

Uso do ORCID como indicador persistente na construção do banco de dados de produção científica da Fundação Oswaldo Cruz

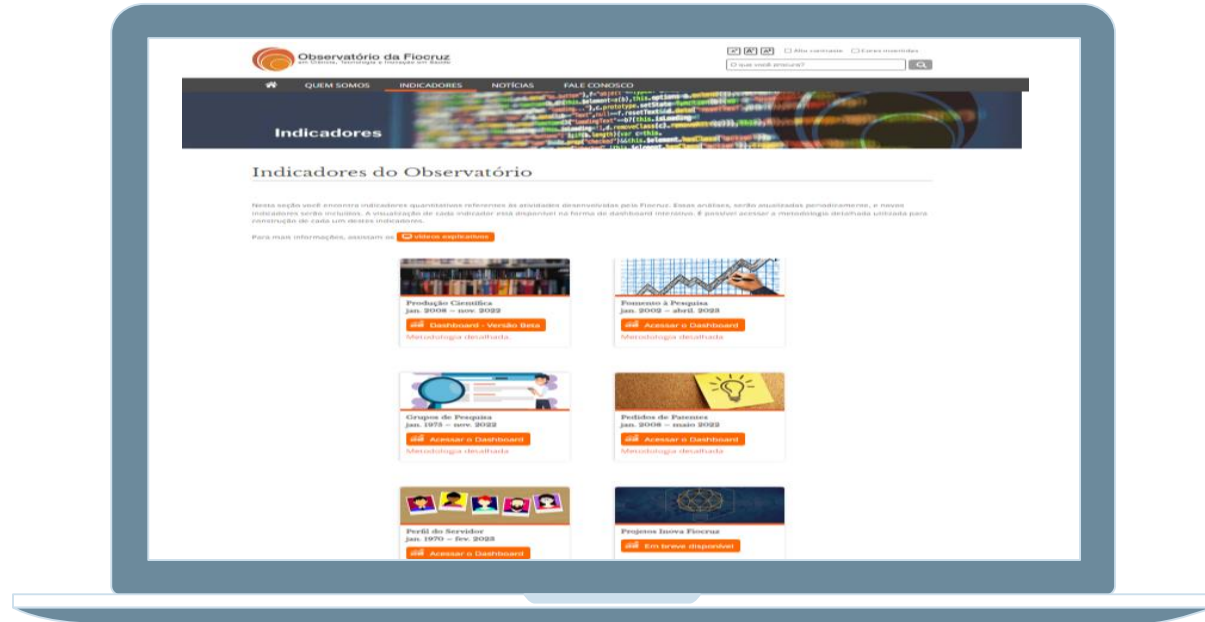
Marcos Wesley Soares Alves - 0009-0005-1133-0069 <marcos.weslwy@fiocruz.br>

Waldeyr Mendes Cordeiro da Silva - 0000-0002-8660-6331 <waldeyr.mendes@fiocruz.br>

Rafaela Lora Grando - 0000-0003-4239-9622 - <rafaela.grando@fiocruz.br>

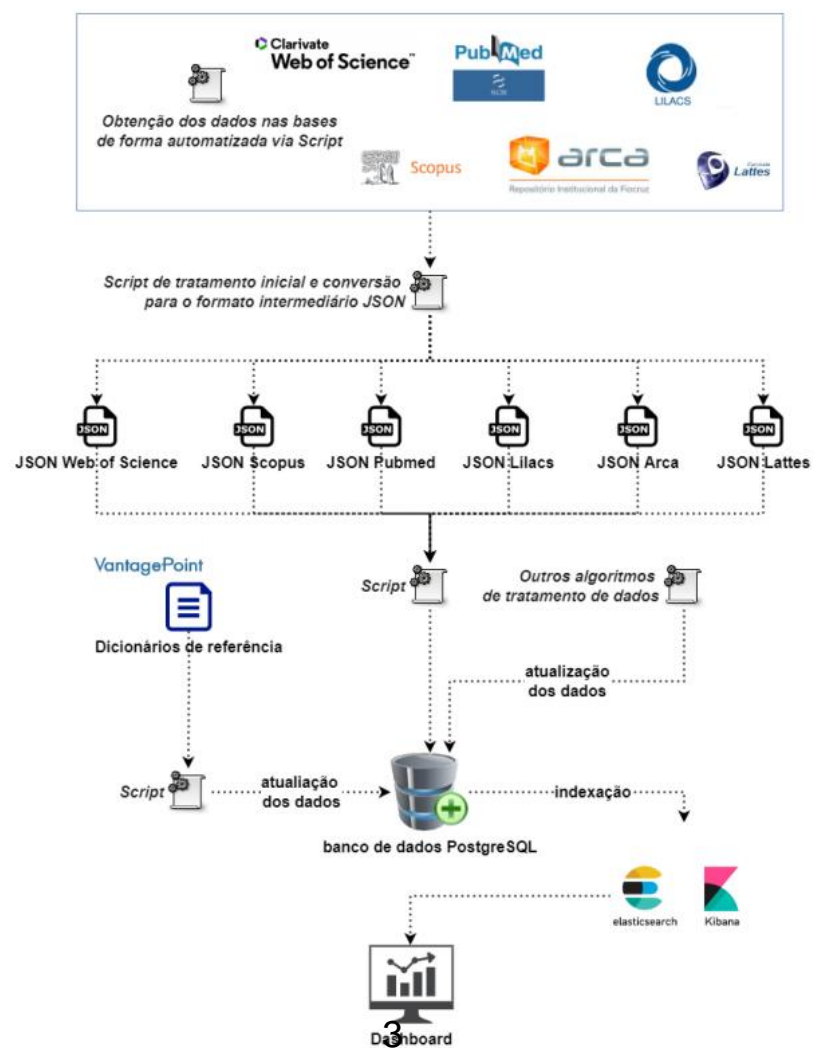
Vanessa de Arruda Jorge - 0000-0002-5298-9311 - <vanessa.jorge@fiocruz.br>

Observatório C,T&I em Saúde da Fiocruz

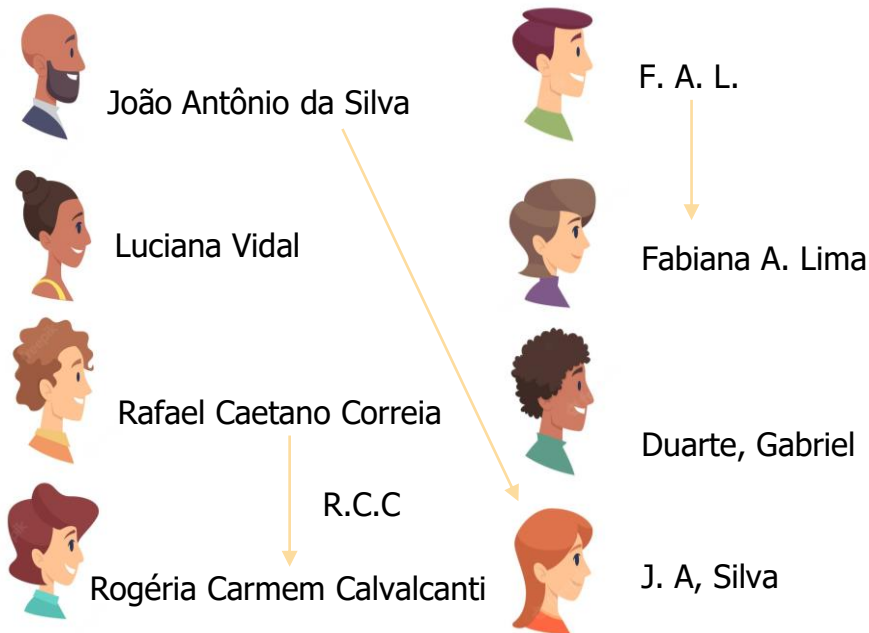


<https://observatorio.fiocruz.br>

FLUXO geral DO PROCESSO de coleta dos dados



Questão – Problemas na padronização dos nomes dos autores



Não há padrão para os nomes vindo das bases: ora estão abreviados, ora por extenso ora um misto destas opções. Ainda podem estar na ordem direta nome sobrenome, ou com o nome do meio e primeiro nome.

```
**Saint Clair dos Santos Gomes Junior  
0 1 ^S\. C\. Gomes$  
0 1 ^Saint -Clair Gomes Junior$  
0 1 ^Saint Clair dos Santos Gomes Junio$  
0 1 ^Saint Clair dos Santos Gomes Junior$  
0 1 ^Saint clair G\. Junior$  
0 1 ^Saint Clair Gomes Junior$  
0 1 ^Saint clair Gomes$  
0 1 ^Saint Clair Gomes-Junior$  
0 1 ^Saint Clair S\. Gomes Junior$  
0 1 ^Saint Clair S\. Gomes-Junior$  
0 1 ^Saint Clair Santos Gomes Junior$  
0 1 ^Saint-Clair Gomes Junior$  
0 1 ^Saint-Clair Gomes-Junior$  
0 1 ^Sant Claire G\. Junior$
```

ESTUDO DE CASO



Aplicação do *Open Researcher and Contributor ID* (ORCID), como um identificador digital persistente para deduplicação de autores vinculados à Fiocruz, no processo de criação do indicador de produção científica do Observatório em C,T& em Saúde.

Objetivo: utilizar o ORCID para aprimorar as atividades de tratamento de dados relacionados aos autores de produções científicas institucionais.

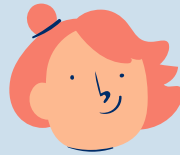
O ORCID é um identificador único de 16 dígitos que pode ser atribuído a cada pesquisador para manter um registro das atividades de pesquisa associando-as a cada identificador.



Padronização dos nomes de autores da produção científica



Servidores
4.436 ativos



Alunos
XXXXX



**Funcionários
terceirizados**
XXXX



**Bolsistas e
Prof. visitante**
XXXX

Metodologia



Usando a API do ORCID no conjunto de dados coletado nas 6 bases, verificamos se os nomes de autores em nosso banco de dados têm alguma correspondência com algum ORCID.

Nos casos positivos, verificamos também se existe correspondência do ORCID com alguma afiliação da Fiocruz.

Se encontramos as mesmas correspondências em mais de um nome do nosso banco de dados, unificamos estes registros, por se tratar da mesma pessoa.

Resumo visual da metodologia

Através de uma consulta sql selecionamos os seguintes campos da coluna pessoa: nome e orcid

O resultado da consulta é então exportado, resultando no arquivo na sheet_authors.xlsx



postgres_db



sheet_authors.xlsx



pipeline_orcid.py

```
Pipeline

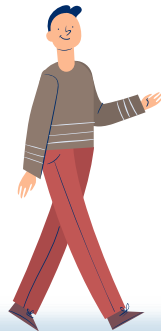
task_1 = get_orcid() (endpoint: https://pub.orcid.org/v2.1/search?q=<nome do autor>)
task_2 = get_source_name_by_employment() (endpoint: https://pub.orcid.org/v2.1/<orcid>/employments)
task_3 = get_organization_by_employment() (endpoint: https://pub.orcid.org/v2.1/<orcid>/employments)
task_4 = update_sheet() # adiciona as colunas orcids, author_by_orcids, organization a planilha autor.xlsx
```

Resultado



new_sheet_authors.xlsx

Agora esse arquivo tem adicionado aos campos originais as colunas: orcids, author_by_orcids e organization



Resultados

Na coleta em que a metodologia foi aplicada:

- 179.789 registros de autores afiliados à Fiocruz ou a Instituições parceiras em nosso banco de dados.
- Deste total, 3.670 registros já tinham ORCID identificado.
- Após a aplicação do método descrito neste trabalho, houve um incremento de 23% na vinculação de ORCID a autores das produções, totalizando 13.956 registros. Houve um ganho de 10.286 novos ORCID identificados.
- Adicionalmente, ao promover o uso do ORCID, o Observatório CT&I em Saúde da Fiocruz contribui para a padronização e a transparência na atribuição de créditos e reconhecimento aos autores de trabalhos científicos, promovendo a integridade da pesquisa e a valorização do trabalho acadêmico



Conclusão

O ORCID como identificador persistente para autores de produção científica existente no banco de dados do Observatório CT&I em Saúde da Fiocruz, tornou-o mais confiável, assegurando que um número maior de autores estará livre de ambiguidades.

Consequentemente, esta atualização do banco de dados, também permite melhorias na identificação de metadados das produções científicas, tais como indicadores de Ciência Aberta.

Como trabalho futuro, o Observatório CT&I em Saúde da Fiocruz pretende usar os demais *endpoints* da API ORCID para melhorar continuamente a qualidade dos dados de seu banco de dados, incluindo outros indicadores, como grupos de pesquisa e patentes.



Observatório em CT&I
Ciência, Tecnologia e Inovação em Saúde

Obrigada!

Vanessa Jorge

vanessa.jorge@fiocruz.br

<https://observatorio.fiocruz.br/>

