DBL: Analysis and Development of an Information System to Manage Survey-Based Studies

Francesco Galbiati

Dipartimento di Elettronica Informazione e Bioingegneria
Politecnico di Milano
Milan, Italy
francesco4.galbiati@mail.polimi.it

Abstract—Research studies often require large amounts of data to be collected, stored, and processed. For this reason, it is essential to have a well-designed architecture to keep the data organized through time and to allow the involved researchers to retrieve and analyze them. Information system technology can provide powerful instruments to meet the requirements of such complex scenarios by automatizing an otherwise tedious and repetitive process, thus aiding both the subjects and the researchers in the swift fulfillment of their respective tasks. This paper addresses the needs just described by presenting a high-level view on the development and implementation of the DBL platform, an information system that will be deployed in the next months and will be used to manage psychological and psycho-social surveys. The discussion will verge on topics such as the requirements of the system, how its various components interact, the involved hardware platforms, its performance, and how an efficient infrastructure could easily allow a small-scale survey-based study to be expanded to a larger set of subjects.

Index Terms—Information system technology, system analysis, software deployment, survey-based research

I. INTRODUCTION AND CONTEXT

A. Background

Over the past few years, engineering communities have increasingly focused on the link between a student's psychological well-being and their academic performance, intending to design projects that empower students in their degree programs. Several studies [1-5] have shown that students' personal growth towards a proactive mindset correlates with higher educational achievements. This type of research usually involves multiple questionnaires, such as psychological, psycho-social, and academic performance surveys, which are administered cyclically throughout the semester to assess the effectiveness of the designed experience with a number of participants involved that can range from hundreds to tens of thousands. Therefore, it is crucial to have a well-organized system for efficiently storing and retrieving data from various surveys or different iterations of the same survey.

In such a context, it's not difficult to understand how a properly implemented information system could provide simple and effective interfaces that could be used by the participants to complete the requested surveys, by the researchers to retrieve and automatically extract analyzed data, and by specialists (such as psychologists) to eventually discuss, during personal meetings with those who took part in the study, the results of the analysis.

B. Information Systems

During the last decades, Information System Technology has become one of the most important tools to collect, store, and distribute large sets of data. The success of this kind of technology can be found in its great capability to adapt to different fields, spanning from education [12-13] to healthcare [14-15], and in the benefits it can bring to the organization of complex real-world scenarios.

To proceed with the discussion it's first necessary to understand the main parts of an information system and how their combined work can have a positive impact on the context in which the system is deployed.

Stair and Reynolds, in their "Principles of Information Systems" [16], identify five fundamental components:

- Hardware: Computing machines and I/O devices.
- Software: The programs that allow the information system to process data.
- Data: The facts collected by the system, that allow it to produce useful data. The component used to store data is the database.
- Procedures: Real-world policies that regulate the functioning of the information system.
- Network: The infrastructure that allows communication between different machines.

The quality of an information system can be assessed by considering different aspects of the system itself and the circumstances in which it is used. One of the most important discussions concerning these aspects is a study conducted by DeLone and McLean [17], who in 1992 described a set of dimensions on which the success of an information system depends. Among them, the most significant ones are:

- The quality of the system, in terms of implementation, performance, reliability, and flexibility.
- The system's impact on the organization of the context in which it is deployed.
- The satisfaction of the users in terms of usability, reliability, and impact of the system on the individual recipients.

In conclusion, during the development of an information system, it is important to consider, aside from the quality and efficiency of the implementation, also how the resources at the disposal of the system can be efficiently used to improve the different dimensions that determine its quality.

C. Personal Awareness and University Career - A NECST-Camp Study

To properly understand the requirements and the design choices of the information system discussed in this paper, we need to introduce the context from which it originated.

"Personal Awareness and University Career" is a research conducted at Politecnico di Milano that involves the students participating in NECSTCamp, an extracurricular program whose objective is the development and improvement of personal awareness in its participants through a series of heterogeneous activities, such as sport (CrossFit and Spartan races), research projects and particular attention to a healthier lifestyle and psychological well-being. The study aims to collect data about different psychological parameters and university career data about its participants, to monitor them and their correlations over time, with the final objective of understanding if better psychological well-being and personal awareness can lead to an improvement of the university career, not only in terms of better performances but also of a more realistic definition of objectives.

The study comprises five psychological questionnaires [6-11], a psycho-social one about past experiences inside and outside the academic context, and a survey on the student's university career, goals, and accomplishments. These questionnaires are administered to the participating students every semester, and each one of them collects useful data about the students that will be used in the analysis process.

II. Problem Statement

Since the start of the "Personal Awareness and University Career" study, its questionnaires have been carried out using Microsoft Forms, and their results have been saved using Excel worksheets, one for each iteration of every survey proposed. After the surveys are completed, the Excel files are manually downloaded and analyzed one by one to produce graphs that highlight the most significant results. The results are then forwarded to a psychologist, who will organize personal meetings with the study participants, to discuss the outcome of the surveys.

This process, besides being very slow, can cause several issues over time: for example, since the study aims, among other things, to compare data from different iterations of the surveys, it becomes difficult to retrieve the correct files that contain the data to analyze. Other issues that may arise over time are the conservation of the stored data since it's difficult to keep well organized a large number of files, and the privacy of the students who compiled the surveys, concerning the data they provided.

Ultimately, a better-organized system to collect, keep, and present the data gathered by the surveys increases the scalability of the study, making it less costly to add new questionnaires or expand the study to include a larger set of students. The DBL information system was designed to address these issues, by providing a durable and flexible infrastructure, able to collect and store the required data automatically, while helping researchers and psychologists during the analysis process.

III. Methodology

The development of the DBL information system was divided into two modules: the first one comprises the functionalities concerning the data collection process, while the second one will focus on the interfaces needed by researchers and psychologists to produce reports from the data provided by the students. The first part is set to be completed by the end of July and deployed in September, with an open testing phase at the end of August. The development of the second part will begin shortly after and will be finished by the end of 2024.

The development process has been carried out in collaboration with the DEIB department's ICT Services personnel and follows, a waterfall approach, which divides the development process into six sequential stages: requirements analysis, system design, implementation, testing, deployment, and maintenance. The first module has been implemented on a temporary testing machine, and once completed its contents will be copied to a production machine, with the same characteristics as the testing one, where it will be made accessible to the public and will begin the open testing phase. The development of the second module will follow the same approach.

From an implementation point of view, the platform's user interfaces were realized with the use of a Bootstrap 5-based template, while the server-side scripting is written in PHP 8.1.

IV. System Description

This section aims to provide the reader with a complete understanding of the functionalities of the DBL platform from different perspectives. The first part will focus on the user's perspective by showing the use cases covered by the system and the interfaces that will be presented to the users (specifically regarding the first module since the second one is still to be developed). The second section will instead focus on a functional perspective and explain how the features described in the High-level View are actuated from the system's point of view. Lastly, the focus will shift to the architecture on which the DBL system will be deployed, providing insights into the hardware and software specifications.

A. High-level View

The features required from an information system are directly dependent on the needs of those who will use it. For this reason, during the requirements analysis, four categories of users were identified:

- **Students**: they are the subjects of the "Personal Awareness and University Career" study, their job is to complete the surveys provided during certain periods. At the end of each semester, they can request a personal meeting with a psychologist to review and discuss the most significant results of the analysis. The students can also be divided into different groups, this is an important feature as it allows a categorization of the students based on arbitrary real-world parameters (such as the degree program or the NECSTCamp level) that can be used during the data processing phase. Another advantage of this functionality

is that it becomes easy to send the surveys only to a restricted number of students without the need to know all of their person codes.

- **Analysts**: students (from Master's Degree or PhD) that can extract the data regarding the students through a dedicated interface and conduct the required analysis processes to extract significant results.
- **Psychologists**: psychologists are the specialists who follow the study and the students involved in it, their job consists of deciding which surveys to carry out, eventually adding them to the system, and returning the results of the surveys during personal meetings.
- **System Administrators**: They are the users allowed to create and manage the groups of students on the platform and send them the survey requests. Each of them is associated with a "school" and will be able to manage the groups belonging to it.

From a high-level perspective, a study like "Personal Awareness and University Career" could be integrated with the system as follows: at the start of a semester, a set of empty groups is created (either manually by a system administrator or automatically by the system), in those groups, the system administrator will insert the students taking part in the study. During the semester, when a questionnaire needs to be sent to the students, the system administrator will use the dedicated interface (fig. 1) to forward a request to the desired groups; the students belonging to them will receive an automatic e-mail notification regarding the activation of a survey.

Requests are immediately effective and have an expiration date, before which the students should complete the survey, and a final date, before which the questionnaire will still be accessible to latecomers. After receiving the e-mail, the students will find the requested survey posted on their DBL home page (fig. 2), from which they will be able to complete it (fig. 3).

As the study progresses, the analysts will be able to access the data provided by the participants and extract the most significant results from a dedicated dashboard, while the psychologists will be able to generate and download a report regarding each of the students taking part in the study.

Fig. 1. Request creation interface, where the system administrator can select as many groups as necessary, the survey to activate, and the relative dates

Fig. 2. The "Your surveys" section on the user's home page, where for each of the questionnaires the platform displays the expiration date and the type, with a button that starts the survey. If the survey has already been completed the button is deactivated

Fig. 3. The appearance of a survey in the DBL platform, where the question's text is displayed first and the set of possible answers below it along with the relative labels

Another functionality implemented in the DBL system is the concept of "school": the groups of students can be divided into larger clusters, called "schools". This feature will allow different organizations in PoliMIi, like a department or one of PoliMi's schools, to use the DBL platform to conduct studies similar to "Personal Awareness and University Career" on a different audience. The system will keep the data from one school isolated from the others and ensure that researchers and system administrators can only access the information regarding their school(s) of reference.

B. Functional View

From the DBL system's point of view, the functionalities described in the previous section are a result of the combined work of the following functional components:

- A **Database** that stores the information necessary to fill the web pages, the data regarding the different surveys and the students, and the answers collected from the surveys. The student's data are stored anonymously in

the database: each person registered in the system is identified by a unique ID, automatically generated from their person code using a hash function.

- The **Application logic**, composed of server-side scripts written in PHP. These programs have two main functions: the first one is to generate the web pages according to the requests sent to the web server by the users. In this case, the script enquires the database, retrieves the necessary contents, composes the page's hypertext, and forwards it to the web server. The second function is to elaborate and insert into the database the data collected from the surveys. The communication between the application logic and the database relies on server-side functions written using PHP's PDO class.

- A **Web Server**, which manages the HTTPS requests coming from the users, retrieves the corresponding web pages from the application logic and sends them to the respective clients as hypertext documents.
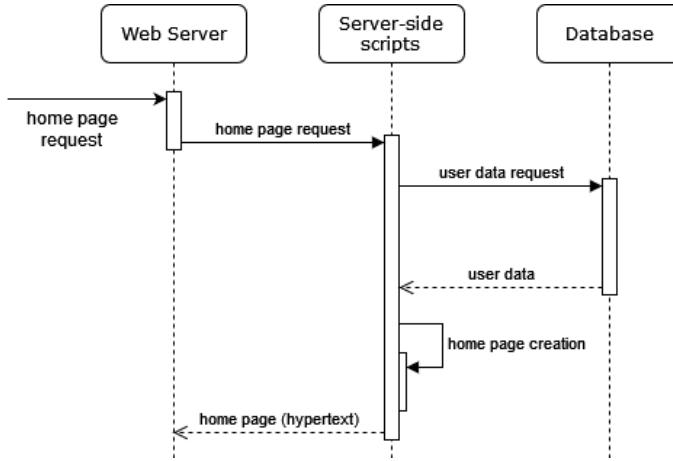

Fig. 4. Communication between the components during the generation of the home page

The previous image shows a simplified diagram that represents the communication between the various components during the generation of a home page. The same sequence of messages is also valid for the generation of all the survey pages.

The DBL platform also relies on the Aunicalogin APIs to manage the login functionalities of the users and on PoliMi's email provider to manage the notifications to the students.

Several security measures have been implemented to prevent attacks that may compromise the system's functioning or the integrity of the data stored in the database, such as input validation, session integrity checks, and authorization controls to access the web pages.

C. Deployment View

From a deployment perspective, the DBL system is hosted on the DEIB department's servers, specifically on two virtual machines, one for testing and one for production, virtualized through the KVM (Kernel-based Virtual Machine) technology,

which converts Linux into a bare-metal hypervisor. The virtual machines have the following characteristics:

- Operating System: Red Hat Enterprise
- Storage: 400 MB SSD
- Web Server Software: Apache Web Server
- DataBase Management System: MySQL
- Scripting Language: PHP 8.1
- Access to the database: Adminer 4.8.1
- Access to the server's file system: FTPS protocol

V. RESULTS

Although the system described in this paper is only in the final phase of its first module and real-world results about its usage are not available yet, we can still draw conclusions regarding some important aspects of the work, like the positive impact that the methodology employed during the development has on the system in its current state and will have on its future instances.

In particular, the parts that have been developed show great results in terms of usability and extensibility. Sending a survey request to a large number of students requires little effort and time thanks to the simplicity of the interface. Moreover, to add a new questionnaire to the platform, a system administrator or psychologist will only have to insert the necessary parameters (the survey's description, the questions, and the possible answers) into the system through the dedicated page. Once a client requests the page of the new survey, the system will automatically generate the user interface; thanks to this feature it is not necessary for any user to write code for the system.

Another result that can be presented at the moment is an assessment of the system's performance concerning the features available at the moment, this analysis will allow the reader to understand the efficiency of the communication infrastructure and the server-side scripts. (Data have been registered using the Firefox Performance Profiler on a laptop computer connected to a private Wi-Fi, to simulate a context as similar as possible to the ones in which the system will be used). A simple, yet interesting, assessment can be carried out by measuring the system's performance when generating the interfaces for the psychological surveys. In particular, we can compare the time necessary for the server-side script to compose the hypertext of each survey (The HTTP response of each survey has been measured ten times):

Survey	Number of items	Average HTTP response
PWB-42	42	98,4 ms
GRIT-8	8	28,8 ms
FFMQ-39	39	70,0 ms
Mindset-8	8	36,1 ms
ZTPI-16	16	32,8 ms

Table 1: results of the survey generation response time

The results obtained from the measurements are expected: in Table 1 it can be noticed that the surveys with more questions

(PWB and FFMQ) take significantly more to be generated time than those with fewer items. During all the measurements, the web page loading time (including the CSS and JavaScript) did not exceed 400 ms, which is much faster than Microsoft Forms. This result is to be expected since the DBL system is a much more specialized platform.

VI. CONCLUSIONS AND FUTURE WORKS

This paper has discussed the development and implementation of the DBL information system, by explaining the context in which it originated, with particular attention to the "Personal Awareness and University Career" study, its methodology, and how its functioning can be improved and extended by an information system. The discussion has then shifted towards the description of the DBL information system itself, providing its requirements and functionalities from different points of view and ending with the presentation of some performance measurements.

The next steps in the development of the DBL platform are an in-depth testing of the functionalities regarding the questionnaires, which will be opened to the public by the end of September, and the implementation of the interfaces for the analysts and the psychologists, which are set to be released in the last months of 2024.

Once the DBL platform is fully deployed it will also be usable by other organizations inside Politecnico di Milano, to help them manage research studies based on data collection from surveys.

REFERENCES

[1] K. J. Reid and D. M. Ferguson, "Do design experiences in engineering build a "growth mindset" in students?," 2014 IEEE Integrated STEM Education Conference, Princeton, NJ, USA, 2014, pp. 1-5, doi: 10.1109/ISECon.2014.6891046.

[2] M. Marques, S. F. Ochoa, M. C. Bastarrica and F. J. Gutierrez, "Enhancing the Student Learning Experience in Software Engineering Project Courses," in IEEE Transactions on Education, vol. 61, no. 1, pp. 63-73, Feb. 2018, doi: 10.1109/TE.2017.2742989.

[3] L. Da Re, R. Bonelli, and A. Gerosa, "Formative Tutoring: A Program for the Empowerment of Engineering Students," in IEEE Transactions on Education, vol. 66, no. 2, pp. 163-173, April 2023, doi: 10.1109/TE.2022.3205962.

[4] A. Danowitz and K. Beddoes, "Mental Health in Engineering Education: Identifying Population and Intersectional Variation," in IEEE Transactions on Education, vol. 65, no. 3, pp. 257-266, Aug. 2022, doi: 10.1109/TE.2022.3182626.

[5] E. Kim and G. J. Strimel, "The Influence of Entrepreneurial Mindsets on Student Design Problem Framing," in IEEE Transactions on Education, vol. 63, no. 2, pp. 126-135, May 2020, doi: 10.1109/TE.2019.2918253.

[6] Baer, Ruth and Smith, Gregory and Hopkins, Jaclyn and Krietemeyer, Jennifer and Toney, Leslie. (2006). Using Self-Report Assessment Methods to Explore Facets of Mindfulness. Assessment. 13. 27-45. 10.1177/1073191105283504.

[7] Ryff, C. D. (1989). Happiness is everything, or is it? Explorations on the meaning of psychological well-being. Journal of Personality and Social Psychology, 57(6), 1069–1081. https://doi.org/10.1037/0022-3514.57.6.1069

[8] Dweck, C. S. (2006). Mindset: The new psychology of success. Random House Incorporated.

[9] Angela Lee Duckworth and Patrick D. Quinn (2009) Development and Validation of the Short Grit Scale (Grit–S), Journal of Personality Assessment, 91:2, 166-174, DOI: 10.1080/00223890802634290

[10] Zimbardo, P. G., and Boyd, J. N. (1999). Putting time in perspective: A valid, reliable individual-differences metric. Journal of Personality and Social Psychology, 77(6), 1271–1288. https://doi.org/10.1037/0022-3514.77.6.1271

[11] Martoni, Monica and Fabbri, Marco and Russo, Paolo. (2023). The Italian Validation of the Zimbardo Time Perspective Inventory and Its Comparison with Three Time Perspective Inventories. International Journal of Environmental Research and Public Health. 20. 2590. 10.3390/ijerph20032590.

[12] A. Hašková, "Information systems as a component part of tertiary education quality assurance," 2016 IEEE 10th International Conference on Application of Information and Communication Technologies (AICT), Baku, Azerbaijan, 2016, pp. 1-5, doi: 10.1109/ICAICT.2016.7991792.

[13] U. Khamdamov, A. Abdullaev, K. Sultanov, and J. Elov, "Models of integration of higher education management information systems," 2021 International Conference on Information Science and Communications Technologies (ICISCT), Tashkent, Uzbekistan, 2021, pp. 1-4, doi: 10.1109/ICISCT52966.2021.9670171.

[14] H. Demirkan, "A Smart Healthcare Systems Framework," in IT Professional, vol. 15, no. 5, pp. 38-45, Sept.-Oct. 2013, doi: 10.1109/MITP.2013.35.

[15] P. -F. Tsai, I. -S. Chen and K. Pothoven, "Development of handheld healthcare information system in an outpatient physical therapy clinic," Proceedings of the 2014 IEEE 18th International Conference on Computer Supported Cooperative Work in Design (CSCWD), Hsinchu, Taiwan, 2014, pp. 599-602, doi: 10.1109/CSCWD.2014.6846912.

[16] Stair R, Reynolds G. Principles of Information Systems. 14th ed. Cengage; 2020.

[17] Delone, William and McLean, Ephraim. (1992). Information Systems Success: The Quest for the Dependent Variable. Information Systems Research. 3. 60-95. 10.1287/isre.3.1.60.