

Repositórios digitais abertos: engajamento, tecnologias e impacto

XXVIII Semana do Livro e da Biblioteca na USP 2025

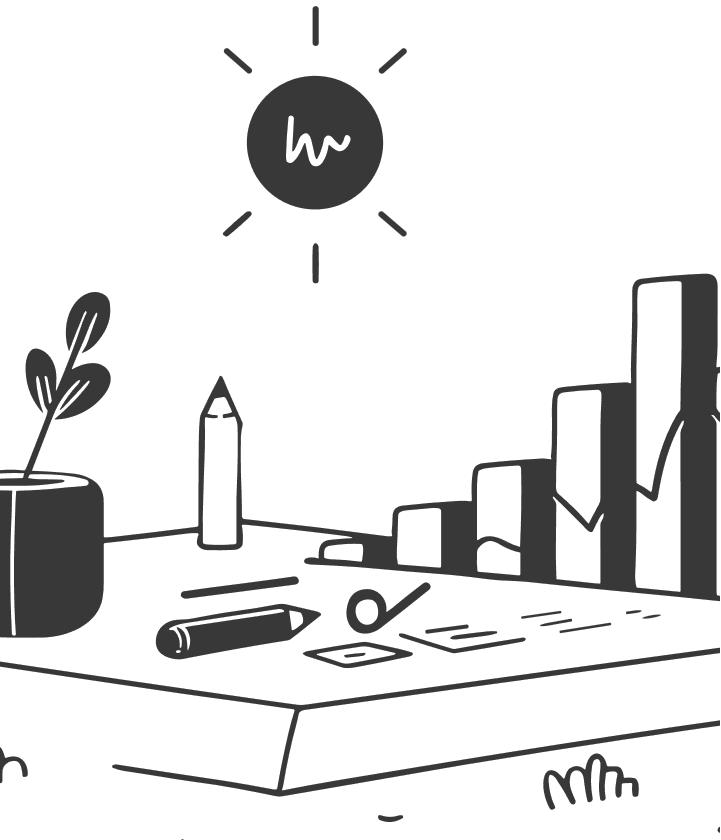
Quatro décadas preservando e compartilhando o conhecimento da USP: o legado das Bibliotecas

21 de outubro de 2025



Washington Segundo, washingtonsegundo@ibict.br





Estratégias de publicação e maximização de impacto

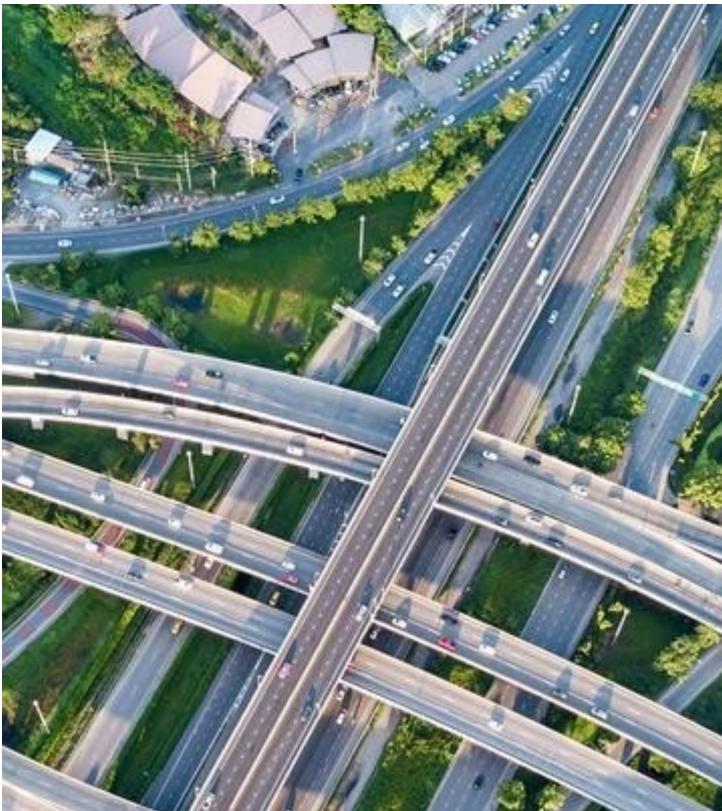
Publicar é apenas o primeiro passo. Estratégias inteligentes de disseminação e gestão da sua presença acadêmica são essenciais para maximizar o impacto da sua pesquisa.

A comunicação científica como infraestrutura

Devemos reconceituar a publicação acadêmica como infraestrutura pública essencial, comparável a estradas, eletricidade ou sistemas de água. Assim como as sociedades investem em infraestrutura física para possibilitar atividades econômicas e sociais, o investimento em infraestrutura de comunicação acadêmica aberta possibilita a pesquisa e a produção de conhecimento.

- ▣ O controle comercial da infraestrutura de comunicação científica cria os mesmos problemas que o controle privado de serviços públicos essenciais: acesso restrito, custos excessivos e priorização do lucro em detrimento do benefício público.

O investimento público em infraestrutura aberta garante que a comunicação acadêmica sirva às comunidades de pesquisa e à sociedade, em vez dos interesses dos acionistas. Isso requer compromisso político e financiamento sustentável comparável a outras prioridades de infraestrutura.



Fonte: UN PRI Infrastructure

A missão do Ibict

O Instituto Brasileiro de Informação em Ciência e Tecnologia (Ibict) é o órgão nacional de referência em informação científica e tecnológica, atuando sob a coordenação do Ministério da Ciência, Tecnologia e Inovação (MCTI).

Missão central: promover o acesso, a organização sistemática e a disseminação ampla da informação científica produzida no Brasil, contribuindo para o fortalecimento da pesquisa nacional e o desenvolvimento científico-tecnológico do país.

Desde sua criação, o Ibict tem sido fundamental na construção de infraestruturas que garantem a preservação e o compartilhamento do conhecimento científico brasileiro com a comunidade nacional e internacional.

Acesso

Garantir disponibilidade ampla da informação científica

Organização

Estruturar e padronizar o conhecimento produzido

Disseminação

Promover visibilidade internacional da ciência brasileira

Contexto da Informação Científica

A comunicação científica passou por uma transformação radical nas últimas décadas, impulsionada pela digitalização e pela internet. Este cenário criou novas oportunidades e desafios para a gestão do conhecimento científico.

Revolução digital

A comunicação científica digital cresceu exponencialmente, transformando como pesquisadores compartilham e acessam conhecimento globalmente.

Infraestrutura necessária

Demanda por plataformas robustas que garantam o acesso aberto e a preservação de longo prazo da produção científica nacional.

Papel dos Repositórios

Os Repositórios Institucionais emergem como ferramentas essenciais para democratizar o acesso e preservar a memória científica brasileira.



O Que são Repositórios Institucionais?



Plataformas digitais

Sistemas organizados para armazenar, preservar e disseminar a produção científica de instituições de pesquisa e ensino.



Acesso Aberto

Garantem acesso público e gratuito ao conhecimento produzido, democratizando a ciência brasileira.



Interoperabilidade

Conectam-se a redes nacionais e internacionais, ampliando a visibilidade da pesquisa brasileira.

Metadados e interoperabilidade



Padrões de metadados
Adoção de esquemas reconhecidos internacionalmente.

Visibilidade global
Indexação em mecanismos internacionais de busca.

Integração de sistemas
Conexão entre diferentes plataformas e repositórios.

Descoberta de conteúdo
Facilita a localização de pesquisas em múltiplas bases.

Rede Brasileira de Repositórios Digitais

RBRD: Articulação nacional

A Rede Brasileira de Repositórios Digitais, coordenada pelo Ibict, representa uma iniciativa fundamental para a articulação de comunidades de prática em todo o país.

Esta rede favorece a troca constante de experiências entre instituições, promove a padronização de procedimentos, metadados e políticas institucionais, e fortalece o ecossistema nacional de repositórios.





A Rede Federada de Repositórios Institucionais de Publicações Científicas (LA Referencia) é uma iniciativa colaborativa que conecta os sistemas nacionais de informação científica dos países da América Latina.

Interoperabilidade regional

Permite que repositórios brasileiros compartilhem metadados e conteúdo com toda a América Latina de forma padronizada.

Visibilidade Internacional

Aumenta significativamente a descoberta e o impacto da produção científica brasileira em plataformas globais.

Coordenação nacional

O Ibict atua como nó nacional, coordenando a participação brasileira e garantindo conformidade com padrões.

Contexto



BrCris

Ecossistema de Informação da
Pesquisa Científica Brasileira



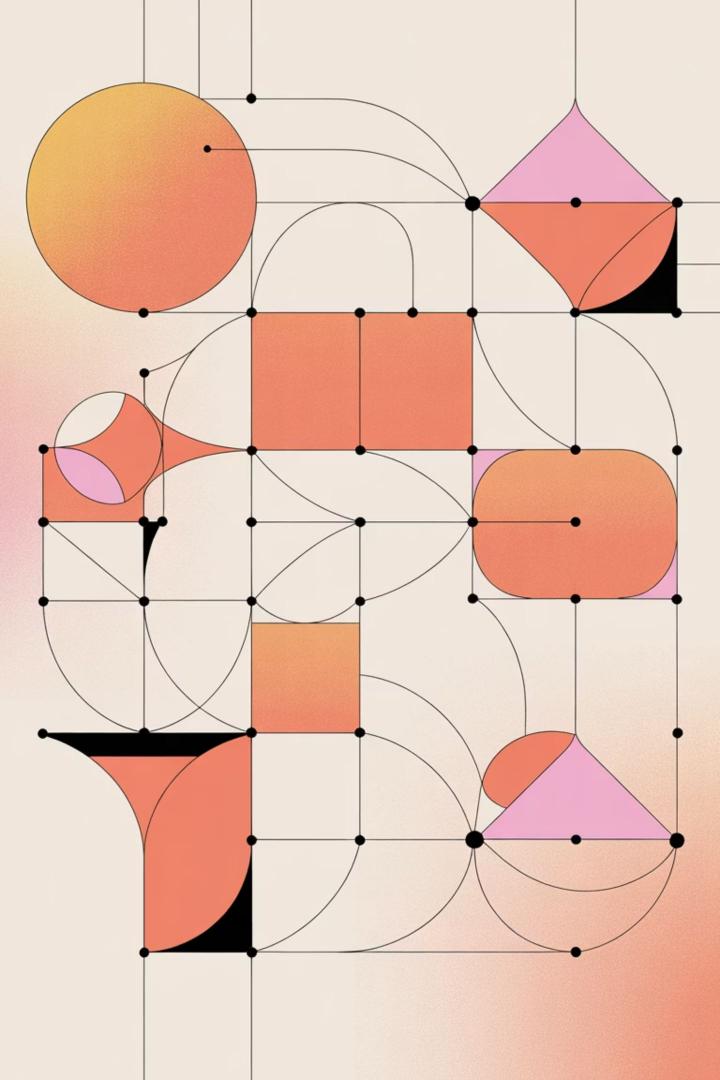
Oasisbr

Portal que agrega mais 6 milhões de
objetos digitais



Laguna

Infraestrutura de informação aberta e
lago de dados



Reprodutibilidade operacionalizada

A reproducibilidade, um pilar fundamental da Ciência Aberta, deixa de ser apenas um ideal metodológico abstrato e se torna um processo concreto operacionalizado através de múltiplas tecnologias integradas.

1

Oasisbr

É um portal agregador que reúne e disponibiliza a produção científica de repositórios digitais brasileiros. Sua missão é promover o acesso aberto e a visibilidade do conhecimento nacional, funcionando como uma plataforma de busca unificada.

2

BrCris & Laguna

é BrCris é uma infraestrutura nacional de informações sobre pesquisa, ciência, tecnologia e inovação. Já o Laguna é um ambiente computacional em nuvem que oferece ferramentas e serviços para analisar os grandes volumes de dados, por exemplo os contidos no BrCris.

Topologia da rede brasileira de repositórios digitais



Instituições produtoras

Universidades, institutos de pesquisa e laboratórios que geram dados científicos por meio de suas atividades de pesquisa.

Repositórios locais

Infraestruturas institucionais para armazenar e gerenciar coleções de dados específicas, adaptadas às necessidades individuais de cada organização.

Repositórios temáticos

Agregadores de dados por área de conhecimento com curadoria especializada, reunindo pesquisas de disciplinas específicas.

Metabuscador nacional

Portal de acesso único e integrado que se conecta a redes internacionais, fornecendo uma interface de pesquisa abrangente.

Laguna: Lago de dados científicos

Fonte de acesso abrangente

Laguna é um projeto de lago de dados científicos concebido como um repositório abrangente que acomoda dados de diferentes padrões e níveis de tratamento, tanto em formatos estruturados como não estruturados.

Fontes de dados FAIR

Os conjuntos de dados são coletados de fontes nacionais e internacionais em conformidade com FAIR, com o OpenAlex servindo como a principal fonte de informações para publicações científicas globais.

Processamento multinível

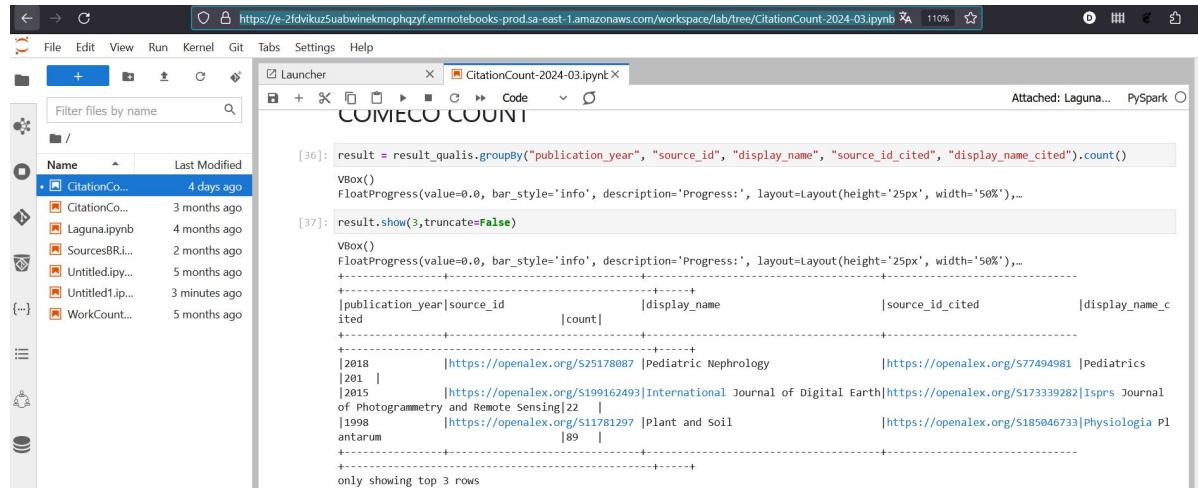
Os dados são processados por meio de múltiplas vias analíticas, permitindo diversas metodologias de pesquisa e suportando vários níveis de refinamento de dados para diferentes necessidades de pesquisa.



Citações em revistas

A análise de citações revela padrões críticos nas redes de comunicação acadêmica. Ao analisar os fluxos de citações entre revistas no conjunto de dados OpenAlex, podemos identificar publicações influentes, grupos de pesquisa emergentes e padrões de transferência de conhecimento interdisciplinar que moldam o panorama acadêmico.

A visualização demonstra relações complexas de citação entre periódicos acadêmicos, com o tamanho dos nós representando o volume de publicações e a espessura das bordas indicando a frequência de citações. Essa análise de rede permite a identificação de periódicos centrais que servem como centros de conhecimento dentro de suas respectivas disciplinas.

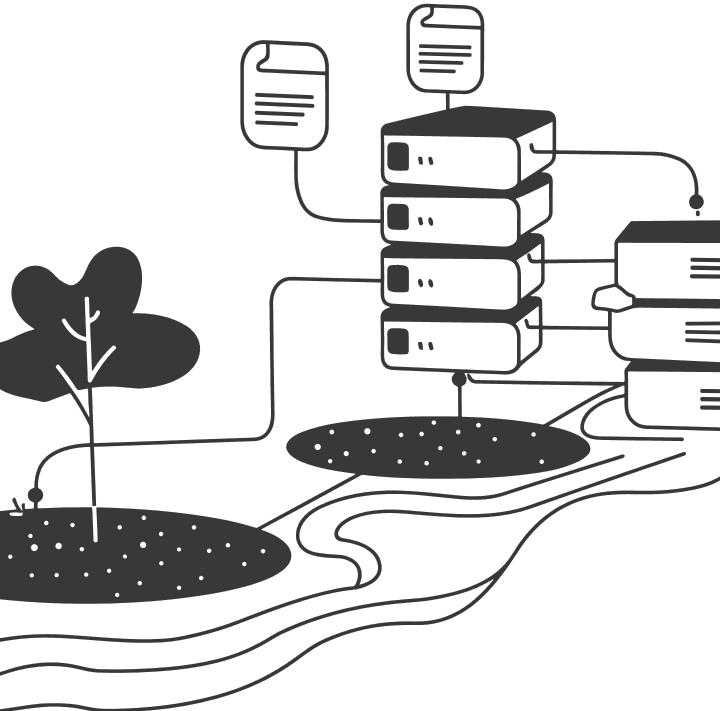


The screenshot shows a Jupyter Notebook interface with a sidebar containing file navigation and a search bar. The main area has a title 'CitationCount' and displays two code cells and their corresponding outputs.

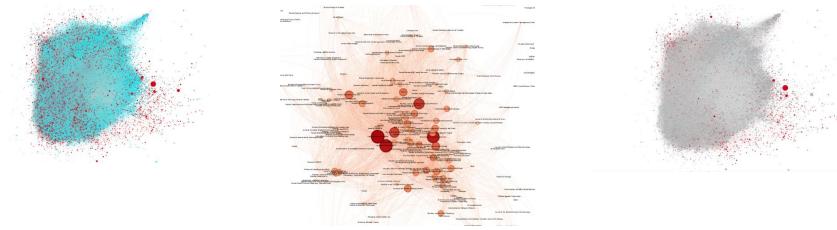
```
[36]: result = result_qualis.groupBy("publication_year", "source_id", "display_name", "source_id_cited", "display_name_cited").count()
VBox()
FloatProgress(value=0.0, bar_style='info', description='Progress:', layout=Layout(height='25px', width='50%'),...
```

```
[37]: result.show(3,truncate=False)
VBox()
FloatProgress(value=0.0, bar_style='info', description='Progress:', layout=Layout(height='25px', width='50%'),...
+-----+
|publication_year|source_id|display_name|source_id_cited|display_name_cited|
+-----+-----+-----+-----+
|[2018|https://openalex.org/S25178087|Pediatric Nephrology|https://openalex.org/S77494981|Pediatrics
|[2015|https://openalex.org/S199162493|International Journal of Digital Earth|https://openalex.org/S173339282|Isprs Journal
of Photogrammetry and Remote Sensing|22|Plant and Soil|https://openalex.org/S11781297|Physiology Pl
|[1998|https://openalex.org/S11781297|Plant and Soil|https://openalex.org/S185046733|Physiology Pl
|antarum|89|
```

only showing top 3 rows



Atração entre citações em revistas médicas brasileiras

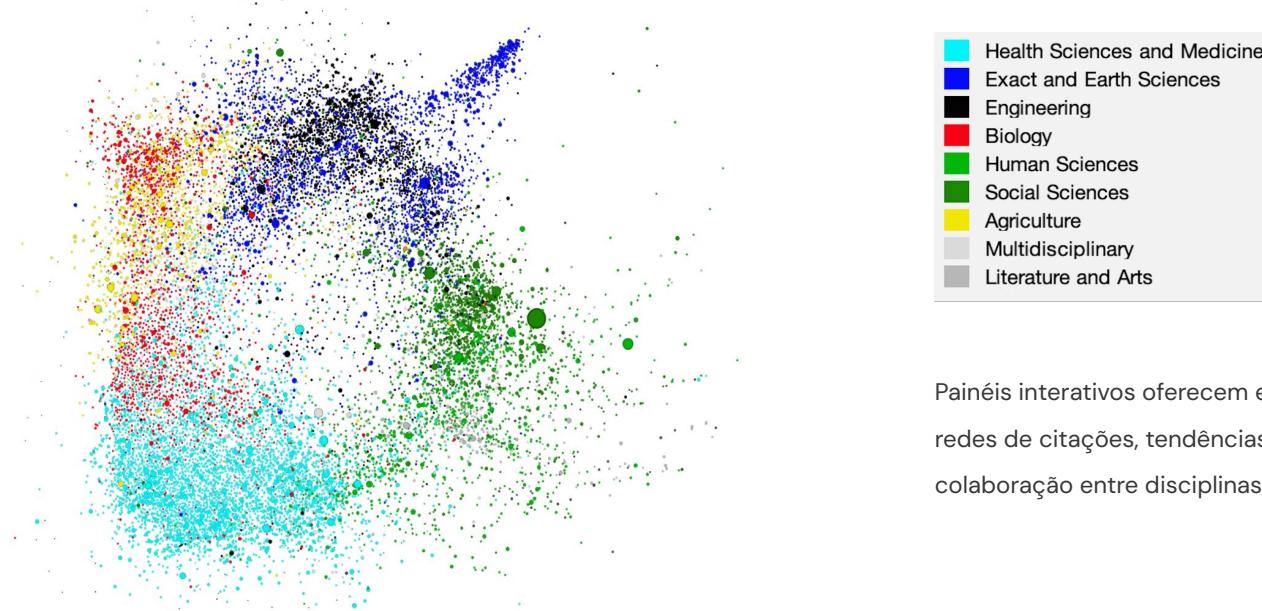


□ **Citação da pesquisa:** de Carvalho-Segundo, Washington & Canto, Fábio & Pinto, Adilson & Sundfeld, Daniel. (2024). Atração entre periódicos brasileiros de medicina análise a partir de dados de citação do OpenAlex. Encontro Brasileiro de Bibliometria e Cientometria.

Esta pesquisa utiliza dados de citações do OpenAlex para analisar padrões de atração entre periódicos médicos brasileiros, revelando redes colaborativas e fluxo de conhecimento dentro do ecossistema de pesquisa médica do país.

Visualizações

Técnicas avançadas de visualização transformam dados biométricos complexos em insights açãoáveis. Essas representações gráficas permitem que os pesquisadores identifiquem rapidamente tendências, padrões e relações em enormes conjuntos de dados acadêmicos que seriam impossíveis de discernir apenas através da análise de dados brutos.



Painéis interativos oferecem exploração em tempo real de redes de citações, tendências de publicação e padrões de colaboração entre disciplinas.

BrCris: uma plataforma para integração, interoperabilidade e análise da produção científica brasileira



O BrCris representa uma visão ambiciosa para uma infraestrutura abrangente de informações sobre pesquisa no Brasil. Ao integrar dados de várias fontes — publicações, pesquisadores, instituições, agências de financiamento e conjuntos de dados —, o BrCris pode fornecer uma visão sem precedentes da produção científica brasileira, mantendo o controle público sobre os dados e análises de pesquisa.

Objetivos principais

Integração: conectar fontes de dados fragmentadas em um ecossistema coerente

Interoperabilidade: possibilitar a troca de dados com sistemas internacionais

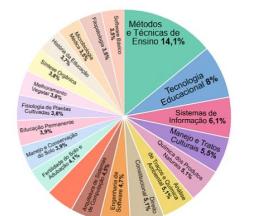
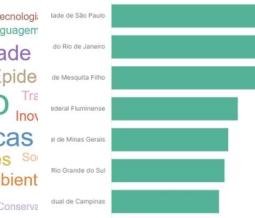
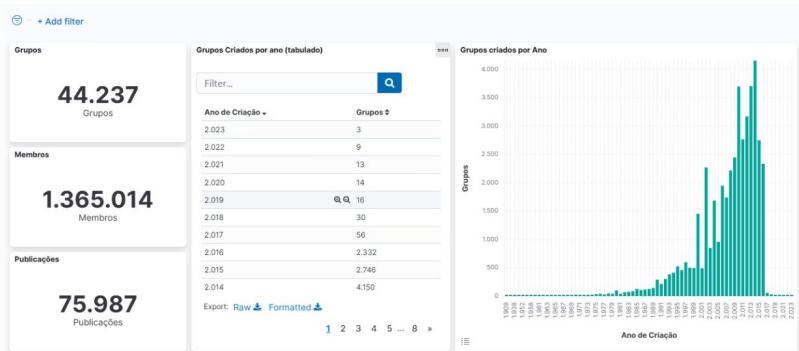
Análise: apoiar avaliações sofisticadas de pesquisa e planejamento de políticas

Transparência: proporcionar visibilidade pública das atividades de pesquisa e financiamento

Soberania: manter o controle brasileiro sobre a infraestrutura de informações de pesquisa

A implementação do BrCris requer compromisso político sustentado, financiamento adequado e coordenação entre instituições e agências. Esse investimento em infraestrutura é essencial para o desenvolvimento científico do Brasil e para desafiar as estruturas globais injustas de comunicação acadêmica. A questão não é se podemos nos dar ao luxo de construir o BrCris, mas se podemos nos dar ao luxo de não fazê-lo.

Painéis interativos e exportação de dados



Filtragem flexível

Diversos filtros em todos os campos da entidade permitem análises específicas e subconjuntos personalizados, permitindo que os pesquisadores aprofundem-se em segmentos de dados precisos relevantes para suas investigações.

Exportação em vários formatos

CSV para planilhas e análise VOSviewer

GraphML para visualização de rede Gephi

RDF para integração de endpoint SPARQL

Redes de colaboração científica

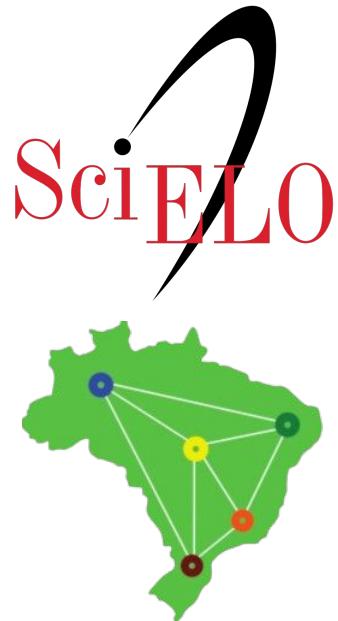
A visualização das redes de colaboração permite:

- Análise das conexões entre pesquisadores
- Identificação de padrões de coautoria
- A descoberta de comunidades científicas
- Mapeamento da interdisciplinaridade
- O reconhecimento de pesquisadores centrais em suas áreas

O gráfico interativo facilita a exploração das relações, permitindo ajustes de filtros e parâmetros para análises personalizadas.



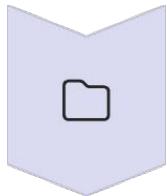
Iniciativas Brasileiras: SciELO e Outras Plataformas



Iniciativas como o **SciELO**, pioneiro desde 1998 e agora consolidado como plataforma nacional de Ciência Aberta com o apoio de CAPES, FAPESP e CNPq, são referências mundiais.

É fundamental mapear e integrar a totalidade das iniciativas existentes em instituições brasileiras, como **Repositórios de Dados e Publicações**, **Portais de Periódicos**, **PGD-BR**, **Rede Moara**, **Plataforma Civis**, **RBRD**, entre outras, para fortalecer o ecossistema da Ciência Aberta no país.

Coleta de dados (período 2000 a 2024)



Repositórios (132)

500.132 registros de **artigos de revista em repositórios**, no **Oasisbr**



OpenAlex Brasil

2.834.342 artigos com pelo menos um autor afiliado a uma instituição brasileira

Metodologia

Coleção de artigos em repositórios

421.243 artigos deduplicados do **Oasisbr**

Coleção OpenAlex Brasil

1.901.216 artigos de revista com **autores somente de afiliação em instituição brasileira**

Metodologia

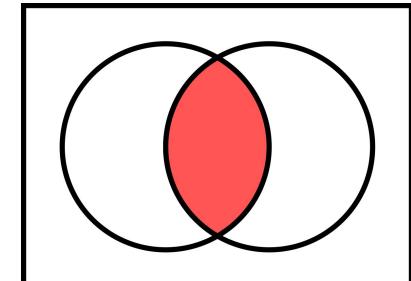
Correspondência de registros

Títulos semelhantes com distância de **Levenshtein** ≤ 5 e Ano de publicação =

Comparação e análise de citações

153.671 artigos presentes em ambos os conjuntos analisados

Comparação de média de citações entre classes



Metodologia de Comparação

A análise bibliométrica proposta estabelece uma distinção fundamental entre duas classes de registros científicos para avaliar o impacto das citações. Artigos presentes:

Classe 1

revistas científicas e repositórios digitais

α (alfa): Total de citações da Classe 1

n: Número total de artigos da Classe 1

Classe 2

somente em revistas científicas

β (beta): Total de citações da Classe 2

m: Número total de artigos da Classe 2

Parâmetros Matemáticos da Análise

- As médias de citações são calculadas como
 $A = \alpha/n$ (Classe 1) e $B = \beta/m$ (Classe 2),
permitindo análises comparativas precisas entre os diferentes tipos de publicação.

$$\text{Crescimento} = (A - B) / B$$

Análise de Crescimento: 4 Dimensões

Análise Temporal



Comparação **por ano de publicação**, revelando tendências evolutivas do impacto de citações ao longo do tempo e identificando padrões sazonais na produção científica.

Análise por h-index



Segmentação **por percentil do h-index da revista**, buscando mensurar a influência do veículo de publicação



Análise por Domínio

Classificação **por domínio científico** segundo a taxonomia OpenAlex, permitindo comparações entre grandes áreas do conhecimento.



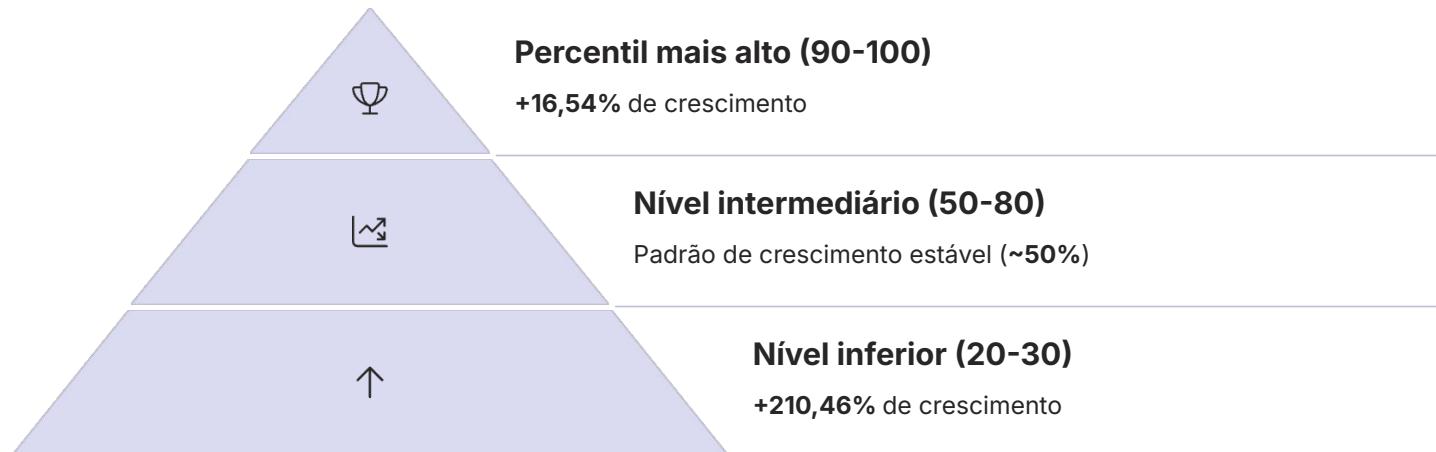
Análise por Campo

Segmentação **por campo específico** conforme classificação OpenAlex, oferecendo granularidade detalhada para especialidades acadêmicas.

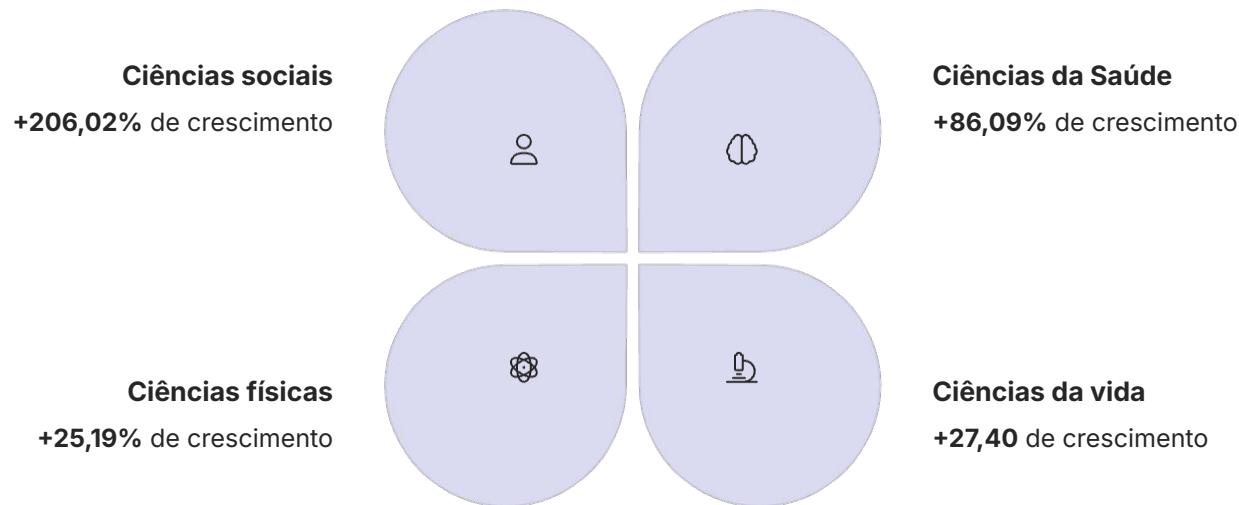
Crescimento por ano de publicação



Crescimento por índice H da revista



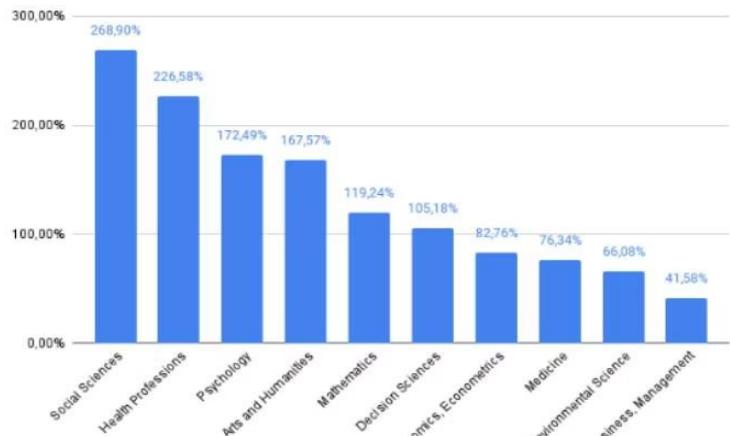
Crescimento por domínio de conhecimento



Crescimento por campo do conhecimento

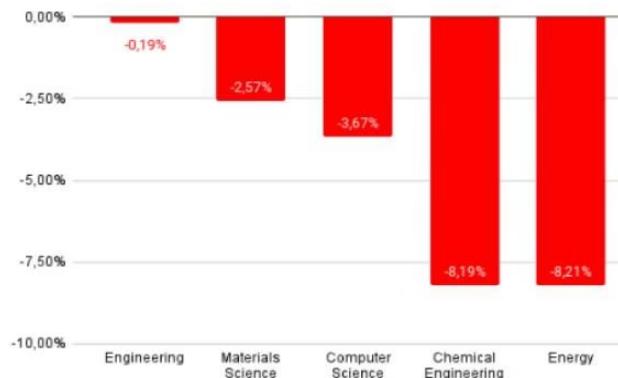
Principais campos de crescimento

- Ciências sociais: **+268.90%**
- Economia: **+82.76%**
- Negócios: **+41.58%**



Campos com crescimento negativo

- Engenharia: **-0,19%**
- Ciência da Computação: **-3,67%**
- Energia: **-8,21%**



Conclusões

Impacto na visibilidade

Os repositórios aumentam a visibilidade dos artigos

Diferenças entre campos

As Ciências Sociais são as mais beneficiadas, enquanto a Engenharia é a menos beneficiada

Impacto da revista

As revistas de nível médio e baixo impacto são as mais beneficiadas

Possíveis canais de amplificação da disseminação

GS, Oasisbr, LA Referencia, OpenAIRE, etc. aumentam a visibilidade

+83.10%

Crescimento médio

Para artigos em repositórios vs. somente em revistas

+206.02%

Ciências sociais

Maior crescimento por domínio

+25.19%

Ciências físicas

Menor crescimento positivo por domínio

-8.21%

Engenharia

Menor crescimento geral por campo

Obrigado!

washingtonsegundo@ibict.br

