Open Science/Open data: Challenges and Best Practices in Scientific data Management

UNIFESP
Ruy R Campos
Why has UNIFESP decided to carry out the initiative to promote the creation of an open repository of research data?
Open Access to Scientific Data will promote:

• Increasing possibilities and create new research/ideas from accessible data;
• Increasing Visibility;
• Increasing possibility of sharing and analyzing data in real time;
• Increasing interaction among researchers from different areas of research (Interdisciplinary Projects);
• Maximizing data reuse by efficient data extraction;
• Increasing transparency of scientific research, an important tool to reduce fraud;
• Ensuring compliance with funding agencies (For example: FAPESP) and Scientific Journal Policies...
What are the main actions already undertaken by UNIFESP?
Dataset Repository
DR UNIFESP

Starting point


WORKSHOP

Sonia Barbosa – Manager of Data Curation and Acquisition for the Harvard Dataverse Project from Murray Research Archive, Harvard University

Corey Davis – Curation Digital Preservation Digital COPPUL, Vancouver, Canadá.

The consortium comprises 22 university libraries located in Manitoba, Saskatchewan, Alberta and British Columbia, plus 17 affiliate members that participate in e-resource licensing only.
COMPOSITION OF THE STUDY GROUP AND IMPLEMENTATION WORK Datasets
Repository - DR UNIFESP - MAY/2017 - Data Science & Product Research

Soraya Soubhi Smaili (Reitora); Maria Angélica Pedra Minhoto (Assessoria de integração acadêmica)
Esper Abrao Cavalheiro (Pró-reitor de pós-graduação e pesquisa) e Ruy Ribeiro de Campos Junior (Pró-reitor adjunto de pós-graduação e pesquisa)
Maria Eduarda S. Puga (Diretora da CRBU), Norma Iwashita Moriama (Diretora Adjunta CRBU)
Flávio Castro de Sousa, Александро Кардосу Карвалхо (Divisão de Tecnologia da Informação)
Diogo Misoguti, Adrêia Akemi Oribe Hayashi (Comitê Gestor do Repositório Institucional)
Lidiane Cristina da Silva (Procuradora Educacional Institucional)
Juliana Almeida dos Santos (Técnica em Arquivos)
CRIAMOS UM GRUPO DE ESTUDO E TRABALHO PARA A IMPLANTAÇÃO DO REPOSITÓRIO DE DADOS DE PESQUISA DA UNIFESP MAIO/2017

Dataset Repository DR UNIFESP

SELECTED SOFTWARE

Open source research data repository software

Enjoy full control over your data. Receive web visibility, academic credit, and increased citation counts. A personal dataverse is easy to set up, allows you to display your data on your personal website, can be branded uniquely as your research program, makes your data more discoverable to the research community, and satisfies data management plans. Want to set up your personal dataverse?
Dataset Repository
DR UNIFESP

GT – FAPESP / INFRASTRUCTURE

Distribuição da rede de dados de pesquisa Unifesp

Datacenter São Paulo

Sevidor Dell R730XD
Sevidor Dell R730XD
Storage Stormwise V3700 28TB
Storage Stormwise V3700 28TB

Datacenter Diadema

Sevidor Dell R730XD
Sevidor Dell R730XD
Storage Stormwise V3700 28TB
Storage Stormwise V3700 28TB

32 Terabytes

AIM: 90 TB

* Plano de expansão

UNIFESP

PRO-REITORIA DE PÓS-GRADUAÇÃO E PESQUISA
Dataset Repository
DR UNIFESP

GT FAPESP –
Definition of 21 metadata for all institutional projects
Institutional platform that contemplates the data management plan for all projects. This will allow establishing the Institutional need for the custody of all research activity carried out at UNIFESP.
Repositorio de Dados de Pesquisa UNIFESP

PILOT PROJECT DEPOSIT DATA CAMPI BAIXADA SANTISTA AND GUARULHOS
I Seminário
Gestão de Dados de Pesquisa

Universidade Federal de São Paulo e Universidade Federal do ABC

Local: Anfiteatro da Reitoria da Universidade Federal de São Paulo
Rua Sena Madureira, 1500 Vila Mariana. São Paulo/SP. Térreo

Datas e Horários:
27 de agosto de 2018 das 8h30 às 16h
28 de agosto NÃO HÁ PROGRAMAÇÃO
29 de agosto de 2018 das 9h00 às 17h
What are the challenges?
Challenges:

- Financial support, how will costs be managed?
- Institutional infrastructure (physical and employee) to support the researcher's data management plan;
- Which data should be retained (final data)?
- For how long should data be maintained?
- Complexity and size of dataset;
- Complexity of different areas of research (Images; some research, e.g. with personal identifiers cannot be shared; can someone steal my idea?)...
The **Federal University of São Paulo** is located in the most populated and wealthiest state of Brazil: São Paulo.
UNIFESP | Campuses

- Baixada Santista
- Diadema
- Guarulhos
- Osasco
- São José dos Campos
- São Paulo
<table>
<thead>
<tr>
<th>Category</th>
<th>Nº</th>
<th>Enrollments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Undergraduate Courses</td>
<td>52</td>
<td>13,000</td>
</tr>
<tr>
<td>Graduate Programs</td>
<td>67</td>
<td>5,340</td>
</tr>
<tr>
<td>Medical Residence</td>
<td>86</td>
<td>1,096</td>
</tr>
<tr>
<td>Multiprofessional Residence</td>
<td>20</td>
<td>561</td>
</tr>
<tr>
<td>Continuing Education Courses (Specialization and Professional Development)</td>
<td>130</td>
<td>2,369</td>
</tr>
<tr>
<td>Social Programs</td>
<td>60</td>
<td>-</td>
</tr>
<tr>
<td>Social Projects</td>
<td>154</td>
<td>-</td>
</tr>
<tr>
<td>Learning Distance</td>
<td>21</td>
<td>9,264</td>
</tr>
<tr>
<td><strong>TOTAL</strong></td>
<td></td>
<td><strong>35,160</strong></td>
</tr>
<tr>
<td>Faculty</td>
<td></td>
<td>1,547</td>
</tr>
<tr>
<td>Staff (UNIFESP + HU)</td>
<td></td>
<td>4,000</td>
</tr>
<tr>
<td><strong>TOTAL</strong></td>
<td></td>
<td><strong>5,547</strong></td>
</tr>
</tbody>
</table>

Source: Assessoria de Pesquisa Institucional, 2018
#3 in Brazil, Federal #1

#4 in Latin America

Publication with high citation index
UNIFESP | Fundamental Principles and Structural Axes

- Ethics
- Instituting Process
- Direct Democracy and Participative Governance
- Strategic Areas for Education, Research, Extension and Continuous Assessment
- Intercampi and Convergent Structure
- Promotion of Social and Environmental Well-being
- Sustainability
- Unity and Diversity
- Quality and Relevance
- Democracy, Transparency and Equity
- Ethics
Thank You