Abstract—The FICA project - Tools for Identifying and Combating Dropout - started at the University of Aveiro in 2015 with the aim to help reduce and prevent dropouts and increase academic success among university students. Within the project a significant amount of data is provided to different University stakeholders to monitor academic issues, however, these data are currently provided in large tables, a format difficult to analyze. In this paper, we present the main aspects of the data, the users and contexts of use. We also propose an approach to allow the visual and interactive exploration of the FICA project data to help monitor the path of the students and identify risk indicators and failure factors that can lead to critical situations such as dropout. A solution developed using the participatory design methodology is presented, detailing all stages of its creation process, from the requirements elicitation based on focus groups and interviews, design and prototype development in Power BI to its evaluation. Some suggestions for future work are also presented.

Keywords—information visualization, FICA, higher education, dropout, academic achievement, power BI, participatory design.

I. INTRODUCTION

The accessibility to higher education has been increasing in recent years, making the next step for most students who graduate from high school their university enrolment. Among other aspects, social changes and the democratization of education are some of the main causes for the explosion of enrolment and the emergence of mass higher education verified in recent decades [1]. The high number of enrolments is also often associated with a significant number of students who, particularly during the first year, end up dropping out of university [2]. These dropouts could derive from factors such as academic failure, entry regime and access to scholarships among others.

In higher education institutions, information systems play a central role in obtaining data from students, but such systems generally focus only on the storage of raw data, without any kind of processing making it difficult for users who analyse the data to detect patterns and reach conclusions. [3].

In this context, the University of Aveiro has created the FICA project - Ferramentas de Identificação e Combate ao Abandono / Tools for Identifying and Combating Dropout - which, since 2015, monitors a series of indicators in order to detect University students who are at risk of dropping out and aims to contribute to prevent and reduce this number. A limitation of this project lies in the way the data obtained is presented to its stakeholders: an extensive set of indicators in the format of Excel tables difficult to analyze.

Starting with a contextualization of this topic and a brief analysis and comparison of related work and some of the most popular visualization tools to date, this paper proposes a solution for the visual exploration of the data generated by the FICA project. The objective is to help users to obtain the answers they need through the data in a fast and interactive way. Given the interdisciplinarity nature of the data and the several types of users involved, a participatory design methodology was followed in order to promote active involvement by all stakeholders so that the end result meets their expectations. The proposed solution was developed using a user-centered design (UCD) approach to foster the accessibility to all its future users, regardless of their profile [4]. The rest of the article is organized as follows: section II briefly reviews related work and visualization platforms that might be used, section III introduces the FICA project main goals and the data as well as the users’ needs, section IV presents the prototype and its evaluation, and finally, section V presents some concluding remarks as well as possible improvements.

II. VISUALIZATION OF ACADEMIC DATA

A. Visual Exploration of Data

The main objective of any visualization is to allow information to be extracted from a large volume of abstract data in a fast, clear and precise way, without need of a great cognitive effort [5]. The communication of information through visual stimulus can be established from various visualization techniques based on different visual metaphors. The final decision about the techniques to be used should only be made after considering factors such as the type of data to be represented, the task to be performed and its cognitive characteristics as well as the questions to be answered by the users [6]. Regardless of the visualization to be created, its main function is to offer greater ease to the user in the execution of a certain task and thus, in order to meet all its requirements, the adoption of a user-centered design methodology is essential [7].
The development of a visual data exploration application should not only focus on the user needs, but also follow a participatory design methodology in which the user participates actively and directly throughout all the development process [4]. From contextualization and definition of the key factors of data to be represented to the deployment of the final solution, there are several phases to be followed. Fig. 1 shows the different steps of the development of a visualization according to Kirk [8].

![Fig. 1. Phases of the development cycle of a visualization application according to Kirk [8]](image)

The first step is to understand what will be seen, by whom, and especially why. It is in the design phase that these questions are answered. This is followed by the creation of mock-ups and/or prototypes that lead to a first version of the solution and which should already demonstrate some of the main features of the final version. Tests are then conducted on the prototypes created, first by the developers themselves and then by users. From the results, if necessary, the solution is refined and consequently new tests are carried out. This process is repeated until the solution is complete and in accordance with the requirements initially defined. Lastly, the final version of the application is released and, although it has already been evaluated during the development phase due to the participatory design methodology, it is subjected to a final evaluation [8]. The normal activity of many institutions, Universities included, produces more and more data, which has lead to the search for new methods of visually conveying their most important aspects in a single screen, and consequently an increasingly frequent use of dashboards. Few [9] defines dashboard as the visual display of the most important information needed to achieve one or more goals, consolidated and organized into a single screen resulting to a clear, accurate and effective communication with its users.

When compared to the process of creating individual views, the design of a dashboard is a task that requires special care because the fact of gathering a huge amount of information in a single place can lead to problems such as exceeding the limits of the screen, lack or excess of accuracy in the representation of data, inappropriate choice of views, inefficiency in highlighting important information, shortage or abundance in use of a certain color, unnecessary decorations, among others. It is essential that the information presented allows the user to obtain knowledge from the data in a clear way without causing any deviations in their focus of interest [10].

Given the characteristics of dashboards, this was the concept used to develop the solution presented in section IV, where for each user profile a summary of the most relevant data is presented, along with the possibility of exploring it in more detail through other visualizations.

B. Visual Exploration Applications of Academic Data

The increase in the number of students in higher education and the appearance and rapid expansion of MOOCs (Massive Open Online Courses) [11], have originated several studies and works that have as main focus the analysis and interpretation of the large amount of academic data generated. Most of these investigations are carried out in fields such as psychology or educational sciences [2], however, other information visualization projects in the context of finances [12] or medicine [13] can also be considered analogous to the academic context being also good references for the development of the proposed solution. Fig. 2 shows some of the academic data exploration applications analysed before starting the development of FICAvi.

![Fig. 2. Interfaces and visualizations of some of the academic data exploration applications analysed (A) [14], (B) [15], (C) [16] and (D) [17]](image)

TECNICOVIS (Fig.2 (A)) is an information visualization application that, based on the extraction and respective preprocessing of a dataset from Instituto Superior Técnico, Portugal, referring to all courses and degrees since 2002, allows the analysis of the academic path of students over the years. It presents a set of interactive visual representations that enable the identification of possible problems in the different courses and programs, in particular a matrix composed of columns referring to the course, subject and the different semesters over the years where each cell represents the percentage of students approvals in relation to the number of enrolments, a scatter plot that allows to establish the correlation between the various attributes and finally a line chart with the professors evaluation over the years. There is constant interconnectivity between all views by highlighting certain elements according to the selection made [14].

Considering data of three years of the Faculdade Cenecista de Osório, Brasil (about the quantity of resources accessed
by the students), some visualizations were developed (Fig. 2 (B)) that, with the help of data mining algorithms, allow some prediction on student success. The user interface is composed of two main views, where at the top, in the form of a table lens, a summary of the data loaded is categorized and in the remaining space of the page, through a pie chart and a graph of stacked columns, information about the access to various resources by students is shown [15]. The layout of the visual elements and the different colours presented, associated to the type of resource, are some of the characteristics that promote an easy interpretation by the users.

In order to monitor and help support decision making regarding future investigations, the Autonomous Popular University of the State of Puebla (UAPAEP), Mexico, has developed a system - SAPI: Publications Management System for Researchers - which gathers all data from publications made by its researchers in previous years and can be accessed by different users with different objectives. These data are shown to users in the form of multiple interactive visualizations that are part of the reports provided to them as response to the requests made to SAPI [16]. Fig.2 (C) presents some of these visualizations, including a bar chart showing the amount of approved and rejected publications in the last years, a pie chart with the distribution of the total number of publications by department and a line graph showing the number of approved publications by category.

With the purpose of detect and understand changes in the flow of students that may justify the considerable dropout rates verified, the Budapest University of Technology and Economics (BME), Hungary, has developed a visualization tool that allows to monitor its more than 30,000 students through Sankey diagrams (see Fig.2 (D)). Created with Google Charts, the diagrams are composed of several nodes with a unique name identified by the semester and current status - active if the student is attending the plan or passive in case of having the studies on pause - and several bands that correspond to the path of the various students. The width of the bands is proportional to the number of students as well as their color which varies according to the path. [17].

Although they represent different datasets, it was possible to detect various patterns in the visualization techniques used in these works, particularly in the mechanisms of selecting and filtering, which allows to interact with the various visualizations and makes possible to highlight certain elements or eliminate information that could deflect the focus of analysis. The colour and the existence of a time dimension were also other characteristics used in all the works fundamental to make possible the extraction of information from the data.

C. Visualization Tools

There is currently a huge variety of resources for data visual exploration, varying from visualization libraries such as D3.js [18], which due to their low-level language offer great flexibility in creating new visualizations, to Business Intelligence (BI) platforms, mainly used in the development of dashboards and reports that contribute to the management and planning within an organization by visualizing, analyzing and detecting patterns in large amounts of data [19].

In order to choose the right framework to be used in this work, and considering some of the most popular and reputable visualization tools [20], a benchmarking study on some essential attributes in this kind of platforms and useful within the FICA project was done:

- **Interactivity**: the visualizations created allow interaction with the user;
- **Desktop Client**: existence of a desktop version of the tool where it is possible to develop the visualizations;
- **Web Client**: possibility to create online visualizations and publish them on the web in order to share them with other users;
- **API**: existence and availability of an API;
- **Data Modeling**: ability to manipulate and connect data from multiple sources; combine data from the relationship between tables;
- **Dashboard**: possibility to display multiple visualizations on a single screen;
- **Community**: existence of an active community and online tutorials/courses.

<table>
<thead>
<tr>
<th></th>
<th>D3</th>
<th>Gephi</th>
<th>Spotfire</th>
<th>Tableau</th>
<th>Excel</th>
<th>Power BI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Open Source</td>
<td>✓</td>
<td>✓</td>
<td>x</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Interactivity</td>
<td>✓</td>
<td>✓</td>
<td>x</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Desktop Client</td>
<td>x</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Web Client</td>
<td>x</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>API</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Data Modeling</td>
<td>x</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Dashboard</td>
<td>x</td>
<td>x</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Community</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
</tbody>
</table>

Table 1. Comparison of the visualization tools considered according to the set of attributes defined

The results of this comparison, shown in Table 1, indicate a significant difference between D3 and Gephi and the rest of the tools, mainly due to the fact that they are both aimed at users with programming skills.

Excel, despite respecting almost all the attributes of the other tools, is a solution more oriented to the analysis in depth of the data in its raw state. The other three frameworks, Spotfire, Tableau and Power BI have similar functionalities and all constitute intuitive BI solutions with drag & drop systems that allow their use by any user with basic knowledge in the visualization area, however only Power BI has an open source component that offers its users the possibility to access the source code of existing visual elements and contribute with new visualizations, developed mainly in D3.

In addition, the types of licenses and prices presented by the various tools were also analysed.

Among the BI tools, Power BI is the only one that offers, even limited in functionalities, a free desktop version with no trial time. Also the cost of upgrading to the pro version is by a good margin the lowest of the three options. For these reasons the proposed solution - FICAVIS - was developed using Power BI.
III. THE FICA PROJECT

As already mentioned in section I, FICA is a program developed by the University of Aveiro to monitor a set of attributes of its students in order to detect possible situations of risk of dropout and intervene rapidly to support their academic success [21].

A. The data from FICA

Collected by the University Information and Communication Technology Services (sTIC), through various sources that include the integrated academic management system, the university portal (PACO) or the rectoral services, the data is sent to the various stakeholders every month in the form of large Excel files composed of thousands of entries corresponding to each one of the students which, due to their complexity, are difficult and time-consuming to analyse.

Every month, for each student, there are dozens of attributes characterizing their personal information, context or academic background, etc. Among these, FICA gives special focus to a set of seven indicators that are generally associated with risk situations:

- **Achievement**: Academic success rate below 50%;
- **Fees**: Students with late tuition fees;
- **Scholarship**: Students who requested and did not obtain a scholarship;
- **Entry grade**: Entry score below 120/200 points;
- **Attendance**: Students with at least one course with assiduity below 50%;
- **SWLS Scale** [22]: Students who have a low life satisfaction level according to the SWLS questionnaire;
- **SPANÉ Scale** [23]: Students with negative experience according to the SPANE questionnaire.

There are different types of users with different access to the FICA data: Program Directors have access to data of all students in the program, Department Directors have access to all programs in their department and finally the Rectory has access to data of all university students.

B. Learning the user needs

The first phase of the whole life cycle of a system must be focused on the needs and expectations of the users who will guide the whole development process. In this work a user-centred design approach was used where the requirements analysis phase starts precisely with the study of the users and their needs [24].

Given the exploratory nature of information visualization systems and the increasingly frequent use of focus group in the requirements gathering and evaluation of visualization prototypes, this was the methodology adopted to complete this stage [25]. Focus group is a qualitative research technique that consists of conducting interviews within small groups, formed by people with shared interests, guided by open-ended questions that privilege the interaction between the various participants and aims to study their opinions and perceptions about a product or service allowing the researcher to predict the reactions expected by the general population [25].

During this work, two focus group sessions were conducted, plus a videoconference interview. We involved some experts in the field, familiar with the FICA data, in order to understand what questions the solution to be developed should answer. The focus group participants included four professors, two program directors, one vice-director of different programs and a professor of a first year course with a large number of students and significant failure rate. The videoconference was held with an advisor to the rectory concerning the FICA project.

Although with different granularity, it was clear that all the participants had some knowledge about the FICA data, and the program directors knew them in depth and pointed out as major problems the overall complexity, the existence of fields with similar information and the difficulty in monitoring the path of a particular student throughout the academic year, leading to a huge waste of time in its analysis.

In the case of program directors, their main objectives were to examine the performance of students, particularly at the end of each semester, to detect potential risk situations from the identification of non-compliance with some indicators, to correlate attributes such as attendance, entry score, among others, with students’ performance or to obtain information about the number of dropouts. In the end, the set of questions to answer that emerged from the focus groups and interviews was the following:

- What was the average grade of the students in the program?
- What was the distribution of student achievement in the program?
- Is there any difference between student achievement at the end of the first and second semesters?
- How was a particular students’ path throughout the academic year?
- Which students do not satisfy a certain indicator?
- How does a specific attribute influence the performance of students in a program?
- How many students have cancelled or suspended enrolment in a given month?

IV. FICAvis

The focus of FICAvis is to allow the several attributes present in the FICA data to be analyzed in a clear and interactive way in order to answer the questions defined in the previous sections. The proposed solution, developed using PowerBI, consists in a report [26] for each FICA user profile: Rectory, Department Director and Program Director, offering a multi-perspective view into the dataset composed by several visualizations that allow its users to interact and get insight from the data.

A. Data Pre-Processing

Regardless of their profile, the Excel files containing the FICA data reach their users with an identical structure (same columns and attributes) allowing Power BI to combine them into a single logical table.
This transformation is fully automated by Power BI itself, which records in the Applied Steps section all the actions performed during the process leaving room for an easy possible reformulation or addition of new intermediate steps in future. Also, any change concerning the data source, for example the addition of data regarding a new month, is immediately considered by Power BI ensuring constant synchronization of the dataset.

With all the Excel files combined in a single table, some preprocessing steps are performed, such as removing irrelevant columns, changing the data type of some attributes, normalizing the data or creating attribute-value pairs for the seven main FICA indicators.

In addition to the operations performed on the global table, using some formulas from the DAX function collection [26], a calendar type table containing columns with information for the day, month and year, is created that assumes a one-to-many relationship with the global table, allowing the analysis of FICA data with a time dimension.

B. Development of visualizations

As mentioned, each report consists of multiple pages that are composed of several visual elements, related within its context, that allow to answer user questions.

In order to make the analysis of the data a fast and interactive process, all the pages developed present an identical structure composed by three main sections as can be seen in Fig. 3: a top bar that displays, in addition to the page title, information about the program or programs being shown, such as name, type of degree, grades average or number of students, constantly updated as a result of the user’s interactions with the rest of the page; a sidebar with various data filtering options and a button to reset the page to its initial status; and finally, a set of interconnected visualizations that allow user interaction with the data. Whenever possible, color is also used to help get insight from the data. Each visual element has a '?' symbol that shows detailed information about it on mouse over. Most pages also display a table with the list of students, resulting from the current interaction, that allows drillthrought to obtain more detailed information of a specific student.

In total there are eight different pages in the report. The first one the user comes across is the Home Page. Here the user can see a list of all the report pages sorted according to their importance resulting from the focus groups, complemented by a brief description of their content and an arrow link that allows navigation to the respective page.

The remaining pages are composed of several visual elements, set in a particular context, that allow to analyze and answer several questions based on the data.

The Overview of student achievement page is exclusive to the users who have access to data from more than one program. It gives the user an overview of the performance of all students of those programs. In order to represent several attributes of a large amount of hierarchical data, an interactive treemap was used that aggregates all the data in different levels of detail: education field, degree and program. The size of each treemap rectangle codes the number of students and its color is mapped based on the average achievement of that group. By selecting a particular group it is possible to navigate to any of the other pages which are immediately configured with the data of these students.

The remaining pages, available for all different user profiles are the following:

- **Monthly Tracking**: Presentation, with monthly granularity, of students’ behavior in relation to the main indicators of FICA - achievement, fees, scholarship, entry grade, attendance and satisfaction and quality questionnaires - and the risk situations generally associated with them.
- **Monthly Comparison**: Monthly comparison of the main indicators of FICA, enrolment cancellations (Fig.4 (A,D)) and evolution of student achievement in the selected programs over the years.
- **Overall Performance**: Overall perspective of student performance and achievement in the selected programs, including its distribution by achievement rate, average grade, enrolment status (Fig.4 (B)) or gender.
- **Achievement by Indicator/Profile**: Detailed comparison of student achievement by different attributes, specifically by the main FICA indicators - fees (Fig.4 (C)), attendance, scholarship or entry grade - and by their profile - access type, phase and option’s entry and frequency regime.
- **Personal Information**: Detailed information of a student personal or academic context and tracking of various attributes such as the evolution of attendance to the different courses enrolled or compliance with the main FICA indicators throughout their academic path.

Different evaluation methods were applied during the implementation of FICAvis, namely an analytical method, heuristic evaluation, and an empirical method, usability tests. The heuristic evaluations were carried out in two distinct phases of the implementation process, first an initial prototype was evaluated by an experienced evaluator and then a more advanced version was evaluated by six evaluators with knowledge in usability, allowing correct potential usability issues. The usability tests were carried out by domain experts and in
addition to suggesting improvements, provided very positive feedback from all stakeholders who have shown interest in using FICAVis frequently in the future as an alternative to the format currently used.

Fig. 4. Examples of visualizations present in the different pages

V. CONCLUDING REMARKS

The data from the FICA project may provide important information concerning the University students allowing to detect and adopt measures to prevent various risk situations leading to dropout. However, their complex format means that some problems are not always identified in time, leading to irrecoverable situations. Through the use of visualizations it is possible to foster insight with much more clarity and speed allowing a more constant and active action by responsible entities. The solution proposed was developed, since the very beginning, with the active participation of expert users with different goals and needs, privileging a user-centered approach and promoting participatory design.

The next step will be the integration of FICAVis into the University Services allowing the visual exploration of the data by the users. This will involve connecting FICAVis with the FICA project database and defining the data security layer according to the different user profiles and privileges.

ACKNOWLEDGMENT

The authors are grateful to all who have contributed in any way, particularly to Prof. Joaquim Madeira, António Neves, Leonor Teixeira and Augusto Silva, as well as to Dr. Carolina da Costa, Telmo Rocha and Dr. Loraine Nazaré for their valuable contributions. This work was partially supported by IEETA - Institute of Electronics and Informatics Engineering of Aveiro, funded by National Funds through FCT, in the context of the project [UID/CEC/00127/2019]

REFERENCES


