

Ana M B Pavani, Member IEEE

apavani@puc-rio.br

Pontifícia Universidade Católica do Rio de Janeiro
Rio de Janeiro, Brazil

Abstract

This work has the objective of examining Brazilian ETD programs in terms of how they participate in the international scenario and also of their adoption of current trends – research data available with the works (that used them), alternative media files along with conventional text and the use of social media for the repositories. An overview of the ETDs programs is presented in the introduction; it has the objective of presenting the differences among them. Only ETD Programs that contribute to the Brazilian national consortium are considered.

Keywords: **open access; institutional repositories; research data**

01. Introduction – Some Facts and Some Numbers

ETD programs are well established in Brazil. BDTD – Biblioteca Digital de Teses e Dissertações (<http://bdttd.ibict.br/>), the Brazilian national consortium, was created 15 years ago. Three universities were involved with the initial project in 2001 – two of them in the Southeastern part of the country and the third in the South. These two regions are the most developed of Brazil. These three institutions still have their ETD programs and were joined by over 100 others. The institutions are located in all regions of the country; they are public, private and communitarian universities. Some are large and traditional, and others are small and new.

Currently, there are 105 member institutions and the number of metadata records hosted by the union catalog is almost 424K (approximately 70% in the M level); data available on Jul 01-03, 2016. It is important to observe that there are institutions that have ETDs but are not members of the consortium; these are not considered in this work.

The programs are also quite different from one another. One obvious difference is the numbers of ETDs they have. Table 1 shows the numbers of programs counted by numbers of ETDs on their repositories. Observe that only 90 contribute with metadata records to BDTD. It clearly indicates that the number of institutions with over 5,000 ETDs is 19 (approximately 21.1%) and the number of institutions with under 1,000 is 39 (approximately 32.2%). Figure 1 presents the same data in a histogram.

Numbers of ETDs	Numbers of Programs
> 60,000	1
40,001 – 50,000	1
30,001 – 40,000	1
20,001 – 30,000	2
10,001 – 20,000	5
5,001 – 10,000	9
3,001 – 5,000	10
2,001 – 3,000	5
1,001 – 2,000	17
501 – 1,000	9
101 – 500	22
≤ 100	8
Total	90

Table 1 – Numbers of ETD programs counted by the numbers of ETDs they offer.

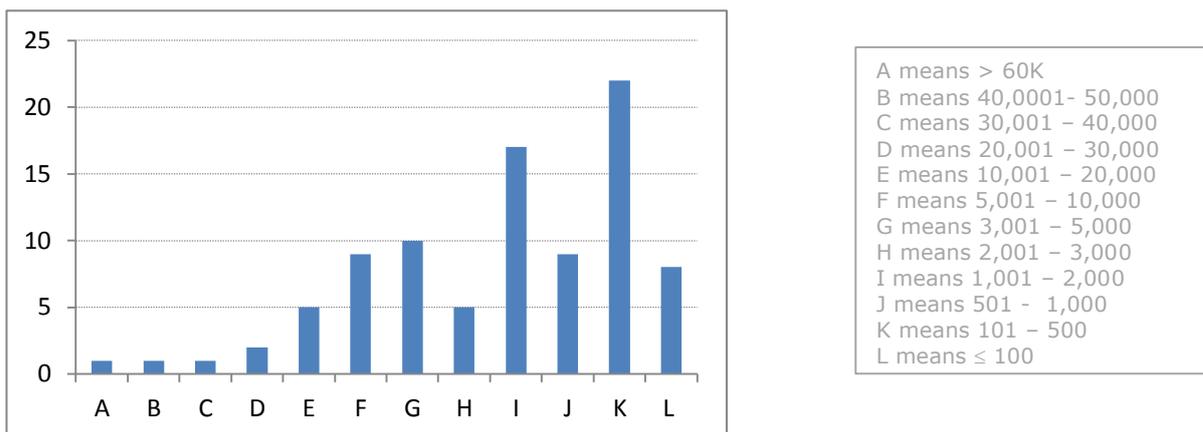


Figure 1 – Numbers of ETD programs counted by the numbers of ETDs they offer.

McMillan [01] gathered data through a survey that was run with 298 institutions in May/June 2015. The results showed that 83% of the institutions have fewer than 10,000 ETDs. In Brazil, as indicated in table 1, 88.89% of the institutions have fewer than 10,000 ETDs. The histogram in slide 4 of McMillan’s presentation shows that highest frequency happens for repositories with more than 1,000 and less than 5,000 ETDs. In the Brazilian case, this occurs for repositories with 101 to 500 ETDs. So data indicates that the profile of ETD programs that contribute to BDTD is different from the profile of the institutions surveyed by McMillan.

This work addresses the similarities and the differences among repositories; in some aspects a special consideration will be given to the 19 that host more than 5,000 ETDs. They are located in different regions of the country but there is a high concentration in the SE (42% of the Brazilian population lives in this region). The percentages of the population in the NE and the SO are, respectively, 14.3% and 26.0%. Table 2 shows the numbers by region. Population data are estimates for 2016 [02].

Region	> 5,000 Programs	%	All Programs	%
SE	10	52.63	42	46.67
SO	5	26.32	19	21.12
NE	3	15.79	17	18.89
Federal District	1	15.26	4	4.44
NO	0	0.00	4	4.44
CW(*)	0	0.00	4	4.44
Total	19	100.00	90	100.00

Table 2 – Numbers of ETD programs by region of the country.

(*) CW is the Central Western region of Brazil.

It is interesting to observe that while only approximately 68% of the programs are in the SE and SO regions, the percentage of programs with over 5,000 ETDs in the same regions is almost 80. Another important characteristic is the concentration of records in the top 19 institutions (20%) – they hold almost 334 K of the almost 424 K records available on the union catalog; this represents almost 79%.

The next sections of this work address some characteristics of the ETD programs with special emphasis on the largest 19. The characteristics are on the repositories and how they are organized, and of their visibility in the international scenario. Information and data were gathered from BDTD, the ETD programs websites and international organizations that foster Open Access activities. It is important to remark that data obtained from BDTD were not reliable since they conflicted with corresponding data on the institutions systems.

02. Main Characteristics of the Repositories

This section is divided in subsections each one devoted to one aspect of the analysis of the ETD programs.

a. ETD Systems x IRs

One of the differences among programs is on the way institutions make they graduate research results available to the public. Currently, many have Institutional Repositories; some have one repository for both (ETDs and other works) while other have separate systems. At the same time many still use a system called TEDE – Sistema de Publicação Eletrônica de Teses e Dissertações that is customized to Brazilian ETDs.

TEDE is based on ETD-db developed by Virginia Tech. A new version of TEDE, called TEDE 2, was developed based on DSpace and institutions are migrating their collections to the new version. It is suitable for ETDs only. Both versions were developed by IBICT – Instituto Brasileiro de Informação em Ciência e Tecnologia (<http://www.ibict.br/>), an organization of the Brazilian Federal Government that runs BDTD. Along with the products, a solution to migrate from one to the other and coaching on how to do it were offered. In this work TEDE2 is referred as IR-ETD. IBICT also customized DSpace for a wider use – ETDs and other types of resources.

Examination of the solutions used by the 90 programs yielded the results shown in table 3.

Types	Numbers of Programs	Percentage
IR	36	40.91
IR-ETD	27	30.68
TEDE ^(*)	25	28.41
Total	88^(**)	100.00

Table 3 – Numbers of ETD programs by the type of system they use.

(*) Means there is another system based on ETD-db counted along with TEDE.

(**) Indicates that total number in table 3 is 88 due to the fact that one system is an OPAC and another was down and could not be examined.

When a similar analysis is performed for the largest 19 programs, the percentages are 68.42 (IR), 10.53 (IR-ETD) and 21.05 (TEDE). Among the largest programs the percentage of IRs is higher. It is expected that other TEDE users will migrate to IR-ETD. BDTD and TEDE had a very important role in the development of digital libraries and digital publishing in Brazil – they introduced the culture in most Brazilian institutions and probably this next step, IRs, will be stimulated by BDTD and the two customizations of DSpace.

b. Research Data

The theme of this symposium is “Data and Dissertations, with a focus on the handling of data produced by PhD students”. NDLTD – Networked Digital Library of Theses and Dissertations (<http://www.ndltd.org/>) has been awarding innovative ETDs for many years. In 2006, one of the Innovative ETD Awarded was granted to Evan Dorn for his PhD Dissertation at the California Institute of Technology; he added research data to his work (<http://www.ndltd.org/a/ndltd.org/ndltd/ndltd-awards/innovative-etd-awards/awards/evandorn>).

Degkwitz [03] classified texts with added data as “enriched publications” and considered that data can have different types of contents and be in different digital formats.

The topic is so important that Lippincott [04] addressed it in her speech, entitled E-Research: the Role of Data Sets in ETDs, during ETD 2015 in India.

Considering the international scenario some examples of research data repositories consortia maybe cited: (1) ANDS – Australian National Data Service (<http://www.ands.org.au/>) – is an Australian national initiative or research data projects; (2) DataONE – Data Observation Network for Earth (<https://www.dataone.org/>) – is an international initiative “providing access to data across multiple member repositories, supporting enhanced search and discovery of Earth and environmental data”; and (3) re3data (<http://www.re3data.org/>) – a registry of research data repositories with over 1,500 entries.

A search on BDTD identified two Brazilian institutions that make research data available from their repositories, though data sets do not have be related to ETDs. A third institution, a member of both re3data and DataONE, has three separate repositories – one for research data, one for ETDs and the third for articles, books, etc. It is possible that other institutions make research data available too, but the systems they use do not make this information easy to be found.

Unfortunately, McMillan’s survey did not ask about research data associated with ETDs and how this data are managed. So, there is no way to compare with international institutions.

c. OpenDOAR, ROAR and SHARE

Open Access and Open Access repositories are quite popular in Brazil. This is a fact that can be observed by the number of Brazilian repositories that are listed in international directories.

OpenDOAR – The Directory of Open Access Repositories (<http://www.opendoar.org/>), ROAR – Registry of Open Access Repositories (<http://roar.eprints.org/>) and SHARE (<http://www.share-research.org/>) are well known nongovernmental organizations devoted to make open research and other resources discoverable, accessible and shared. They host directories of Open Access repositories.

The examination of the three directories searching for the 90 Brazilian ETD programs yielded the results shown in table 4.

Directory	Numbers of Programs	Percentage
OpenDOAR	40	44.44
ROAR	33	36.67
SHARE	1	1.11

Table 4 – Numbers of ETD programs registered on OpenDOAR, ROAR and SHARE.

It is important to remark that ROAR had a serious system problem many months ago and there is a probability its data are not reliable.

Considering the 90 programs, SHARE (<http://www.openarchives.org/>) is obviously the least used. SHARE harvests metadata records using the OAI-PMH – Open Archives Initiative Protocol for Metadata Harvesting. The total number of metadata records on the repositories that contribute to SHARE is over 7.6 M. Repositories are from different countries in the world but most of them are in the United States.

OpenDOAR and ROAR have many more Brazilian repositories but they are not members of BDTD. Many of them are do nor belong to universities or other institutions of higher education but are of government agencies (libraries anr archives) that publish their collections in Open Access.

d. Webometrics

Twice a year, the Spanish Cybermetrics Lab publishes the Ranking of Web Repositories (<http://repositories.webometrics.info/en>). The last ranking, dated January 2016, classified 2,205 institutional repositories from all over the world. Among the 90 ETD programs, 25 were ranked – three were among the top 10 and 9 were among the top 500 (top 25%).

e. NDLTD

“The Networked Digital Library of Theses and Dissertations (NDLTD) is an international organization dedicated to promoting the adoption, creation, use, dissemination, and preservation of electronic theses and dissertations (ETDs). We support electronic publishing and open access to scholarship in order to enhance the sharing of knowledge worldwide. Our website includes resources for university administrators, librarians, faculty, students, and the general public. Topics include how to find, create, and preserve ETDs; how to set up an ETD program; legal and technical questions; and the latest news and research in the ETD community.” (<http://www.ndltd.org/about>)

NDLTD runs an international union catalog of metadata records (<http://search.ndltd.org/>) with over 4.4 million records from all over the world and in more than 10 languages. An institution does not have to be a member to contribute to the union catalog.

Among more than one hundred NDLTD members, only two are Brazilian. BDTD is not a member.

f. Statistics

One of the hardest problems to get information and data about ETD programs is the lack of a minimum set of common statistics. Institutions organize their repositories in many different ways, according to their organizational structures and cultures. This is expected.

The problem is that there is no set of minimum information to external users. One of the problems faced during data gathering for this work was the identification of the numbers of ETDs in each program and the breakdown between M and D levels. Some programs that use DSpace organize their collections using the Graduate Programs as the communities – this requires that many searches be performed and most results do not separate M and D. Even if this organization is used, the settings of the software allow to count by document type; it is just that this is not used or maybe displayed.

g. Social Media

In ETD 2015, works related to current trends were presented. An interesting work, addressing social media and its impact in promoting scholarly communications, in general, and ETDs, in particular was presented by Greenberg, Narang and Sengupta [05]. They explore the idea of introducing AltMetrics (<http://altmetrics.org/manifesto/>) usage to ETDs and analyze worldwide solutions that host collections.

Due to the importance of assing the impact of ETDs, the programs were examined in terms of using social media. Table 5 shows the results when the 90 programs are considered.

Networks	Numbers of Programs
FaceBook	14 ^(*)
Twitter	13 ^(**)
LinkIn	2
Google+	6

Table 5 – Numbers of ETD programs using different social media networks.

(*) Seven required the user to be logged to an account and one was empty.

(**) Four required the user to be logged to an account.

Many programs use more than one network. In some cases, the accounts belong to the library and not only to the ETD programs.

An interesting remark is that BDTD is not present iin any social network.

03. Final Comments

This initial analysis of Brazilian ETD programs and how they relate to the trends in the international scenario indicates that there is an aligment, even though it is not widespread over all members of the national consortium.

In Brazil there is a strong presence of the Federal Government in all aspects of society and ETDs are no different. BDTD is run by an Federal Government Institution. Maybe, this fact can be used in favor of Open Access and dissemination – suggest to the managers of the consortium that they inform and stimulate institutions to make Research Data available, register repositories on directories, use social media and enhance information about their collections on the websites of the repositories.

04. References

- [01] McMillan, Gail
Digital Knowledge Repositories: What the ETD Survey Reveals
 Presentation in the DELNET (Developing Library Network) Panel Discussion Digital Knowledge Repositories: Emerging Issues, Trends and Technologies" at the India International Centre in New Dehli, India, 2015
 Available <https://vtechworks.lib.vt.edu/handle/10919/71317>
 Accessed on June 19, 2016
- [02] IBGE – Instituto Brasileiro de Geografia e Estatística
População – Projeção da População do Brasil e das Unidades da Federação

<http://www.ibge.gov.br/apps/populacao/projecao/>

Accessed on June 19, 2016

- [03] Degkwitz, Andreas
Enhanced Publications Enhance the Potential of Digital Media
Proceedings of ETD 2015 – 18th International Symposium on Electronics Theses and
Dissertations, pp. 51-59, India, 2015
- [04] Lippincott, Joan
E-Research: The Role of Data Sets in ETDs
Abstract of the speech
Proceedings of ETD 2015 – 18th International Symposium on Electronics Theses and
Dissertations, p. 530, India, 2015
- [05] Greenberg, Charles J., Narang, Sageeta and Sengupta, Sebanti
Exploring Altmetrics for ETDs
Proceedings of ETD 2015 – 18th International Symposium on Electronics Theses and
Dissertations, pp. 77-84, India, 2015