Big Data and Open Data in Education, evaluation of the scope of existing initiatives. Case study Faculty of Technology and Applied Sciences of the National University of Catamarca

Claudia M. Herrera[1], Carolina I. Chayle[1], Ana C. Pauletto[1], María A. Barrera[1], and Nadir Morales Apaza[1]

[1] Department of Computer Science, School of Technology and Applied Sciences, Universidad Nacional de Catamarca, Argentina. https://tecno.unca.edu.ar
herrera.claudia.mabel@gmail.com

Abstract. Data is the fuel of the 21st century. Increasingly our economies are driven by this input. Data has the potential to create social value, something that higher education institutions are beginning to rely on, as evidenced by the initiatives of public access to their data and the various advantages and benefits provided by big data in improving educational management, the development of customized curricula, monitoring the academic performance of students, as well as in the generation of digital repositories that are the product of years of academic, teaching and research activity. The present research aims to evaluate the scope of existing initiatives on the subject of Big Data and Open Data in education, in the Faculty of Technology and Applied Sciences of the National University of Catamarca (UNCA), producing an effective contribution with preliminary results for the research project called "ICT in the Service of Open Data: Current Situation, Conceptualization and Initiatives for Opening Public Information" which is being carried out in the Department of Informatics.

Keywords: Big Data · Open Data · Education · UNCA

1 Introduction

In recent years, an information revolution has begun worldwide that is directly related to free access to data. Open Data responds to an initiative already installed in several countries and that pursues the publication of data sets that governmental organizations, companies, institutions, among others, have in their possession, so that they can be disclosed and potentially reused by society in general, and in particular, by entities that can add value to such data and thus develop beneficial products for the community.[1]

However, it is impossible to process and analyze huge data repositories with conventional database and analytical tools. This is only possible with the use of the Internet and tools that allow the collection of information, this massive data analysis

is known as Big Data.

The concepts of big data and open data are related, but they are not the same. "Big data" or big data refers to such a massive set of data that it generally requires specialized software to analyze it and extract patterns. In contrast, "open data" refers to the possibility that any person or company can reuse a data set to analyze trends, generate new applications or create businesses, among other uses. For this to happen, data must be accessible in two complementary ways: it must be available and in a format that allows it to be digitally reused. In other words, while big data is defined by its quantity, open data is defined by its quality. Open data does not necessarily have to be "big data" to have an impact. [2]

In addition, the application of Big Data in the field of education requires much deeper research that can provide a broad scope of understanding about the functioning and abilities of higher education institutions in their four primary missions: research and data analysis, teaching, training and administration; through these elements, for example teachers can customize each of their lectures to improve the performance of their students. [3]

This research aims to evaluate the scope of existing initiatives on the subject of Big Data (BD) and Open Data (OD) in education, within the Faculty of Technology and Applied Sciences (FTyCA) of the National University of Catamarca (UNCA), producing an effective contribution with preliminary results for the research project called "Las Tic al Servicio del Dato Abierto: Situación Actualización, Conceptualización e Iniciativas de Apertura de Información Pública" which is executed in the Department of Informatics.

2 Big Data

Although there is no commonly accepted definition of DBs, we can say that they are data that can be defined by a combination of the following five characteristics [4]:

- *Volume:* where the amount of data to be stored and analyzed is large enough to require special considerations.
- *Variety:* where the data consists of multiple types of data, potentially from multiple sources; here we should consider structured data stored in tables or objects for which metadata is well defined, semi-structured data such as documents or other similar where metadata is contained internally (e.g. XML documents) or unstructured data, which may be photographs, video or any other form of binary data.
- *Velocity:* where data is produced at high rates and running on "stale" data is not valuable.
- *Value:* when the data has a perceived or quantifiable benefit to the company or organization using it.
- *Veracity:* where the accuracy of the data can be assessed.

All these features today are undoubtedly provided by new technological trends, such as cloud computing and the Internet, without the latter this would not be possible. The versatility and magnitude with which data is handled through these tools facilitates and shortens the distances of time and space, which in itself is a great benefit for educational institutions and the social impact before a society in constant

advance. The benefits of using DB techniques are quite broad. Two main groups of benefits emerge: cost savings; and competitive advantage. The potential of BD depends on the sectors of activity where it is used. There have been significant advances in science through the adoption of BD, particularly in astronomy, biology and bioinformatics. [4]

3 Open Data

OD (Open Data) are those data that are freely available for use, reuse and redistribution. These data must comply with the following premises [5]:

- *Availability and access:* Data should be available at a reasonable cost and in a convenient and modifiable form.
- *Reuse and redistribution:* The format of the data will allow for reuse, redistribution and integration to other datasets.
- *Universal participation:* Everyone should be able to use, reuse and redistribute information without restrictions.

In December 2007, the Opengov Data organization published eight principles that should be considered by governments when publishing data. These principles do not indicate which data should be public, but rather specify the conditions for public data to be open. [5]:

- *The data must be complete. All data shall be available.* Data is stored electronically including documents, databases, transcripts, audio and video recordings, images, etc. Public data is data that is not subject to existing privacy, security or privilege limitations, which is governed by other laws.
- *Data should be primary.* Data are published as extracted from the source, at the best possible level of granularity, not in aggregated or modified forms.
- *Data must be timely.* Data are made available as soon as necessary to preserve the value of the data.
- *Data must be accessible.* Data are available to the widest range of users and for the widest range of purposes.
- *The data shall be machine processable.* The data shall be reasonably structured to permit automatic processing of the data.
- *Access should be non-discriminatory.* The data are available to anyone, without registration.
- *The format of the data should not be proprietary.* The data are available in a format over which no entity has exclusive control.
- *The data must be license-free.* The data are not subject to any copyright, patent, trademark or trade secret regulations. Reasonable privacy, security and privilege restrictions may be allowed as they are governed by other laws.

Tim Berners-Lee proposed a categorization of the degree of openness of data based on stars, regarding how open and usable data an institution can offer (see Fig.1). [6]

- *1 star:* Data must be available on the web in any format and under an open license to be considered open data.
- *2 stars:* Data must be structured and in a format that can be interpreted or

processed by machines.

- *3 stars:* Same as above but in a non-proprietary format.
- *4 stars:* All of the above plus the use of w3c recommended standards (RDF - SPARQL) to identify things.
- *5 stars:* All of the above plus linking to the data of others and thus providing context.

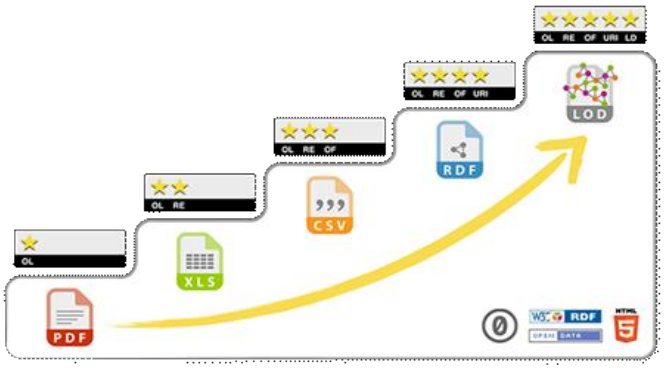

Fig. 1. 5-star development scheme for open data.

4 Big Data and Open Data in Education

BD in the educational world refers, above all, to the enormous amount of data and information we have about the educational system itself, the academic performance of students, the training needs of teachers, or the impact of educational measures taken in the field of training to improve the quality of educational processes in times of globalization, interculturalism and multilingualism. Thus, one of the key features of BD in the academic context is linked to learning analytics, which are understood as processes of continuous and procedural assessment that address the changes and evolution of certain school factors or variables. Thus, the potential of BD can be specified in the following aspects [7] Among the main educational methods derived from BD analytics (whose origins date back to the year 2000), and its integration with the new smart devices and web technology that today are already being applied in the educational field, are the following [8]:

- *Adaptive learning:* knowledge data, which will facilitate analytics to create the tailored curriculum. This method emphasizes the areas in which students have more difficulties, in order to adapt to their way and pace of learning and create a personalized, differentiated and tailored teaching path for each student.
- *Competency-based education:* today's competency-based education platforms, derived from the power of DB analytics under learning metrics, empower student independence while allowing teachers to assess progress. One example is Mastering.
- *Flipped classroom and blended learning:* A typical platform that uses these educational methods is Moodle. The knowledge bases generated by these types of

platforms serve to complement the data lakes of an educational institution and thus facilitate learning analytics.

- *Gamification:* This tool helps teachers enhance classroom learning in an efficient and fun way. In particular, it enriches specific behaviors and boosts students' active participation by issuing rewards in real time. This platform delivers tracking reports to inform students' progress to both parents and teachers.
- *Mobile learning:* Currently, these devices promote learning through educational content, such as electronic textbooks or interactive courses that help improve school performance. They also allow the application of assessment or placement tests that speed up and simplify the student evaluation process. These devices are one of the most important pillars for gathering information about students and the educational processes they attend, thus increasing the possibilities of performing analytics with BD technology and identifying new areas of opportunity to improve education.

These methods, which allow education to reach a larger student population and meet the continuous demands of students (who are increasingly demanding a more flexible and personalized education), are providing very promising results and have changed the way in which students relate to each other, to their teachers and to educational institutions.

For their part OD can be used in higher education, as follows [9] :

- Collaborating with researchers on real research projects.
- Promoting collaboration among students from diverse disciplines by creating learning activities based on sustainable development goals.
- Encouraging activities in which students help their local communities solve real-life problems.

Likewise, the use of OD can contribute to the development of critical thinking, research, teamwork and citizenship skills. So, when working with students, we can incorporate OD into teaching activities by guiding learning through [9]:

- Identify and describe the learning outcomes for the planned activities.
- Identify the portals that will generate the data.
- Identify and clearly describe the challenges that students might face.
- Provide training materials for the software that students will need to analyze the data.
- Support students in communicating their findings to local communities.

5 Methodology

The research is within the paradigm of non-experimental research design, of the exploratory type, observing the phenomenon as it occurs without manipulation of variables or comparison of groups of subjects. [10]. A descriptive and comparative study was carried out. A search strategy was proposed considering the characteristic aspects of each of the terms related to the object of study, i.e., BD and OD in education that allow the evaluation of existing initiatives in the Faculty of Technology and Applied Sciences (FTyCA) of the National University of Catamarca (UNCA). With respect to obtaining information, samples were taken through direct observation

of the FTyCA website.

6 Preliminary results

The FTyCA showed a considerable amount of data published in the different domains that articulate the university (institutional, academic, teaching, research repositories, archives). Specifically, the following were observed: a research repository, publication of government and academic acts, among others. In other words, actions corresponding to the first steps in OD were evidenced when opening small and simple data sets, which are part of a larger data set. All published data have an open format but the level of openness corresponds to 1 star. This is because most of the published data are in a format that does not allow their reuse, for example in .PDF, .DOC, .ODT, .TXT, among others; or a scanned image. Furthermore, these data are available free of charge without the need to register or request access. On the other hand, only some of the data published can be downloaded from the URL where they were found and comply with the "Updated Data" evaluation index, which guarantees users that the data are renewed every year.

In relation to the observation and exploration of the aforementioned aspects of BD, the FTyCA has the Moodle virtual education platform in which applications of competency-based education, flipped classroom and blended learning, and gamification were found. It also offers teachers and students the possibility of applying Mobile Learning through the application of this platform, which is also offered by the academic unit. Although different statistics are available on student learning in terms of the number of passing and failing students, student desertion, enrollment dropout, among others, no evidence was found of the application of BD for adaptive learning in customized study plans that take into account these statistics, learning habits, knowledge, weaknesses and strengths of each student.

7 Conclusions

Based on the study conducted and the results obtained, the following comments can be made:

- To advance in the opening of data it is necessary to consider the criteria for selecting data for opening provided by the Undersecretary of Public Innovation and Open Government of the Ministry of Modernization of the Nation in the "Open Data Kit".
- Note that data openness focuses on non-personal data that are not subject to specific restrictions.
- To give greater impetus to the use of big data in the field of FTyCA.
- To train human capital for the application of big data technology in the academic unit.

Open data is important for the university because it encourages greater transparency in the information it produces and efficiency in its actions. For society,

these open data are important because they provide information about the different activities carried out in the university environment as well as the careers taught, the number of students, the number of graduates, research, etc.. Likewise, they allow to know the current problems such as student desertion and which are the actions that are being carried out to solve them.

It would also be very beneficial for the Faculty of Technology and Applied Sciences to take advantage of the benefits offered by Big Data to move from the massive collection of data and the mere generation of statistics, to the positive and integrative use of educational information to achieve greater quality and management of the education system, and achieve academic excellence; To this end, the curricula can be adapted with the knowledge of the strengths and weaknesses of the students, through the statistics produced, in order to reduce dropout and dropout rates; to bring students closer to employment; and to produce research and contributions to the local public and private sector according to the needs.

References

1. Gobierno de Chile, Unidad de Modernización y Gobierno Digital, Minsegpres. Norma Técnica para Publicación de Datos Abiertos en Chile. 2013.
2. Straface Fernando. Big data y datos abiertos: cómo crean valor gobiernos y empresas. Cronista, 2019. https://www.cronista.com/columnistas/Big-data-y-datos-abiertos-como-crean-valor-gobiernos-y-empresas-20190310-0026.html#:~:text=Es%20decir%2C%20mientras%20que%20big,data%22%20para%20generar%20un%20impacto. Visitado última vez 14 de abril de 2022.
3. Heredia Mayorga, H. Big Data en la educación. Artículo de Investigación. Horizontes. Revista de Investigación en Ciencias de la Educación. Volumen 4 / No. 16. ISSN: 2616 - 7964. ISSN-L: 2616 - 7964. pp. 523 - 535. Octubre - Diciembre 2020. https://doi.org/10.33996/revistahorizontes.v4i16.134
4. Tejada-Escobar, F., Murrieta-Marcillo, R., Villao-Santos, F., Garzón-Balcázar, J. Big Data en la Educación: Beneficios e Impacto de la Analítica de Datos. Revista Científica y Tecnológica. UPSE. Vol. 5, Nº 2 (2018), 88-96 (Enero – Junio 2019).
5. AGESIC - Agencia para el Desarrollo del Gobierno de Gestión Electrónica y la Sociedad de la Información y el Conocimiento. Datos Abiertos de Gobierno. Guía Básica de Apertura y de Reutilización. Buenas Prácticas. Versión 1.0 – 2012.
6. Tim Berners-Lee. 5 Estrellas Datos Abiertos. https://5stardata.info/es/ visitado última vez el 14 de abril de 2022.
7. Matas, A., Leiva, J.J. & Nuñez, N. El Big Data desde la perspectiva de una muestra de estudiantes de universidades del norte de Perú. Propósitos y Representaciones, 8(1), e376. 2020.
8. Salazar Argonza, J. Big data en la Educación. Revista Digital Universitaria. UNAM. Vol. 17, Núm. 1. ISSN 1607 - 6079.2019.
9. Atenas, J., Raffaghelli, J. Los Datos Abiertos en la didáctica universitaria. Herramientas para el desarrollo del pensamiento crítico. Workshop Red Argentina de Educación Abierta. 2020.
10. Rivas-Rebaque B., Gértrudix-Barrio F., Cisneros de Britto J.C. La percepción del docente universitario ante el uso y valor de los datos abiertos. Educación XX1, vol. 22, núm. 2, pp. 141-163. Universidad Nacional de Educación a Distancia. España. 2019.