

Ciência de Dados aplicada ao Arca – Repositório Institucional da Fiocruz: relato de experiência

Luciana Danielli
Claudete Fernandes
Bibliotecárias

Coordenação RI Arca

Fundação Oswaldo Cruz - Fiocruz
Instituto de Comunicação e Informação Científica e Tecnológica em Saúde – ICICT

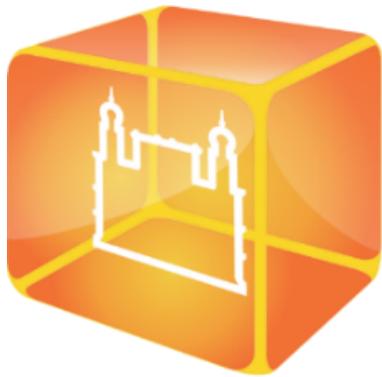
Rio de Janeiro, 25 de abril de 2019.



Fonte foto: Freedigitalphotos.Net

“Os repositórios digitais (RDs) são bases de dados online que reúnem de maneira organizada a produção científica de uma instituição, armazenando arquivos de diversos formatos. Resultam em uma série de benefícios tanto para os pesquisadores quanto para as instituições ou sociedades científicas, proporcionando maior visibilidade aos resultados de pesquisas e possibilitando a preservação da memória científica da instituição. Os Repositórios podem ser institucionais ou temáticos” (INSTITUTO BRASILEIRO..., 2016).

- ✓ Gerenciamento da produção científica;
- ✓ Acesso à informação científica;
- ✓ Crescimento da visibilidade dos trabalhos disponibilizados;
- ✓ Institucionalização da produção intelectual produzida;
- ✓ Confiabilidade das informações;
- ✓ Preservação Digital dos documentos;
- ✓ Disponibilização da Produção intelectual num único lugar;
- ✓ Participar do Movimento Acesso Livre;
- ✓ Integração com outros sistemas de informação;
- ✓ Acesso aos dados estatísticos.



arca

Missão: Reunir, hospedar, preservar, disponibilizar e dar visibilidade à produção intelectual da Fundação Oswaldo Cruz.

<http://www.arca.fiocruz.br/>



Repositório Institucional da Fiocruz

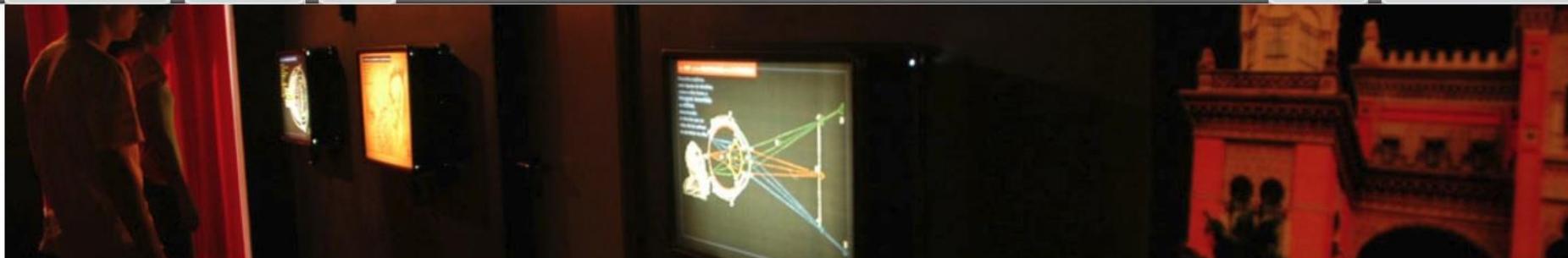
[Página inicial](#)

[Navegar](#) ▾

[Ajuda](#)

[Idioma](#) ▾

[claudete.que...](#) ▾



[Autoarquivamento](#)

[Área Pessoal](#)

[Comunidades & Coleções](#)



Novos documentos depositados

[RSS 1.0](#)

[RSS 2.0](#)

[RSS](#)

The physician in the face of death in the emergency room

OBJETIVO: Analisar a forma como médicos, enquanto parte de um grupo sociocultural, lidam com diferentes tipos de morte, em um serviço de pronto socorro metropolitano.

MÉTODOS: Trata-se de uma etnografia realizada em um dos maiores serviços de pronto socorro da América Latina. A coleta dos dados deu-se ao longo de nove meses de observação participante e entrevistas com 43 médicos de diferentes especialidades – 25 homens e 18 mulheres, entre 28 e 69 anos. RESULTADOS: À análise, guiada ...



ARCA

- ▶ Sobre
- ▶ Equipe
- ▶ Termos de uso
- ▶ FAQ
- ▶ Fanpage Arca

Acesso Aberto

- ▶ O que é
- ▶ Glossário
- ▶ Direitos Autorais
- ▶ Política

Visualização de Dados



▶ Veja os dados do Arca de forma dinâmica

Fonte: <http://www.arca.fiocruz.br/>

- ✓ 2009 - Projeto do Laboratório de Informação Científica e Tecnológica em Saúde – LICTS/ICICT;
- ✓ 2011 - O Arca é lançado oficialmente como Repositório Institucional da Fiocruz;
- ✓ 2011 - Desenvolvimento do primeiro Manual de preenchimento de metadados para entrada dos documentos no RI;
- ✓ 2014 - Instituída a Política de Acesso Aberto ao Conhecimento na Fiocruz, através da Portaria 329 que estabelece a seguinte estrutura de governança: Comitê da Regulação da Política de Acesso Aberto ao Conhecimento; Comitê Gestor do Repositório Institucional Arca; e Núcleos de Acesso Aberto ao Conhecimento (NAACs);
- ✓ 2014 – A Política tem caráter mandatório inicialmente no que se refere ao depósito no Repositório Institucional Arca das dissertações e teses defendidas nos Programas de Pós-graduação da Fiocruz e dos artigos científicos publicados em periódicos;
- ✓ 2014 - Lançado o Plano Operativo que define responsabilidades, infraestrutura e fluxos para operação e manutenção do RI e para as coleções mandatórias - teses, dissertações e artigos;
- ✓ 2014 – O Arca se torna um Indicador Global de Desempenho para o Governo Federal. A avaliação do Indicador é realizada pelo crescimento anual de depósitos no RI;

- ✓ 2015 – Atualização do Manual de preenchimento de metadados para entrada dos documentos no RI (3. ed.) <http://www.arca.fiocruz.br/handle/icict/13257>;
- ✓ 2015 - Desenvolvimento Manual de preenchimento de metadados no Arca - Recursos Educacionais <http://www.arca.fiocruz.br/handle/icict/16420>;
- ✓ 2017 - Criação do Grupo de Estudos Dspace (com a participação dos Bibliotecários e Analistas) visando estudo das novas versões e implementações do Software;
- ✓ 2018 – Projeto de Ciência de Dados aplicada à Saúde para Visualização dos Dados no RI Arca;
- ✓ 2019 – Em desenvolvimento o plano de preservação digital do Arca;
- ✓ 2019 - Atualmente o RI Arca possui 26 comunidades (abril/2019);
- ✓ 2019 - Total de 27.728 objetos digitais (abril/2019).



Luciana Danielli
Chefe da Seção de Informação



Claudete Fernandes
Coordenadora do RI Arca



Éder Freyre
Técnico de Biblioteca



Raphael Belchior
Bolsista de
Biblioteconomia



Angelo
Analista de
Sistemas -
Desenvolvimento



Vinicius
Analista de
Sistemas -
Desenvolvimento



Lucas
Analista de
Sistemas -
Infraestrutura



- ✓ O RI Arca é coordenado pelo Instituto de Comunicação e Informação Científica e Tecnológica em Saúde (ICICT) e encontra-se organizado internamente em comunidades que correspondem às unidades técnico científicas da Fiocruz;
- ✓ Cada comunidade possui informações sobre o gestor, homepage, coleções e quantitativo;
- ✓ **Botão para realização de Autoarquivamento;**
- ✓ Interoperabilidade com sistemas das bibliotecas Fiocruz/Importação direta dos catálogos das Bibliotecas para o Arca;
- ✓ Inclusão da ferramenta Altmetric – métricas alternativas (Twitter, Facebook, etc.);
- ✓ Participação em redes e diretórios de RIs - OASIS.BR (IBICT), LaReferencia – (Rede de Repositórios da América Latina), COAR (Confederation of Open Access) e formatação de parceria com a Universidade do Minho (Portugal) (<https://repositorium.sdum.uminho.pt/>);
- ✓ **Inclusão de cláusula nos editais do PIBIC, PIBITI e INOVA da Fiocruz** com a informação “Apresentação do comprovante de depósito da produção intelectual no Repositório Institucional ARCA (www.arca.fiocruz.br), para fins de disponibilização pública em acesso aberto, nos termos e de acordo com o estabelecido na Política Institucional de Acesso Aberto da Fiocruz”.



- ✓ Realização de Reuniões com Unidades da Fiocruz e instituições externas;
- ✓ Realização de palestras nas Unidades da Fiocruz e em Instituições externas;
- ✓ Realização de treinamentos com Unidades da Fiocruz e Instituições externas;
- ✓ Fortalecimento do povoamento no RI (autoarquivamento, migrações, importações);
- ✓ Mobilização permanente (atendimento presencial, por telefone e-mail);
- ✓ Apresentação de trabalhos sobre o RI Arca em Congressos e Seminários;
- ✓ Participação no Projeto PIPDT – interface para DSpace – segundo critérios de usabilidade e acessibilidade;
- ✓ Inclusão de artigos sobre Zika produzidos por pesquisadores da Fiocruz (Emergência Zika 2016);
- ✓ Fanpage criada em 2015 com + 1.100 seguidores (abril/2019);
- ✓ Participação na Rede Sudeste de Repositórios, criada em 2017, que congrega 43 Instituições de Ensino e Pesquisa visando troca de experiências e disseminação do conhecimento. **Nos dias 28 a 30 de maio será realizado o “I Encontro da Rede Sudeste de Repositórios Institucionais”. Neste evento será realizada a palestra: “Ciência de Dados e Repositórios: identificando soluções” - que será ministrada pelo servidor Jefferson Lima.**

- ✓ Atualmente está na v. 4.7 caminhando para a v. 6.3;
- ✓ 1Tb de espaço usado;
- ✓ Máquina virtual com 8 processadores Intel Xeon E3-12xx v2 (Ivy Bridge);
- ✓ 16Gb de Memória RAM;
- ✓ 3 instâncias (Produção, Homologação (testes) e Treinamento);
- ✓ Plano de Backup.

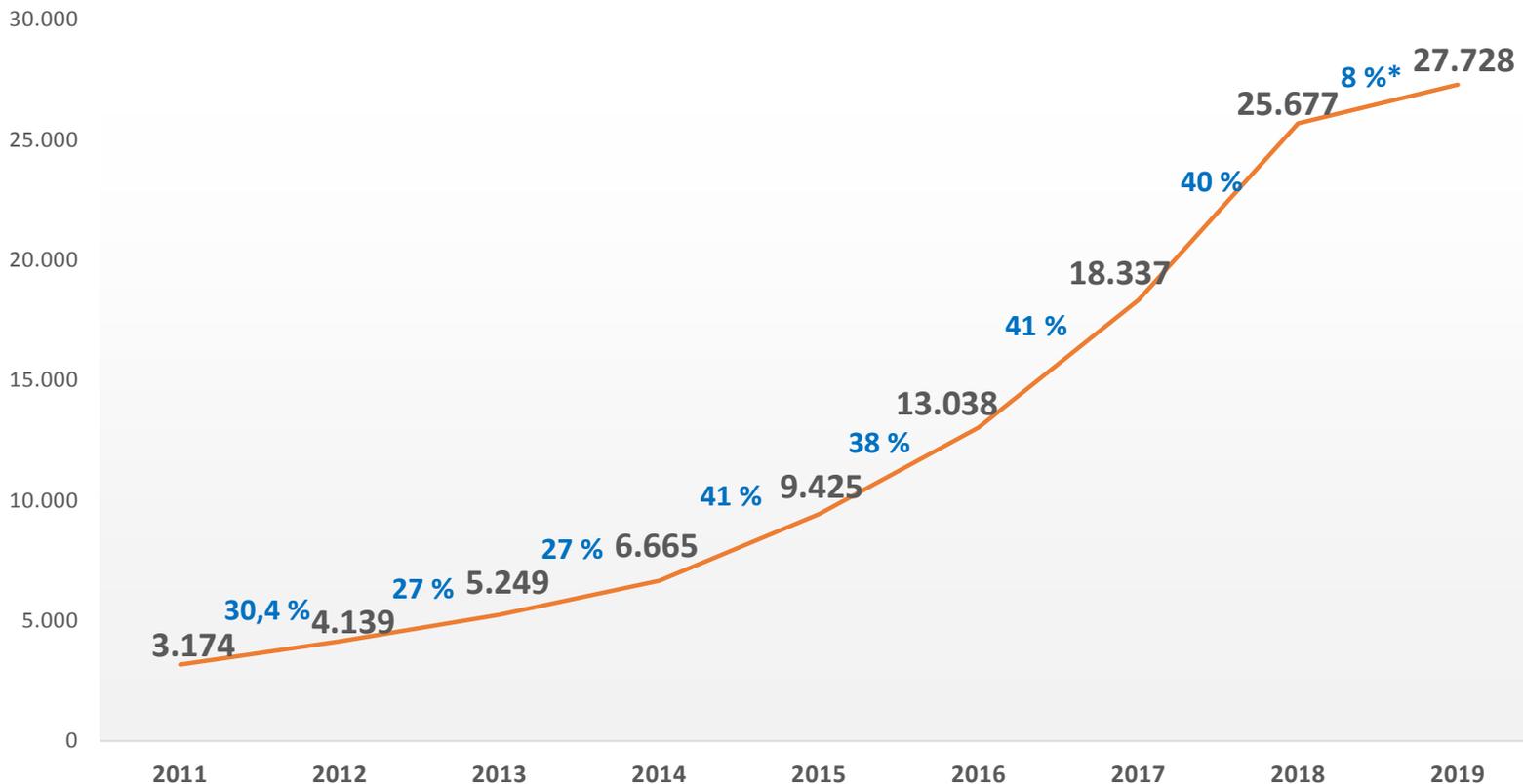


- ✓ Repositório de dados de pesquisa
 - Projeto da VPEIC
 - Arca – realização de experiência piloto desenvolvida na Comunidade do Instituto Oswaldo Cruz, com a inserção de dois dados de pesquisa.
- ✓ Desenvolvimento do Plano de Preservação Digital do Arca;
- ✓ Controle de Autoridade – assunto, nome;
- ✓ Estabelecimento de índices de citação.





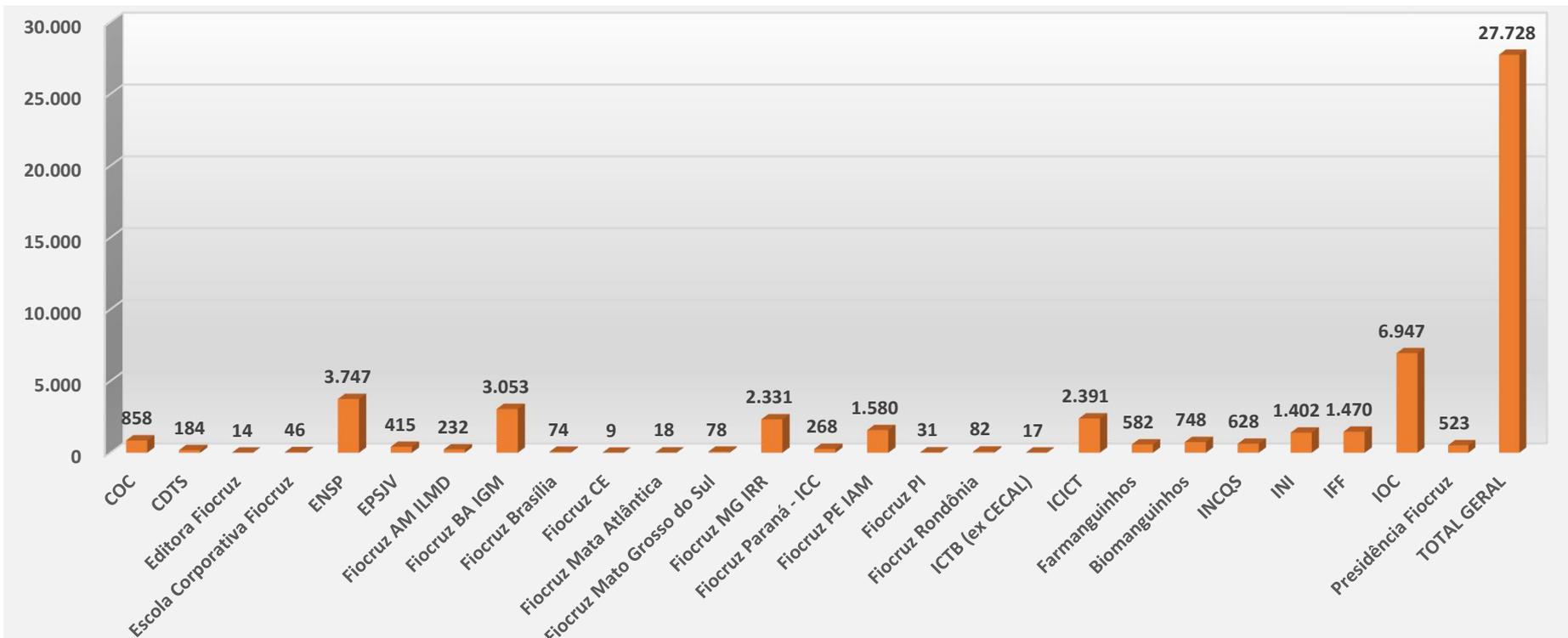
Crescimento Anual Arca*



*Atualizado até 19/04/2019

Fonte: Fiocruz

Crescimento Arca por Unidade - 2011-2019*



*Atualizado até 19/04/2019

Fonte: Fiocruz

Parceria RI Arca e Plataforma de Ciência de Dados aplicada à Saúde



Ciência de Dados aplicada ao Arca

Descrição:



O Arca é o Repositório Institucional da Fundação Oswaldo Cruz (Fiocruz) e sua função é reunir, hospedar, disponibilizar e dar visibilidade à produção intelectual da Instituição; visa estimular a mais ampla circulação do conhecimento, fortalecendo o compromisso institucional com o livre acesso da informação em saúde, além de conferir transparência e incentivar a comunicação científica entre pesquisadores, educadores, acadêmicos, gestores, alunos de pós-graduação, bem como a sociedade civil.

Objetivos do Projeto de Pesquisa, Inovação e Desenvolvimento Tecnológico:

- curadoria de dados: identificar inconsistências no preenchimento dos metadados do Arca, por meio da classificação automática utilizando *machine learning*, e consequente correção, visando qualidade das informações e dos dados extraídos;
- recuperação da informação e visualização de dados: oferecer uma plataforma de exploração interativa para extração e visualização de dados, utilizando filtros e combinações de dados contidos no Arca, como quantidade de produção por tipo de material, por unidade da Fiocruz, assunto, ano, entre outros, e que possam ser manipulados pelas diferentes unidades representadas no repositório institucional.

Equipe:

Claudete Fernandes de Queiroz e Luciana Danielli - Coordenadoras do Projeto

Pesquisadores e técnicos da Plataforma de Ciência de Dados aplicada à Saúde (PCDaS/icict)

Fonte: <https://bigdata.icict.fiocruz.br/ciencia-de-dados-aplicada-ao-arca>



Estabelecimento de parceria entre a equipe do RI Arca e a equipe do Laboratório de Ciência de Dados da Fiocruz em 2018, que culminou no Projeto “Ciência de Dados aplicada ao Arca” e estabeleceu os seguintes objetivos:

- ✓ **Curadoria de dados:** identificar inconsistências no preenchimento dos metadados do Arca, por meio da classificação automática utilizando *machine learning*, e consequente correção, visando qualidade das informações e dos dados extraídos, facilitando o trabalho de curadoria iniciado em 2015 pela equipe do RI Arca;
- ✓ **Recuperação da informação e visualização de dados:** oferecer uma plataforma de exploração interativa para visualização e extração de dados, utilizando filtros e combinações de dados contidos no Arca, e que possam ser manipulados pelas diferentes unidades representadas no Repositório Institucional;
- ✓ **Visualização dos dados:** Coleções mandatórias do Repositório: Artigos, Dissertações e Teses.



- ✓ Tendo em vista que a alimentação no RI Arca é descentralizada, sendo realizada através das diversas Unidades e de suas Bibliotecas, além do recurso de autoarquivamento, se tornou fundamental o monitoramento da qualidade dos dados preenchidos através da curadoria digital. Dentro deste contexto, em 2015, estabeleceu-se um plano de ação para dar início ao trabalho de curadoria digital no RI Arca que tinha como principal objetivo firmar padrões visando a organização das informações e dos objetos digitais dentro do RI;
- ✓ O crescimento exponencial no número de depósitos, notadamente, após o estabelecimento da Política de Acesso Aberto ao Conhecimento no ano de 2014 até os dias atuais, tornou-se necessária e fundamental a utilização de mecanismos que facilitassem a curadoria digital, a recuperação e a visualização do conteúdo disponibilizado, permitindo assim, obter um panorama da produção científica institucional, tendo em vista que os RIs são, também, instrumentos de gestão.



- ✓ A parceria firmada entre o RI Arca e a equipe do Laboratório de Ciência de Dados ajudou a complementar uma lacuna que existia no que se refere a curadoria digital e na gestão dos registros disponibilizados. Além disso, foi possível abordar de forma prática grandes quantidades de dados em diferentes formatos por meio de estratégias e técnicas relacionadas a Ciência de Dados;
- ✓ A utilização do software Kibana e Elasticsearch para a visualização de dados de forma dinâmica, ofereceu uma plataforma de exploração interativa para extração e mineração de dados. O software permitiu a utilização de filtros e combinações de dados contidos no Arca, como produção por tipo de material, Unidades da Fiocruz, assunto, autor, ano e direito autoral;
- ✓ Outro fator importante para a realização do projeto foi que seria possível estabelecer estratégias para a coleta, gestão e correção dos metadados descritos nas diferentes tipologias.



- ✓ Implantação de uma rotina sistêmica no trabalho de curadoria dos dados no RI Arca, de forma que os gestores das Unidades técnico científicas da Fiocruz pudessem visualizar as informações a partir da extração dos registros relevantes;
- ✓ Identificação das inconsistências no preenchimento dos metadados, visando a correção dos dados, de forma padronizada;
- ✓ O sistema também possibilitou apresentar uma nuvem de tags com os assuntos mais indexados no Repositório, destacando assim, a importância da indexação e do papel do Bibliotecário na gestão das informações relevantes para o campo da Saúde e Pesquisa dentro da Fiocruz.

Visualização de Dados

Em parceria com o Projeto de Ciência de Dados aplicada à Saúde, estamos desenvolvendo duas linhas de atuação:

- curadoria de dados, com o objetivo de melhorar a qualidade dos dados na base,
- recuperação e visualização de dados.

No quadro abaixo estão disponíveis os dados referentes às coleções mandatórias – artigos científicos, teses e dissertações dos programas da Fiocruz até fevereiro/2019.

Instruções

Add a filter +

Itens por ano



Obras

22.320

Total

Tipologias

- Artigo
- Dissertação
- Tese

Licenças

- open access
- restricted access
- closed access

Assuntos



Autores

Autor	Count
Bastos, Francisco Inácio	307
Pinkusfeld Monteiro	
Grinstejn, Beatriz	300
Reis, Mitermayer Galvão dos	258
Soares, Milena Botelho Pereira	240
Barral, Aldina Maria Prado	231
Martins Filho, Olindo Assis	221
Barral Netto, Manoel	198
Castro Filho, Bernardo Galvão	182
Szwarcwald, Celia Landmann	172
Ko, Albert Icksang	157

Export: Raw Formatted

1 2 3 4 5 ... 500

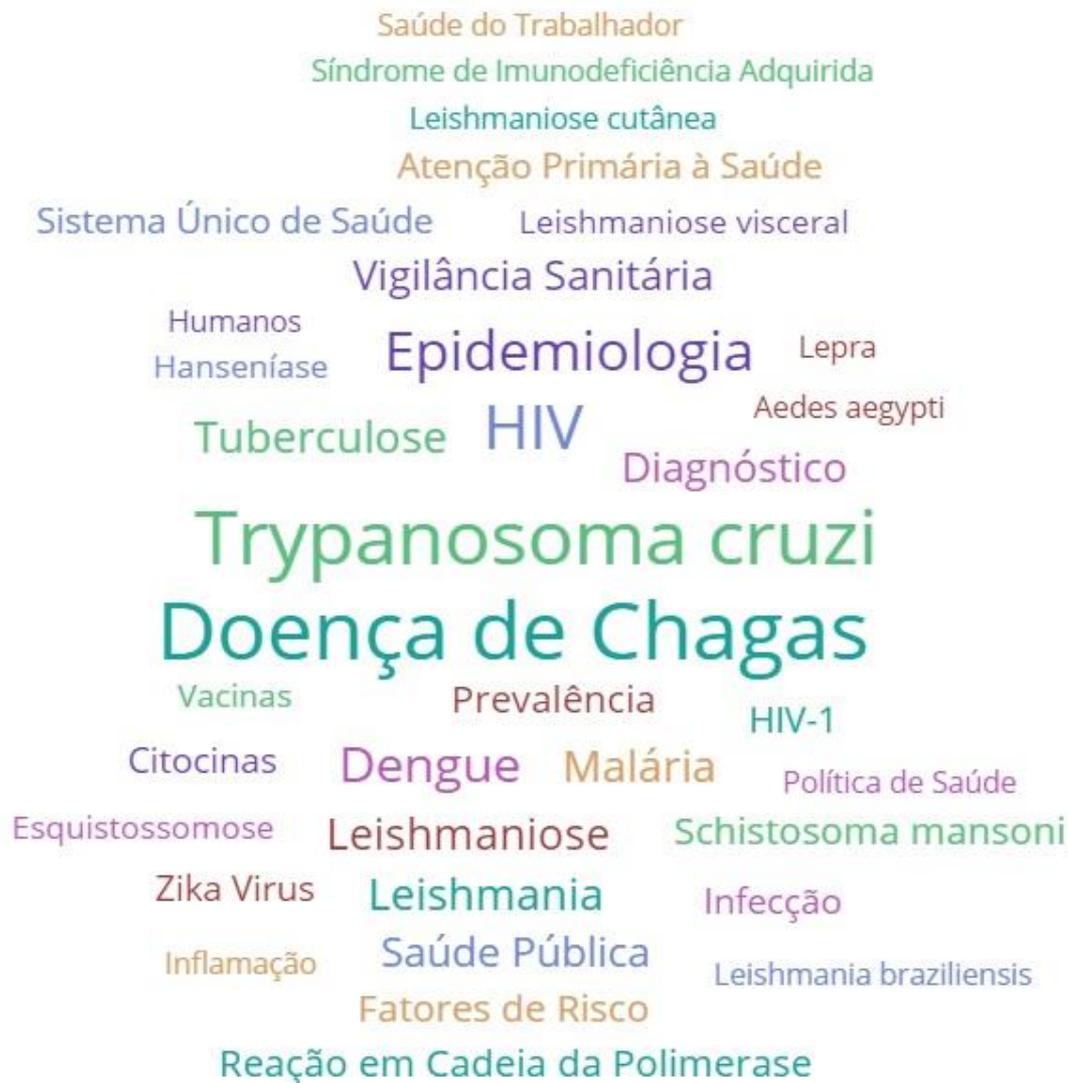
Unidades

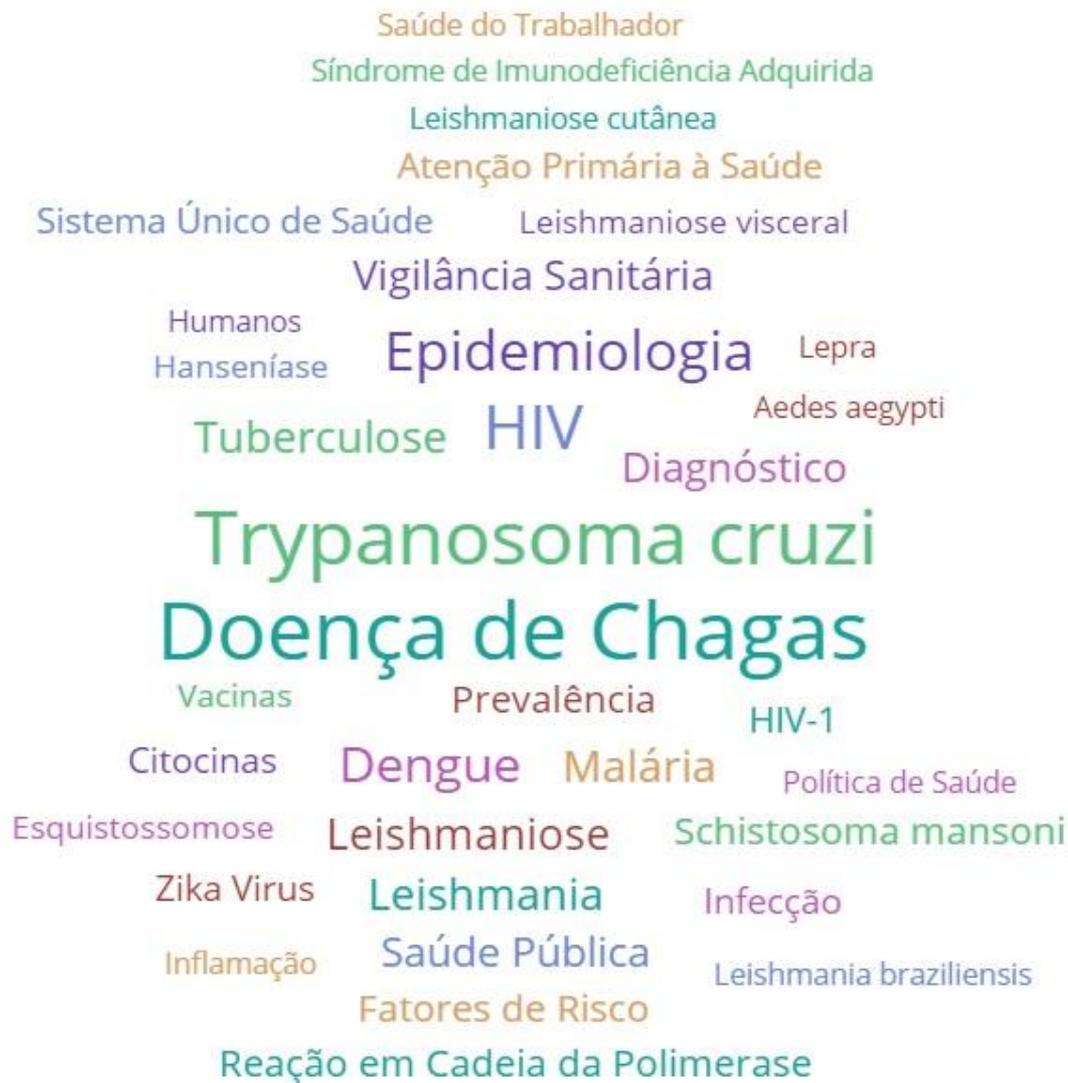
Unidade	Total
IOC	6.772
ENSP	3.464
BA	2.768
MG	1.919
INI	1.347
IFF	1.279
PE	1.124
ICICT	856
COC	776
INCQS	489
EPSJV	271
PR	251
AM	225

Fonte: <https://www.arca.fiocruz.br/terms/visualizacaodedados.jsp>

Apresentação PCDas 25/04/2019

21





Nossos sinceros agradecimentos!

Equipe Arca

Contato: repositorioarca@fiocruz.br

Telefones: (21) 3865-3271 / 3865



Fonte: Google



Instituto de Comunicação e Informação Científica e Tecnológica em Saúde

www.facebook.com/fiocruz.icict

[twitter.com/@Icict_fiocruz](https://twitter.com/Icict_fiocruz)

www.youtube.com/videosaudefio

www.icict.fiocruz.br