An experience of Brazilian educational graduate platform

Abstract – Brazil has a vigorous program for graduate studies with more than 6.400 courses of various knowledge areas all over the country. These courses have to send annually its data of professors, students, dissertations, projects, etc. to Brazilian Federal Agency for Support and Evaluation of Graduate Education (CAPES), who is responsible for collecting and processing data through Sucupira platform. Nowadays, CAPES is in the process of developing and implementing a Current Research Information System (CRIS) for improve the quality of data received and, for consequence, available.

Index Terms - Graduate, Evaluation, Sucupira platform, CRIS.

1. INTRODUCTION

The Brazilian Program for Graduate Studies (PNPG) was officially organized at the end of 1960s, formalizing existing initiatives already going on in universities and research centers (Guimarães & Humann, 1989).

PNPG was designed as a route for accelerating the formation of human resources suitable to supply the urgent need for qualified personnel capable of improving the quality of teaching and strengthening the research activity at universities and other institutions. At the same time, it was expected that the PNPG would contribute to the technological development of the country by supplying it with well-trained scientists and technologists, thus meeting the needs of both public and private industrial sectors and research centers for human resources dedicated to R&D (Guimarães & Humann, 1989).

Since then, Brazilian graduation improved in training of university professors; integration of research developed concern with performance and quality; improvement of the evaluation system; combating the asymmetries and the impact of graduate activities in the productive sector and in the society, resulting in the incorporation of innovation in the SNPG and in the inclusion of societal parameters (CAPES, 2010).

In last four-early evaluation process (2013-2016), CAPES, public agency responsible for the periodic evaluations of Brazilian graduation programs, assigning a 7-point scale score based on criteria such as publishing in high-prestige journals, social impact, producing patents and international exchange in research (Neto, Willinsky, & Alperin, 2016), evaluated 4.175 programs from 49 Areas of Evaluation[1] over the country. Tables 1 and 2 shows some overall numbers.

Table 1
Evolution of number of graduate degrees in Brazil between evaluations

Degrees	2013-2016	% Growth
Master	187.971	59
Doctoral	72.454	94
Professional Master	32.513	193
Total	292.938	76

Table2
Evolution of scientific production in graduate in Brazil between evaluations

Publications	2013-2016	% Growth
Articles	846.981	89
Books	303.457	80
Technical	1.601.780	82
Total	2.752.218	84

[1] The scientific division are categorized in Colleges, Great Areas and Areas of Evaluation. The lists can be accessed in: http://www.capes.gov.br/avaliacao/sobre-as-areas-de-avaliacao

In 5th version (2011 – 2020) of PNPG, Brazil and CAPES have to keep improving in some topics as internationalization and development/using of new indicators for evaluation. Then, it was urgent to build a single integrated system who allow usability, integration, transparency, comparison and big data processing. The platform is called Sucupira in honor of Professor Newton Sucupira (1920-2007), who legally defined graduation in Brazil (Boaventura, 2009).

Therefore, this article aims to present the actual stage of development of Sucupira platform system, which was developed to solve these problems. Also, to introduce planning of new improvements for become a CRIS system.

2. METHODS

This technical paper used some qualitative methods such as documentary analysis of guides and internal reports related to Sucupira platform.

The object of this study is Sucupira platform, which is an online platform developed in Java programming language (version 7), running over a container JBOSS 6.1 EAP. There are 8 machines in the infrastructure park controlled by two internal balancers.

Among other technologies, it uses EJB 3, JSF 2 e Hibernate for communication with database. Data was saved in Oracle Database 11g, having two balancing instances.

Data and analysis were made with SAS Enterprise Guide 7.1.

3. DISCUSSION

Phases of development of Sucupira platform are presented below in chronological order.

I. Phase 1 - Implementation

The phase 1 of new system development occurred between 2013 and 2016, at same time of graduate four-early evaluation cycle.

This first cycle of development intentioned to aggregate many other electronic educational systems from CAPES such as Coleta, WebQualis, APCN, SNPG, etc. in just one, that would be able to collect all data from teaching and research of graduation programs in universities and other institutions, and also have tools for improve analyses. Figure 1 shows schemes and standards of platform developed in Java's programming language with Oracle database:

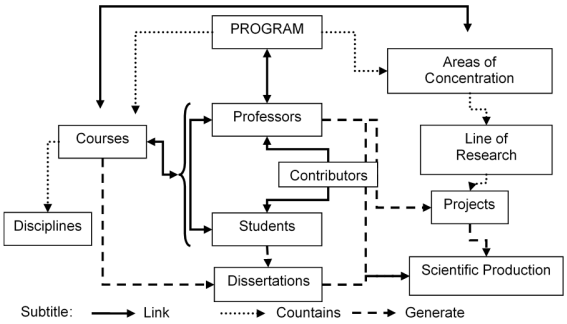

Figure 1. Identification schemes and standards of Sucupira platform. Reprinted from *Sucupira Coleta's Guide*, version 3.5.3, 2018. Retrived from https://sucupira.capes.gov.br/

Looking at schemes, it is possible to notice that Sucupira platform, based on Brazilian graduate model, has levels of hierarchy, where graduation program are the highest and courses, professors, students, projects, scientific production, etc. are above. Program can belong to one or more institutions/universities and institutions/universities can have graduation programs of different Areas of Evaluation. For example, University of Brasilia (UnB) has graduation programs of administration, engineering, law, etc.

This model assumption previous internal and external interoperability, with exchange and reuse of research information from existing resources, like with CV Lattes and CAPES scholarship payment systems, which already occurs.

Besides that, all connections and integrations provide many data, as showed in tables 1 and 2. And analysis of big data is made with SAS Entreprise Guide 7.1 and flag panels with SAS Visual Analytics.

II. Phase 2 – CRIS

The current phase 2 of development intends to enlarge Sucupira's model to become a current research information system - CRIS. According to Bevan and Harrington (2011), a CRIS is intended to ensure a single and accurate source of research information for effective planning, monitoring, reporting and communication.

In addition to that, Wilkinson et al. (2016) defines that good data management is not a goal in itself, but

rather is the key conduit leading to knowledge discovery and innovation, and to subsequent data and knowledge integration and reuse by the community after the data publication process.

With this in mind, it was necessary to project a new concept improving connections of Sucupira platform with other systems (Figure 2).

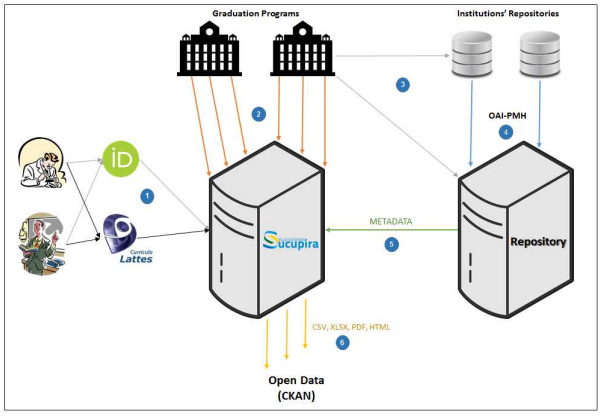

Figure 2. CRIS model.

Explaining dots 1 to 6:

1) Nowadays, professors fulfill their CV Lattes (Brazilian national curriculum system) and Sucupira's collects them through interoperability with an algorithm to identify second, third, etc. authors from metadata production imported. Table 3 shows big data related to this situation.

Table 3
Number of authorships in scientific productions

Year	Number of authorships
2013	2,312,001
2014	2,454,076
2015	2,487,750
2016	2,469,929
Total	9,723,756

ORCid presents a new possibility of data collection and recognition of authors, and is available through Brazilian consortium of institutions (CAPES, CNPq, IBICT, CONFAPs and Scielo);

2) At same time than process 1, graduation programs fill their data in Sucupira. It already occur with some interoperability with, for example, Brazilian Revenue and Post Office systems. So far, adoption of CRIS intend to reduce manual entrance of data;
3) Instead of graduation programs send files and metadata to institutions repositories after fulfill Sucupira, as already happen, they will do it before, including in Repository;
4) Repository of graduation products for evaluation will interoperate with repositories from universities and other organizations, receiving not only data from dissertations and thesis, but also from books, chapters, technical production, etc. in a DSpace technology;
5) In opposite direction of current flow, after receiving metadata from institutions repositories, these data will be listed in Sucupira platform to link with and for correct registration to graduation program;
6) After Sucupira platform collects data from different inputs, there is a process of analyzing and checking (not cleaning) data, providing FAIR Data - Findability, Accessibility, Interoperability, and Reusability in CKAN-CAPES: http://dadosabertos.capes.gov.br

In addition to point 6, metadata from 987.255 dissertations and thesis since 1987, and tens of thousands of data from graduation programs, courses, professors, students, scientific production, etc. since 2004, therefore from prior electronic educational systems to Sucupira, are being analyzed and checked and some of them already are available in CKAN-CAPES.

All these steps are fundamental to evolve from a data management to an information management system (CRIS), with guarding the quality and integrity of datasets, impacting dissemination and knowledge transfer from transparency.

4. CONCLUSION

In summary, it is possible to separate two stages of Sucupira's development: first from 2013 to 2016 and second from 2017 onwards, both coinciding with four-early evaluation period. First phase was completed successfully, incorporating business models of several other systems and presenting new challenges, as expansion of interoperability and management of big data.

In current phase 2, apart from new model of Figure 2, it will be necessary to adapt CERIF theoretical model to consider graduation programs as a unit.

Also, there are some Brazilians scientific projects that don't receive external funding, so coding will have to consider these cases.

In doing so, current effort of evolution Sucupira's model as a CRIS system will provide public view and transparency of data as a accountability of Brazilian public investments and a better way to help in finding a course for graduation studies and/or improving research.

5. REFERENCES

Bevan, S. J. & Harrington, J. (2011). Managing research publications: lessons learned from the implementation of a Current Research Information System. *Serials*. 24(1), 26–30. DOI: http://doi.org/10.1629/2426.

Boaventura, E. M. (2009). Newton Sucupira, a pós-graduação e a universidade. *A construção da universidade baiana: objetivos, missões e afrodescendência*. Salvador: EDUFBA.

Coordenação de Aperfeiçoamento de Nível Superior. (2010). *Plano Nacional de Pós-Graduação (PNPG) 2011-2020 Volume I*. Retrived in February 19, 2018, in http://www.capes.gov.br/images/stories/download/Livros-PNPG-Volume-I-Mont.pdf

Guimarães, J. A., & Humann, M. C. (1995). Training of human resources in science and technology in Brazil: The importance of a vigorous post-graduate program and its impact on the development of the country. *Scientometrics*, 34(1), 101-119.

Neto, S. C., Willinsky, J., & Alperin, J. P. (2016). Measuring, Rating, Supporting, and Strengthening Open Access Scholarly Publishing in Brazil. *Education Policy Analysis Archives*, 24 (54), 1-20.

Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.W., Santos, L. B. S., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., Gonzalez-Beltran, A., Gray, A. J. G., Groth, P., Goble, C., Grethe, J. S., Heringa, J., C 't Hoen, P. A., Hooft, R., Kuhn, T., Kok, R., Kok, J., Lusher, S. J., Martone, M. E., Mons, A., Packer, A. L., Persson, B., Rocca-Serra, P., Roos, M., van Schaik, R., Sansone, S. A., Schultes, E., Sengstag, T., Slater, T., Strawn, G., Swertz, M. A., Thompson, M., van der Lei, J., van Mulligen, E., Velterop, J., Waagmeester, A., Wittenburg, P., Wolstencroft, K., Zhao, J. & Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientifi Data*, 1-9.