

## Scientific Data Management on a Dataverse Network at IBICT

**Authors:** Miguel Ángel Márdero Arellano and Alexandre Faria De Oliveria, (Coordinator & Technological Solutions Coordinator, Brazilian Network of Digital Preservation Services, CARINIANA, IBICT<sup>1</sup>)  
**Email:** miguel@ibict.br / alexandreoliveira@ibict.br

### 15.1. INTRODUCTION

In the 20th century, at the time when the State was the main driving force for the management of information and documentation resources and services, the creation of intermediary information systems in Brazil became a matter of strategic importance. As a result of this, it was necessary to construct a scientific technical infrastructure, as well as to train qualified personnel in the management of the production of, access to and preservation of information in science and technology. Brazil's government conferred upon the Brazilian Institute of Information in Science and Technology (IBICT) the responsibility to become the standard-bearer for core competencies in the process of treatment of, access to, and dissemination of information.

The Cariniana is a distributed preservation network, funded by IBICT, committed to national and international cooperation, promoting the management and dissemination of digital preservation practices and developing a sustainable digital preservation programme to support Brazilian universities and research centers' needs and requirements. In 2012 IBICT recognized the need to address digital preservation issues, and it adopted the LOCKSS (Lots of Copies Keeps Stuff Safe<sup>2</sup>) approach as suitable for the needs of the Cariniana network. Its main focus concerns open access publications in Brazil. The network preserves journals, and doctoral theses, and it is just starting to cover scientific data to be deposited in a research data repository. The experimental phase, using LOCKSS open source, covered a year in 2013, and was supervised by LOCKSS staff from the University of Stanford.

In 2015, the implementation of Cariniana's Dataverse<sup>3</sup> repository added significant new services to the digital preservation network which will help specialized libraries' staff to deal with the demand from researchers for a trusted space for their datasets. IBICT is making available a repository for research data that is responsible for long term preservation and good archival practices, while researchers can share, keep control of, and receive recognition for their data. In addition, the repository supports the sharing of research data with persistent data citation and enables reproducible research.

### 15.2. WHAT MOTIVATED THE REPOSITORY OF SCIENTIFIC DATA AT THE CARINIANA NETWORK?

The Cariniana Network resulted from the need to create a digital preservation service of Brazilian electronic documents to ensure continuous access to these documents throughout time. The creation of the project for the preservation of research data was based on the idea that the more copies of a document that are stored in different places, the safer they will be. First, a centralized storage structure is used; then the content goes through distributed computer resources, with the participation of institutions that support electronic documents.

<sup>1</sup> IBICT: <http://www.ibict.br/>; accessed 5 February 2017.

<sup>2</sup> LOCKSS: <https://lockss.org/>; accessed 5 February 2017.

<sup>3</sup> IBICT Dataverse Network: <http://repositoriopesquisas.ibict.br/>; accessed 5 February 2017.

Initially, the activities were carried out jointly with the University of Brasilia. In the first phase, the Dataverse network is being used for the addition and storage of research documents from individuals, institutional projects, and electronic journals. After that, the possibility of integration with the LOCKSS platform of the partner institutions will be used as the preservation repository of the stored material. Offering digital preservation services includes integrating the scientific data content of the connected institutions into a unified pattern; these mechanisms must facilitate the automation of processes of identification, storage, validation, and conversion of the content into new digital formats.

IBICT started a pilot project in 2015 and one of its objectives is to be a valuable contributor to the development of research data repositories in Brazil. The Cariniana Dataverse network is developing information products to promote the practice of digital curation at institutions with important collections in digital format. Coordinated by IBICT, the Dataverse repository is used by the Cariniana team to help network partners become proficient at using methods of insertion and storage of electronic documents in research data repositories.

**IBICT** Instituto Brasileiro de Informação em Ciência e Tecnologia

POWERED BY THE **Dataverse Network** v. 3.8.2

**IBICT Dataverse Network** Create Account Log In

Search this Dataverse Network Search

Advanced Search Tips

O IBICT está disponibilizando um repositório de dados de pesquisa que cuida da preservação a longo prazo e de boas práticas de arquivamento, para que os pesquisadores possam compartilhar, manter o controle e obter o reconhecimento de seus dados. O repositório suporta o compartilhamento de dados de pesquisa com citação de dados persistentes, permitindo que eles sejam reproduzidos. O Dataverse é um grande repositório aberto de dados de todas as disciplinas, criado pelo Institute for Quantitative Ciências Sociais da Universidade de Harvard. O repositório Dataverse do IBICT fornece um meio disponível gratuitamente para depositar e descobrir conjuntos de dados específicos arquivados por colaboradores das instituições participantes da rede Cariniana.

**Dataverses**  
30 Dataverses

A **Dataverse** is a container for research data studies, customized and managed by its owner.

**RECENTLY RELEASED DATERVERSES**

Marcos Aurélio Gomes	07/12/2016
Datasets de Conferências	10/11/2016
Datasets de Teses	10/11/2016
Datasets de Dissertações	10/11/2016
Guilherme Ataide	08/11/2016

View More >

**Faculdade de Tecnologia Dataverses**  
Dataverses: 1  
Repositorio de pesquisa da Faculdade de Tecnologia da UNB.

**Studies**  
131 Studies, 399 Files, 83 Downloads

A **study** is a container for a research data set. It includes cataloging information, data files and complementary files.

**RECENTLY RELEASED STUDIES**

Gestão do conhecimento	19/01/2017
Viabilidade técnico-econômica de plantas solares	19/01/2017
Resoluções ANEEL	19/01/2017
Modelagem e simulação de sistemas CPV-T	19/01/2017
Revista Brasileira de Energia Solar	19/01/2017

View More >

**MOST DOWNLOADED STUDIES**

Normas do Tamanho da Categoria para 905 palavras da Língua portuguesa by Janczura, Gerson Américo, Castilho, Goiara Mendonça, Oliveira, Nelson Rocha	22
Gestão dos Dados Científicos dos Pesquisadores dos Programas de Pós-Graduação em Ecologia do Nordeste Brasileiro by Anjos, Renata, Dias, Guilherme	11
I Resumo de Atividades Rede Driade 2016 by Marderó Arellano, Miguel Angel	9
O documento na Ciência da Informação: em busca da construção de um objeto a partir da concepções da Documentação e da Diplomática by Rabello, Rodrigo	8
Normas de associação livre para 1004 palavras da Língua portuguesa by Janczura, Gerson Américo, Castilho, Goiara Mendonça, Keller, Victor Nahuel, Oliveira, Nelson Rocha	7

View More >

re3data.org  
REGISTRY OF RESEARCH DATA REPOSITORIES  
<http://doi.org/10.17616/R3K36C>  
IBICT Dataverse Network

Figure 15.1: Homepage of IBICT Dataverse Network Portal

Dataverse is a large repository open to data from all disciplines and hosted by the Institute for Quantitative Social Science at Harvard University. The Dataverse repository at IBICT provides free-of-charge an available means to deposit, find, and access specific datasets that are being archived by researchers from the participating organizations. It will act as a steward of digital content, is open for data deposits from our institutions' affiliated partners, and it shares content with all their researchers and librarians.

The Dataverse repository includes a relatively simple self-service ingest workflow for researchers; it also has the ability to share with trusted groups of researchers prior to publication, and it helps them fulfil Data Management Plan requirements. The Cariniana team was interested in Dataverse because it can be easily installed and maintained, and it can be brought online with a relatively small staff.

Nonetheless, the main reason Cariniana chose to make a Dataverse repository available to its partner institutions was the ability to integrate it with other systems; that is, LOCKSS and Archivematica<sup>4</sup> for distributed and local long-term preservation, OJS<sup>5</sup> for data publication, and DSpace<sup>6</sup> for interoperability.

### 15.3. WHAT SCIENTIFIC DATA IS ARCHIVED?

Thanks to the technical cooperation agreements established for OJS journals, the Dataverse repository had allowed initial collaboration and support on the implementation of a scientific data preservation service. The target of the service is content from institutional and individual projects. IBICT Dataverse's member service provides an individual space for deposit of archives or datasets for researchers, or a community of researchers and institutions. All the data that is being archived is automatically identified, linked, and supported with access mechanisms.

The datasets in Dataverse are considered a structural archive with standardized metadata to maximize its compatibility and retrieval. This helps researchers meet the requirements of the funding institutions for verification of research project data. All the metadata records are made available for research, and the service allows uploading of datasets identified by the author or institutional owner. Information from the dataverses can be used by local libraries in helping users with better-informed answers to their queries. Furthermore, Cariniana is collaborating with its partners' institutions' libraries in their management decisions regarding researchers' data that needs to be processed by digital curation and archiving processes.

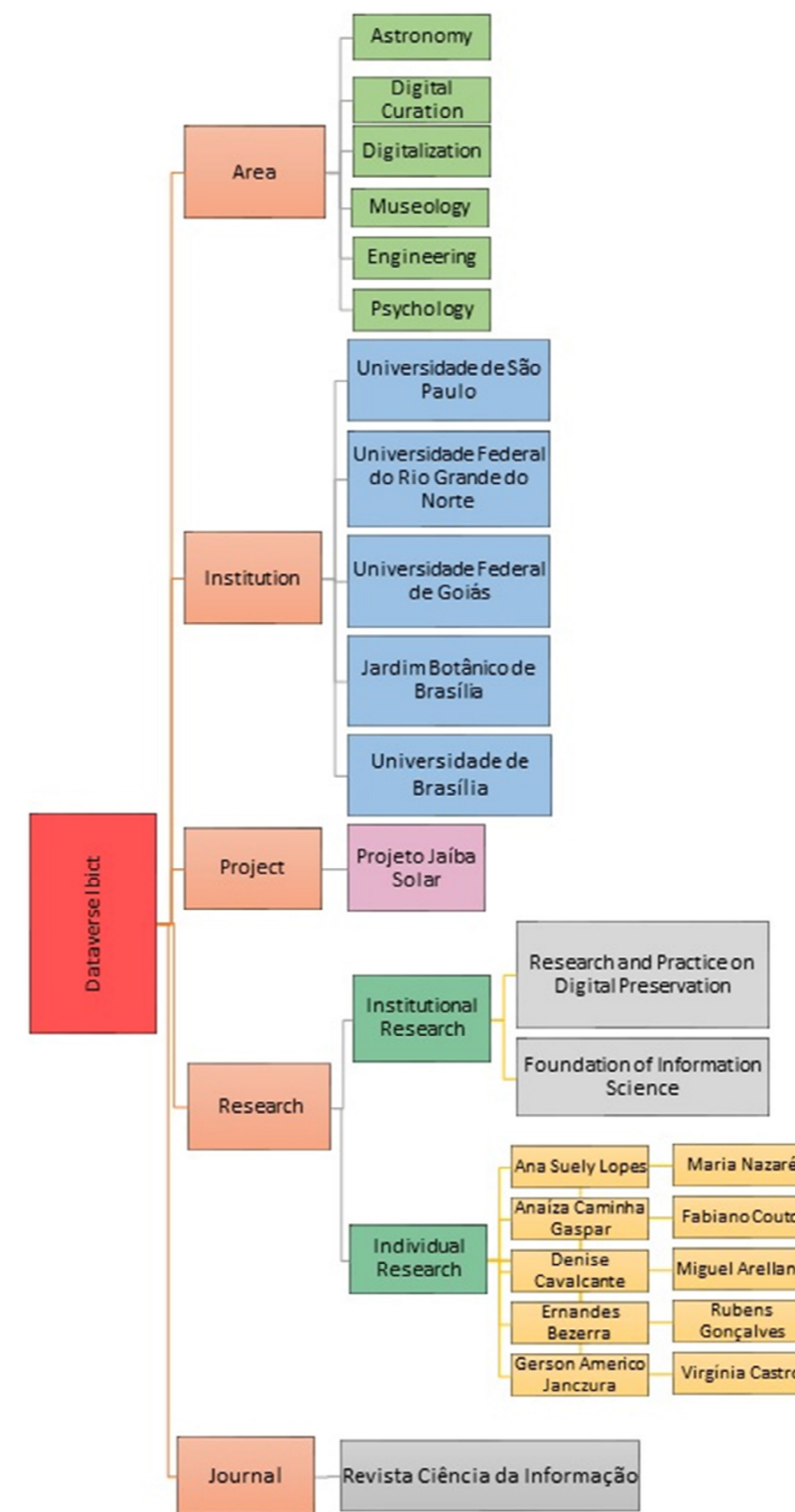


Figure 15.2: Organogram of IBICT Dataverse Network

Currently, the repository is being used by twenty-nine Cariniana collaborators, from six institutions, who have deposited data as members of the network. Researchers may store any type of digital data on any subject specialism. The Dataverse repository hosts 89 studies uploaded on 23 dataverses from institutional and individual research projects.

<sup>4</sup> Archivematica: <http://archivematica.org>; accessed 5 February 2017.

<sup>5</sup> PKP: <https://pkp.sfu.ca/ojs/>; accessed 5 February 2017.

<sup>6</sup> DSPACE: <http://dspace.org>; accessed 5 February 2017.

#### 15.4. WHAT COMES NEXT?

The repository requirements and policies on access, privacy, and reuse need to be well defined. Cariniana staff are establishing a curator service that will help organize data for preservation. A partner institution is planning financially to support the translation of the system's user interface into Portuguese, and the Cariniana team has produced a user manual. The migration of Dataverse to version 4.0 will be accomplished in the current year. This procedure will be combined with an institutional strategy at IBICT for secure backups and the assignment of a persistent URL.

IBICT's International Technical and Scientific Support Committee is establishing guidelines and recommendations for the repository. The Cariniana network is now discussing and planning further collaborative action and an integration of efforts. Its members consider Dataverse a trusted repository that allows for the long-term persistent access to scientific data, with mechanisms that incorporate full identification and the prevention of digital obsolescence.

IBICT considers it fundamental to establish efficient management of data for the development of scientific research, with quality in all aspects related to organization, documentation, archiving, and sharing of scientific information. As the project evolves, new challenges arise, revealing new approaches to the scientific information digital workflow. In Brazil there are still relatively few research data repositories, but the impact of the contributions from the field of Library and Information Science is growing.



This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 654139.