

How can Plan Ceibal Land into the Age of Big Data?

Martina Bailón*, Mauro Carballo*, Cristóbal Cobo†, Soledad Magnone*,
Cecilia Marconi*, Matías Mateu* and Hernán Susunday*

* Plan Ceibal
† Fundación Ceibal

Montevideo, Uruguay

Emails: {mbailon, mcarballo, ccobo, smagnone, cmarconi, mmateu, hsunday}@ceibal.edu.uy

Abstract—In 2007, Plan Ceibal became the first nationwide ubiquitous educational computer program in the world based on the 1:1 model. It is one of the most important programs implemented by Uruguay’s Government to minimize digital divide and is based upon three pillars: equity, learning and technology. As of 2007, Plan Ceibal has covered all public schools, providing every student and teacher in kindergarten, primary and middle school with a laptop or tablet and internet access in the school. To date, Plan Ceibal has close to 700,000 beneficiaries, each with their own device. Since 2011, the Plan has focused on providing the learning community with a wide range of digital content to enhance the teaching and learning process, most notably Learning Management Systems, Mathematics Adaptive Platform, remote English teaching and an online library. Today, Plan Ceibal operates and integrates a large scale of databases fed by a number of management and educational activities. This abundance of data presents a great challenge and a large opportunity to exploit and transform mass data into rich information. The main goal of this article is to describe the most relevant data sources and present an ongoing data analysis research grounded by a case study. In addition, this paper suggests next steps required to implement a learning analytics strategy within Plan Ceibal. If well exploited, this evidence based data can be used to support and improve the current technology and learning educational policies.

Keywords—Plan Ceibal; Big Data; Learning Analytics.

I. INTRODUCTION

The amount of data in the world has exploded; the analysis of large data sets is expected to become a new platform for new business, underpinning new waves of productivity growth, innovation, and consumer surplus [1]. This rapidly increasing amount of information, due to the expansion of social computing and the Internet has been coined as Big Data, and this time period described as the age of Big Data [2]. It is expected that Big Data will have a significant impact in the field of education, the design of curricula and the study of general patterns of teaching and learning activities [3][4].

Learning analytics is established as a proper field of knowledge [5]. The use of predictive modeling and other advanced analytic techniques help target instructional, curricular and support resources, to enhance the achievement of specific learning goals [6]. In 2006, the Government of Uruguay deployed Plan Ceibal as its main strategy to introduce Information and Communication Technologies on a large scale across the entire education system [7][8]. Plan Ceibal has successfully provided a device to all teachers and students in public schools from kindergarten to middle school, as well as internet access in all educational centers and outdoor areas. As stated by Fullan et al. [9], the delivery of computers (in some cases

tablets) and internet connection represent the first stage of Plan Ceibal. From 2011, a second stage arrived with the introduction of a strong initiative to deliver digital content to teachers and students. The Plan has focused on providing the learning community with digital content to enhance and personalize the teaching and learning processes. More specifically Learning Management Systems, Mathematics Adaptive Platform, remote English teaching, online library with digital and media content, amongst many others. As of 2013, the third stage of Plan Ceibal is being overwhelmed by sustainability and quality aspects of its large-scale development and set of platforms, resources and services for educational community.

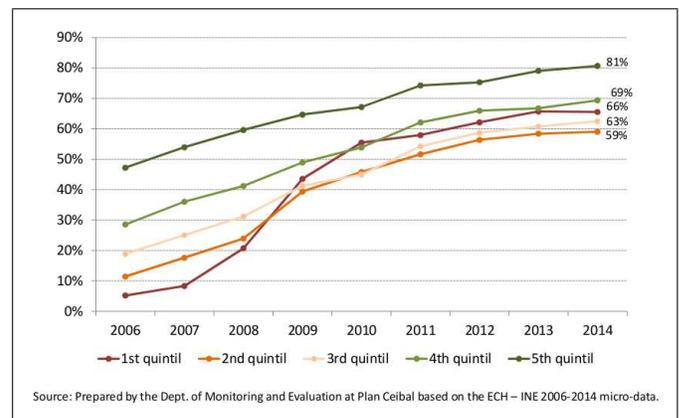


Figure 1. Personal computer access by quintile of per capita income. Whole country, percentage of households. Source: Monitoring and Evaluation Department, Plan Ceibal

The distribution of over 700,000 laptops and tablets, as well as the deployment of internet connections for schools and communities, facilitated higher levels of social inclusion. Moreover, it also provided equity via reduction of the ‘digital gap’ between the Information and Communication Technologies “haves” and “have nots” in all socio-economic contexts [10]; see Figure 1.

The decreased digital gap has enabled Plan Ceibal to focus on improving the quality of education by integrating technology in classrooms, schools and student’s households.

Plan Ceibal is currently pursuing a consistent strategy for the improved implementation of the Data Exploitation stage, outlined in the value chain model of Data Analytics in Figure 2. This stage is one of many Plan Ceibal is simultaneously working on.

This paper provides an initial exploratory review to understand the data sources available in the existing databases

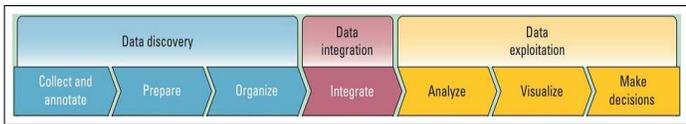


Figure 2. Value chain model of Data Analytics

in Section 2. In Section 3 the key questions to integrate and exploit these sources are presented. In Section 4 a case study that correlates the use of Adaptive Maths Platform (PAM) with infrastructure performance as well as social and demographical variables is described. Section 5 presents some preliminary results and Section 6 discusses the future associated to Learning Analytics and the Plan.

II. DATA SOURCES

Plan Ceibal is currently addressing many obstacles related to Big Data. Today, Plan Ceibal manages national scale databases that are necessary to pursue a number of management (deployment and management of devices and wifi network) and educational activities (teachers and students use of the tablet, laptops, platforms, websites, contents, etc.). The matrix in Table I shows some of the most important sources of information available as well as the volume, variety and frequency of data generation:

TABLE I. MATRIX OF THE MAIN DATA SOURCES

Source	Dimension	Frequency	Unit	Size
CRM	Socio-demographic Features	Annual	User ID	260.000 Primary students, 130.000 Secondary students
	Human and Physical Infrastructure	Annual	School ID	700.000 beneficiaries, 3000 endpoints
	Support Service	Hourly	User ID	3680 tickets a day
PAM, Adaptive Maths Platform	Performance	Daily	User ID	95.000 users in 2015
Tracker System Monitoring	Sys-Computer Usage	Daily	User ID	50 schools with 5000 users
Zabbix Monitoring System	IT Infrastructure Performance	Hourly	School ID	building 3000 facilities
CREA 2 LMS	Performance	Hourly	User ID	85.000 users in 2015

For instance, the *Tracker System for Statistically Monitoring Computer Usage* enables us to know what the most popular applications and activities are among students, how much time they spent on the computers, at what time of the day, and drive behavior patterns throughout the weeks. From this specific data source, we know that the main student activity is web browsing, followed by video camera applications, followed by drawing application is associated to 83% of students involved in Plan Ceibal’s initiative between 6-11 years of age, in the primary school system.

It is worth mentioning that in the customer relationship management (CRM) database, Plan Ceibal records the devices owned per user. The PAM database provides information regarding the intensity of its usage by students and teachers.

Zabbix database is the School’s Network monitoring system and CREA 2 is the learning management system supported by Ceibal in schools based on a social network philosophy.

The different data sources are vital to understand the impact and use of the laptops on the national scale[11].

Regarding the size of generated data, Table II shows the amount of daily information generated in some of the components of the matrix.

TABLE II. SIZE OF DAILY GENERATED DATA

Source	Size (Mega Bytes)
Zabbix	200
CRM	4
Tracker	6
PAM activity	10
Internet activity	150
Total	370

Regarding the structure of the different data sources, some of the most relevant issues today are:

- Lack of Integration: there is very limited interoperability among databases.
- Lack of a common processing and visualization framework: a unique interface for processing and visualization is necessary.
- Lack of traceability: the bulk is generated from different databases, providers, interfaces and the unit of analysis depends on the database (school, device, end user or classroom based).

Most of the variables and data generated through these databases are collected and processed for managerial and operational needs. Simultaneously, a business intelligence (BI) system is being built on top of all databases for management purposes.

III. KEY QUESTIONS

What are the key parameters, significant variables and required data sources to include in the “data integration” and “data exploitation” stages? These questions can be described as:

- How can we improve integration of the different data sources in a more comprehensive and meaningful way?
- How to enable interoperability and consistency between information and variables retrieved from different data sources (i.e: the unit of analysis in some cases are schools, classrooms or individual based information)?
- What are the more reliable analytical techniques to identify strong correlations amongst key variables?
- How can the integration of the different data sources be applied to better understand ways of improving institutional and pedagogical strategies?

IV. CASE STUDY

A. Motivation

The primary objective of Plan Ceibal is the use of the technological development to boost pedagogic tools. In 2012, Ceibal took a step forward by acquiring an adaptive platform for mathematics (in Spanish PAM) that allowed the beneficiaries to exercise and learn algebra, geometry, etc. That platform allowed teachers to monitor and control the classroom’s performance in PAM.

To improve the conditions for the use of PAM, Plan Ceibal made a strategic decision towards deploying an appropriate connectivity network and the acquisition of more modern devices. In order to provide students with high performance experiences via the new equipments a High Performance Network (HPN) is being deployed since 2014 in every urban school.

B. Methodology

The first hypothesis driving the investigation is:

- *The more powerful network infrastructure will facilitate a higher amount of exercises completed by students in PAM.*

The second hypothesis is:

- *The social-demographic features (metropolitan vs. interior urban, and socio-cultural context) affect the use of PAM.*

C. Research Questions

- 1) To what extent does network performance correlate with PAM use? And what are the most reliable techniques and data sources to explore this correlation?
- 2) To what extent does the social-demographic features mediate and moderate the relationships between High Performance Networks and PAM intensity of use?

D. Universe and sample

The universe are schools and students belonging to those schools that had its High Performance Network deployed during 2014. That represents 100 schools with 13800 students from 4th to 6th level.

From the schools a **random** and **stratified** by socio-demographic context sample is taken: **18 schools** with **3,823 students**.

The time frame will be data collected in PAM for the 2014 school year.

V. PRELIMINARY RESULTS

Regarding first hypothesis, an increase of 35.6% active PAM users has been detected between t_0 and t_1 , being t_0 the period of time in 2014 the school had its initial wifi/internet access and t_1 the period of 2014 in which the school had its HPN installed. This is illustrated in Table III

Regarding the second hypothesis, two analyses have been made:

- 1) Comparison of average of usage between t_0 and t_1 by **location**, i.e, interior urban (IU) vs. Montevideo metropolitan Area (MVD) shown in Figure 3.

TABLE III. NUMBER OF PAM ACTIVE USERS BEFORE AND AFTER HPN DEPLOYMENT.

	Before HPN	After HPN
# PAM active users	806	1093
# activities	53179	67523

- 2) Comparison of average of usage between t_0 and t_1 by **socio-cultural context**, i.e, favorable vs. very unfavorable contexts(data not shown).

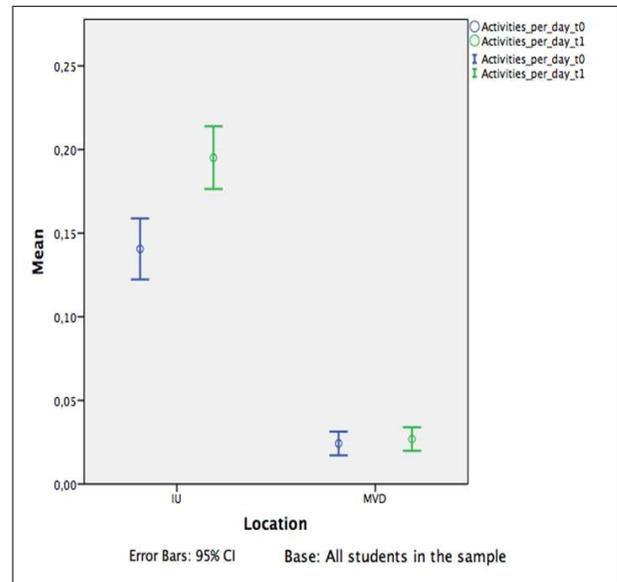


Figure 3. Intensity of PAM use vs. location

The comparison by location highlights that interior urban usage is one order of magnitude bigger than metropolitan use. On the other hand, there is a significant difference between t_0 and t_1 in the urban areas of the provinces whereas in Montevideo there is no significant variation. This can be seen in Figure 3.

The comparison by social-cultural context demonstrates first that unfavorable contexts have approximately one order of magnitude lower intensity of use. Secondly, when compared t_0 and t_1 there is statistical significance in unfavorable contexts only.

VI. DISCUSSION AND FURTHER RESEARCH

From a general perspective, there is a clear need for a more comprehensive information cartography of all the available information resources, as well as a better understanding of how the integration of different data sets can help Plan Ceibal to create better learning experiences, services, tools and public goods specialized in education. The main questions that arise for further research are:

- How to recognize the critical indicators that are more strongly correlated with student’s performance?
- How to improve the data traceability of beneficiaries, as well as a better understanding of how the different technologies are being used?

- Will these new forms of tracking students behavior transform how we study the relationship between learning and human-computer interaction, or narrow the palette of research options and alter what ‘#edtech research’ means?
- Will the deployment of these analytic techniques usher in a new wave of privacy incursions and invasive research initiatives? [12]
- How to provide channels and platforms that provide effective feedback for the beneficiaries, i.e., teachers, students, parents. etc.?
- What are the possible strategies to transit from holistic and statistical approaches, like location and socio-cultural context, to a student’s level approach?

Given the rise of Big Data as a socio-technical phenomenon, we argue that it is necessary to reflect on these matters and consolidate good practices in the emerging field of learning analytics in order to monitor the influence of technologies in the education ecosystem.

The ongoing research shows some preliminary results about correlation between High Network Performance and PAM intensity of use as well as correlation to independent variables like location and socio-cultural contexts.

The next steps will delve deeper into the challenges faced by Plan Ceibal during it’s process of transforming and exploiting the vast amounts of data generated by the organization. We highlight once again, the development of technical and institutional capabilities will allow Plan Ceibal to provide an improved framework towards learning analytics, evidence-based decision making and personalized learning.

ACKNOWLEDGMENTS

The authors would like to thank Daniel Castelo, Claudia Brovetto, Leonardo Castelluccio, Fiorella Haim and Juan Pablo González for their valuable comments. They also want to thank Plan Ceibal for the support of this initiative.

REFERENCES

- [1] J. Manyika et al., “Big data: The next frontier for innovation, competition, and productivity,” McKinsey Global Institute, Tech. Rep., 2011.
- [2] S. Lohr, The age of big data. New York Times., 2012, no. 11.
- [3] Sharples, M. et al., “Learning design informed by analytics,” Open University, Tech. Rep., 2014.
- [4] “7 Cool Things to Help You Understand Big Data for Education,” 2015, URL: <http://tabletsforschools.org.uk/7-cool-things-to-help-you-understand-big-data-for-education/> [accessed: 2015-06-04].
- [5] R. Ferguson, “Learning analytics: drivers, developments and challenges,” International Journal of Technology Enhanced Learning, vol. 4(5), 2012, pp. 304–317.
- [6] C. Bach, Learning analytics: Targeting instruction, curricula and student support. Office of the Provost, Drexel University., 2010.
- [7] Plan Ceibal and ANEP, Plan Ceibal in Uruguay. From Pedagogical reality to an ICT roadmap for the future. UNESCO, 2011.
- [8] E. Severin and C. Capota, 1 to 1 Models in Latin America and the Caribe, Panorama and Perspectives. BID, 2011.
- [9] M. Fullan, N. Watson, and S. Anderson, Ceibal: Next steps. Michael Fullan Enterprises, 2013.
- [10] A. L. Rivoir, Ed., Plan Ceibal and Social Inclusion. Interdisciplinary perspectives. Plan Ceibal and University of the Republic, 2011.
- [11] N. S. I. INE, “Uruguay’s national census, 2011,” 2015, URL: <http://www.ine.gub.uy/censos2011/index.html/> [accessed: 2015-06-19].

- [12] D. Boyd and K. Crawford, “Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon,” Information, communication and society, vol. 5, no. 15, 2012, pp. 662–679.