Scientific data repositories: the USP experience

Pró-Reitoria de Pesquisa: 
Sylvio Canuto
Antonio Mauro Saraiva

Superintendência de TI: 
João Eduardo Ferreira
Fátima L. S. Nunes

2018
Challenge: multidisciplinary area

Collaborative work

Information Technology

Research

PostGraduate

Libraries
Challenge: data diversity

Heterogeneous data (text, image, audio, video)
Distributed data
Specificity of the data in each different area
Challenge: storage (the biggest)

safety, space availability, access availability (24/24 x 7/7), persistence, specialized human resources
Challenge: cultural change and tools

Repositories with intuitive and minimal interfaces, with standardized and integrated data
Easy for any user
9

Actions Taken
USP Actions Made

1

Institutionalize efforts and actions

Creation of a multidisciplinary Work Group to:
- evaluate technologies
- suggest policies
- conduct a pilot study
USP Actions Made

Evaluate software alternatives

Open source? External proprietary? Internal development?

Our option:
Open source platforms
USP Actions Made

3

Define technical team

Tasks: study possible solutions, study market tendencies

Where and how:
Scientific articles
Sites of more experient organizations
Platforms installation
USP Actions Made

4

Define platform architecture

Centralized? Distributed? Hybrid?

Our choice:
Distributed repositories
Centralized retrieval
USP Actions Made

5

Define, prepare and make available repositories platforms

Platforms installation:
Platforms customization
Test phase with volunteers

Pró-Reitoria de Pesquisa
Superintendência de Tecnologia da Informação
USP Actions Made

Define, prepare and make available storage infrastructure

Our choices:
- USP Cloud
- 2 Petabytes up to now
- Redundant platforms
- Expert cloud team

Scientific data repositories: the USP experience
USP Actions Made

7

Standardize a Data Management Plan tool

Our choice:
DMPTool.org
(templates customization)
USP Actions Made

Build a retrieval tool

Integrate metadata of repositories in a centralized platform (metasearcher)
USP Actions Made

Build a site with concepts and information for researchers

Conceptual information related to digital repositories

Links to computational tools
USP Ongoing Actions
USP Ongoing Actions

1. Define a structured pilot study

- Select about 40 research groups from different areas
- Ask for using the 3 repositories and answer a survey about preferences
USP Ongoing Actions

2

Prepare the infrastructure for the future

Study of storage strategies for the future

Define a modular structure that can evolve

Define infrastructure independent of repository platforms
Next steps

• Choose a repository platform (if possible).

• Study the repositories use for planning the future.

• Define policies:
  • Who can use?
  • How much time?
  • How much space?
  • How to ask for access?
  • Who pays?
What have we learned (so far)

- Storage is the biggest challenge (space, time).
- It is almost impossible to centralize repositories, but it is possible to centralize search and retrieval.
- There is no consensus about repository platforms.
- Different rules and different metadata are necessary for different researchers. However, a limited set of common attributes (Dublin Core) must be defined to start the process.
USP Technical Team

Prof. João Eduardo Ferreira
Profa. Fátima L. S. Nunes
Prof. Adilson Gonzaga
Mauro Cesar Bernardes
Marino Hilario Catarino
Diego Araújo
Edmar Martineli
Rodrigo Muller de Carvalho
Scientific data repositories: the USP experience

Pró-Reitoria de Pesquisa: Sylvio Canuto
Superintendência de TI: João Eduardo Ferreira
Antonio Mauro Saraiva
Fátima L. S. Nunes

2018