

Curadoria e Preservação de Dados Científicos para compartilhamento – a Unicamp e o REDU

Claudia Bauzer Medeiros
Instituto de Computação – Unicamp
Comissão de Gestão de Dados de Pesquisa



Agradecimentos

- Valéria dos Santos Martins
- Amarilis Correa
- Laura Rezende

CGDP - UNICAMP



V SINPRED
12 a 14
de Maio
2021

Pontos principais

- Ciência Aberta – UNESCO e preservação digital
- Rede de repositórios do estado de São Paulo
- REDU – UNICAMP -> desafios na preservação de dados

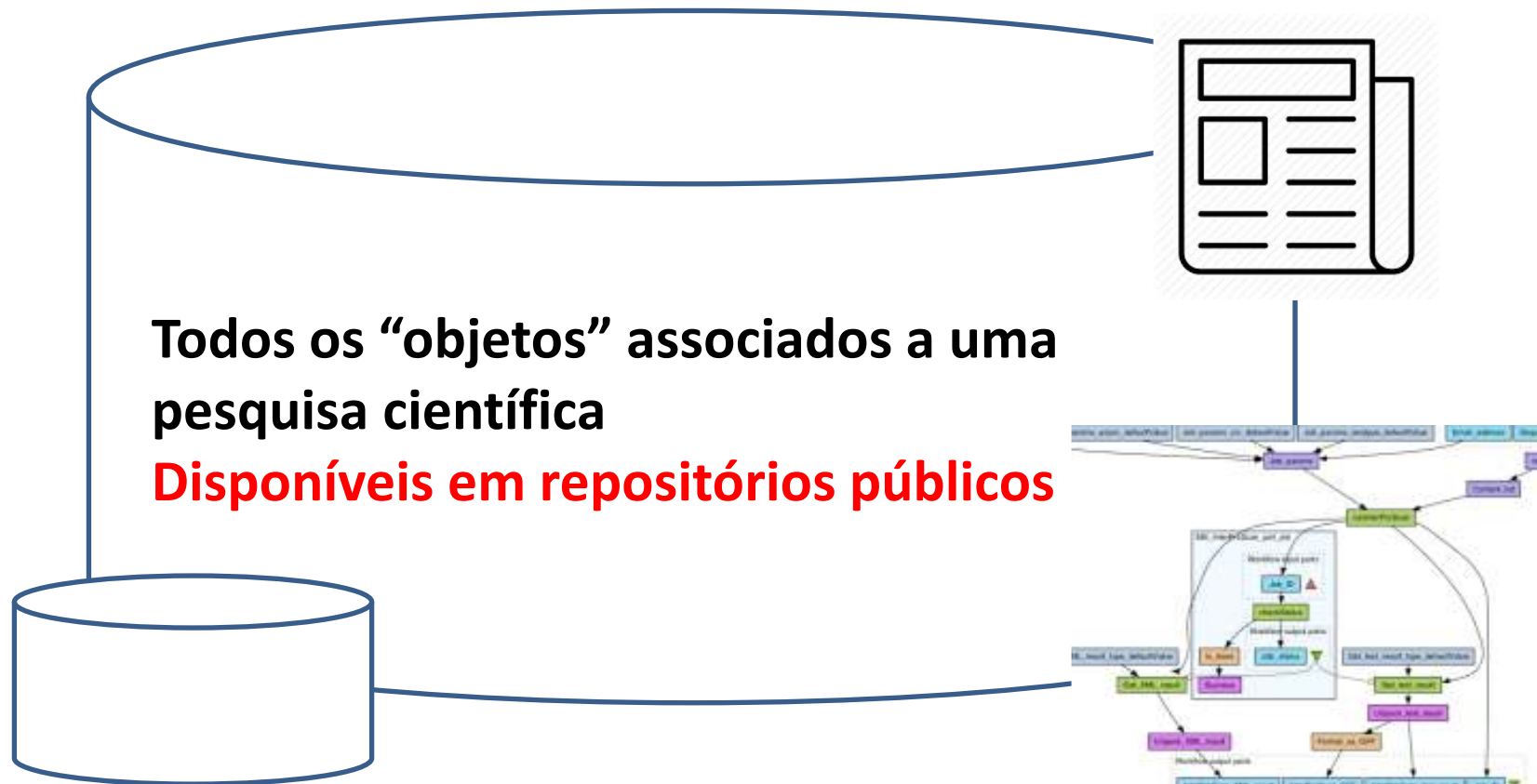
Preservação de dados

- A disponibilidade de dados abertos e repositório institucional passou a ser um **critério de avaliação de instituições**
 - **Internacional**
 - **Estabelecimento de convênios ...**

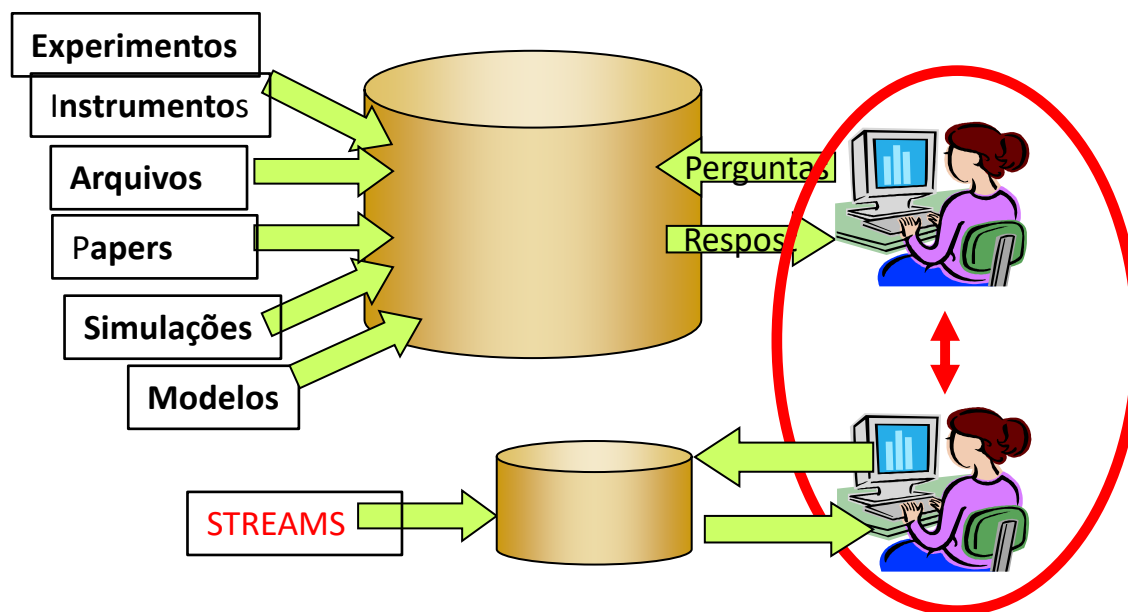
CIÊNCIA ABERTA – UNESCO E PRESERVAÇÃO DIGITAL

@Claudia Bauzer Medeiros

Ciencia Aberta



Open Science -> colaboração



Colaboração por meio dos “objetos” associados à pesquisa



United Nations
Educational, Scientific and
Cultural Organization

Organisation
des Nations Unies
pour l'éducation,
la science et la culture

Organización
de las Naciones Unidas
para la Educación,
la Ciencia y la Cultura

Организация
Объединенных Наций по
вопросам образования,
науки и культуры

منظمة الأمم المتحدة
للتربية والعلم والثقافة

联合国教育、
科学及文化组织

**INTERGOVERNMENTAL MEETING OF EXPERTS (CATEGORY II)
RELATED TO A DRAFT UNESCO RECOMMENDATION ON OPEN SCIENCE**

Online, 6-7 and 10-12 May 2021

**DRAFT TEXT OF THE UNESCO RECOMMENDATION
ON OPEN SCIENCE**

**PROVISIONALLY ADOPTED
(AS OF 11 MAY 2021)**

@ClaudiaBauzerVieira

Unesco

- **(iv) Investing in human resources, training, education, digital literacy and capacity building for Open Science**
- 20. Open Science requires investment in capacity building and human capital. Transforming scientific practice to adapt to the changes, challenges, opportunities and risks of the 21st century digital era, requires targeted research, education and training in the skills required for new technologies and in the ethos and practices of Open Science. Member States are encouraged to consider the following:
 - (a) Providing systematic and continuous capacity building on Open Science concepts and practices, including broad comprehension of the Open Science guiding principles and core values as well as technical skills and capacities in digital literacy, digital collaboration practices, data science and stewardship, **curation, long-term preservation and archiving**, information and data literacy, web safety, content ownership and sharing, as well as software engineering and computer science.

REDE DE REPOSITÓRIOS SP E REDU

@Claudia Bauzer Medeiros

Rede repositórios dados pesquisa SP



48 campi

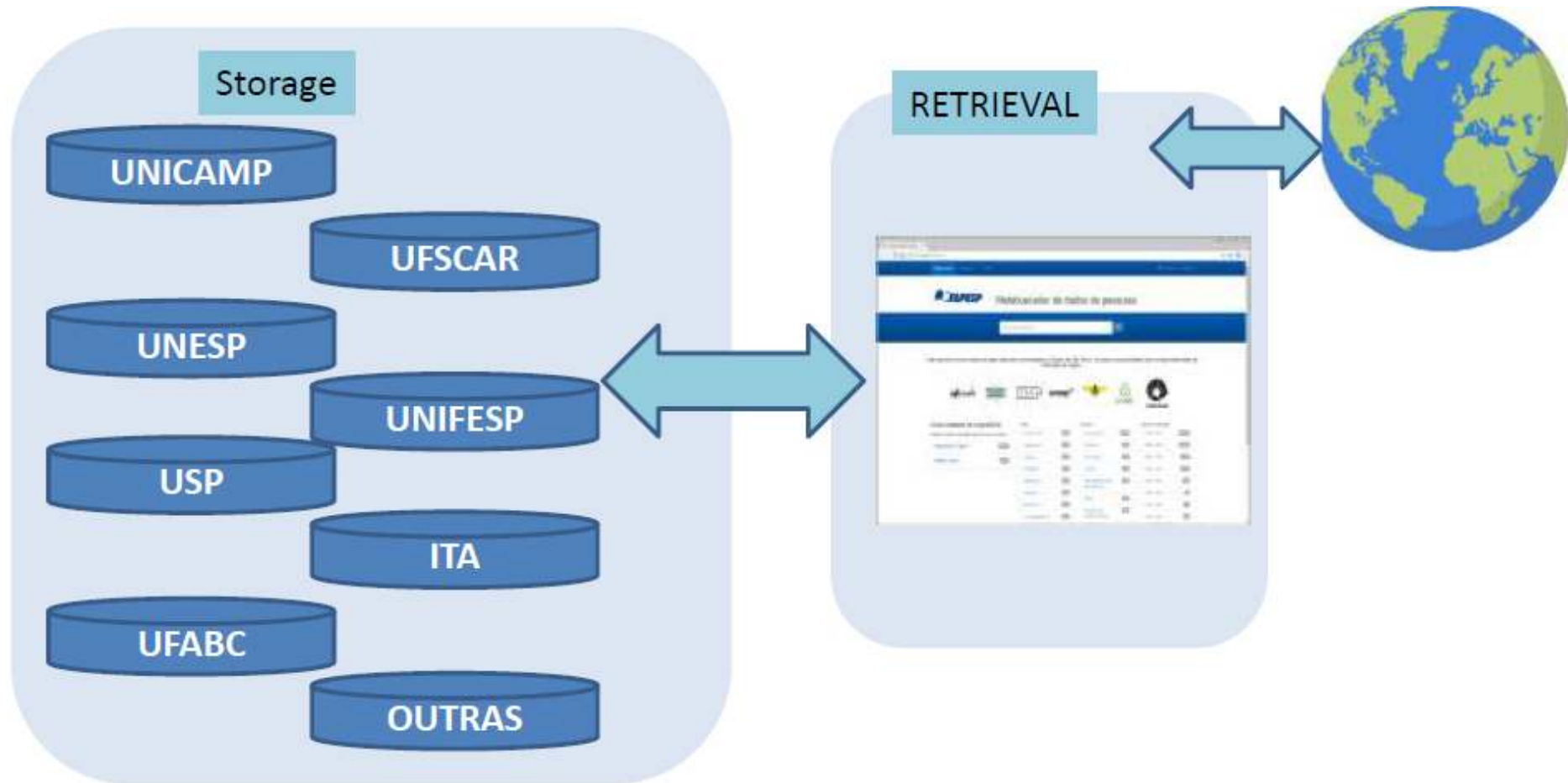
11,5 mil docentes

170 mil alunos

+ pesquisadores Informática na Agricultura

Rede de repositórios

- Cada participante mantém seu repositório
- Interface de metadados única (metabuscador)



metabuscador.uspdigital.usp.br



Research data metasearch

Search DSpace



This repository gathers research data from academic institutions in the State of São Paulo. The responsibility for the data provided is exclusive of who made them available.

This website, software and associated repositories were created to comply with the [FAPESP Data Management Policy](#)



Institutions

Choose a institution to browse its collections.

USP - Universidade de São Paulo	15
EMBRAPA - Empresa Brasileira de Pesquisa Agropecuária	270
FAPESP COVID-19 Data Sharing/BR	3

Discover

Author

Barbedo; Jayme Garcia Arnal	93
Halfeld-Vieira; Bernardo de Almeida	56
Oliveira; Saulo Alves Santos	19

Subject

Computer and Information Science	5
Earth and Environmental Sciences	4
coronavirus	3

Date issued

2018	270
2019	13
2020	13

@Claudia Bauzer Medeiros

REDU - Unicamp

- Repositorio de dados de pesquisa da UNICAMP (REDU)
- Comissão de Gestão de Dados de Pesquisa (CGDP)

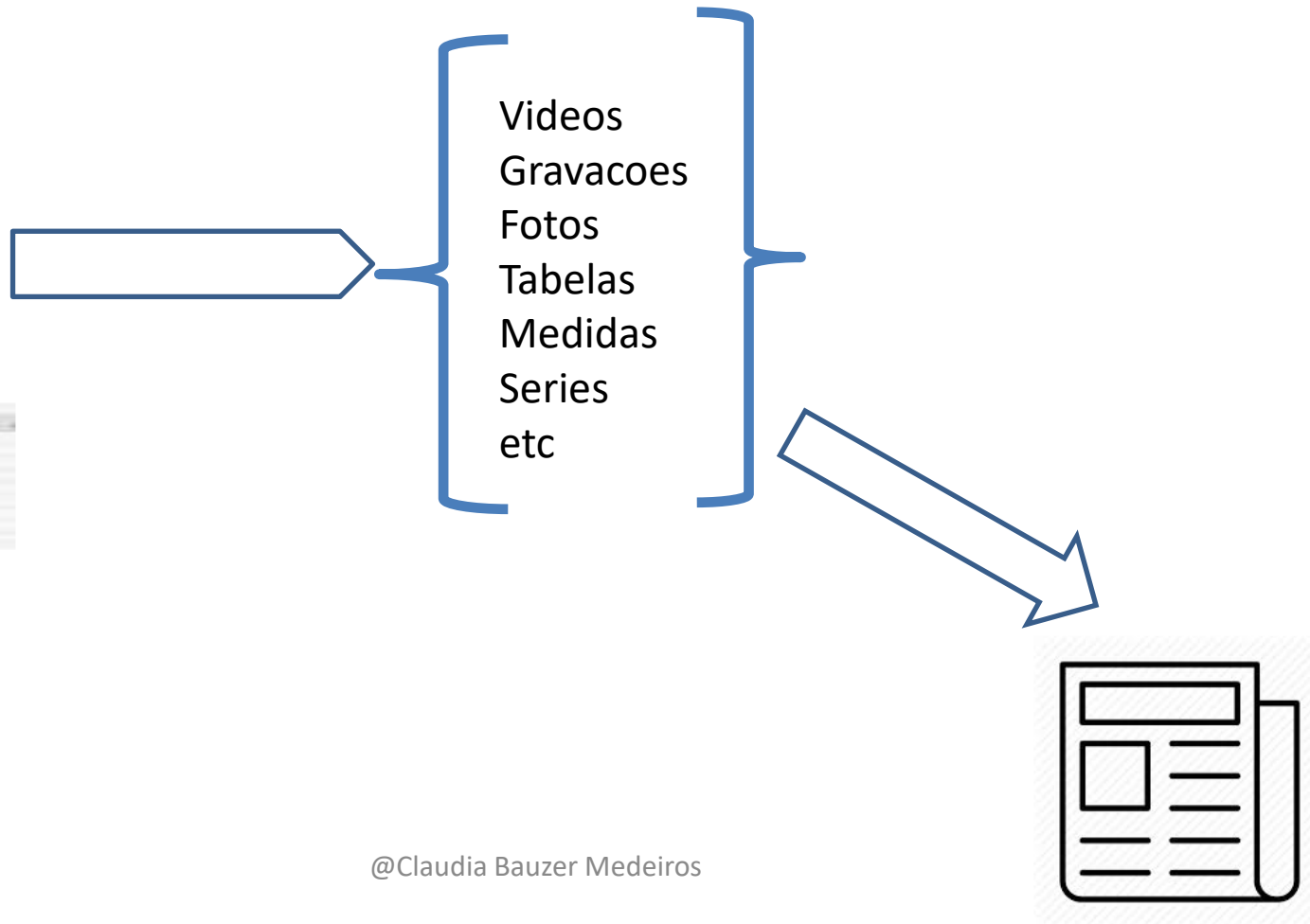
IMECC, FCM, IFCH, COCEN, SBU, CCUEC

(Exatas, Biológicas, Humanas, Engenharias Multidiscp, Biblioteca, Centro Computação)

O que são meus dados de pesquisa

A screenshot of a data table with columns for dates and numerical values.

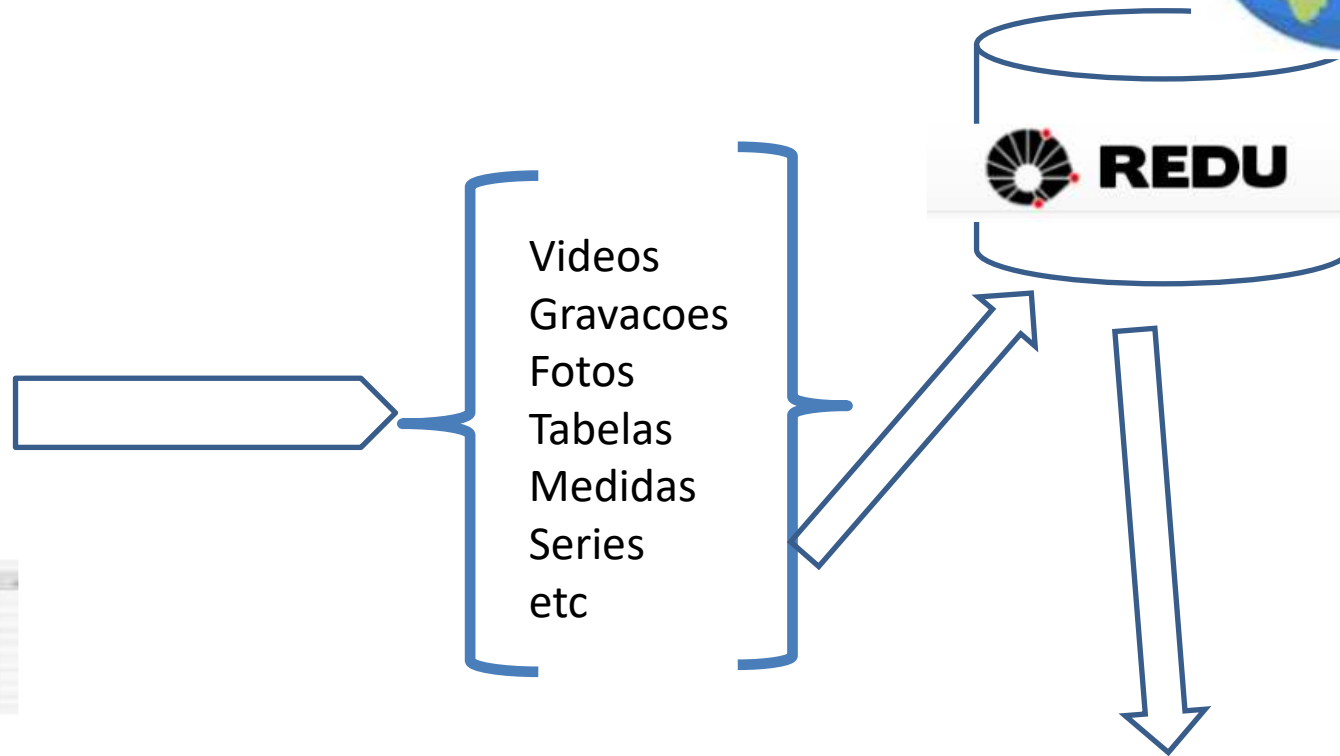
	Jan-18	Feb-18	Mar-18	Apr-18	May-18
Estado	65,0	61,0	58,0	61,0	59,0
Previsão	60,0	60,0	60,0	60,0	60,0
Atual	990,0	990,0	990,0	990,0	990,0
Máximo	100,0	100,0	100,0	100,0	100,0
Mínimo	200,0	200,0	200,0	200,0	200,0
Média	500,0	500,0	500,0	500,0	500,0
Total	1.000,0	1.000,0	1.000,0	1.000,0	1.000,0



Como e onde armazenar



Trimestre	Jan-18	Fev-18	Mar-18	Abr-18	Mai-18
Vendas	65,0	61,0	58,0	61,0	60,0
Produtos	60,0	58,0	55,0	58,0	57,0
Aluguel	100,0	100,0	100,0	100,0	100,0
Manutenção	200,0	200,0	200,0	200,0	200,0
Outros	100,0	100,0	100,0	100,0	100,0
Total	1.065,0	1.069,0	1.067,0	1.069,0	1.066,0



Produção Intelectual



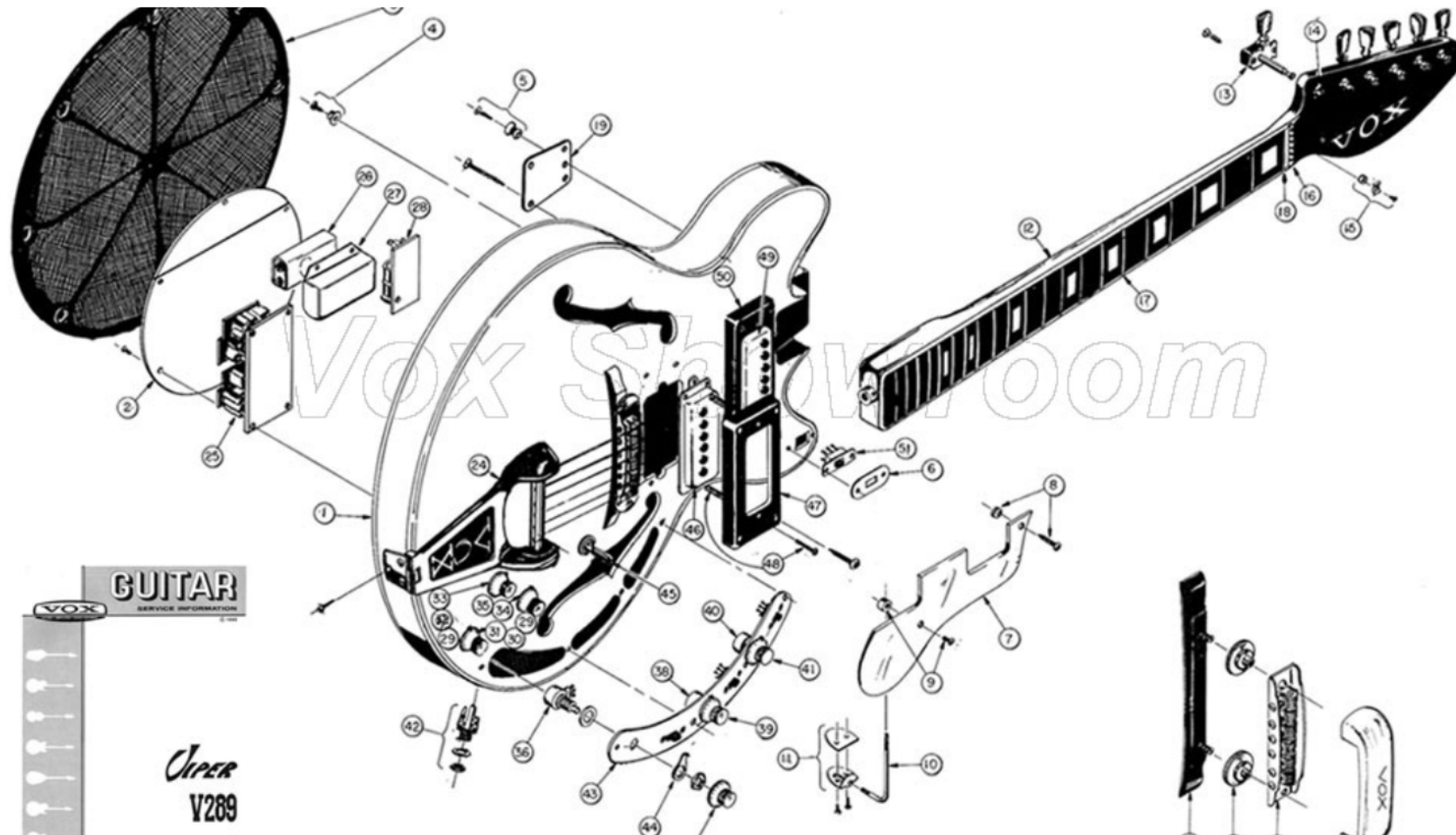
DESAFIOS PARA A PRESERVAÇÃO

@Claudia Bauzer Medeiros

Desafios

- Variedade de tipos de dados e volume
- Variedade de metadados
- Questões específicas de cada domínio
- O que significa dados abertos?
- Preservação para auditoria, reprodutibilidade e reuso

Reuso e compartilhamento



<http://www.voxshowroom.com/us/guitar/viper1.html>

Reuso e compartilhamento



Colagem – Pablo Picasso – Natureza Morta com Cadeira de Palha

<https://es-es.facebook.com/280698011956458/posts/natureza-morta-com-cadeira-de-palha-1912-colagem-de-pablo-picasso-museu-picasso-/1797343096958601/>

Dados digitais abertos?

- Qualquer pessoa **E SISTEMAS** podem
 - Descobrir se os dados existem
 - Descobrir como (e se) obtê-los
 - Compartilhar “tudo”? Não necessariamente
 - **METADADOS ABERTOS**

Restrições = segurança, confidencialidade, ética, propriedade intelectual

Metadados para artigos

- Autor
- Título
- Resumo
- Palavras-chave
- Ano de publicação
- Instituição
- URL

Metadados para dados

- Autor
- Título **do conjunto de dados**
- Resumo **do conteúdo**
- Palavras-chave
- Ano de publicação
- Instituição
- URL **ou PID**
- **Versão, direitos de acesso,**

**PRESERVAR PARA COMPARTILHAR
EXIGE PENSAR METADADOS**

[Artes e Humanas](#)[Biomédicas](#)[Ciências Aplicadas](#)[Exatas](#) Find[Advanced Search](#)  **Dataverses (5)**  **Datasets (36)**  **Files (168)****Dataverse Category****1 to 10 of 41 Results** Support data and documentation on meta-analysis on semantic search

May 6, 2021 - Tecnológica

Gottardi, Thiago, 2021, "Support data and documentation on meta-analysis on semantic search",

 Datasets (30)  Files (168)**Dataverse Category**

Department (5)

Publication Year

2021 (20)

2020 (11)

2018 (5)

2019 (5)

Author Name

Gradwohl, André Leon Sampaio (3)

Liu, Jianna (3)

Meneguello, Rachel (3)

Leonardi, Gislaine Ricci (2)

Lima, Ana Luísa Fogarin de Sousa (2)

[More...](#)**Subject**

Computer and Information Science (15)

Medicine, Health and Life Sciences (8)

Social Sciences (8)

Earth and Environmental Sciences (7)

Support data and documentation on meta-analysis on semantic search 

May 6, 2021 - Tecnológica

Gottardi, Thiago, 2021, "Support data and documentation on meta-analysis on semantic search", <https://doi.org/10.25824/redu/89NUBJ>, Repositório de Dados de Pesquisa da Unicamp, V2


This package contains all output data and documentation associated with a meta-analysis study to understand semantic search issues and mechanisms. when applied to search for data, papers or articles. This analysis was conducted on the entire contents of 9 digital libraries - ACM...

Lexical expansion in brazilian portuguese 

May 5, 2021 - Artes e Humanas

Sabatini, Fernando Luis;Sampaio, Thiago Oliveira da Motta, 2021, "Lexical expansion in brazilian portuguese", <https://doi.org/10.25824/redu/WUAMI3>, Repositório de Dados de Pesquisa da Unicamp, V2, UNF:6:eT6h1OEvApkgJpKh0Fkpdg== [fileUNF]

Lexical Expansion is a recurring theme in historical linguistics and sociolinguistics. As far as we know, the present study is the first attempt to pursue the subject through an experimental approach. The present dataset report the results of a lexical decision test applied to 3...

Effect of ascorbic acid topical application on the skin by high-frequency ultrasound 

May 4, 2021 - Biomédicas

Leonardi, Gislaine Ricci;Aiello, Laura Moretti;Vergilio, Mariane Massufero, 2021, "Effect of ascorbic acid topical application on the skin by high-frequency ultrasound", <https://doi.org/10.25824/redu/SMWRR4>, Repositório de Dados de Pesquisa da Unicamp, V1, UNF:6:uQz3EyWsYP5ghR5QnKHv3Q== [fileUNF]


The general aim of the study is to evaluate the performance of a formulation containing ascorbic acid (AA) using the 50 MHz high-frequency ultrasound skin image analysis method. For this 25 healthy women with visible signs of facial aging, between 35 and 60 years old, were includ...



Maps of permanent protection areas and land use and occupation in the high stroke of the Jaguari hidrographic basin Version 1.0

Santos, Natasha Marques de Paula, 2021, "Maps of permanent protection areas and land use and occupation in the high stroke of the Jaguari hidrographic basin", <https://doi.org/10.25824/redu/0XBUVW>, Repositório de Dados de Pesquisa da Unicamp, V1

 Cite Dataset ▾

 Learn about [Data Citation Standards](#).

Description

The data correspond to maps about permanent protection areas and land use and occupation of this areas in the high stroke of Jaguari hidrographic basin. These results appear with more details in the research "SPATIALIZATION OF THE PERMANENT PROTECTION AREAS OF THE HIGH STROKE OF THE JAGUARI HYDROGRAPHIC BASIN" (FAPESP number 2020/05304-1). Topographic maps of the IBGE (1:50 000 scale) and images from the Sentinel 2A satellite of the Copernicus Project - European Space Agency (ESA) dated July 25th, 2020 were used as database. Besides that, the projection Universal Tranverse Mercator (UTM) Zone 23 S and Datum SIRGAS 2000 were used as spacial reference. The maps are available in png and tiff format and there are maps with the whole area of the high stroke of Jaguari river as well maps cut out according the area of IBGE's topographic maps. All the images have a name, subtitle according the map, scale, spacial reference, datum, database used, autor, and funding agency. The subtitle in the maps about permanent protection areas were summarized in numbers according the kind of preservation protection areas. It can be seen with more details in the total research. (2021-04-26)

Subject

Earth and Environmental Sciences

Keyword

Permanent protection areas, Land use and land occupation, Jaguari hidrographic basin

Desafios

- Variedade de tipos de dados e volume
- Variedade de metadados
- Questões específicas de cada domínio
- O que significa dados abertos?
- Preservação para auditoria, reprodutibilidade e **reuso**

redu.unicamp.br

REDU Search

Log In

Username/Email

Password

Log In

[Forgot your password?](#)

Other options:

You can convert a Dataverse account to use one of the options above. [Learn more.](#)

OBRIGADA
CMBM@IC.UNICAMP.BR

