

Business Intelligence

Project n°8 - Data Analyst

Tran Billy

07–2019

Contents

| | | |
|-----------|---|-----------|
| 1. | <i>Context of the project</i> | 2 |
| 1.1 | Ministry of Europe and Foreign Affairs | 2 |
| 1.2 | The need | 2 |
| 1.3 | Presentation of the team and tools available | 3 |
| 1.4 | My mission | 3 |
| 2. | <i>Development of DANVISA application</i> | 3 |
| 2.1 | The DANVISA Project architecture | 4 |
| 2.2 | DEV : Development phase | 5 |
| 2.2.1 | Configurations | 6 |
| 2.2.2 | Studio Company | 6 |
| 2.2.3 | Dashboard editor | 11 |
| 2.3 | UAT Testing | 12 |
| 2.4 | PROD: Production phase | 12 |
| 3. | <i>Documentation</i> | 13 |
| 3.1 | Documentation - User | 13 |
| 3.2 | Documentation - Technical | 13 |
| 4. | <i>Conclusion</i> | 13 |

The data in this report has been anonymized due to confidentiality.

1. Context of the project

1.1 Ministry of Europe and Foreign Affairs

The Ministry of Europe and Foreign Affairs (MEAE) is the French administration responsible for implementing France's foreign policy and ensuring relations with foreign states. It includes more than 15,000 agents and collaborates with all countries in several fields: diplomacy, culture, security, etc.

Its main missions:

- Represent France, defend and promote its interests in all areas
- Act for security and human rights
- Take action to organize globalization for sustainable development
- Administer and protect the French abroad

The Information Systems Department (ISD) assists the MEAE in the digital modernization. The ISD has five main missions. One of these missions is the Mission Strategy and Architecture Information Systems (MSA) – the department where I work.

MSA participates in the definition of the architecture and the coherence of the information systems, as well as in the establishment of the programming of the direction in liaison with the contracting authorities of the ministry. It also participates in the definition of the technical standards to be followed by the ministry and the technological control of the information systems.

1.2 The need

As part of the redesign of the statistical tool, the Directorate of Information Systems (DSI) has chosen to modernize the statistical tool used to generate its statistical reports for the Sub-Directorate of Visas (SDV). These statistics are intended for public use as well as for internal use to help general management of the activity.

The tool used was BusinessObjects (BO) as well as Excel tables (multiple Dynamic Cross Tables). For nearly ten years, no investment has been made in the update of the BO software: the latter then became more and more obsolete (loss of rights, desire to modernize ...) no upgrades version computer hardware. So, the ISD decided to change and move to Digdash¹.

¹ Digdash is a Business Intelligence software that allows to visualize and interact with its data by curves, sectors, maps, bars, lists and with filters, sorts, selections. The software connects to all types of data sources and is able to process very large volumes of all formats (SQL, OLAP cubes, multidimensional databases, Excel files ...)

1.3 Presentation of the team and tools available

During this mission, I was surrounded by the entire team of the SDV who deals with statistics and who could answer my business questions, and by the head of the implementation of the tool at the DSI, Sandrine Lorenzi who is also my tutor. In addition, on two occasions, Dirgham Noui, a Digdash expert came on site to improve the performance and maintenance of the application that was developed. Regarding the tools used, I had access to the development environment, UAT testing and production of Digdash, SQL Server Management Studio, as well as various office tools (Microsoft Office pack).

1.4 My mission

The purpose of my mission at MEAE was to recreate on Digdash the reports generated by the Ministry from BO. These reports take the form of tables consisting of figures and percentages showing the number of visas requested, issued and refused according to various parameters (year, country, type of visa, reason for stay, etc.) as well as various rates of change compared to previous years.

Users working with Excel tables from the beginning, it was also a mission to change old ways of working - users switching from one software to another. It was necessary to adapt to the existing, while proposing new alternatives made possible by Digdash without the change being too drastic. In parallel, I created a simple and clear dashboard allowing a new governance and visualization of the data at their disposal.

2. Development of DANVISA application

To carry out my mission, on the advice of my tutor, I first established a methodology for the design of the reports. For that I first adapted the world of visas with the team of the SDV. Then, I trained myself on the Digdash software: follow-up of training, reading of documentation and discovery of the existing one.

I was able to attack the technical part. It was necessary to understand the computer architecture of the visas before creating the Excel tables and the dashboard. Given a large number of reports to generate, it was necessary to prioritize these reports. We agreed that this mission in the Ministry would last nine months, spread over my work schedule.

2.1 The DANVISA Project architecture

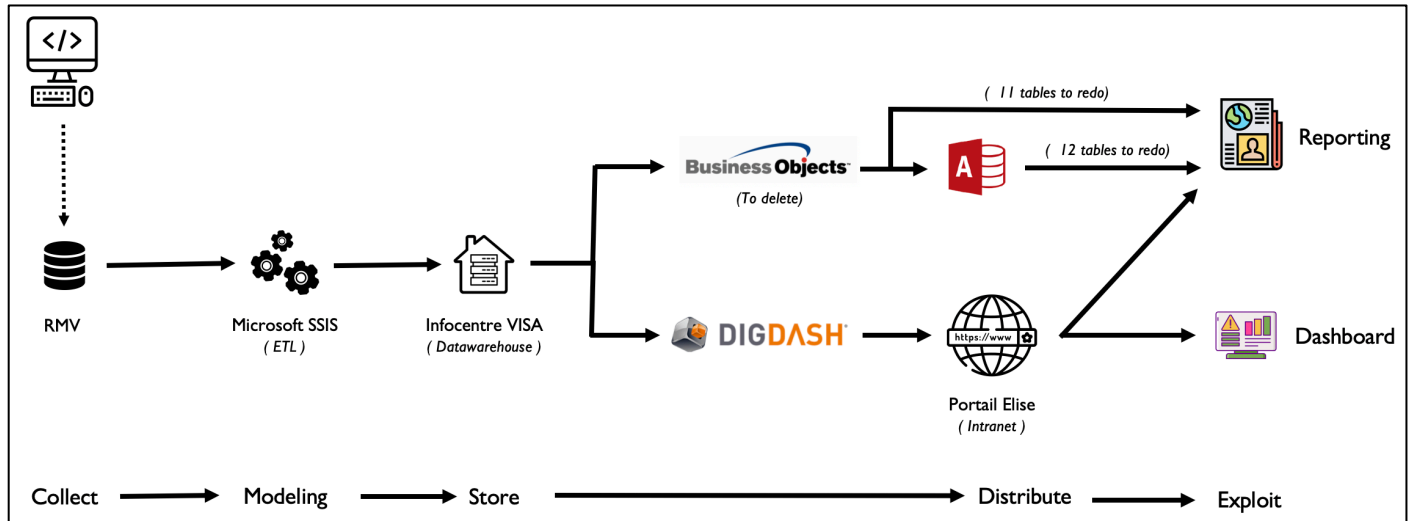


Figure 1 – The DANVISA Project architecture

Collection

There are many sources of information in the information system: most of them consulates, then service providers, border posts, customs, prefectures. The data is processed and aggregated in the RMV (Global Visa Network, gradually replaced by France Visas).

Storage and modeling

Microsoft SSIS, an Extract Transform Load (ETL) software, is used to extract raw data from the RMV, then restructure it, and finally load it into the VISA Infocentre, a data warehouse, also called a data warehouse².

Distribution

In order to obtain a synthetic view of the object of the advanced analysis, the Business Intelligence Digdash application uses the VISA Infocentre as a source of information. The representation of the Department's activity situation - Digdash graphics and tables - for visas is in the form of a decision-making dashboard available on the Élise portal, the departmental intranet.

² The mission of the data warehouse is to filter, cross and reclassify information to make relevant data available to analysis generators. In other words, the data warehouse is a clone of existing data at the operational level of the ministry: the data is modeled and prepared for analysis or for performance issues.

Exploitation

This information will be used to verify the alignment of the department's strategy and assist the organization's management. Thus, the Business Intelligence makes it possible to determine values for the performance indicators met in the organization.

Mission

The Business Objects application is powered by the Infocentre every morning at 5am. Every 05th of the month, the SDV uses the data of the Infocentre, via BO, on Access. From Access, the SDV generates twelve pivot tables. It is possible to integrate the data from the Infocentre, via Digdash, on Access and let the SDV generate their Pivot Tables as they usually do. However, the General Management of Trades also wishes to no longer use Access to focus its activity on a single software: Digdash. Also, every month, the ministry sends ten reports - also from BO - to the various posts abroad. Finally, the SDV uses BO to generate a table of ratios giving information of each embassy and consulate present in the different countries.

So, to get rid of BO, you have to recreate on Digdash ...

- The twelve tables at the exit of Access
- The table of ratios
- The ten monthly reports.

... while providing a dashboard for new data governance.

2.2 DEV : Development phase

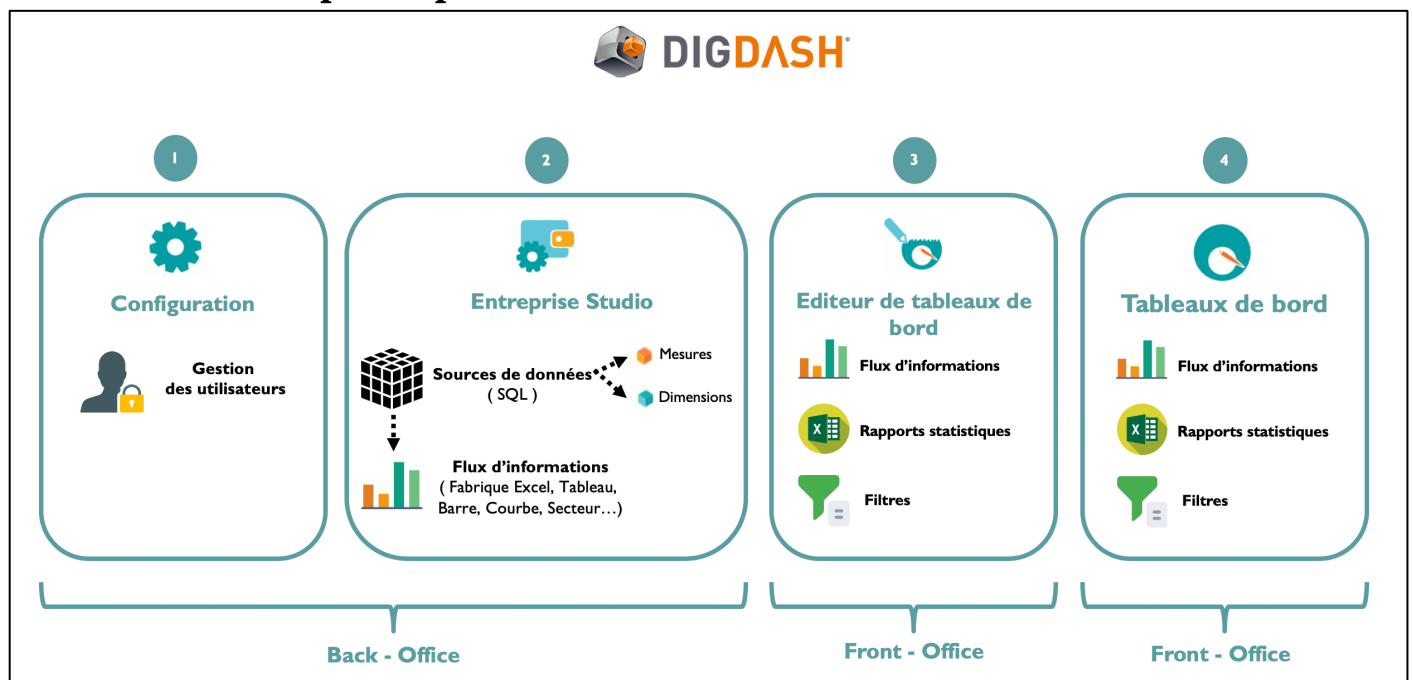


Figure 2 – The stages of the development phase

The development phase is done in three steps: the configuration of rights, the creation of data sources and information flows (tables and graphs) via the Studio Enterprise, the creation of the dashboard).

2.2.1 Configurations

At first, I had to create a user account on Digdash with rights to access the Studio Company, the Dashboard Editor and dashboard DANVISA project.

2.2.2 Studio Company

2.2.2.1 Creating the data source

2.2.2.1.1 SQL query

A data source must be available on which the information flows will be based. This data source must contain all the information present in all the final tables. An upstream study of these allows to have a clear vision of all the data to be put together and used. Digdash makes it possible to connect to different data sources of different formats (database, Excel files, csv ...). For this mission, I connected directly to our datawarehouse, the VISA Infocentre. After testing my queries on SQL Server, I performed them on Digdash. First of all, in order to reduce the loading time of the data during the construction of the application, only the data from 2016 to 2019 have been entered.

```
select
    'Demandés' as 'Modèle',
    dem.ds_date_depot_doss as 'Date',
    serv.sve_pn_dsg_pays as 'Pays',
    serv.sve_dsg_service as 'Poste',
    typ_visa.tv_abv_typ_visa as 'Type de Visa',
    typ_visa.to_dsg_typ_regl_doss as 'Désignation Type Règlement de dossier',
    typ_visa.nd_abv_typ_dem_visa as 'Type de demande',
    typ_visa.ms_dsg_motif_sejour as 'Désignation Motif de Séjour',
    circu.sc_dsg_visa_circul as 'Désignation Visa Circulation',
    sex.pr_abv_sexe as 'Sexe',
    age.ta_l_libelle as 'Age',
    terr.tr_dsg_typ Terr as 'Désignation Type territoire',
    restr.rt_dsg_nature_restrict as 'Restriction territoriale',
    natio.pn_dsg_nationalite as 'Nationalité',
    dem.sc_visa_circul as 'Code Visa Circulation',
    dem.rt_restr_terr_dem as 'Code Nature restriction',
    nat.pn_abv_pays as 'Région autorisée',
    nat.pn_dsg_typ_pays as 'Désignation type pays',
    dem.au_id_en_represent as 'Id en représentation',
    cat.cvt_dsg_cat_typ_regl_doss as 'Type de long séjour',
    nv_nb_visadem as 'Nb visas'

from nbvisa_dem dem
    left join msv_service serv on dem.lc_id_depot=serv.sve_lc_id
    left join msv_typ_regl_doss typ_visa on typ_visa.to_id=dem.to_id
    left join msv_territoire terr on terr.te_id=dem.te_id
    left join msv_circulation circu on circu.sc_visa_circul=dem.sc_visa_circul
    left join msv_tranche_age age ON dem.pr_age=age.ta_id
    left join msv_sexe sex ON dem.pr_sexe=sex.pr_sexe
    left join msv_restriction_territoire restr ON dem.rt_restr_terr_dem =restr.rt_nature_restrict
    left join msv_pays_natio natio ON dem.pn_id_natio_actu=natio.pn_id
    left join msv_pays_natio nat on nat.pn_id=dem.au_id_en_represent
    left join msv_categorie_visa_to cat on cat.to_id = dem.to_id

WHERE YEAR(dem.ds_date_depot_doss) >= '2016'
```

Figure 3 – Part of the SQL query

Once the data is retrieved, the called data model, also an OLAP multidimensional cube, must be configured. Digdash assigns dimension or measure properties to columns in the table from the data source.

2.2.2.1.2 Dimension Management

The dimensions can take different types:

- Dimension (Time) that is associated with the variable: Date
- Dimension (Geographic) that is associated with the variables: Country and Post
- Dimension that is associated with other variables

Once the dimensions types associated, the following configurations were made:

- the format: for example for the Date: YY / MM / DD
- sorting: for example for age groups: 0 - 15 years, 15 - 20 years, 20 - 25 years ...

Finally, to preserve the original structure of the tables, I had to create levels of hierarchies in order to group fields in an over-category that did not exist before. For example, on the "Type of Circulation Visa" dimension, I added a manual grouping: 1-year, 2-year, 3-year, 4-year and 5-year traffic visas in a "Visas Circulation" group and visas out of circulation in a second group "Visas Out of Circulation". This grouping makes it possible to navigate in the figures according to the level of detail desired by the user. This manipulation allowed me to save time in the realization of my charts and graphs since at their creation I could directly select the dimensions with the levels and sub-levels.

For the "Reason for Stay" and "Regulatory File Type" dimensions, the complexity was often to determine which field belonged to which level of hierarchy. Indeed, they exist more than fifty grounds of stay and more than 150 types regulation of file. In addition, some fields had changed their name over the years (eg creation of talent passport visas in 2016), new categories had appeared and others had been removed. It was therefore sometimes difficult to navigate among the different categories.

2.2.2.1.3 Management of measures

At the exit of the SQL query, only one measure is present: the number of visas "Nb visas". I often had to add "calculated measures" in order to obtain additional information. These calculated measures are based on existing dimensions and measurements.

For example, to obtain the number of visas issued in the last year, I created a calculated measure based on:

- The measure "Nb visas"
- The "Model" dimension filtered on "Delivered"
- The "Date" dimension filtered over the previous year

Similarly, the implementation of some measures was very complex because they were not simple measures of sum or average. For example, in a table showing the number of visas requested per country (sorted in descending order), to get the rank of the country in this ranking, I had to create a new calculated measure in Javascript.

In addition, Digdash makes it possible to create objectives on measurements. I mainly used these goals to associate a color with a range of values of my measure. For example, for evolution rates, when the

evolution rate was positive the percentage was displayed in green, in red when it was negative and in black when it was stable.

Once these dimensions, hierarchies, and calculated measures are parameterized, the data source is ready to be used to generate information flows.

2.2.2.2 Information flow

2.2.2.2.1 Copyboards

The first part of my mission focused on the reproduction of Excel tables generated by the Sub-Directorate of Visas, so that they are generated automatically on Digdash.

To create my reports, I mainly used the chart types "table" and "crosstab" in order to keep the same organization and structure of data. As a statistical report contains several tables, it is necessary - for each statistical report - to group the tables used in a single flow of information called Excel factory. A few times, I used Excel macros to have the desired structure impossible to obtain only with Digdash.

2.2.2.2.2 Graphics

In parallel with the statistical reports that I recreated, I also generated more meaningful and intuitive graphics that visually represented the information in the tables: bars, maps, curves, gauges, sectors ... In other words, the goal was to put in shape the data in a much more ergonomic format than a table of hundreds of lines to facilitate analysis and save time processing data.

2.2.2.2.3 Visualization parameters

It was during this step that I took care of setting up the layout of the data (colors, styles, font, display, ...) in order to create aesthetic and visually appealing graphics for the user. I have established a graphic charter so that everything is harmonized.

I have also configured my graphics for the creation of dashboards. I used Digdash's "Interaction" feature. This function makes it possible to add an action on the graph: the user can navigate in the hierarchies and "zoom" on the level of details. For example:

- on the evolution chart showing the number of visas issued per year, it is possible to have the details of the number of visas issued per month then per day by clicking the associated dates.

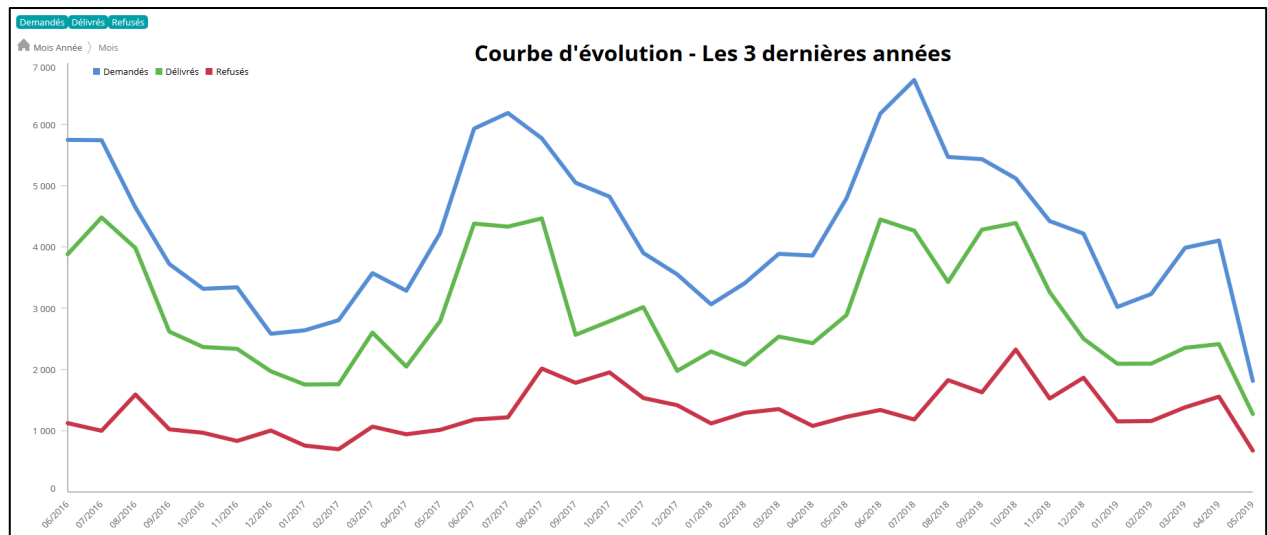


Figure 4 – Graph showing the number of visas issued in a certain country

- a map was designed with interaction on the geographical dimension (zoom from the continent to the countries and then to the posts) with a color gradient depending on the number of visas requested / issued / refused.

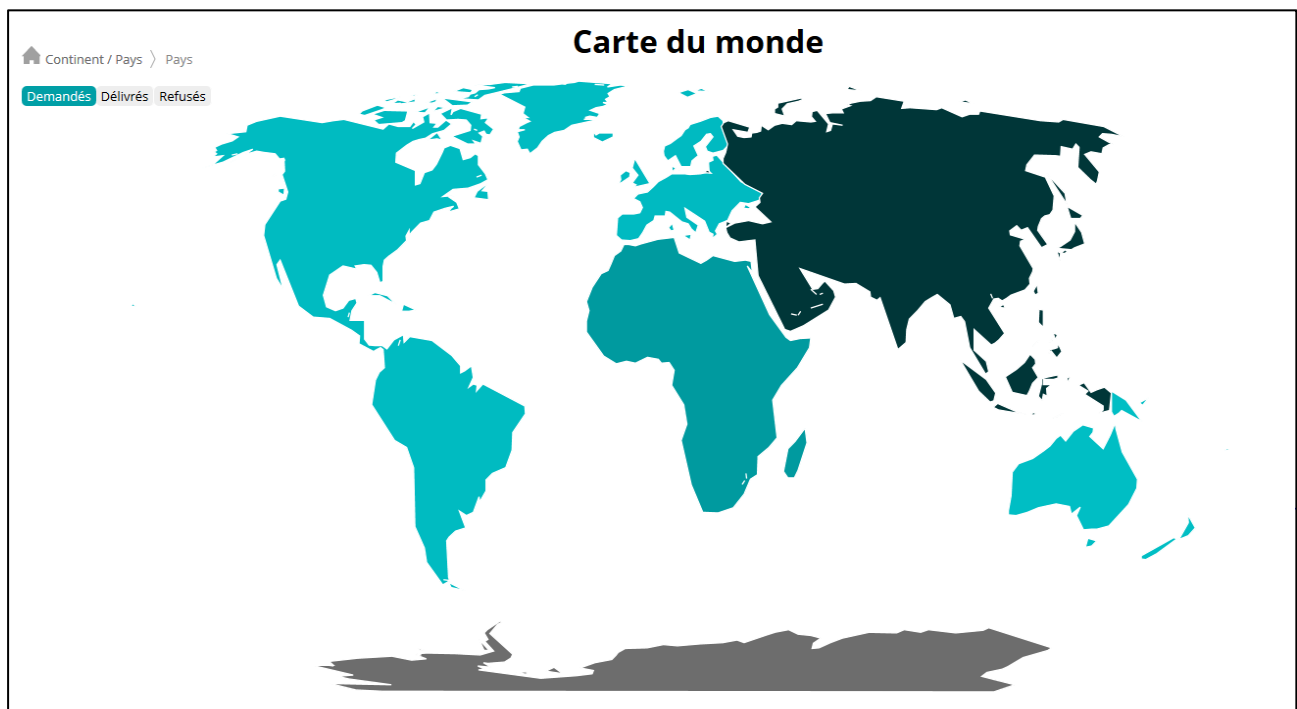


Figure 5 – World Map of Visa Activity

The visualization of the data was very important, it went from the choice of the graph (the most relevant) to the implementation of the user interaction. I thought a lot about the most logical system but also the most complete and the richest for the transmission of information. It required putting myself in the shoes of an inexperienced user to make the graphics easily understandable and accessible to all. For example:

- bar graphs to rank the countries with the highest number of visa applications

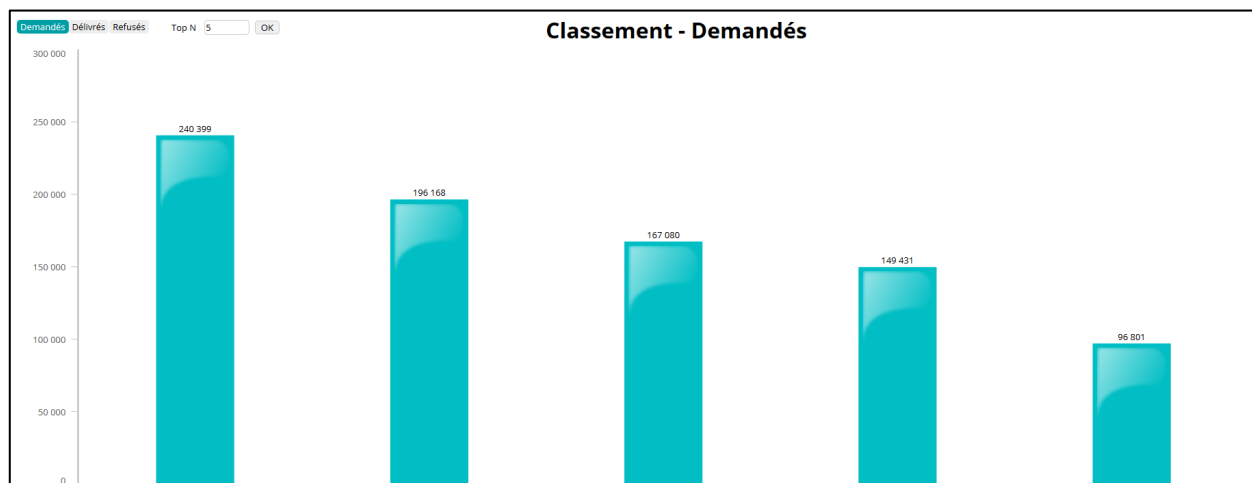


Figure 6 – Bar graphs: Ranking of countries by number of visas requested

- Pie charts highlight the proportion of visas requested by posts in a country

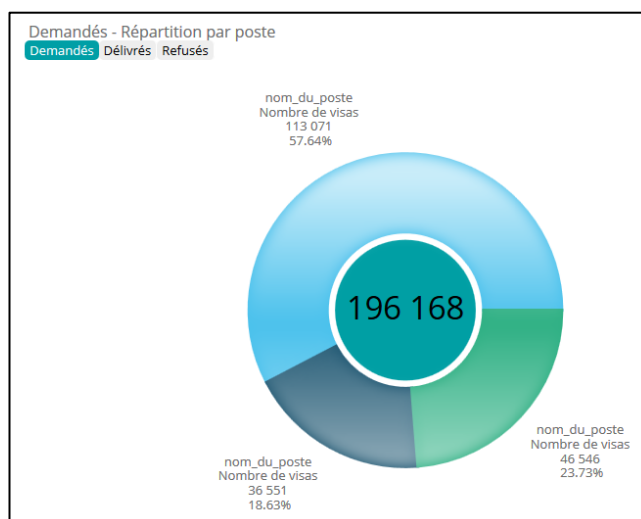


Figure 7 – Sector graphs: Proportion of visas requested by post in a country

2.2.2.2.4 Audit of the data obtained

It was during the configuration of my graphics that I checked that Digdash digits corresponded well to those present in the original tables. This step was very important because if my data were wrong, the tables would lose all its interest.

The construction of my information flows being done, I was able to move to the last stage of the development phase: the edition of the dashboard.

2.2.3 Dashboard editor

The first tab of the dashboard is the home page. It is composed of:

- Icons redirecting to the favorite tabs so that they can be accessed with one click. Then, so that trades have a quick view of activities, I posted the key figures: number of visas requested, issued and refused, and the rate of refusal from January 1 of the current year to the current day. Finally, I added links to documents "user guides" that I wrote to guide trades in the handling of the dashboard.

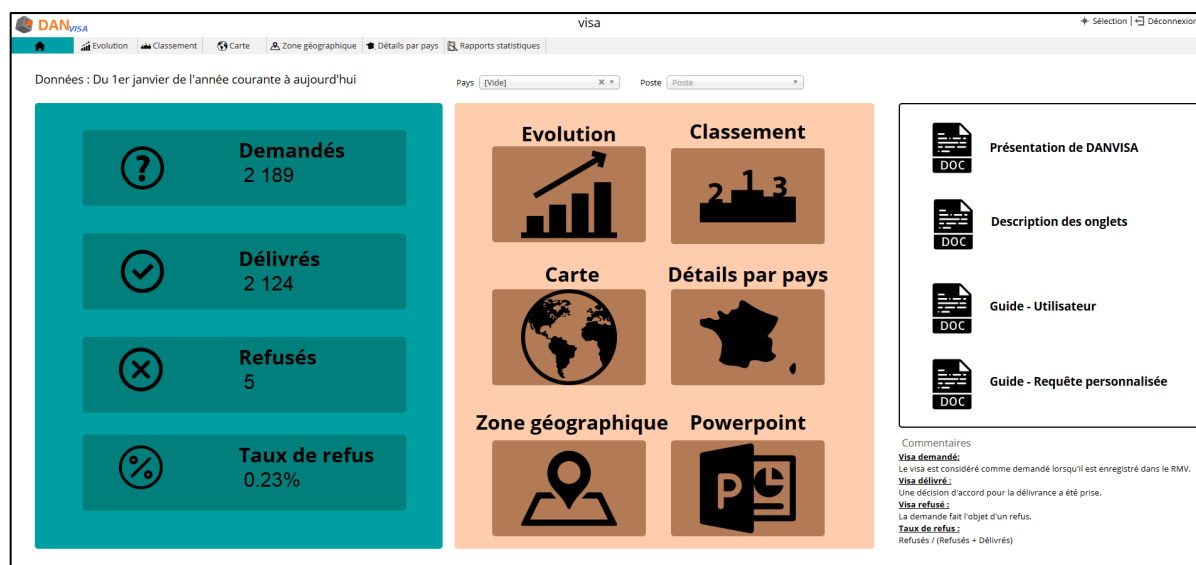


Figure 8 – Home tab of the dashboard

I distinguish two types of graph tabs:

- The tabs composed of the favorite graphs: evolution curve, classification of countries in columns, map of the world, details by country, geographical area ...

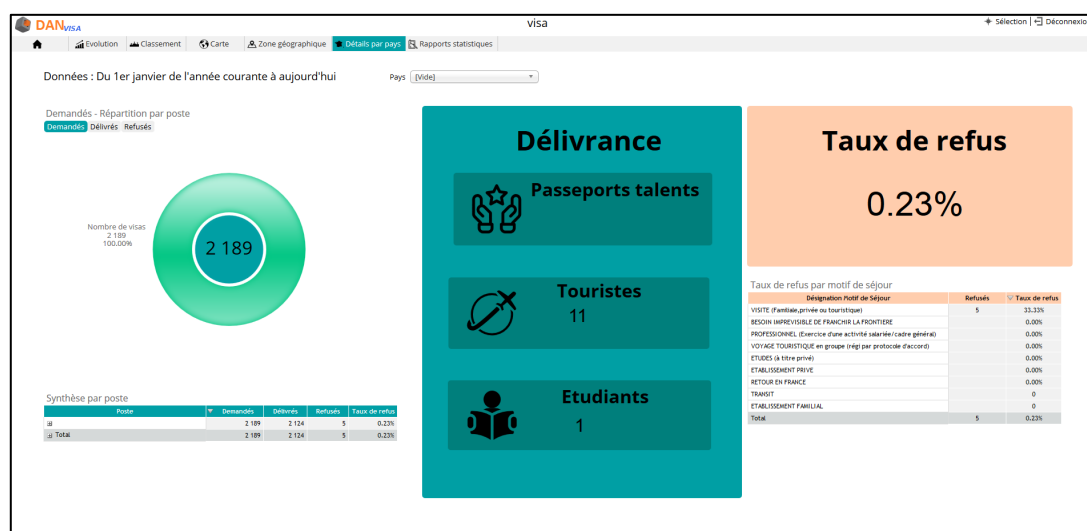


Figure 9 – Tab: Details by country

- Tabs composed of Excel factories (statistical reports to reproduce) and associated graphs.

Filters are present in each tab so that users can have, if they wish, finer data. Other additional elements are also added: a legend of the colors used, a panel of "Filtered elements" which informs about the activated filters on the tab concerned ...

The objective of the development phase is to build the chosen solution according to the validated specifications. Once the development is completed and validated by the trades, the next step is the User Acceptance Testing (UAT Testing).

2.3 UAT Testing

The objective of the test phase is to validate the conformity of the application built and its capacity to put into production.

In this phase, I transposed all my previous work without any data sources, information feeds, and unused features. This was a first necessary cleaning of the job. The transfer of the work is by transfer of backup of the development environment, or by manual configuration if necessary.

Following this transfer, the main problem identified was the performance of the application. Indeed, loading graphics on the dashboard was very long. The first cause that came to my mind is the presence of more than twenty Excel factories. But this presence is necessary. Not knowing how to improve the performance, Sandrine - my tutor - and I called on a Digdash expert - Dirgham Noui - to carry out an audit of the project. The problem comes from the multiplication of data sources for the creation of graphics. Indeed, I had created specifically for some graphics. In order to improve the performance, the ideal is to have only one data source that groups all the information used and on which all the information flows would point.

So I did a second cleanup in order to significantly reduce the number of data sources. On top of that, Dirgham communicated to me some good practices to use in each phase of development. Once the changes were made, the job was ready to be deployed in production.

2.4 PROD: Production phase

A backup of the UAT testing was done and transferred to the production environment. Thus, the office of the SDV has access to the application DANVISA via the portal Élise, the intranet of the ministry.

3. Documentation

3.1 Documentation - User

Digdash is for agents of the SDV. So that everyone can use it and generate exactly the same tables as before, I created tutorials explaining how to use Digdash:

- presentation of DANVISA
- the list of available dashboards and their description
- a guide for configuring customizable graphics
- a general guide for users

I have detailed these sheets to the maximum so that they are accessible to all and started to train them to use Digdash.

3.2 Documentation - Technical

In parallel I also produced technical documentation for Digdash buyers for the SDV project:

- The exhaustive list of the nomenclatures used and their description
- The technical difficulties encountered and their solution
- A maintenance guide for statistical reports
- A tree of user rights

4. Conclusion

DANVISA has become the main statistical management tool of the Visa Sub-Directorate. It is a data visualization tool that extracts and formats data from the visa management information system.

It is used by agents of the Postal Organization, who regularly transmit statistics to certain offices of the Ministry of the Interior and the Ministry of Europe and Foreign Affairs.

It replaces the functions of several obsolete applications used previously: Business Object, Microsoft Access and Excel TCDs (Dynamic Cross-Tabulations). DANVISA is developed from a French software, Digdash Enterprise, specialized in data exploitation and visualization. Reliability of the data, ease of use, speed of execution, capacities of evolutions, DANVISA brings a consequent improvement in the production of the visas statistics.

Thus, this first mission to the Ministry of Europe and Foreign Affairs is for me a real introduction to the world of data in the company. I could see and contribute to all the concepts of business intelligence: recovery, modeling and storage, distribution and exploitation of data. In addition, all stages of an IT project were covered: need analysis, development phase, testing, production, training and maintenance.

Moreover, one of the main obstacles that I encountered does not come from the machine, but from the people. Indeed, changing from one software to another is never easy for users who do not want to go out of their way. To cope with this problem, it was necessary to accompany them throughout the implementation of the application DANVISA so that the latter fully meets their criteria. In the end, the professions of the Sub-Directorate of Visas were very satisfied with this tool which brings a new vision of their data governance. Being able to visualize their data and generate their statistical reports in a few clicks saves them considerable time.