queries, the second is a data exploration interface that allows the conversion of data tables into rich visual insights.

[2] Open Government Data. http://www.oecd.org/gov/digital-government/open-government-data.htm
[3] Lightweight Directory Access Protocol. http://foldoc.org/ldap
[4] Superset Github repository. https://github.com/apache/incubator-superset
[5] https://superset.incubator.apache.org/sqllab.html

3 Data & Analytics Framework (DAF)

DAF has a complex architecture which integrates different components [25]. **Fig**. **1** shows a simplified view of two relevant related characteristics: the interoperability between the components mediated by the use of microservices and the use of docker container[6] technology that isolates the components for greater security.

The *Dataportal* is the main point of access to the DAF and its functionalities. It is characterized by a public section [26] and a private section [27]. In the public section (accessible via https://dataportal.daf.teamdigitale.it/) anyone can browse the *data stories*[7] and dashboards associated with the data in the national catalog. The private section is accessed only after login, thus allowing only accredited users to exploit the functionality of querying, analyzing and sharing data.

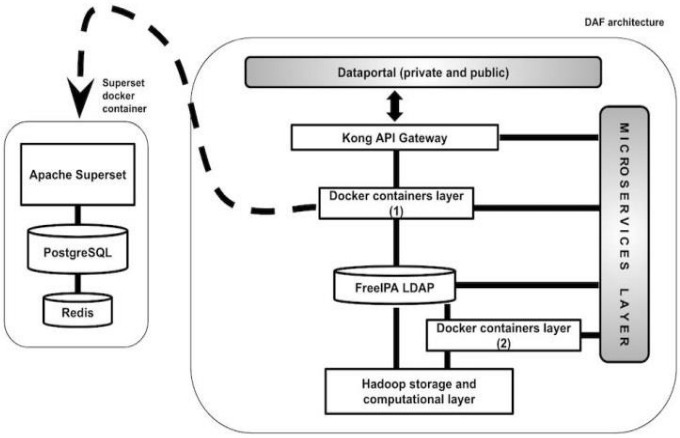

Fig. 1. Logical architecture of the DAF

The Dataportal communicates with the rest of the system through the *Kong API Gateway* [28] and the *MICROSERVICE LAYER*. The docker container layer represented in point 1 of **Fig. 1**, manages the analysis, cataloging and display of data. This layer encapsulates some docker containers, including *Superset* [29], *Metabase* [18], *Jupyter* [30] and *CKAN* [31]. The latter implements a component called a harvester, which allows the DAF to collect all datasets within the DAF. In addition, CKAN performs data catalog functions and allows the download of datasets. *Jupyter* is instead a very useful tool for data scientists, as it allows to perform operations such as data cleaning and transformation, numerical simulations, statistical modeling, machine learning and to run Scala and Python applications on the big data platform of DAF thanks to integration with Apache Spark. On the left side of **Fig. 1**, the *Superset* docker container is shown: as you can see, in this container there are also databases as *PostgreSQL* [17], used to store the tables and *Redis* [32] to manage the cache. Centralized authentication is guaranteed by the *FreeIPA LDAP* [33], an open source solution for integrated identity

[6] https://www.docker.com/resources/what-container
[7] Data stories are an extension of the dashboards, which allow you to express what you can see from the views.

management. The point 2 of **Fig. 1** shows a second layer of docker containers consisting of the platforms *OpenTSDB* [34], *Livy* [35] and *Nifi* [36]. Finally, in the lower part we find the *Hadoop storage and computational layer* [37] that contains the entire storage platform provided by Hadoop (HDFS [38], Kudu [39] and HBase [38]).

The microservices layer, described in the previous study [40], provides the semantic microservices and allow the implementation of semantic technologies, namely the standardization, production, and publication of LOD. These processes can be achieved thanks to the presence of *OntoPiA* [10] a network of remote ontologies and vocabularies, published on Github. *OntoPiA* allows the DAF to provide the catalog of controlled vocabularies. Moreover it favors the function of semantic tagging and the reuse of controlled ontologies and vocabularies by companies and PAs. The network is based on the standards of the Semantic Web and it is aligned with the so-called Core Vocabulary [41] of the European Commission's ISA2 program.

4 Building dashboard process

The process of creating a dashboard in the DAF, inspired by the KDD model [42], is structured in five steps. The first two phases are described in section 4.1 and concern the selection (step 1), the analysis and verification of the quality of the dataset (step 2). In section 4.2 the remaining three steps are presented: configuration of local tables (step 3), creation of slices (step 4) and creation of the dashboard and data story (step 5).

4.1 Selection and dataset analysis (step 1-2)

In the first step you look for the dataset to analyze. This activity can be carried out in two ways: through the *public portal* [26], as anonymous visitor, or from the *private portal* [27]. Using the search form you can search with keywords or browse through the categories that describe the domain. Once the dataset has been identified, we move on to the second step consisting in carrying out a first analysis of the data to evaluate a part of the quality characteristics foreseen by the ISO / IEC 25024 [43] and described in action no. 9 of the *National guidelines for the enhancement of public information*[8] .

4.2 Superset’s tables configuration and visualization (step 3-5)

The last three steps of the DAF dashboard creation process are mainly realized in Superset. Before proceeding to the realization of the slices, it is necessary to modify the settings of the fields (or columns in the Superset jargon) of the table. This means that it will be necessary to verify the correct assignment of the field types (INT, BIGINT, CHAR, DATETIME etc.), establish the dimensions on which to perform the aggregations and define the relative metrics.

[8] https://lg-patrimonio-pubblico.readthedocs.io

Once the step of settings of local tables is finished, it is possible to proceed with the realization of the first slice (step 4). Superset has no less than 34 different types of graphics[9].

The last step is the realization of the dashboard which consists of a personalized composition of the individual slices.

5 Case study

The case study we created in the DAF on February 4, 2018 during the *Open Sardinia Contest*[10], concerns the development of a dashboard and the related data story based on the Sardinia Region's tourism dataset. Specifically, we describe the 5 steps of the process for creating a dashboard illustrated in the previous paragraph. In particular, the phenomenon to be studied and the analysis of the relative dataset will be introduced (see paragraph 5.1), the whole development activity in Superset will be illustrated (see paragraph 5.2), concerning the configuration of the dataset, the realization of the slices and the dashboard. In paragraph 5.3 the case of data story associated with the dashboard is presented.

5.1 Sardinia Tourism dataset

The tourism industry in general occupies an important place in the economy of a country and tourism activities are a potential source of employment, so it is good to know the volume of tourism and its characteristics. This is important for the local government to answer questions such as:

- What is the origin of tourism between June and September?
- What is the tourist period preferred by the Germans in the Olbia area?
- What is the accommodation capacity in the south of the island?
- Which types of accommodation are present?

This allows the local government to decide how and where to spend public money. For example, with regard to the first question, if a high percentage of German tourism is highlighted, the Mayor could establish the installation of tourist road signs in German or provide German language courses for City employees. Regarding the second question, if you find that the German tourism is concentrated mainly in the months of September and June, the local administrator may choose for example to enhance tourist services in the areas and months involved, extending the opening hours of offices or municipality health districts.

In the open data catalog of the Sardinia Region (published on dati.regione.sardegna.it) it is possible to find the datasets related to the *movements of the clients in the hospitality establishments in the years between 2013 and 2016*. The data derive from the communications for statistical and obligatory purposes by law, which the accommodation facilities do to the Sardinia Region. Finally, the Region transmits the data to

[9] https://superset.incubator.apache.org/gallery.html#visualizations-gallery

[10] Contest dedicated to open data, promoted by the Sardinia Region and held in the period from 16/10/2017 to 21/01/2018. http://contest.formez.it/

ISTAT (The Italian National Institute of Statistics), according to the current legal requirements.

The following datasets (in CSV format) are those selected for the case study:

- (1) Tourist movements in Sardinia by municipality;
- (2) Tourist movements in Sardinia by province;
- (3) Tourist movements in Sardinia by macro type of accommodation facility;
- (4) Capacity for accommodation facilities in Sardinia.

The CSV (1-3) collect the *arrivals* (number of hosted customers) and the *visit duration* (number of nights spent) of tourists in Sardinia, divided by the tourist's origin and type of accommodation. The CSV (4) collects the capacities of the receptive structures of the Sardinia Region. The capacity measures the consistency in terms of number of accommodation facilities and related beds and rooms.

Through data portal it is possible to perform a series of operations including downloading the dataset (in JSON format [44] and limited to 1000 records), obtaining the endpoint API, access the analysis tools and display dashboards related. It is also possible to access to CKAN [31], a module integrated in DAF which allows to see all metadata associated to the current dataset.

5.2 Sardinia Tourism dashboard

After selecting the dataset, it is possible to open Superset very intuitively by clicking on a button on the web interface. At this point we have reached the third step of the process of creating the dashboard described in paragraph 4. In this phase the Superset tables must be configured for a specific slice. As an example, let's consider the type of slice *Table view* named “Arrivi e presenze totali per provenienza turista”, one of slices created for this use case.

The datasource's configuration form of *Table view* (see **Fig. 2**) is divided into three tabs (*Detail*, *List Columns* and *List Metrics*) and allows the modification of the table parameters. The initial tab shows some basic information such as the table name and the associated slices. The second tab displays the fields in the table, for example, province, macro-typology, arrivals, etc. The first and third columns respectively contain the name of the field (e.g. province, macro-typology, arrivals, etc.) and the type of data contained (STRING, INT, DATETIME, etc.).

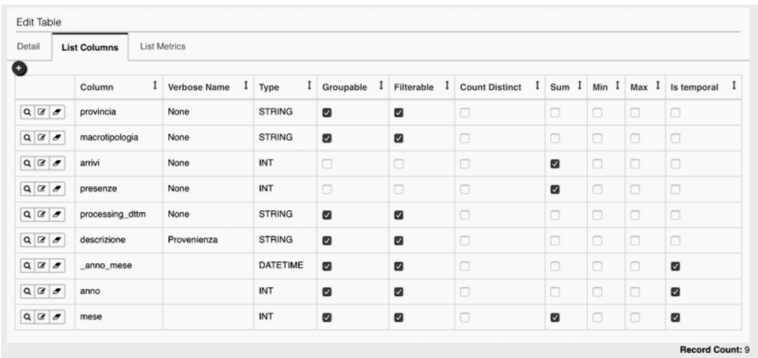

Fig. 2. Edit table form

Superset automatically assigns the correct properties in relation to the type of data declared, but you can decide to change them or increase the details. Special attention should be paid to DATETIME columns. Formatting and orders by day, month and year can change. Superset allows you to customize them thanks to all the combinations that can be created with the management of dates in Python. In the specific case of the table in **Fig. 2**, the format "% Y-% m-% d% H:% M:% S" was used for the calculated field "_anno_mese", ie a timestamp useful for a time based analysis on month / year. The "_anno_mese" field is not a standard field, but is dynamically generated by the function "Expression = concat (cast (year as string), '-', cast (month as string), '-01 00:00:00')"and from the option "Datetime Format =% Y-% m-% d% H:% M:% S": this sets the timestamp format according to the Python datetime string pattern. Here the syntax can vary depending on the database used: the DAF uses Apache Impala [45] the engine for Hadoop SQL queries [37]. If you use SQLite [46] (the default Superset database) the setting to use would be "Expression = year || '-' || month "and" Datetime Format =% Y-% d ". This happens because SQLite accepts a date format of type Year + Month that does not exist in Impala. The only way to get the same result in Impala is to use a timestamp format and concatenate the "year" and "month" fields with the string '-01 00:00:00'.

Now let's consider the "year" field: on this the “Is temporal” option has been set which tells Superset to treat the values to describe variable phenomena over time. On this same field, "Datetime format =% Y" was set, in order to tell Superset to treat the values as years. If in the ingestion phase (as happened in this case) the field should be of an incorrect type (for example INT) it is possible to use the option "Database expression = date_part ('year', '{}')" to perform the conversion from INT to DATETIME. The Impala function "date_part" in addition to casting from INT to DATETIME, extracts the year from the timestamp.

Once the dimensions have been established, ie the fields on which to group, the metrics (SUM, AVG, etc.) to be applied to these groupings must be defined (third and last tab of **Fig**. **2**). For this specific case we created "Origin" field and the "sum_arrivi" and "sum_presenze" metrics were defined and assigned using the SUM(arrivals) and SUM(visits-duration) functions. The result is a “Table view” with the list of arrivals and visits duration divided by geographical area of origin of the tourist.

After completing the table configuration phase, we move on to the fourth step of creating the slices. Superset has no less than 34 different types of graphics. Now suppose we want to create a graph that shows the average number of tourists in Sardinia and at the same time a graph that shows the trend of visit duration over time. This is achievable thanks to a slice view as *Big number with trendline*. Using the "_anno_mese" field on the abscissa axis and on the ordinates the "_presenzamedia" metric will get the result as shown in the last graph on the right of the **Fig. 3** (point 1).

Many other options can be applied on the slice: for example, in the *Filters section* you can add one or more values to exclude in the results. For more complex filters, you can use *Where clause* option, which allows you to write conditions directly in SQL. There is also the possibility to create complex tables (such as pivot tables) or create dynamic selectable filters that can be used directly by the user on the dashboard.

Once we have finished the first slice, we will already be able to proceed with the creation of the dashboard (fifth step of the process), which initially will naturally contain only one slice. The layout of the Superset dashboards is very flexible: slices can be easily resized by drag and drop. Between one slice and another you can insert a box containing text or html to better describe the graph. For the realization the *tourist movements dashboard in Sardinia from 2013 to 2016* (**Fig. 3**) were used 13 slices. For reasons of space, in **Fig. 3** only part of the realized dashboard[11] is shown.

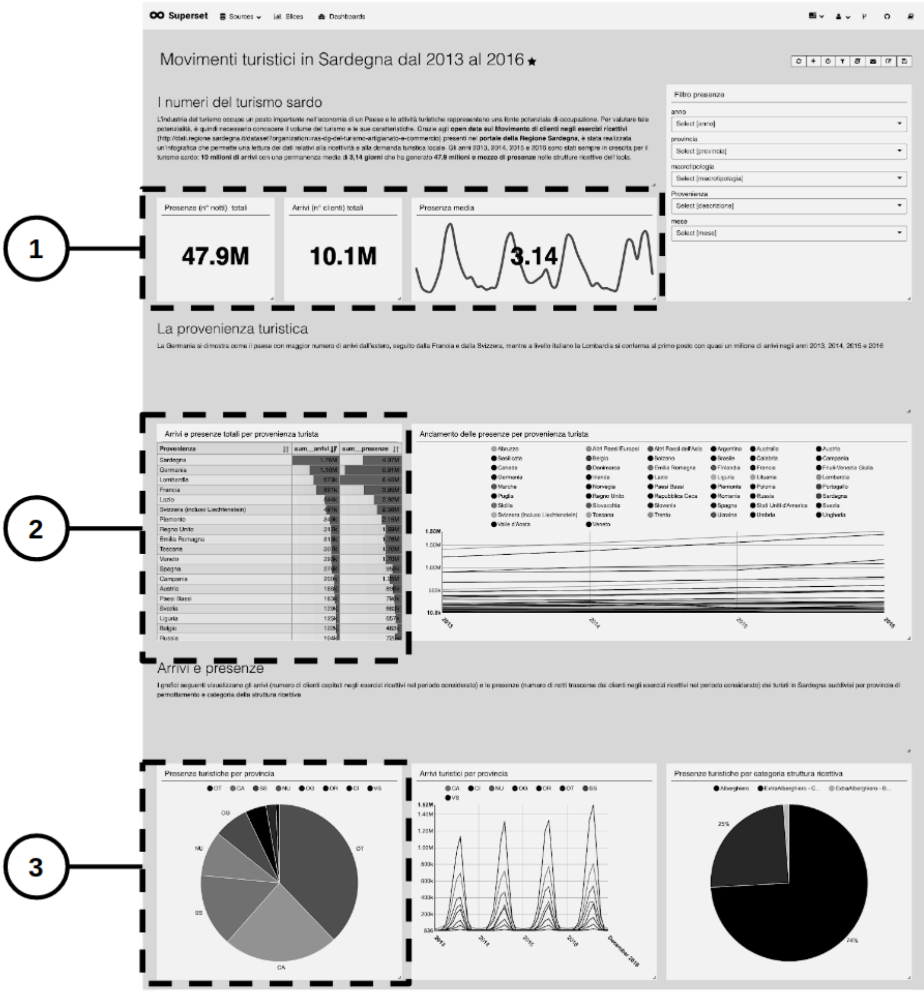

Fig. 3. A section of the dashboard of tourist movements in Sardinia

[11] The complete dashboard is accessible online at http://bit.ly/storia-turismo-sardo-story

5.3 Sardinia tourism data story

With the data story we try to provide an interpretation of a phenomenon from the data. This narration was published in the *Public Data Portal* and can be accessed from the *Menu > Community > Story*.

Thanks to the filters the user can analyze the data by year, province, macro-typology, origin and month. For example, analyzing the datasets as a whole (i.e. without setting filters) it is possible to observe how Sardinian tourism has grown steadily both in terms of visit duration and arrivals during the four years of survey (2013-2016): 10 million arrivals in total with an average stay of 3.14 days which generated 47.9 million of nights in the accommodation facilities of the island (see **Fig. 3** point 1).

From the point of view of the foreign tourist, the most important numbers are those of the Germans (5.91 million) and the French (3.95 million), while the major visit duration from Italy, as shown in the **Fig. 3** (point 2), they are those of the residents of Lombardy (6.45 million) and Lazio (2.92 million). These data, filtered for 2016, are compatible with an ANSA article [47] that stated that 2016 "was a record year for Sardinian tourism: 2.9 million arrivals with an average stay of 4.6 days which generated 13.5 million of nights in the accommodation facilities of the Island ".

As shown in **Fig. 3** (point 3) tourism is mainly distributed in the provinces of Olbia-Tempio, Cagliari and Sassari with preference for hotel facilities in 74% of cases.

Conclusion

In this document we presented the shortcomings of the some data visualizing tools on the market with respect to the potential of open source tools integrated into the DAF, a project requested by the Italian Government to overcome the difficulties of channeling data stored in local public administrations into a single container (data lake). We have therefore introduced the DAF architecture and the semantic functionality of OntoPiA. In particular, we introduced Superset, a data visualization tool which has a central role in the creation of dashboards. Finally, a dashboard use case was presented using some datasets from the Sardinia Region present in the DAF. The use case represents not only one of the first experimental uses of the DAF but also the demonstration of how, following a process in only 5 steps, it is possible to extract information from large data collections with a simple tool available not only to the PA, but also to businesses and ordinary citizens. Currently the DAF is still being tested and there are critical issues, such as the reluctance of some PAs who still do not want to provide their data and accept the arrival of the new platform. But in general the number of public and private subjects who are using it is growing and the production release is now imminent as it is scheduled for December 2019.

References

1. AgID Advance digital transformation. https://avanzamentodigitale.italia.it/it/progetto/open-data. Accessed 10 Aug 2019

2. Temiz S, Brown T (2017) Open data project for e-government: case study of Stockholm open data project. Int J Electron Gov 9:55. https://doi.org/10.1504/IJEG.2017.10005479
3. Drakopoulou S (2018) Open data today and tomorrow: the present challenges and possibilities of open data. Int J Electron Gov 10:157. https://doi.org/10.1504/IJEG.2018.10015213
4. Team Transformation D GitHub - italia/daf: Data & Analytics Framework (DAF). https://github.com/italia/daf. Accessed 10 Sep 2018
5. (2019) Digital Transformation Team. https://teamdigitale.governo.it. Accessed 1 Jun 2018
6. HDFS Architecture Guide. https://hadoop.apache.org/docs/r1.2.1/hdfs_design.html. Accessed 12 Sep 2018
7. Desouza KC, Jacob B (2017) Big Data in the Public Sector: Lessons for Practitioners and Scholars. Adm Soc. https://doi.org/10.1177/0095399714555751
8. Gomes E, Foschini L, Dias J, et al (2018) An infrastructure model for smart cities based on big data. Int J Grid Util Comput 9:322. https://doi.org/10.1504/IJGUC.2018.10016122
9. European Interoperability Framework (EIF). https://ec.europa.eu/isa2/eif_en. Accessed 10 Dec 2018
10. Agid & Team Digitale (2018) OntoPia. https://github.com/italia/daf-ontologie-vocabolari-controllati. Accessed 10 Nov 2018
11. Apache Superset. https://superset.incubator.apache.org. Accessed 19 Aug 2018
12. Sabotta C (2015) Introducing Microsoft BI Reporting and Analysis Tools. In: Microsoft. http://msdn.microsoft.com/en-us/library/d0e16108-7123-4788-87b3-05db962dbc94
13. Tableau (2011) Free Data Visualization Software. Tableau Public
14. Google Data Studio. https://datastudio.google.com/. Accessed 15 Sep 2018
15. Plotly (2017) Modern Visualization for the Data Era - Plotly. https://plot.ly/
16. MySql. https://www.mysql.com/i. Accessed 15 Sep 2018
17. PostgreSQL (2014) PostgreSQL: The world’s most advanced open source database. In: http://www.postgresql.org/. http://www.postgresql.org/
18. Metabase. https://www.metabase.com/. Accessed 16 Sep 2018
19. Redash. https://github.com/getredash/redash. Accessed 15 Sep 2018
20. Graves A, Hendler J (2013) Visualization tools for open government data. In: Proceedings of the 14th Annual International Conference on Digital Government Research - dg.o ’13. p 136
21. Ronacher A Flask MicroFramework. http://flask.pocoo.org/. Accessed 10 Sep 2018
22. Python. https://www.python.org/. Accessed 15 Sep 2018
23. NVD3 Project. http://nvd3.org/. Accessed 15 Sep 2018
24. Dierendonck R van, Tienhoven S van, Elid T (2015) D3.JS: Data-Driven Documents
25. Team Transformation D Data & Analytics Framework (DAF) - Developer Documentation. https://daf-docs.readthedocs.io/. Accessed 15 Sep 2018
26. DAF - Public Dataportal. https://dataportal.daf.teamdigitale.it. Accessed 5 Jan 2019
27. DAF Private Dataportal. https://dataportal.daf.teamdigitale.it/#/login. Accessed 5 Jan 2019
28. Kong. https://konghq.com/. Accessed 15 Sep 2018
29. Apache Software Foundation, Apache Superset Contributors (2018) Apache Superset

— Apache Superset documentation. https://superset.incubator.apache.org/
30. Jupyter Project (2016) Project Jupyter | Project. http://jupyter.org/about.html
31. (2018) CKAN. https://github.com/ckan/ckan. Accessed 12 Sep 2018
32. Redis. https://redis.io. Accessed 15 Sep 2018
33. FreeIPA LDAP. https://www.freeipa.org/. Accessed 19 Sep 2018
34. StumbleUpon (2012) OpenTSDB - A Distributed, Scalable Monitoring System
35. Apache Livy. https://livy.incubator.apache.org/. Accessed 19 Sep 2018
36. The Apache Software Foundation (2017) Apache NiFi. In: Website
37. Apache hadoop. https://hadoop.apache.org/. Accessed 3 Feb 2018
38. HBase A, Cafarella M, Cutting D, HBase A (2015) Apache HBase. In: Http://Hbase.Apache.Org/. https://hbase.apache.org/. Accessed 12 Sep 2018
39. Apache Kudu. https://kudu.apache.org/. Accessed 11 Sep 2018
40. Fallucchi F, Petito M, De Luca EW (2019) Analysing and Visualising Open Data Within the Data and Analytics Framework. pp 135–146
41. (SEMIC) SIC Core Vocabularies. https://joinup.ec.europa.eu/collection/semantic-interoperability-community-semic/core-vocabularies. Accessed 1 Dec 2018
42. Fayyad U, Piatetsky-Shapiro G, Smyth P (1996) The KDD process for extracting useful knowledge from volumes of data. Commun ACM. https://doi.org/10.1145/240455.240464
43. ISO/IEC ISO/IEC 25024:2015 Systems and software engineering — Systems and software Quality Requirements and Evaluation (SQuaRE) — Measurement of data quality. In: Int. Stand. ISO/IEC 250242015. https://www.iso.org/obp/ui/#iso:std:iso-iec:25024:ed-1:v1:en. Accessed 14 Sep 2018
44. Sriparasa SS (2013) JavaScript and JSON Essentials
45. Apache Impala. https://impala.apache.org/. Accessed 15 Sep 2018
46. SQLite. https://www.sqlite.org/. Accessed 15 Sep 2018
47. Ansa (2018) Turismo: 2,9mln arrivi in Sardegna 2016. http://bit.ly/AnsaTurismo2016. Accessed 15 Dec 2018