

Construindo uma Infraestrutura Aberta de Dados de Pesquisa no Brasil



MINISTÉRIO DA
CIÊNCIA, TECNOLOGIA
E INOVAÇÕES



10 de agosto de 2022



Infraestrutura Informacional **A**berta

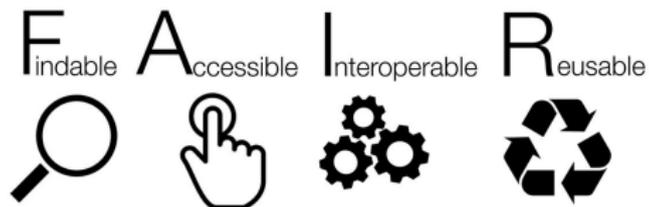
- Acesso
- Segurança
- Preservação
- Tecnologia
- Capacitação



Princípios **FAIR**

Dados tem que ser:

- Encontráveis
- Acessíveis
- Interoperáveis
- Reutilizáveis



Princípios **TRUST**

Aplicar os princípios de:

- Transparência
- Responsabilidade
- Foco no usuário
- Sustentabilidade
- Tecnologia adequada



Portal Brasileiro de Publicações e Dados Científicos em Acesso Aberto

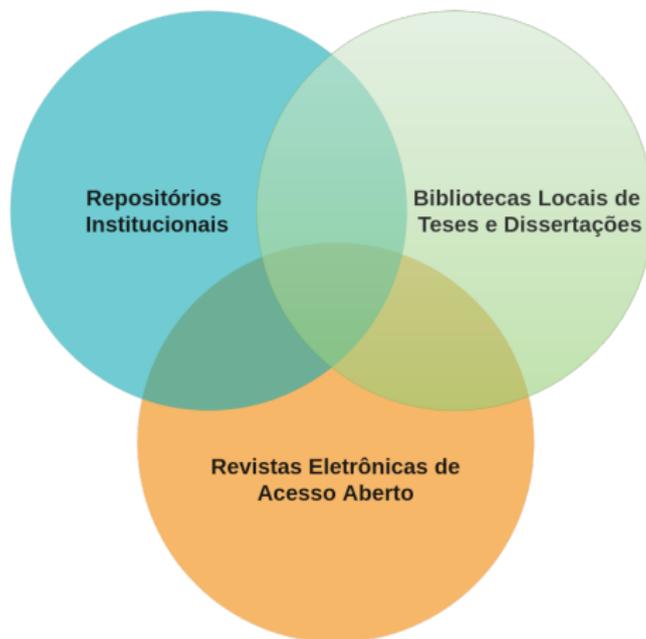
<<https://oasisbr.ibict.br>>





- Bibliotecas locais de TDs
- Repositórios Institucionais
- Revistas Eletrônicas

Rede com mais de **700** instituições que agregam mais de **1400** fontes e aproximadamente **4.000.000** docs.



RI s	TD s	Revistas	RDP	RCAAP	PDA CAPES
110	130	1,3K	10	1	1
1,12M	770K	805K	800	670K	825K

Rede LA Referencia



LA Referencia

Red de repositorios de acceso abierto a la ciencia

- [Argentina](#), [Brasil](#), [Chile](#),
[Colômbia](#), [Costa Rica](#), [Equador](#),
[El Salvador](#), [México](#), [Perú](#), [Uruguai](#), [Panamá](#),
[Espanha](#) e [Portugal](#)

Diretrizes e Infraestruturas Internacionais

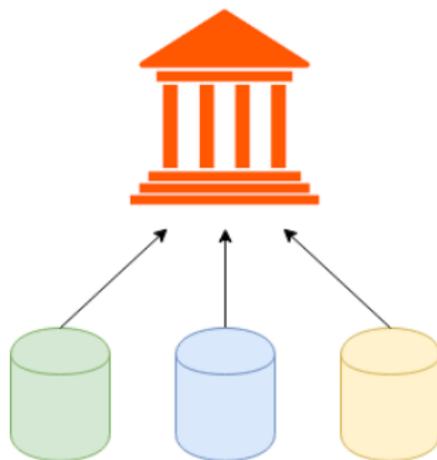


**EUROPEAN OPEN
SCIENCE CLOUD**



Relevância de um Repositório de Dados de Pesquisa

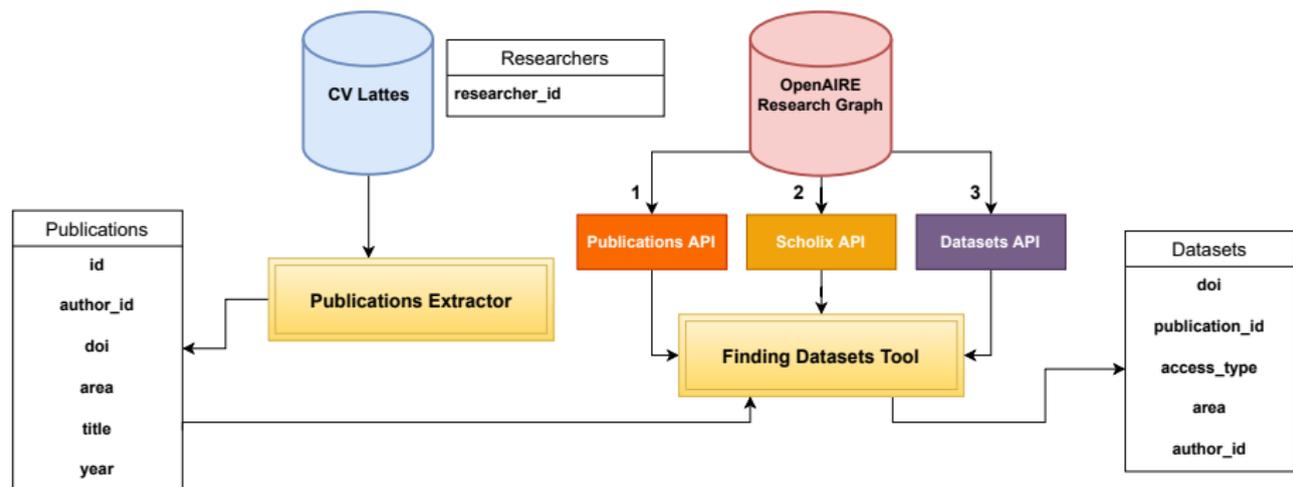
- **compartilhar e reutilizar**
- **curadoria**
- **visibilidade**
- **gestão**
- **preservação**
- **economia de recursos**



Onde eu deposito os dados gerados na minha pesquisa?

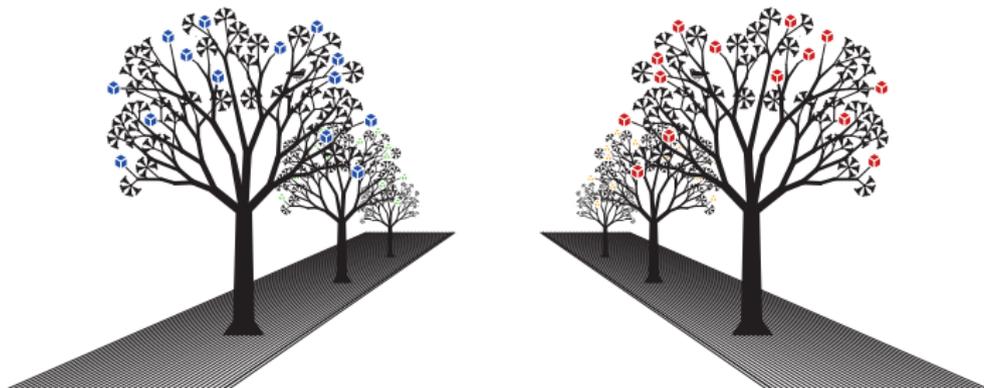


Dados de Pesquisa



- Pesquisadores com bolsa produtividade **CNPq 1A - 13.000 datasets**
- Pesquisadores que participaram de Projetos do tipo **INCT- 2.500 datasets**

LattesData/CNPq e Aleia/IBICT



LattesData/CNPq <<https://lattesdata.cnpq.br>>

The screenshot shows the homepage of the LattesData/CNPq repository. At the top, there is a dark blue header with the CNPq logo on the left and the text 'Deposite os seus conjuntos de dados' on the right. Below the header is a navigation menu with links for 'Pesquisa', 'Sobre', 'Guia do usuário', 'Suporte', 'Idioma Padrão', and 'Iniciar sessão'. The main content area features the LattesData (CNPq) logo on the left and a decorative background image of a network graph on the right. A paragraph of text describes the repository's purpose: to collect, store, and disseminate scientific data sets from CNPq-funded researchers. Below this is a search bar with the placeholder text 'Digite sua pesquisa' and a magnifying glass icon. Underneath the search bar are four blue buttons with white text and numbers: '23 Conjunto de DataVerses', '15 Conjunto de Dados (DataSets)', '55 Arquivos', and '410 Projetos'. A large blue button labeled 'Explorar' is positioned below these buttons. At the bottom of the page, a dark blue footer contains the copyright information 'Copyright © 2022, CNPq/IBICT' on the left and the development details 'Desenvolvido por The Dataverse Project v. 5.11.1 build 1069-02e3e92' on the right.

CNPq **Lattes Data** **Deposite os seus conjuntos de dados**

Pesquisa - Sobre Guia do usuário Suporte Idioma Padrão - Iniciar sessão

Lattes Data (CNPq)

O repositório LattesData tem como objetivo reunir, armazenar e divulgar os conjuntos de dados científicos de pesquisadores beneficiários do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq). Futuramente, será também disponível para depósito de dados de pesquisa oriundos de pesquisadores não financiadas pelo CNPq, mediante acordos celebrados com suas instituições no âmbito do Consórcio Nacional para Ciência Aberta (CoNCiência).

Digite sua pesquisa

23 Conjunto de DataVerses **15** Conjunto de Dados (DataSets) **55** Arquivos **410** Projetos

Explorar

Copyright © 2022, CNPq/IBICT

Desenvolvido por **The Dataverse Project**
v. 5.11.1 build 1069-02e3e92

Aleia/IBICT

<<https://aleia.ibict.br>>

[Pesquisa ▾](#)[Sobre](#)[Guia do usuário](#)[Suporte](#)[Idioma Padrão ▾](#)[Inscreva-se](#)[Iniciar sessão](#)

Bem-vindo ao repositório de dados de pesquisa do Ibiict

[13 Conjunto de DataVerses](#)[2 Conjunto de Dados \(DataSets\)](#)[3 Arquivos](#)[0 Downloads](#)[Explorar](#)

Copyright © 2022

Desenvolvido por  **DataVerse**
v. 5.11.1 build 1069-02e3e92

CoNCiênciaA



Consortium lead



Organization 1
(Also Consortium Lead role)



Organization 2



Organization 3



Repository



Repository



Reposito



Repository



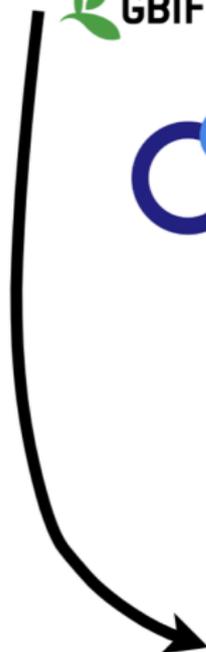
Incubando Repositórios de Dados

- RNP, IBICT & CNPq **chamada** - 4 instituições brasileiras selecionadas:
UFG, UFRA, UFC & CBPF





LA Referencia
Red de repositorios de acceso abierto a la ciencia



The
Dataverse[®]
Project

?

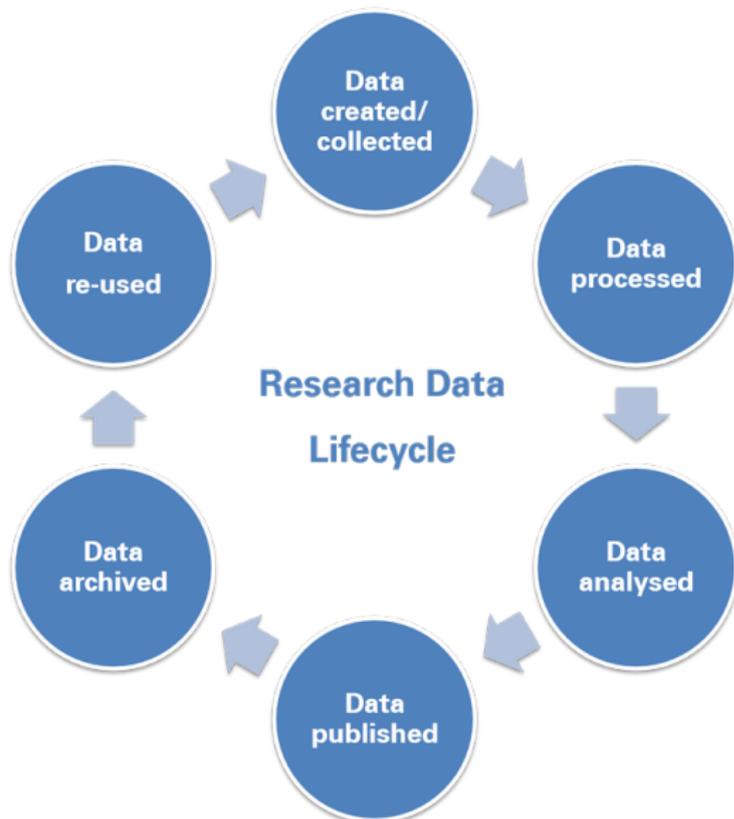
O papel do Ibict

- Suporte para a criação de repositórios de dados de pesquisa
- Desenvolvimento, adaptação e difusão de diretrizes
- Metabuscarador nacional de conjuntos de dados
- Extensão do  oasisbr
nó da rede LA referencia

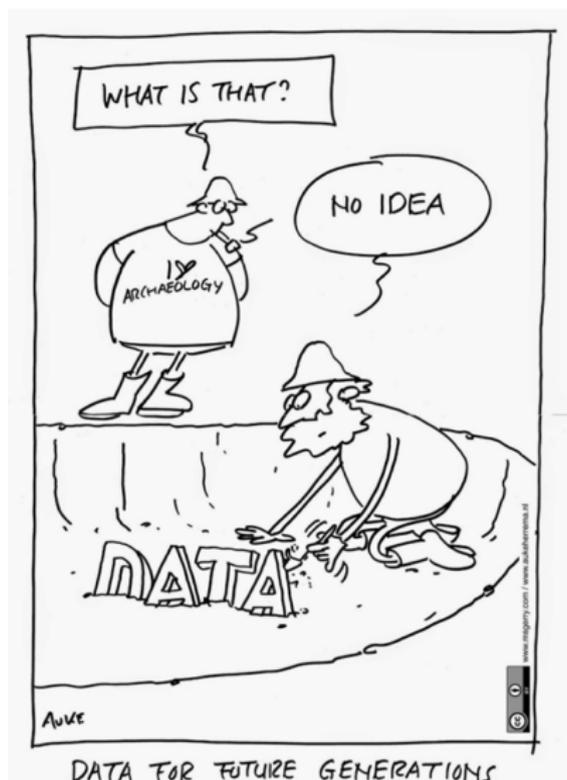
O papel das agências de fomento



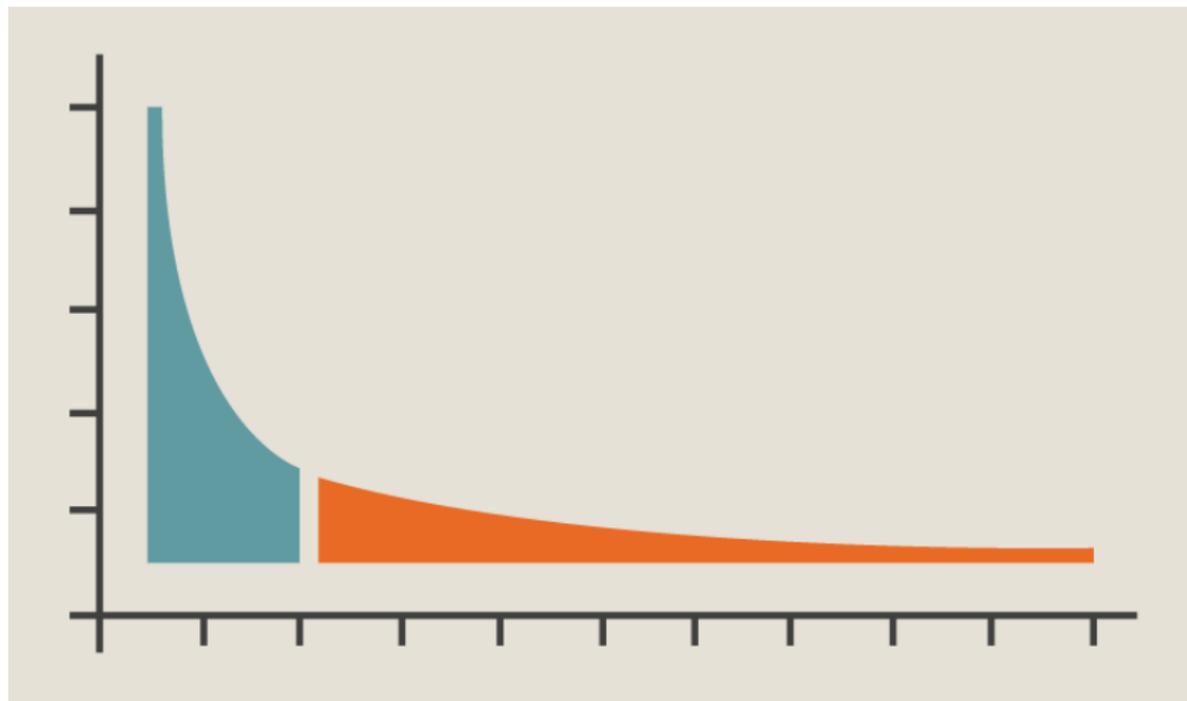
O ciclo de vida dos dados de pesquisa



Os dados depositados tem que estar bem documentados



Long tail



OpenAIRE Guidelines

<https://guidelines.openaire.eu/en/latest/data/index.html>

- 1 Identifier (M)
- 2 AlternateIdentifier (O)
- 3 RelatedIdentifier (MA)
- 4 Creator (M)
- 5 Contributor (MA/O)
- 6 Title (M)
- 7 Publisher (M)
- 8 PublicationYear (M)
- 9 Subject (R)
- 10 Date (M)

OpenAIRE Guidelines

- 11 Language (R)
- 12 ResourceType (R)
- 13 Size (O)
- 14 Format (O)
- 15 Version (O)
- 16 Rights (MA)
- 17 Description (MA)
- 18 GeoLocation (O)

4 Creator

```
1 <creators>
2   <creator>
3     <creatorName>Miller, John</creatorName>
4   </creator>
5   <creator>
6     <creatorName>Smith, Jane</creatorName>
7     <affiliation>OpenAIRE</affiliation>
8     <nameIdentifier nameIdentifierScheme="ISNI"
9                       schemeURI="http://www.isni.org">
10      1422 4586 3573 0476
11    </nameIdentifier>
12  </creator>
13 </creators>
```

Por que Dataverse?

■ Por que usar **Dataverse** e não **DSpace**?



Pontos Observados	Programado Na linguagem	Ferramenta de pesquisa	Interface gráfica	Banco de Dados	Comunicação De interfaces	Processo de instalação	Formatos atendidos
Dataverse	JAVA	Solr	PrimeFaces E Bootstrap	PostgreSQL	OAI -PMH	Fácil, documentação Bem detalhada.	SPSS, STATA, R, XLSX, CSV
Invenio	Flask (Python)	Elasticsearch	WSGI, e Aipo	PostgreSQL	API REST	Complicada, Instalação incompleta	IIIF Image API support
DSpace	JAVA	Solr	JSPUI, XMLUI	PostgreSQL	OAI -PMH	Fácil, Documentação Bem detalhada.	PDF, XML, txt, asc, MARC, JPEG, JPG, GIF, PNG, TIFF, AIFF, RTF, Postscript
CKAN	Python	Solr	ckan plugins	PostgreSQL	Not enabled	Fácil, Documentação Bem detalhada..	XML and JSON CSV, Excel, XML, PDF, RDF

Grupo brasileiro de usuários do Dataverse

<https://groups.google.com/g/dataverse-brasil>

The screenshot displays the Google Groups interface for the 'Dataverse Brasil' group. On the left, there is a navigation sidebar with options like 'Nova conversa', 'Meus grupos', 'Grupos recentes', 'Todos os grupos', and 'Grupos favoritos'. The main content area shows the group name 'Dataverse Brasil' with 234 participants and a search bar. Below this, a list of messages is displayed, including:

- Message 1: From 'eduard...@furg.br', subject 'Fwd: [rda-brazil] Archimatica para preservação de dados do Dataverse -- Dr. Eduardo ...', timestamp '13:28'.
- Message 2: From 'eduard..., washington...', subject 'Fwd: [rda-brazil] Dataverse com armazenamento compatível com S3 -- Muito bacana, Ed...', timestamp '13 de ago.'.
- Message 3: From 'washington..., ... eduard...', subject 'Aplicação "local_pollys.jar" ocupando 100% do processamento -- Foi exatamente o que ...', timestamp '22 de jul.'.
- Message 4: From 'wtonr...@gmail.com', subject 'Fwd: [rda-brazil] Disponibilizado pacote em Português para o Dataverse 4.20 -- ----- F...', timestamp '24 de jun.'.
- Message 5: From 'washingtonsegundo', subject 'Aula sobre Interoperabilidade e Tecnologias Aplicadas a Repositórios de Dados de Pesquis...', timestamp '20 de jun.'.
- Message 6: From 'washingtonsegundo 2', subject 'Venham trabalhar com a gente! -- Em qui, 18 de jun de 2020 07:23, Washington Luís R. d...', timestamp '18 de jun.'.
- Message 7: From 'washingtonsegundo 2', subject 'Palestra sobre Tecnologias de Suporte a Dados de Pesquisa FAIR com o Prof. João Moreir...', timestamp '17 de jun.'.
- Message 8: From 'washingtonsegundo', subject 'Live realizada sobre Repositórios de Publicações e Dados de Pesquisa -- Caros, Segue li...', timestamp '12 de jun.'.
- Message 9: From 'washingtonsegundo', subject 'Dataverse Community Meetings -- Caros, As informações sobre os eventos Dataverse C...', timestamp '2 de jun.'.
- Message 10: From 'wtonr...@gmail.com', subject 'Fwd: [Dataverse-Users] #Dataverse2020 Registration and Schedule -- Início da mensag...', timestamp '28 de mai.'.
- Message 11: From 'washingtonsegundo 2', subject 'Vem trabalhar com a gente! -- Temos diversas vagas para realização de trabalhos em pr...', timestamp '26 de mai.'.

Topologia da rede

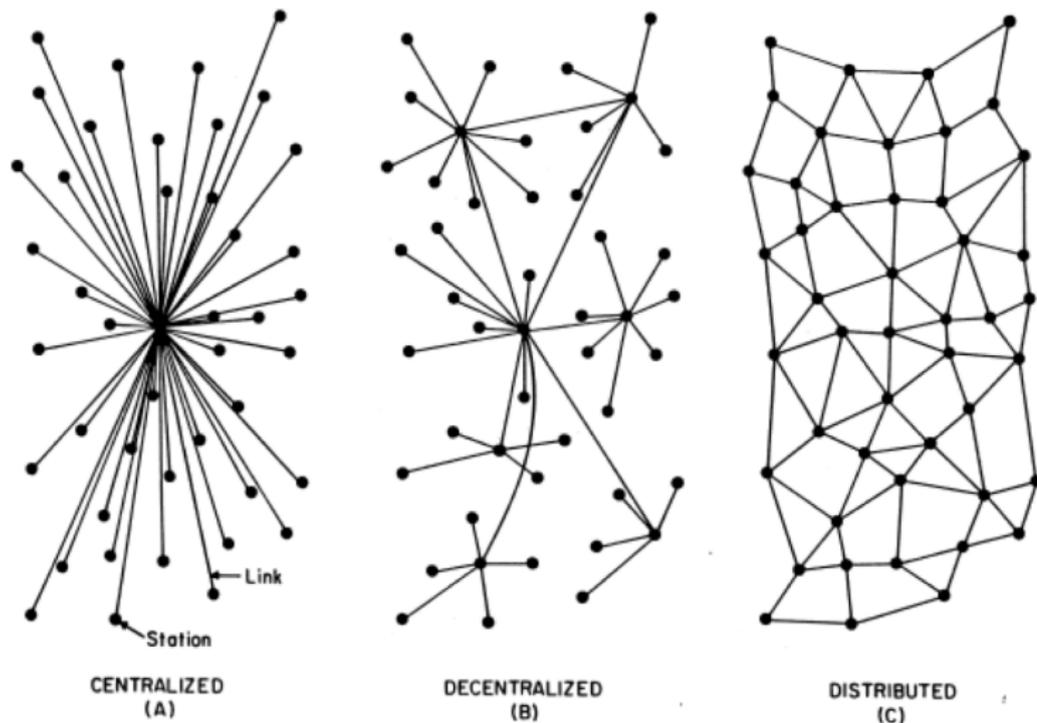


FIG. 1 - Centralized, Decentralized and Distributed Networks

5. Pubfair architecture

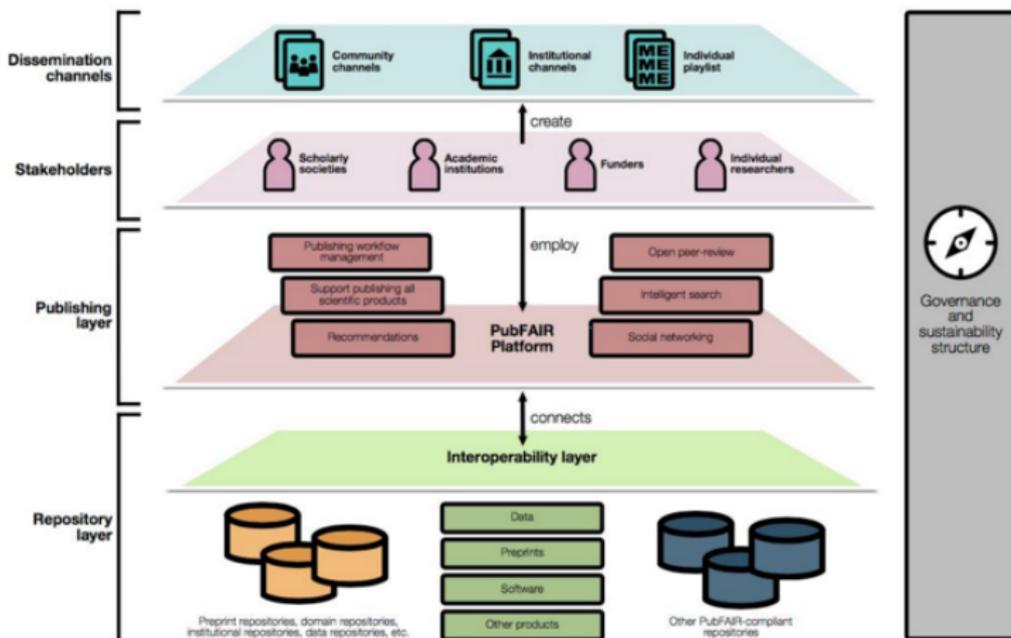
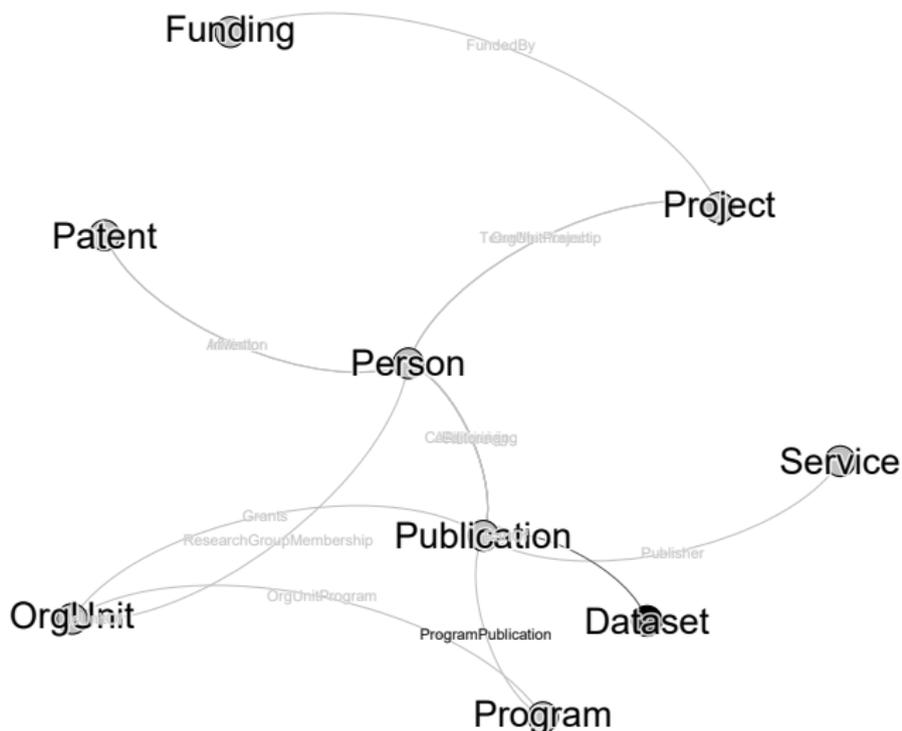
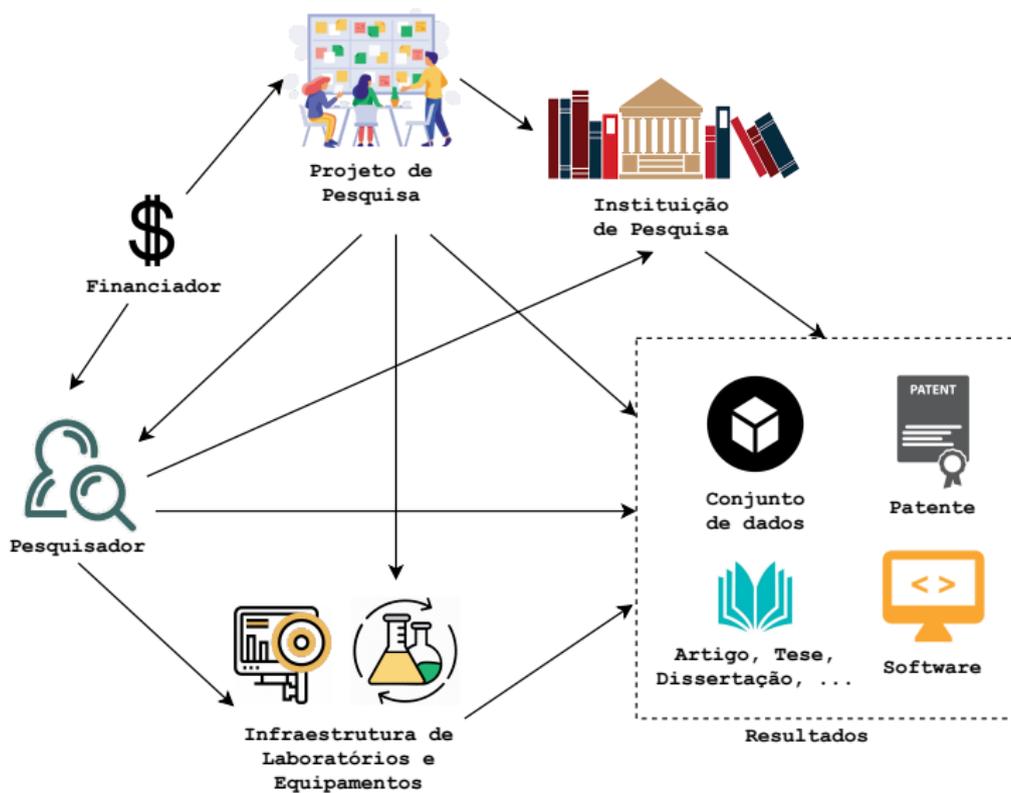
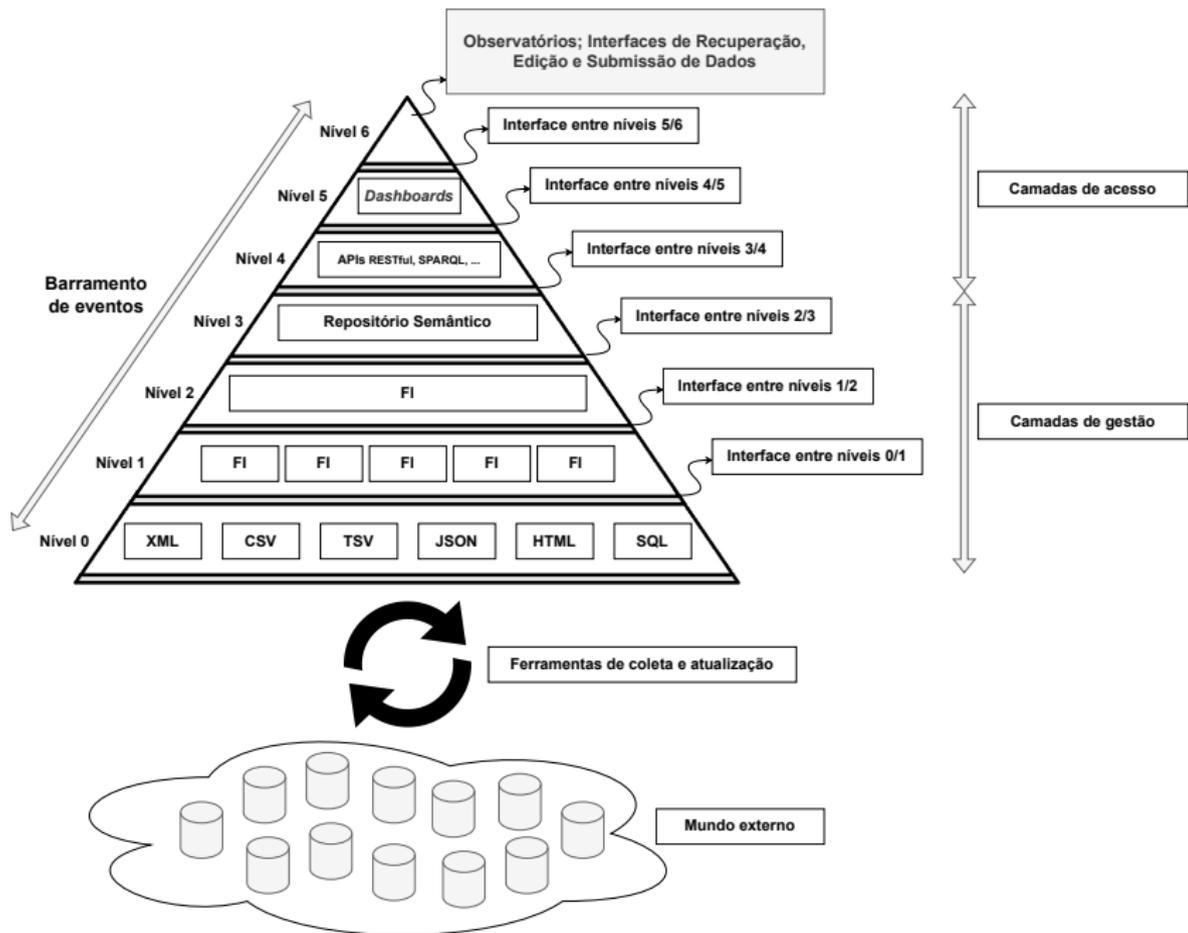


Figure 1 - Pubfair high level architecture

Ecossistema dos Dados







Conclusões

- **Cooperação** entre instituições com o objetivo de fomentar a criação de repositórios de dados e troca de experiências
- Criação de **política** para o repositório
- Criação de materiais de **capacitação e treinamento**
- Coleta de **metadados** dos dados de pesquisa pelo **Oasisbr**

Reflexão

Temos que nos **unir** para conseguirmos ter uma
Lei de Ciência Aberta, em **nível nacional**.

Necessitamos **convencer**
Deputados e **Senadores** a respeito da
importância de termos uma Lei!

