

A PUBLICAÇÃO DE DADOS EM HUMANIDADES DIGITAIS COM FAIR: *datasets* de pesquisa do projeto MPO compartilhados no CKAN da UNIRIO

Cláudio José Silva Ribeiro

Prof. Dr. PPGB e DPTD, Universidade Federal do Estado Rio de Janeiro (UNIRIO), Brasil
claudio.ribeiro@unirio.br

Martha Tupinambá de Ulhôa

Profa. Dra. PPGM, Universidade Federal do Estado Rio de Janeiro (UNIRIO), Brasil
mulhoa@unirio.br


Objetivo

Motivados pela proposta de compartilhamento de dados, o nosso objetivo é publicar e disseminar os conjuntos de dados de pesquisa segundo as recomendações apresentadas pelos princípios FAIR (*Findable, Accessible, Interoperable e Reusable*) na plataforma CKAN da Unirio.



Introdução – Projeto MPO

- Recorte de dados contendo notícias sobre música em periódicos do Sec XIX
- Registro em Banco de Dados (Projeto MPO)
- Disponível para consulta no endereço: <http://musica-sec-xix.unirio.br/>



Música em Periódicos Oitocentistas

[BUSCA](#) [SOBRE](#) [PROJETOS](#) [LOGIN](#)

Projeto de pesquisa / Proyecto de investigación

Periódicos

Gênero musical / Género musical

Pessoa, profissão / Personas, profesión

Cidade / Ciudad

Consulta livre(palavras chave) / Consulta libre(palabras clave)

Ano / Año

Redefinir

Buscar

RIO DE JANEIRO.

Transcrição / Transcripción	Título	Data
José Joaquim Lopes faz ciente ao público, ter chegado proximamente de terras estrangeiras, com uma peça de grande gosto, na qual apresenta várias figuras: esta peça tem relógio de sala e um grande realejo, com uma máquina de figuras, as...	Gazeta do Rio de Janeiro (RJ) - 1809 a 1822	25/05/1808
Nas Gazetas de Londres de 14 do mez passado, vem o seguinte paragrapho. “Hontem, certo numero de [p.79] Cavalleiros Portuguezes, não menos distinctos por seu patriotismo, que por sua graduação, se ajuntaram para celebrar, com um esplen...	Correio Braziliense ou Armazem Literário (Londres, ING) - 1808 a 1822	01/06/1808
Antonio Pinto Santos, por estar próximo a partir para a Europa, faz venda da metade, que lhe pertence da casa da Opera do Rocio. Quem a quizer comprar deverá dirigir-se a Fructuoso Carneiro Leão, morador da Rua dos Pescadores nas casas ...	Gazeta do Rio de Janeiro (RJ) - 1809 a 1822	11/01/1809
Acha-se residindo na Corte hum sujeito que se offerece para ensinar por casas particulares a ler, escrever, e contar, por ser para isso summamente habil. Quem precisar utilizar-se do prestimo do dito sujeito, poderá deixar o seu nome, e...	Gazeta do Rio de Janeiro (RJ) - 1809 a 1822	10/02/1809
Vende-se hum Piano forte muito bom, quem o quizer comprar, falle na rua Direita nas loges das casas N.o 15.	Gazeta do Rio de Janeiro (RJ) - 1809 a 1822	26/04/1809
Na Freguezia de Santa Rita, Beco dos Quarteis, n. 5, assiste hum Professor de Grammatica Portugueza e Latina; vindo há pouco tempo de Lisboa; em cuja casa ensina com a perfeição possível a Grammatica da lingua Portugueza , com a sua Ort...	Gazeta do Rio de Janeiro (RJ) - 1809 a 1822	21/06/1809
Plano do attaque sobre o exercito inimigo em sua marcha entre Aspern e Essling, para Hirschenstetten. (p. 65) [...] As tropas estavam possuidas de um enthusiasmo geral: alegres cantigas de guerra, acompanhadas por musica Turca, resoávam...	Correio Braziliense ou Armazem Literário (Londres, ING) - 1808 a 1822	01/07/1809
Logo que o Excellentissimo Governador do Rio Grande de S. Pedro soube da faustíssima Restauração de Portugal, a fez annunciar com a maior pompa no dia 02 de Fevereiro pelo Senado da Camera da Villa de Portoalegre, precedido de hum	Gazeta do Rio de Janeiro (RJ) - 1809 a 1822	22/07/1809

Premissas

- Novas práticas da ciência têm resultado em um aumento da geração e valorização dos dados de pesquisa e do interesse do compartilhamento desses dados de maneira que seja possível a sua reutilização por outros pesquisadores.
- É possível afirmar que pesquisadores utilizam boa parte do tempo das suas investigações fazendo a coleta, organização e provendo a garantia de qualidade dos dados, ao invés de utilizar o tempo na própria pesquisa.
- No entanto, observa-se um debate entre os pesquisadores apontando questões sobre problemas no reuso dos objetos, acarretando uma dificuldade na reprodução de experimentos, a chamada “crise da reprodutibilidade”.
- Os princípios *FAIR* surgem com a intenção de possibilitar que os dados sejam mais facilmente localizáveis, acessíveis, interoperáveis e reusáveis, oferecendo recursos para o seu gerenciamento e facilitando o compartilhamento, o uso e o reuso dos dados de pesquisa.

Premissas

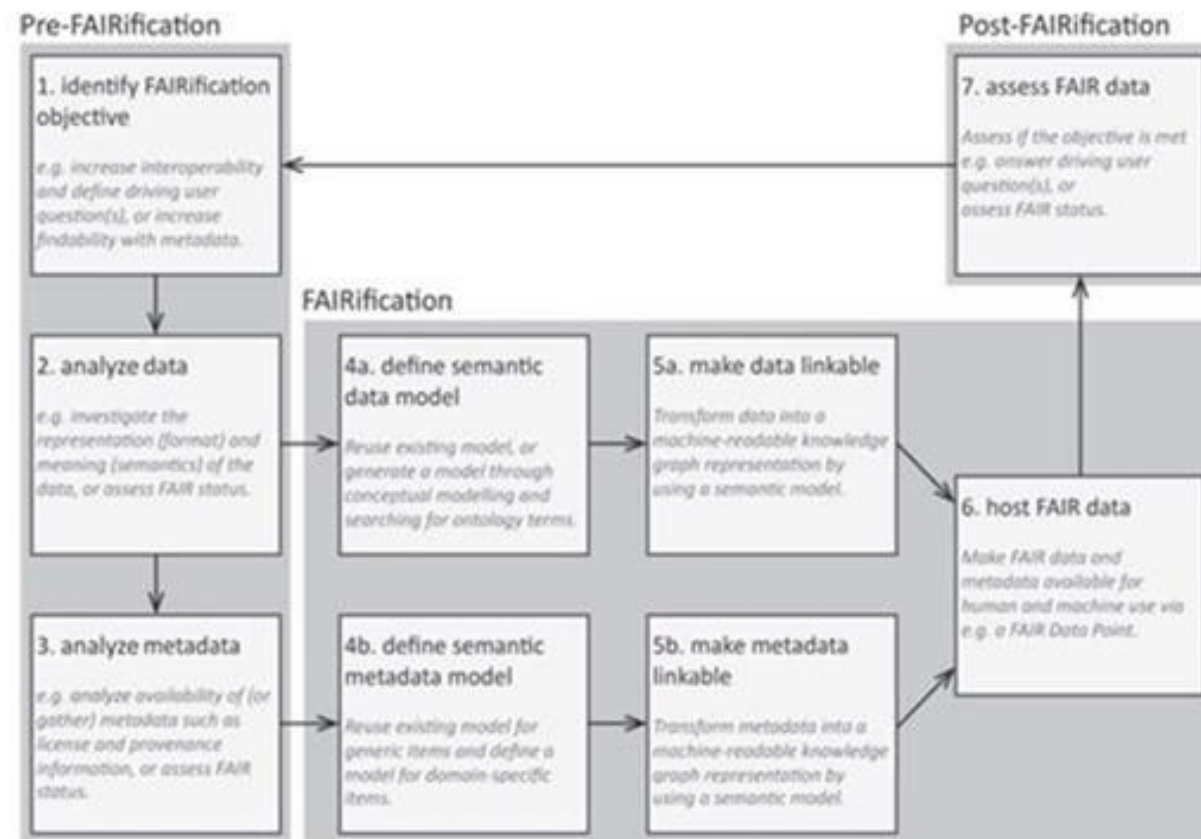


Figura 1. Fluxo de FAIRificação (JACOBSEN et al., 2020)

Trajetória do projeto: publicações

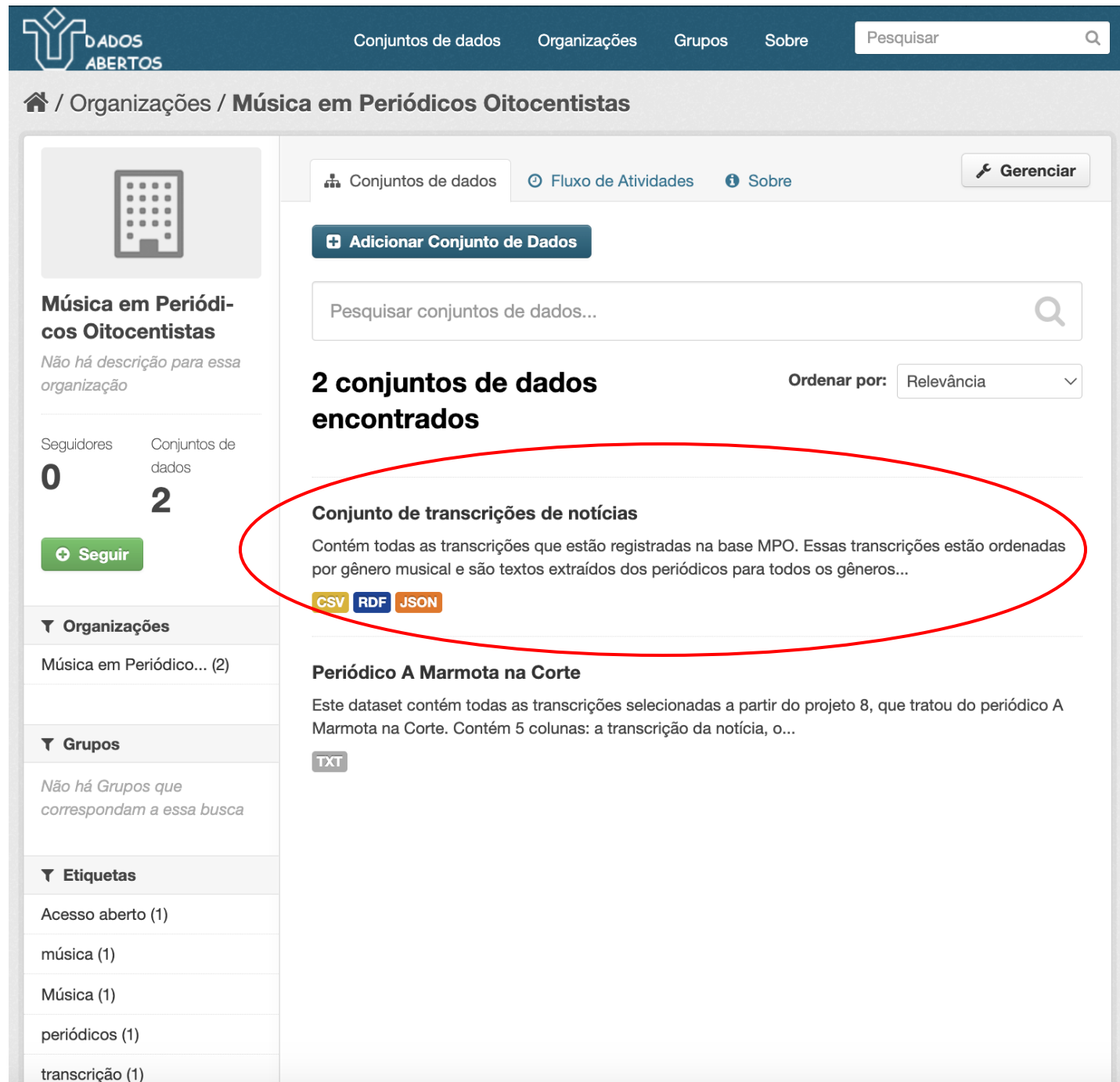
Tema	Referência
Planejamento	Ribeiro et al. (2023)
FAIRificação	Ribeiro e Ulhoa (2023)
Reúso	Ribeiro e Ulhoa (2024a)
Publicação e FAIRness	Ribeiro e Ulhoa (2024b, no prelo)

Métodos e Materiais

- Pesquisa aplicada
- Exploratória
- Revisão de literatura (ISKO, Portal de Periódicos da Capes, Brapci e GoogleAcadêmico)
- Estudo de caso (base MPO)
 - O dataset utilizado foi recuperado a partir do software *OpenRefine* como apresentado em Ribeiro e Ulhôa (2023).
 - Publicação do conjunto de dados no CKAN-Unirio – Já instalado e disponível em <http://dados.unirio.br>

Resultado - Conjuntos de Dados

Disponível em: <http://dados.unirio.br/>



The screenshot displays the OpenAIDoc website interface. At the top, there is a navigation bar with the logo 'DADOS ABERTOS' and links for 'Conjuntos de dados', 'Organizações', 'Grupos', and 'Sobre'. A search bar is located on the right. Below the navigation bar, the breadcrumb trail reads 'Organizações / Música em Periódicos Oitocentistas'. The main content area is divided into two columns. The left column features a header with a building icon, the title 'Música em Periódicos Oitocentistas', a description 'Não há descrição para essa organização', and statistics for 'Seguidores' (0) and 'Conjuntos de dados' (2). A green 'Seguir' button is present. Below this, there are sections for 'Organizações', 'Grupos', and 'Etiquetas'. The 'Organizações' section lists 'Música em Periódico...' (2). The 'Grupos' section states 'Não há Grupos que correspondam a essa busca'. The 'Etiquetas' section lists 'Acesso aberto (1)', 'música (1)', 'Música (1)', 'periódicos (1)', and 'transcrição (1)'. The right column shows a search bar with the text 'Pesquisar conjuntos de dados...'. Below it, the results are displayed as '2 conjuntos de dados encontrados', ordered by 'Relevância'. The first result, 'Conjunto de transcrições de notícias', is circled in red. It includes a description: 'Contém todas as transcrições que estão registradas na base MPO. Essas transcrições estão ordenadas por gênero musical e são textos extraídos dos periódicos para todos os gêneros...'. Below the description are three buttons: 'CSV', 'RDF', and 'JSON'. The second result, 'Periódico A Marmota na Corte', includes a description: 'Este dataset contém todas as transcrições selecionadas a partir do projeto 8, que tratou do periódico A Marmota na Corte. Contém 5 colunas: a transcrição da notícia, o...'. Below this description is a 'TXT' button.

DADOS ABERTOS

Conjuntos de dados Organizações Grupos Sobre

Pesquisar

/ Organizações / Música em Periódicos Oitocentistas

Conjuntos de dados Fluxo de Atividades Sobre Gerenciar

Adicionar Conjunto de Dados

Pesquisar conjuntos de dados...

2 conjuntos de dados encontrados Ordenar por: Relevância

Conjunto de transcrições de notícias

Contém todas as transcrições que estão registradas na base MPO. Essas transcrições estão ordenadas por gênero musical e são textos extraídos dos periódicos para todos os gêneros...

CSV RDF JSON

Periódico A Marmota na Corte

Este dataset contém todas as transcrições selecionadas a partir do projeto 8, que tratou do periódico A Marmota na Corte. Contém 5 colunas: a transcrição da notícia, o...

TXT

Música em Periódicos Oitocentistas

Não há descrição para essa organização

Seguidores 0 Conjuntos de dados 2

Seguir

Organizações

Música em Periódico... (2)

Grupos

Não há Grupos que correspondam a essa busca

Etiquetas

Acesso aberto (1)

música (1)

Música (1)

periódicos (1)

transcrição (1)



Resultado - Conjuntos de Dados

Disponível em: <http://dados.unirio.br/>

DADOS
ABERTOS

Conjuntos de dados

Organizações

Grupos

Sobre

Pesquisar

/ Organizações / Música em Periódicos ... / Conjunto de transcrições ...

Conjunto de transcri-
ções de notícias

Seguidores

0

+ Seguir

Organização

Música em Periódicos Oitocentistas

Não há descrição para essa organização

Social

Google+

Twitter

Facebook

Licença

Creative Commons Atribuição

OPEN DATA

Conjunto de dados

Grupos

Fluxo de Atividades

Gerenciar

Conjunto de transcrições de notícias

Contém todas as transcrições que estão registradas na base MPO. Essas transcrições estão ordenadas por gênero musical e são textos extraídos dos periódicos para todos os gêneros musicais que foram catalogados pelo projeto MPO.

Dados e recursos

Transcrições de notícias

O conjunto de dados contém 2 colunas. A primeira com o nome do gênero musical...

Explorar

Modelo semântico

Esquema semantico serializado em RDF/XML

Explorar

Transcrições de notícias

Arquivo JSON com as transcrições de notícias

Explorar

Acesso aberto

Música

Informações Adicionais

Campo	Valor
Fonte	http://musica-sec-xix.unirio.br
Autor	Martha Tupinambá de Uihôa e Claudio José Silva Ribeiro
Mantenedor	Claudio Jose Silva Ribeiro
Versão	1.0
Estado	active
Última Atualização	17 de Setembro de 2024, 15:20 (UTC-03:00)
Criado	29 de Julho de 2024, 11:22 (UTC-03:00)



Resultado - Conjuntos de Dados

Disponível em: <http://dados.unirio.br/>

Conjuntos de dadosOrganizaçõesGruposSobrePesquisar

/ Organizações / Música em Periódicos ... / Conjunto de transcrições ... / Transcrições de notícias

Transcrições de notícias

Ir para recurso

URL: <http://dados.unirio.br/dataset/cb1be48e-5412-456b-a214-ce6c6f0eee70/resource/16324207-54ff-4544-80fb-0812789b81e2/download...>

O conjunto de dados contém 2 colunas. A primeira com o nome do gênero musical e a segunda com o trecho que foi extraído.

Explorador de Dados

Embutir

GridGraphMap1000 records1-100Search data ...Go »Filters

nom;.....;transcricao;
Aria;EATRO S. JÉO PEDRO DE ALC. QNTARA.;+
;espet.ulo que deveria ter lugar domingo 28 do corrente fica transferido para segunda, Aifeira 29
;.bado
;s bilhetes vendem-se em casa da beneficiada
.....+
.....
Aria;EATROS De S. Janu.rio. Domingo
Aria;RESENTES PARA FESTAS;.....;+
;"om este t./#tulo sair.fo quatro bel./#ssimas pe./Bas de m./jsica
Aria;"omino Noir.Jesus sauee enfin. Grande e mui linda .ria do 3-f ato da mesma pe./Ba: acaba de se imprimir
Aria;e Santa Theresa
Aria;e Santa Theresa.Ter./Ba-feira
Aria;ecita de Assinatura. Quarta-feira 19 do corrente representar-se-fo a muita aplaudida trag./dia em 5 atos: Othelo
Aria;e Santa Theresa.Ter./Ba-feira
Aria;e Santa Theresa.Ter./Ba-feira
Aria;"Progresso retroativo da Religi./fo no nosso pa./#s [...] uma rid./#cula .ria intitulada o Rom./ntico Estragado
Aria;. JANUARIO;.....;+
;omingo
;/IRF./ÉO PORTUGU./s;.....;+
;o fim da pe./Ba
;MAN./çACO;.....;+
;s bilhetes de camarotes e cadeiras acham-se no escrit./rio do teatro;.....;+



Próximas atividades para viabilizar o reuso com FAIR

Implementar os modelos semânticos – em andamento

Testar as APIs do CKAN para reuso

A fase de pós-FAIRificação: uso do FairDataPoint

Próximas atividades

A fase de pós-FAIRificação: uso do FairDataPoint

- (1) Em andamento: ajuste no CKAN para estender suas interfaces com as interfaces do FairDataPoint – problemas de atualização de versão do CKAN (versão 2.6.2 para a versão 2.10.5);

<https://www.fairdatapoint.org/>
<https://home.fairdatapoint.org/>

Próximas atividades

A fase de pós-FAIRificação: uso do FairDataPoint

- (2) Caminho alternativo: uso de aplicação de referência do FairDataPoint e criar um mapeamento entre os esquemas de metadados definidos no FairDataPoint e os metadados no CKAN.

<https://www.fairdatapoint.org/>

<https://home.fairdatapoint.org/>

FAIRness e Monitoramento do acesso aos dados

Considerações finais

- Observou-se que processo de transformação de dados de pesquisa para torná-los *FAIR* ainda é uma iniciativa a ser explorada em diferentes áreas.
- Delimitar a coleção de dados é tarefa fundamental quando se trabalha com os registros armazenados em um banco de dados relacional.
- Agilidade em obter resultados de investigações anteriores com o reuso e com uso de outras ferramentas pode impulsionar estudos em áreas diversas. (Ribeiro e Ulhôa, 2024a)

Obrigado !

A PUBLICAÇÃO DE DADOS EM HUMANIDADES DIGITAIS COM FAIR:
datasets de pesquisa do projeto MPO compartilhados no CKAN da UNIRIO

Cláudio José Silva Ribeiro

Prof. Dr. PPGB e DPTD, Universidade Federal do Estado Rio de Janeiro (UNIRIO), Brasil
claudio.ribeiro@unirio.br

Martha Tupinambá de Ulhôa

Profa. Dra. PPGM, Universidade Federal do Estado Rio de Janeiro (UNIRIO), Brasil
mulhoa@unirio.br