

## **COMPARTILHAMENTO DE DATASETS DE PESQUISA DO PROJETO MÚSICA EM PERIÓDICOS OITOCENTISTAS (MPO): UM PROTÓTIPO COM O USO DA SOLUÇÃO CKAN <sup>[1]</sup>**

**Cláudio José Silva Ribeiro, Universidade Federal do Estado Rio de Janeiro (UNIRIO), Brasil, <https://orcid.org/0000-0002-9571-1707>**

**Martha Tupinambá de Ulhôa, Universidade Federal do Estado Rio de Janeiro (UNIRIO), Brasil, <https://orcid.org/0000-0002-6886-1267>**

### **RESUMO**

O movimento do acesso aberto está sendo impulsionado pela iniciativa Open Science, onde, em essência, os objetos de pesquisa e resultados obtidos em investigações vão além da produção documental em C&T, pois passam a incorporar também os dados que são utilizados no desenvolvimento dessas pesquisas. A disponibilização desses dados deve ser implementada de forma a possibilitar que outras iniciativas de pesquisa possam reutilizá-los no contexto de novas investigações. Este relato explora o uso do banco de dados do projeto Música em Periódicos Oitocentistas (MPO) e apresenta o protótipo com a iniciativa de compartilhamento de dados desse projeto. O desenvolvimento do protótipo para compartilhamento seguiu as recomendações apresentadas pelos princípios FAIR (Findable, Accessible, Interoperable e Reusable), pois a adoção desses princípios promove a melhoria no compartilhamento de dados de pesquisa. A metodologia utilizada combina métodos e abordagens epistemológicas, avaliando os dados de forma qualitativa e utilizando a revisão bibliográfica para reunir os aspectos teóricos envolvidos, além da aplicação prática por meio de estudo de caso para a composição dos conjuntos de dados. Fazendo uso do processo de FAIRficação (FAIRfication process) sobre conjuntos de dados do projeto MPO, selecionados a partir do periódico A Marmota na Corte, foram definidas estruturas de recursos para elucidar o conjunto de dados (dataset) depositado no software CKAN - Comprehensive Knowledge Archive Network. Os resultados do protótipo apresentam a estrutura de dois conjuntos de dados que contém registros com descrições textuais de notícias sobre música no século XIX. Os conjuntos de dados estão publicados sob licenciamento Creative Commons Attribution License (cc-by) e hospedados na plataforma CKAN, possibilitando que o seu conteúdo possa ser reutilizado por pesquisadores interessados na temática.

**Palavras-Chave:** Ciência Aberta; Música em Periódicos Oitocentistas; FAIR; FAIRficação; Reuso; CKAN.

### ***EL USO COMPARTIDO DE CONJUNTOS DE DATOS DE INVESTIGACIÓN DEL PROYECTO MÚSICA EN PERIÓDICOS OCHOCENTISTAS: UN PROTOTIPO CON EL USO DE LA SOLUCIÓN CKAN***

### **RESUMEN**

El movimiento del acceso libre se está impulsando por la iniciativa de la Ciencia Abierta, que presupone, básicamente, que los objetos de investigación y resultados obtenidos a través de las investigaciones van más allá de la producción de documentos en Ciencia y Tecnología, puesto que pasan a incorporar también los datos que se utilizan en el desarrollo de esas investigaciones. La disponibilidad de esos datos debe aplicarse a fin de posibilitar que otras iniciativas de investigación puedan reutilizarlos en el marco de nuevas investigaciones. Este informe analiza el uso del banco de datos del proyecto Música en Periódicos Ochocentistas (MPO) y presenta el prototipo que pone en marcha la iniciativa del uso compartido de datos de ese proyecto. El desarrollo del prototipo para el uso compartido siguió las recomendaciones formuladas por los principios FAIR, porque adoptar dichos

principios fomenta la mejora del uso compartido de datos de investigación. La metodología empleada mezcla métodos y enfoques epistemológicos, evaluando los datos cualitativamente y empleando la revisión bibliográfica hacia reunir los elementos teóricos implicados, además de su aplicación práctica mediante un estudio de caso que compone los conjuntos de datos. Aprovechando el proceso de FAIRificación sobre los conjuntos de datos del proyecto MPO, seleccionados de la publicación periódica *A Marmota na Corte*, se definieron estructuras de recursos con el fin de dilucidar el conjunto de datos depositado en el software CKAN - Comprehensive Knowledge Archive Network. Las conclusiones obtenidas con el prototipo reúnen a la estructura de dos conjuntos de datos que contienen registros con descripciones textuales de noticias sobre la música en el siglo XIX, además de los recursos que describen la estructura semántica de los conjuntos de datos. Los conjuntos de datos se publican bajo licencia Creative Commons Attribution License (cc-by) y se alojan en la plataforma CKAN, lo que posibilita la reutilización de su contenido por investigadores interesados en la temática.

**Palabras-Clave:** Ciencia Abierta; Música en Periódicos Ochocentistas; FAIR; FAIRificación; Reutilización; CKAN.

***SHARING RESEARCH DATA FROM DATABASE OF THE MUSIC IN 19TH CENTURY PERIODICALS (MPO) PROJECT: A PROTOTYPE USING CKAN SOFTWARE***

**ABSTRACT**

The open access movement is being promoted by the Open Science initiative, where, in essence, the research objects and results produced by investigations go beyond documentary production in S&T, as they also incorporate the data used in developing this research. The availability of this data must be implemented in such a way as to enable other research initiatives to reuse it in the context of new investigations. This report explores the use of the database of the *Música em Periódicos Oitocentistas* [Music in 19th Century Periodicals] (MPO) project and presents the prototype with the project's data-sharing initiative. The development of the sharing prototype followed the recommendations presented by the FAIR principles (Findable, Accessible, Interoperable and Reusable), as adopting these principles improves the sharing of research data. The methodology used combines methods and epistemological approaches, evaluating the data in a qualitative way and using a literature review to bring together the theoretical aspects involved, as well as practical application through a case study to compose the data sets. Using the FAIRification process on datasets from the MPO project, selected from the periodical *A Marmota na Corte* [The Marmot in Court], resource structures were defined to elucidate the dataset deposited in the CKAN software. The results of the prototype show the structure of two datasets containing records with textual descriptions of news about music in the 19th century. The datasets are published under the Creative Commons Attribution License (cc-by) and hosted on the CKAN platform, allowing their content to be reused by researchers interested in the subject.

**Keywords:** Open Science; Music in 19th Century Periodicals; FAIR; Reuse; CKAN.

---

## 1 INTRODUÇÃO

O movimento do acesso aberto na produção em Ciência e Tecnologia (C&T) vem impulsionando o desenvolvimento de pesquisas. A discussão sobre o compartilhamento dessa produção incrementa e molda novos conhecimentos. Nessa direção é possível afirmar que a melhoria dos processos

de disponibilização, distribuição e uso de informação contribui para uma redução nas diferenças e desigualdades de acesso à informação pela sociedade.

A iniciativa de Open Science trouxe alguns desafios para os processos de

disponibilização, distribuição, uso e agora também o reuso da informação. Os objetos de pesquisa e resultados produzidos por investigações vão além da produção documental em C&T, pois passam a incorporar também os dados que são utilizados no desenvolvimento dessas pesquisas.

A ciência orientada aos dados de pesquisa está em processo de consolidação, envolvendo a atuação de instituições de ensino e pesquisa, com o apoio agências de fomento. No Brasil, o compartilhamento de dados começou a ser praticado em 2017 (Henning et al., 2018). Entre as iniciativas encontram-se o plano de Dados Abertos do Instituto Brasileiro em Ciência e Tecnologia (Ibict), além de projetos como o CarpeDIEN (IEN-Comissão Nacional de Energia Nuclear), REDAPE (Embrapa), Arca Dados (Fiocruz), PBIOData (INPA) e Ibict Dataverse.

Apesar de existir na atualidade uma maior valorização na disponibilização de dados científicos para fins de transparência e reprodutibilidade (Curty & Aventurier, 2017), há indícios que podem ser compreendidos como uma dificuldade em reproduzir experimentos. Esses indícios apontam para ausência no reuso de objetos e a capacidade de reprodução das pesquisas científicas (Baker, 2016; Fanelli, 2018; Nürst et al., 2018). É surpreendente a afirmação de Baker (2016) indicando que 50% dos entrevistados em sua investigação relataram que não conseguiram reproduzir o próprio experimento.

Com o intuito de apoiar o processo de compartilhamento, o uso e reuso de dados de pesquisa é que surgem os princípios FAIR (acrônimo para Findable, Accessible, Interoperable e Reusable). A intenção é que dados sejam localizáveis, acessíveis,

interoperáveis e reusáveis, tanto por seres humanos quanto por mecanismos automatizados. A adoção dos princípios FAIR promove a melhoria no compartilhamento de dados de pesquisa, propondo de forma destacada “a necessidade desses objetos [de pesquisa] manterem uma contextualização precisa para transmitirem informação e conhecimento ao longo do tempo” (Henning et al., 2018).

O campo empírico deste artigo engloba os registros de notícias e comentários sobre música em periódicos do século XIX contidos no banco de dados Música em Periódicos Oitocentistas (MPO). Iniciado em 2002, o MPO tem servido de suporte para pesquisas sobre gêneros musicais, bem como sobre músicos, a imprensa musical, o teatro musical e a ópera, em periódicos brasileiros dos anos 1800. Foi criado inicialmente apenas com um objetivo de natureza prática, ou seja, otimizar e centralizar a coleta realizada em cada subprojeto que era desenvolvido no contexto destes periódicos.

Apresentamos, a seguir, o desenvolvimento do protótipo para compartilhamento de datasets do projeto MPO. A plataforma adotada para o protótipo foi a CKAN (Comprehensive Knowledge Archive Network), um sistema de gerenciamento de dados de código aberto utilizado para disponibilização, compartilhamento e reuso de dados governamentais brasileiros (Ribeiro e Almeida, 2011). Apesar de existirem esforços no Brasil para a adoção de soluções específicas para o compartilhamento de dados de pesquisa, a decisão de utilizar o CKAN neste projeto foi baseada nos princípios da agilidade, viabilidade e economicidade, na medida em que já existe infraestrutura na Unirio capaz de absorver as demandas do protótipo<sup>[2]</sup>.

### ***1.1 Uma Breve Contextualização sobre as Iniciativas na Área de Musicologia***

A maioria das iniciativas informatizadas em música se referem a dados musicais em vários formatos, principalmente partituras. Para Novaes, Tolare e Moreira

(2023), por conta da especificidade do domínio da música, as várias bibliotecas estão desenvolvendo ferramentas e projetos auxiliares, para permitir uma melhor

caracterização do material musical. Inicialmente, as pesquisas relacionadas a bibliotecas estavam voltadas para discutir e propor soluções para o acesso, a representação e a recuperação informacional. A partir de 2014, considerando a evolução tecnológica, os projetos auxiliares passaram a incluir a ideia de big data. Com o aumento exponencial das informações e dados musicais na web, houve a necessidade de estabelecer comunicação e acesso entre diferentes dados da área de música, evidenciando a presença não somente dos elementos que caracterizam a noção de big data, como também de linked data e linked open data.

Na continuação da sua revisão, Novaes, Tolare e Moreira destacam o MusicBrainz, uma enciclopédia de dados abertos centrada em metadados de gravações musicais comerciais. Outra iniciativa no domínio da música são as bibliotecas de código

aberto como a Early Music Online (EMO). A padronização de entradas de assuntos usada é a lista de autoridades da Biblioteca do Congresso norte americano, abrigados junto ao Arquivo Internacional de Autoridade Internacional (VIAF). Observe-se que o setor de música da Biblioteca Nacional do Rio de Janeiro utiliza a mesma lista de autoridades com acréscimos.

Para além destes esforços, Novaes, Tolare e Moreira também registram as iniciativas Music to Knowledge (M2K) e Music Information Retrieval. O M2K é considerado importante para nosso projeto, por promover a padronização dos dados em música, entre os quais a classificação de gênero musical, central para o refinamento do conteúdo do MPO. Já no Brasil, podemos destacar o Portal Música Brasilis<sup>[3]</sup>, criado em 2009, com o objetivo de promover o resgate e difusão de repertórios brasileiros de todos os tempos e gêneros.

## 2 PROJETO MPO E O DESEJO DE COMPARTILHAR CONTEÚDO SEGUNDO PRINCÍPIOS FAIR

O projeto MPO reúne e trata uma coleção de registros sobre notícias que podem ser caracterizadas como dados de pesquisa. Em conformidade com Sayão e Sales (2020, p. 32), a coleção de recortes de jornal é formada por “dados brutos coletados diretamente por um instrumento ou um sensor e agregados a partir de múltiplas fontes”. Há que se observar que, semelhante a várias outras áreas do conhecimento, a noção de dado no domínio da música assume várias características, sejam eles dados musicais, a exemplo de partituras ou gravações sonoras, ou dados sobre música, como no nosso caso. Ressalte-se, no entanto, que apesar de termos dados brutos (as notícias e anúncios), assim como múltiplas fontes (as várias coleções de periódicos), a plataforma da Hemeroteca Digital Brasileira (HDB)<sup>[4]</sup> não contempla a possibilidade de extração automática dos dados. O MPO é, portanto,

uma etapa intermediária na extração dos dados, coletados artesanalmente e agora sendo organizados para adequá-los a uma utilização mais ampla.

A ideia de compartilhar dados para reuso por outros projetos passa pelo uso de uma estrutura tecnológica que permita a compreensão e o trabalho com os conjuntos de dados, seja de forma manual ou automática por mecanismos computadorizados. Contornar a dificuldade em consumir datasets de forma automatizada poderá viabilizar o reuso da preciosidade que é o conteúdo das notícias sobre música durante o século XIX, em periódicos brasileiros.

Tornar estes dados passíveis de reuso é um passo importante na direção da Ciência Aberta no campo das Humanidades.

### 2.1 Ciência Aberta

O surgimento de uma nova ciência colaborativa e que envolve a sociedade na

práxis do cotidiano investigativo vem acarretando mudanças nos modelos de

produção e comunicação do conhecimento. Isso tem gerado transformações que viabilizam a democratização do acesso, o uso e reuso dos dados e das informações científicas. A Ciência aberta surge neste contexto de expansão das fronteiras científicas e de ampliação da participação social, com um conjunto de práticas e iniciativas que se desenvolvem de modo aberto, representando um novo paradigma na comunidade da Ciência e Tecnologia.

A Ciência Aberta suplanta os moldes de pesquisa tradicional, promovendo não só o acesso às publicações, mas também aos dados de pesquisa. Proporciona a abertura de todo o ciclo da investigação com o compartilhamento, a colaboração e o possível reuso, desde a reunião de insumos de pesquisa, os métodos para análise, o ferramental utilizado e os resultados. Dessa forma, expressa o desenvolvimento de uma ciência que abarca várias práticas e ações, como: acesso aberto, dados abertos, ciência cidadã, revisão por pares aberta, códigos abertos, cadernos de laboratório abertos, recursos educacionais abertos e redes sociais científicas (Packer e Santos, 2019; Portugal, 2016).

Diversas instituições e pesquisadores têm buscado incrementar o fomento ao acesso aberto e à abertura dos dados gerados pelas pesquisas científicas. Ações com alcance global que promovem a adoção desta nova filosofia, incluem a Open Archives Initiative (OAI), que surgiu como uma das primeiras iniciativas sobre acesso aberto e originou importantes declarações, como a Declaração de Budapeste, a Declaração de Bethesda e a Declaração de Berlim (Silva e Alcará, 2009).

## **2.2 Compartilhamento de Dados e Interoperabilidade**

Como observado anteriormente, no contexto brasileiro, esforços decorrentes da LAI como a Infraestrutura Nacional de Dados Abertos (INDA), o Projeto de Interoperabilidade no Governo – ePING (Brasil, 2019) e o Portal de Dados Abertos Brasileiros podem ser considerados como precursores que

Embora o Brasil ainda não possua uma política pública orientada para a Ciência Aberta, existem algumas iniciativas em direção ao fortalecimento da abertura de dados. O Plano de Ação Nacional sobre o Governo Aberto e a Lei 12.527/2011, conhecida como Lei de Acesso à Informação (LAI), asseguram o direito fundamental de acesso à informação de maneira transparente, clara e em linguagem de fácil compreensão, preconizando que a publicidade dos dados e das informações é a regra e o sigilo é a exceção (Santos; Almeida e Henning, 2017). Os fundamentos adotados na LAI podem ser considerados como impulsionadores para a Open Science e se coadunam com os princípios do compartilhamento de dados no contexto da Ciência Aberta.

O Brasil passou a ser membro do Open Government Partnership (OGP). Esse esforço global visa incentivar a transparência dos governos e as práticas de compartilhamento de dados gerados pela administração pública. A filiação do Brasil ao OGP incrementou os processos para disseminação, compartilhamento e reuso das informações produzidas na esfera governamental (Packer e Santos, 2019; Ribeiro e Velho, 2021).

Em todo o mundo são registradas diversas discussões que tratam do acesso e da abertura dos dados a fim de que eles sejam compartilhados e alcancem o máximo potencial de seu reuso. Os fundamentos adotados na LAI podem ser considerados como impulsionadores para a Open Science e se coadunam com os princípios do compartilhamento de dados no contexto da Ciência Aberta.

contribuem para o movimento do compartilhamento de dados na Ciência. Esses movimentos surgiram em decorrência da parceria do Governo Aberto, disponibilizando os conjuntos de dados sobre as ações de governo (Ribeiro e Velho, 2021).

Estes esforços indicam que o Brasil está evoluindo na perspectiva de promover o acesso livre, aberto e o compartilhamento dos dados de governo. No entanto, quando nos deparamos com a diversidade e complexidade dos dados em Ciência e Tecnologia, é necessária a adoção de elementos que melhorem a qualidade dos dados e facilitem a sua gestão.

Adicionalmente, parte-se do pressuposto que em um ambiente heterogêneo e distribuído é essencial que dados, serviços e computadores interajam tão autonomamente quanto possível. Para que esse grau crescente de autonomia seja atingido é necessário que esses participantes possam interagir em diversos níveis. A pesquisa na área da ciência da computação proporcionou a superação de diversos desafios neste processo de interconexão. Da impossibilidade de compartilhamento de dados e programas em ambientes computacionais distintos, passou-se à possibilidade crescente de troca de dados por meio do uso de protocolos de comunicação e formatos de representação de dados. Para dar

### 2.3 Princípios FAIR

Como resultado de uma conferência denominada *Jointly designing a data FAIRPORT*, realizada em 2014, surgiram princípios que buscavam solucionar problemas no tratamento, bem como na sua aplicação e utilidade, para o crescente volume de dados de pesquisa gerados no contexto da e-Science (Henning et al., 2018;

conta desse processo de interligação é que se moldou o conceito de interoperabilidade.

Dentre os variados tipos de interconexão destacamos a interoperabilidade semântica, pois essa é utilizada para solucionar os problemas advindos da necessidade de combinar diferentes espaços de informação, provendo informação de maneira unificada. Nesse sentido, percebe-se que as dificuldades e obstáculos encontrados para convergir os ambientes tecnológicos e padronizar métodos de trabalho tornam complexa a utilização e reutilização dos dados, além de indicar a necessidade de criação de uma infraestrutura global de compartilhamento. Essa infraestrutura possibilitaria a interoperabilidade dos dados, viabilizando a semântica consensual e usando padrões e protocolos já aprovados.

É nesse contexto que surgem os princípios FAIR, pois esses foram propostos e criados como forma de minimizar as dificuldades encontradas na criação e no gerenciamento dos dados.

Wilkinson et al., 2017). Partindo de cada uma das letras do acrônimo FAIR, tem-se a adoção de um conjunto de especificações que demandam um domínio de conhecimentos técnicos e estruturais distintos. O quadro 1 a seguir reúne uma síntese desses princípios:

**Quadro 1: síntese dos princípios FAIR**

<b>Findable (F)</b>	
F1	Os dados e metadados devem ter identificadores globais, persistentes e identificáveis.
F2	Os dados devem ser descritos com metadados enriquecidos.
F3	Os metadados devem incluir clara e explicitamente identificadores dos dados que descrevem.
F4	Os dados e metadados devem ser registrados ou indexados com recursos que ofereçam capacidade de busca.
<b>Accessible (A)</b>	
A1	Os conjuntos de dados e metadados devem possuir identificador que utilize protocolo de comunicação padronizado, para facilitar a recuperação.
A2	Garantir a acessibilidade aos metadados, mesmo quando os dados não estiverem mais disponíveis.
<b>Interoperable (I)</b>	

I1	Os dados e metadados devem ser representados por meio de uma linguagem formal, acessível, compartilhada e amplamente aplicável para a representação do conhecimento.
I2	Os dados e metadados devem usar vocabulários que seguem os princípios FAIR.
I3	Os dados e metadados devem incluir referências qualificadas para outros dados e metadados, possibilitando a conexão semântica entre eles.
<b>Reusable (R)</b>	
R1	Os dados e metadados são descritos com uma pluralidade de atributos precisos e relevantes.

Fonte: Henning *et al.* (2018).

A evolução tecnológica impulsionou o movimento da Ciência Aberta e com isso tornou-se cada vez mais comum e benéfico o compartilhamento de dados e informações. Os princípios FAIR congregam as boas práticas de gestão de dados científicos, envolvendo um conjunto de ideias que possibilitam a melhora da qualidade dos dados de pesquisa e,

consequentemente, seu reuso por homens e máquinas.

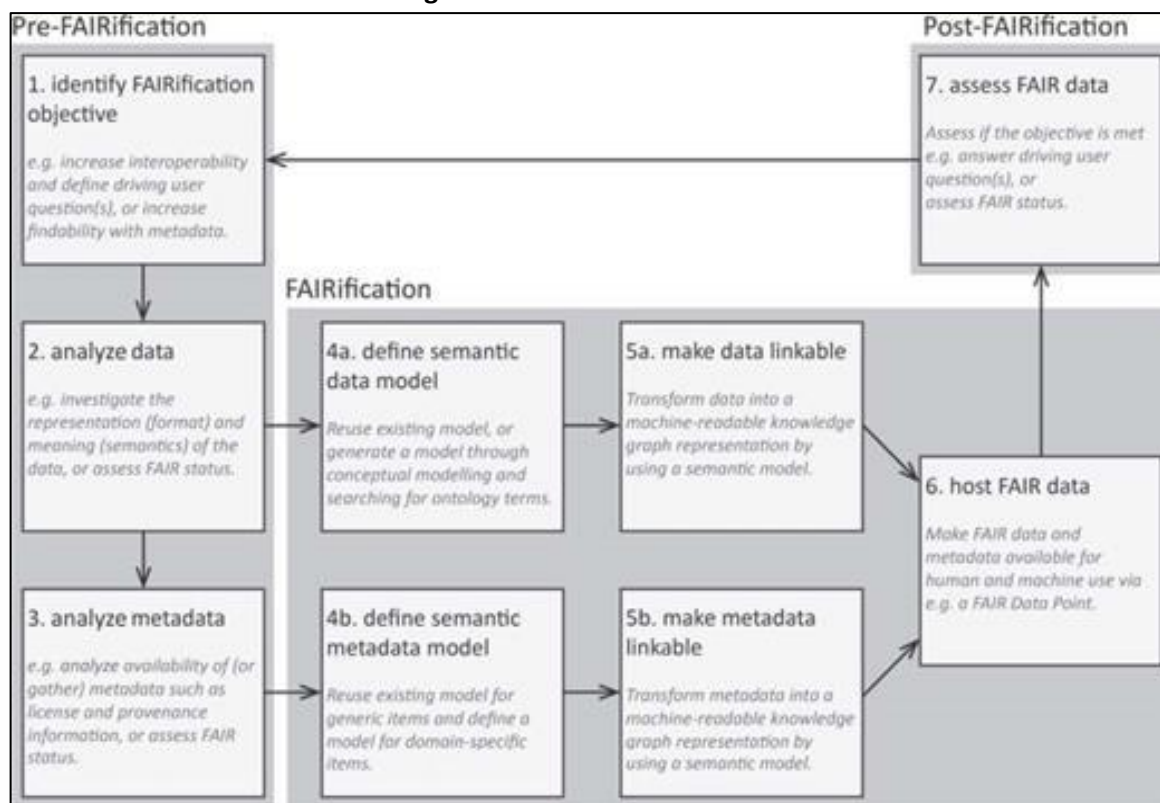
Para adoção de FAIR faz-se necessário seguir o conjunto de orientações que objetivam transformar os conjuntos de dados. Esse processo de transformação é denominado FAIRification Process (Jacobson *et al.*, 2020; University of Twente, 2022).

### 2.3.1 Processo de Tornar os Dados FAIR (FAIRification Process)

O conjunto de tarefas para tornar os dados FAIR foi apresentado por Jacobson *et al.*

(2020) e pode ser visualizado na figura 1 a seguir:

Figura 1: Fluxo de FAIRification



Fonte: Jacobson *et al.* (2020).

Em projeto preparatório para esta pesquisa, Ribeiro et al. (2023) apresentaram um primeiro ensaio que objetivava a interconexão com verbetes da Wikipedia e entradas na Wikidata. Esse esforço pode ser

### **2.3.2 Licenciamento dos Datasets**

A fundamentação para o compartilhamento apresentado na subseção 2.2 nos conduz à temática do licenciamento dos dados. Esse debate aponta para algumas abordagens sobre autorização de acesso, conteúdo e pleno uso desses dados, mas sempre garantindo que haja respeito ao licenciamento original, bem como a possibilidade de garantir o direito à citação de dados (Fenner, Crosas, Grethe et al., 2019).

Nessa direção, podemos enumerar duas possibilidades: o uso de Creative Commons, com suas 6 categorias que permitem distribuir, remixar, adaptar e redistribuir (Creative Commons, [s.d.]); e a Open Data Commons Open Database License (ODC-ODbL), licenciamento específico para o compartilhamento de bancos de dados (Open Knowledge Foundation, [s.d]).

Não é intenção deste trabalho debater sobre as opções de licenciamento de dados, portanto, no contexto do protótipo optamos por seguir os resultados de um levantamento de dados no Portal de Dados Abertos Brasileiros em conjunto com as recomendações obtidas em Rautemberg et al. (2018).

### **2.4 Dados do Projeto MPO**

Os dados do projeto MPO são gerados a partir dos periódicos abrigados na Hemeroteca Digital Brasileira (HDB), da Fundação Biblioteca Nacional do Rio de Janeiro. As consultas na HDB podem ser feitas via internet, por título, período, edição, local de publicação e palavra(s)-chave (tecnologia OCR), mas os resultados da busca retornam no

A consulta ao sítio que contém o banco de dados pode ser feita por todos esses parâmetros, além de palavras ou expressões

considerado como planejamento para o processo de FAIRificação, pois inclusive aponta em suas conclusões para as convergências das tarefas para conectar os dados (linked data) e tornar os dados FAIR (FAIRification Process).

Em levantamento realizado no âmbito do Portal no mês de julho de 2023, há conjuntos de dados que fazem uso das possibilidades relacionadas anteriormente com as seguintes proporções: de um total de 12.204 conjuntos de dados, cerca de 45% utilizam licença Creative Commons e cerca de 36% utilizam licença ODbL, o restante dos 19% utilizam outras modalidades de licenciamento (sem licença atribuída ou licença em Domínio Público).

Rautemberg et al. recomendam que a análise sobre o modelo de licenciamento de dados contemple a verificação das seguintes características: necessidade de reuso dos dados com a possibilidade de alteração do seu conteúdo, mas com rastreabilidade nas mudanças; a necessidade de manutenção da licença atribuída; e, por fim, a necessidade de permissão para uso com fins comerciais.

Após análise dos conjuntos publicados optamos por adotar o licenciamento Creative Commons Attribution License (cc-by), pois essa licença permite a redistribuição e o reuso, mas com a manutenção de créditos ao criador no conjunto de dados no contexto do protótipo.

formato imagem em PDF. Assim, o garimpo nas múltiplas coleções de periódicos, ocorre artesanalmente, sendo as informações armazenadas em tabelas – incluindo título do periódico, local (cidade e estado), número da edição, número da página, data de publicação e conteúdo.

em um campo de consulta livre. Seu aproveitamento por outros pesquisadores é muito precário ainda. Os conjuntos de dados



precisam ser disponibilizados de forma adequada aos critérios e padrões de compartilhamento de dados em rede, incluindo a possibilidade de interoperabilidade. Assim, investigadores em

qualquer parte do mundo terão acesso irrestrito a esses dados, a fim de reutilizá-los em suas pesquisas. As Figuras 2 e 3 apresentam as interfaces de consulta e resultados obtidos em consulta livre do MPO.

**Figura 2: Interface para consulta à base MPO**



Fonte Imagem: Projeto MPO (<http://musica-sec-xix.unirio.br/>). Acesso em: 08 2023.

**Figura 3: Exemplo de resultado obtido em consulta livre**

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 paragraho. "Hontem, certo numero de [p.79] Cavalleiros Portuguezes, não menos distinctos por seu patriotismo, que por sua gradação, 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 possivel a Grammatica da lingua Portugueza , com a sua Ort...	Gazeta do Rio de Janeiro (RJ) - 1809 a 1822	21/06/1809
Plano do ataque sobre o exercito inimigo em sua marcha entre Aspern e Essling, para Hirschenstetten. (p. 65) [...] As tropas estavam possuidas de um entusiasmo geral: alegres cantigas de guerra, acompanhadas por musica Turca, resoavam...	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 faustissima 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

Fonte Imagem: projeto MPO (<http://musica-sec-xix.unirio.br/>). Acesso em agosto de 2023.

O recorte adotado neste protótipo foi um conjunto de notícias extraídas do periódico A Marmota. A Marmota foi um jornal de variedades joco-sério criado, que circulou no

Rio de Janeiro entre 1849 e 1864, recebendo diferentes nomes: A marmota na corte, Marmota fluminense: Jornal de modas e variedades e A Marmota. Um dos subprojetos

contidos no MPO trata exatamente de dados sobre música garimpados nesse periódico.

### 3 MATERIAIS E MÉTODOS

Esta é uma pesquisa que combina diferentes abordagens para atingir o resultado almejado. Pode ser caracterizada como qualitativa, exploratória e estudo de caso (SEVERINO, 2007). A investigação foi desenvolvida em duas etapas: a primeira contemplou a revisão de literatura para o projeto e a segunda a implementação do protótipo proposto.

A revisão de literatura reuniu os fundamentos teóricos e conceituais sobre Ciência Aberta, compartilhamento de dados, adoção dos princípios FAIR e o licenciamento dos datasets.

A segunda etapa foi desenvolvida utilizando métodos que unem atividades ligadas à implantação de bancos de dados, uso de tecnologias da Web Semântica e estudo de caso, conforme apresentado a seguir:

1. Estudo sobre a estrutura e o conteúdo do banco de dados.
2. Formulação das questões de competência e desenvolvimento de queries para responder às

### 4 RESULTADOS

Os datasets foram extraídos da base Música em Periódicos Oitocentistas (MPO), seguindo as etapas do planejamento de FAIRificação (Figura 1). Esses agrupamentos de dados foram propostos segundo a visão de recursos conforme previsto no software CKAN. Foram estabelecidos para o protótipo a disponibilização de 4 recursos: os recursos 1 e 2 são apresentados na subseção 4.1; os recursos 3 e 4, que contém a serialização dos modelos semânticos dos recursos 1 e 2, estão apresentados nas Figuras 6 e 7. O processo de

questões identificadas (Allemang & Hendler, 2011).

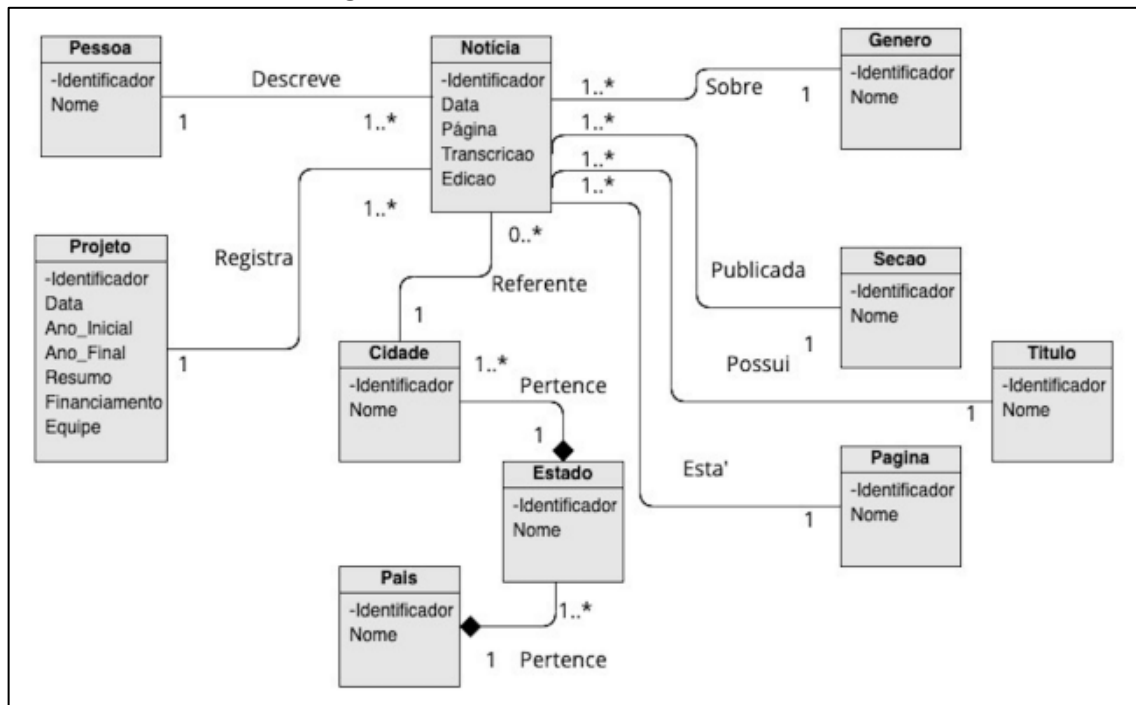
3. Extração dos arquivos e carga no software OpenRefine (University of Twente, 2022).
4. Preparação dos dados e a exportação dos registros para o formato CSV.
5. Preparação e serialização dos modelos semânticos.
6. Definição e atribuição de metadados.
7. Publicação na plataforma CKAN.

Cabe destacar que na tarefa 5, o modelo semântico foi desenvolvido com base nas representações presentes em Ribeiro et al. (2023) com uso de vocabulários RDF/XML em conjunto com Dublin Core (DC), Simple Knowledge Organization System (SKOS), Provenance Data Model (PROV-DM), FRBR-aligned Bibliographic Ontology (FaBiO) e Web Ontology Language (OWL). Já para a tarefa 6, os metadados foram planejados a partir das características da ferramenta CKAN, que faz uso do padrão Data Catalog (DCAT).

confeção dos modelos semânticos não será apresentado neste relato.

Os resultados do ensaio foram obtidos em banco de dados gerado a partir do estudo sobre sua estrutura (Figura 4). Esse banco foi denominado de MPO\_tst. O Dump (recuperação de conteúdo do banco de dados) foi carregado no software PostgreSQL (10.23) em Notebook com o sistema operacional Linux sob distribuição Ubuntu (3.28.1/18.04.2).

**Figura 4: Modelo de dados do banco MPO**



Fonte: Dados de pesquisa (2023).

As subseções a seguir apresentam os resultados da pesquisa.

#### **4.1 Seleção dos Dados e Formação do Dataset**

O início do trabalho foi marcado com a formulação das questões de competência (QC) abaixo relacionadas:

QC1: quais as transcrições textuais reunidas por um projeto?

QC2: como reunir transcrições de um gênero musical?

Para atender às QCs foram formuladas queries em linguagem de consulta SQL. Para a QC1, o arquivo obtido continha as colunas transcrição, projeto\_id, registro\_id, gênero\_id, genero. Estas colunas foram carregadas no software OpenRefine para organização e transformação dos dados. A Figura 5 apresenta um recorte com o conteúdo do dataset 1 (DS1):

**Figura 5: Extrato do dataset carregado no OpenRefine**

OpenRefine dataset1.txt <small>Permalink</small>				
30 rows <span style="float: right;">Extensions Wikibase</span>				
Show as: rows records <span style="float: right;">« first » previous 1 next » last »</span>				
transcrição	projeto_id	registro_id	genero_id	genero
[Progresso retroativo da Religião no nosso país] [...] uma ridícula ária intitulada o Romântico Estragado, na qual o executor da música meteu todas as qualidades e trejeitos e ditos que lhe ocorreram para fazer rir o povo no tempo em que deve chorar; e isto na Semana intitulada da Paixão! E é para estas baboseiras que se pede e que o governo despense contos de réis em ordenados mensais com os cantores vindos da Europa? [...] Link: <a href="http://memoria.bn.br/DocReader/706906/1014">http://memoria.bn.br/DocReader/706906/1014</a> [Crônica]	8	4079	42	Aria
[Charada simples] Aparte cortante de qualquer instrumento agudo e forte [1ª e 4ª] / Derivo de um pronome, e na música assento como [2ª] / Negra tempestade eis que te assalta, Quebrados vão os mastros, rota a enxarcia, Um único bem te resta, sem o qual Colombo / A América não veria. [3ª e 4ª] / Sou resina de odorífero aroma. [Resposta no n. 93, p. 4: GUMELUME]. Link: <a href="http://memoria.bn.br/DocReader/706906/366">http://memoria.bn.br/DocReader/706906/366</a> [Charada]	8	4014	42	Aria
[Charada simples] Aparte cortante de qualquer instrumento agudo e forte [1ª e 4ª] / Derivo de um pronome, e na música assento como [2ª] / Negra tempestade eis que te assalta, Quebrados vão os mastros, rota a enxarcia, Um único bem te resta, sem o qual Colombo / A América não veria. [3ª e 4ª] / Sou resina de odorífero aroma. [Resposta no n. 93, p. 4: GUMELUME]. Link: <a href="http://memoria.bn.br/DocReader/706906/366">http://memoria.bn.br/DocReader/706906/366</a> [Charada]	8	4014	44	Barcarola
[Charada simples] Aparte cortante de qualquer instrumento agudo e forte [1ª e 4ª] / Derivo de um pronome, e na música assento como [2ª] / Negra tempestade eis que te assalta, Quebrados vão os mastros, rota a enxarcia, Um único bem te resta, sem o qual Colombo / A América não veria. [3ª e 4ª] / Sou resina de odorífero aroma. [Resposta no n. 93, p. 4: GUMELUME]. Link: <a href="http://memoria.bn.br/DocReader/706906/366">http://memoria.bn.br/DocReader/706906/366</a> [Charada]	8	4014	65	Cachucha
[Peição/Imo. Sr. Redator da Marmota] As abaixo-assinadas, moças sem ciência, vaidosas no último ponto, persuadidas de que são as mais lindas moças que jamais viu a luz [...] lhe ordenam (não repare que é moeda corrente) [que] componha para a sua Marmota uma vista em que se represente o Passeio Público sem asseio, sem flores, e ali, por mal dos seus pecados sem música aos domingos; pois como a sua Marmota corre todas as praças, ruas, becos, esquinas, corredores, cozinhas e sobrados, e é vista por todos, todos erguerão um brado de indignação, e ouvindo o aquelles a quem compete, ainda que não por outro motivo senão por desmentir e borrar essa pintura, mandaram ir para ali uma música. [...] Link: <a href="http://memoria.bn.br/DocReader/706906/342">http://memoria.bn.br/DocReader/706906/342</a> [Texto irônico; petá]	8	4013	65	Cachucha
Teatro S. Januário / Benefício do ator / ROMUALDO DE NORONHA, Terça-feira, 25 de fevereiro de 1851, haverá o drama em 3 atos / NOVO DESERTOR FRANCÊS. / O interessante e gracioso papel do capitão Valcourt será desempenhado pelo senhor Sr. João Caetano dos Santos; uma excelente banda de música, tropa de linha etc., tornarão mais brilhante este espetáculo. Link: <a href="http://memoria.bn.br/DocReader/706906/522?pesq=musica">http://memoria.bn.br/DocReader/706906/522?pesq=musica</a> [Anúncio. Ator Romualdo de Noronha; Ator João Caetano dos Santos]	8	4035	40	Can-can
[Modinha Brasileira] [...] A simplicidade da música corresponde perfeitamente a singularidade da poesia. [...] Link: <a href="http://memoria.bn.br/DocReader/706914/1437?pesq=musica">http://memoria.bn.br/DocReader/706914/1437?pesq=musica</a> [Romance; modinha]	8	4325	22	Modinha
[DA MÚSICA RELIGIOSA] [...] Música deveria ser a linguagem preferida de uma religião de amor e mistério. A música religiosa é a parte da arte, que se resente mais da perturbação, e agitação que caracterizam a sociedade moderna. Na verdade já não há música religiosa [...] As artes já não tem horizonte, falta-lhe o infinito, seu reino é deste mundo, e eis porque não pode haver música religiosa. Neste estado de cousas, três paridos disputam em França, e mesmo na Europa, a regeneração da Música religiosa. Um queria que se não cantasse nas igrejas. [...] o outro que se ajuntasse ao cantocho uma bela música vocal; sem outro acompanhamento que o do órgão; e o terceiro pensa, que seria absurdo privar a música religiosa dos imensos recursos da arte e da instrumentação moderna. Uma vista d'olhos lançada rapidamente sobre a historia da musica religiosa nos fornecerá os elementos necessários para bem apreciar a questão que acabamos de estabelecer. [...] Link: <a href="http://memoria.bn.br/DocReader/706914/569">http://memoria.bn.br/DocReader/706914/569</a> [Música religiosa]	8	4217	89	Música religiosa
[O PICA-PAU ATREVIDO.] Com este número distribui-se grátis aos Srs. assinantes este lindo fado Mineiro, que tanto tem agradado no teatro de S. Francisco, cantado no 2º ato da ópera do Sr. F. C. da Conceição - o casamento e a mortalha no oahu se calha -. Outras peças de música ficam no preto para serem distribuídas oportunamente. [...] [vende-se avulso a 400 rs]. Link: <a href="http://memoria.bn.br/DocReader/706906/809">http://memoria.bn.br/DocReader/706906/809</a> [Anúncio]	8	4060	25	Ópera

**Fonte: Dados de pesquisa (2023).**

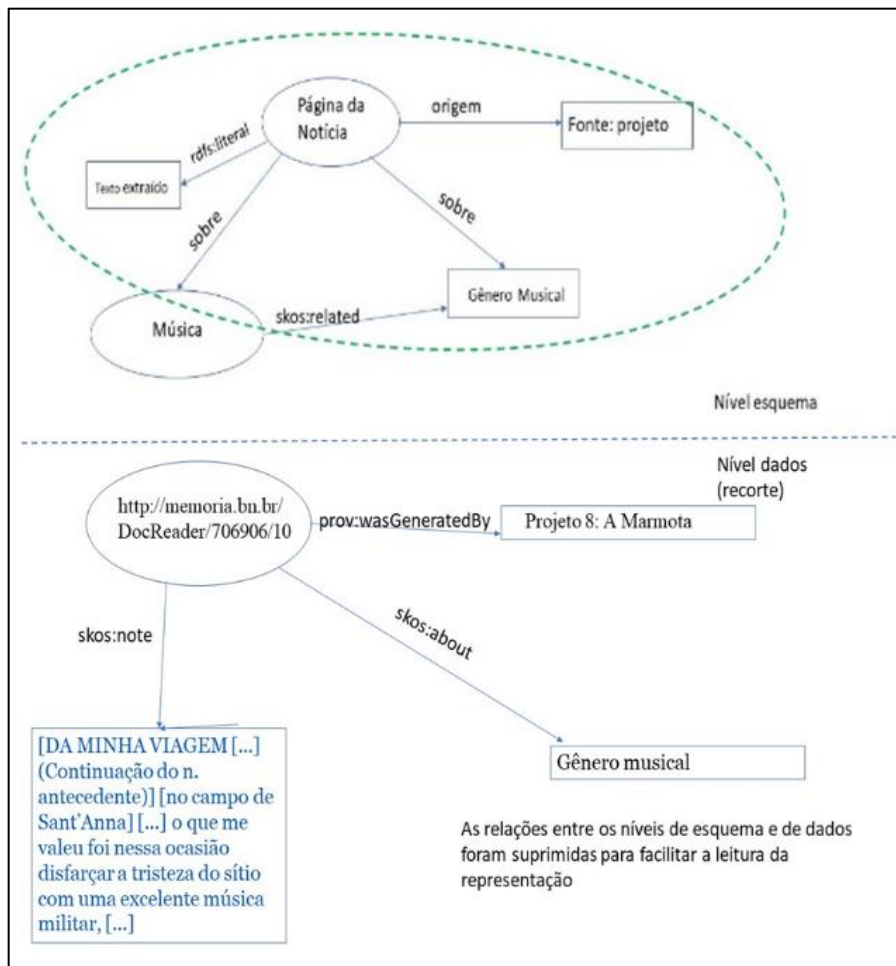
Com os dados carregados no OpenRefine foi possível fazer a análise do conteúdo para ajustes e a posterior exportação do conjunto de dados para o formato CSV (Comma Separated Values).

Para atender à QC2 foi efetuado o mesmo procedimento da QC1 e que resultou

no dataset2 (DS2). As colunas obtidas foram transcrição, gênero\_id e gênero.

O relato apresentado na subseção 2.2 iluminou o caminho na direção de estabelecer um modelo semântico para ambos os datasets. Esse modelo foi obtido a partir dos estudos desenvolvidos em Ribeiro et al. (2023). O grafo RDF gerado é apresentado na Figura 6.

**Figura 6: Grafo RDF**



**Fonte: Dados de pesquisa (2023).**

Conforme previsto na seção 3, no contexto da publicação na Web, cabe esclarecer que para a descrição semântica dos recursos DS1 e DS2 foi necessário a serialização do grafo apresentada no Figura 6. A representação RDF/XML foi adotada para essa tarefa. Para o recorte de dados neste estudo de caso, foram selecionados os predicados `prov:wasGeneratedBy`, `dc:date`, `dc:title`, `fabio:NewsItem`, `skos:about` e `skos:note`.

A serialização corresponde à transformação de parte do grafo apresentado na figura 6 (nível dos dados) em código, permitindo assim que haja processamento e interpretação das informações semânticas da representação gerada. O trecho de código a seguir ilustra o resultado desse processo de serialização (Figura 7):

**Figura 7: Extrato da serialização em RDF/XML**

```
<?xml version="1.0" encoding="UTF-8"?>
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
  xmlns:prov="http://www.w3.org/ns/prov#"
  xmlns:owl="http://www.w3.org/2002/07/owl#"
  xmlns:fabio="http://purl.org/spar/fabio#"
  xmlns:dc="http://purl.org/dc/elements/1.1#">
  <rdf:Description
    rdf:about="http://memoria.bn.br/DocReader/706906/10">
    <rdfs:label>minha viagem</rdfs:label>
    <fabio:NewsItem>[DA MINHA VIAGEM [...] (Continuação do n.
    antecedente)] [no campo de Sant'Anna] [...] o que me valeu foi
    nessa ocasião disfarçar a tristeza do sitio com uma excelente
    música militar, [...]</fabio:NewsItem>
    <dc:date>01/01/1849</dc:date>
    <prov:wasGeneratedBy
    rdf:resource="https://pt.wikipedia.org/wiki/A_Marmota"/>
    <prov:wasGeneratedBy
    rdf:resource="https://www.wikidata.org/wiki/Q16142982"/>
  </rdf:Description>
```

**Fonte: Dados de pesquisa (2023).**

A definição do conjunto de metadados que permite a catalogação dos datasets segue o padrão DCAT que é utilizado pelo CKAN. A estrutura de campos desse conjunto de metadados ficou definida com: Identificador,

data de publicação, data de atualização, título, autor, descrição, assunto, fonte (proveniência), licença atribuída e publicações relacionadas. O quadro 2 reúne os metadados utilizados, bem como suas fontes para carga e ambientes alvo:

**Quadro 2: Escolha de Metadados**

Metadado	Fontes	Ambiente Web
Identificador	BD (bib_registro.Id)	uri(CKAN)
Data de publicação	Data de inserção no CKAN	dc:date
Data de atualização	BD (bib_registro.atualizacao)	dc:date
Título	Título do Dataset	dc:title
Autor	Projeto MPO e projetos associados	prov:wasGenerateBy
Descrição	BD (bib_registro .descricao)	skos:note fabio:NewsItem
Assunto	-	skos:about
Fonte (origem)	BD (bib_registro .projeto_id)	prov:wasGenerateBy
Licença	cc-by	dc:LicenseDocument
Identificador documentos relacionados	Projeto MPO	skos:about

**Fonte: Dados de pesquisa (2023).**

Observe-se que na coluna de Fontes no Quadro 2, a expressão "BD (tabela.atributo)", registra o nome da tabela e atributo do Banco de Dados de periódicos (MPO\_tst).

Após a definição do conteúdo a ser publicado passamos para o uso da plataforma de compartilhamento.

## 4.2 Publicação dos Datasets no CKAN

Existem diferentes iniciativas para promover o compartilhamento de dados de pesquisa no contexto brasileiro, mas ainda há muitos desafios para serem vencidos. Para além das dificuldades tecnológicas, que podem ser resolvidas de forma mais objetiva, há aspectos de governança, direito de acesso, mudanças comportamentais e políticas para implementação que demandam tempo para sua solução (Faustino Junior et al., 2022).

Como já mencionado anteriormente, a plataforma adotada por este protótipo foi o CKAN, pois esta solução é um open-source data management system utilizado para

disponibilização, compartilhamento e reuso de dados no contexto dos dados governamentais brasileiros. Apesar de existirem esforços no Brasil para a adoção de soluções específicas para o compartilhamento de dados de pesquisa, a decisão de utilizar o CKAN neste projeto de pesquisa foi baseada nos princípios da agilidade, viabilidade e economicidade, na medida em que já existe infraestrutura na Universidade capaz de absorver as demandas do protótipo. A liberação dos datasets 1 e 2 com a respectiva publicação na plataforma CKAN está planejada para o final de outubro de 2023.

## 5 ANÁLISE E DISCUSSÃO

Como relatado na introdução deste relato, o projeto MPO foi iniciado em 2002 e o cenário do desenvolvimento da Ciência passou por diferentes transformações neste período. A proposta inicial de coletar notícias foi sendo incrementada, pois a necessidade de reunir os registros e reutilizá-lo no contexto de trabalhos investigativos conduzidos por membros da equipe impulsionou o uso de ferramentas tecnológicas mais adequadas para o desenvolvimento de pesquisas colaborativas e em rede.

Nesse sentido, a primeira contribuição que pode ser destacada foi a estrutura de armazenamento dos dados, pois a obtenção do modelo conceitual do Banco de Dados (BD) (figura 4) como modelo semântico inicial foi fundamental para a definição dos conjuntos de dados e a formulação das questões de competência apresentadas na seção 4.1.

Na ausência de um plano de Gestão de Dados (Veiga et al., 2022), pode-se depreender que a estrutura de armazenamento proposta e hospedada, em rede e plataforma Web, mas no ambiente corporativo da Unirio implementa algumas das características desejadas para o compartilhamento de dados. Aspectos ligados ao formato dos dados, seus metadados, características para acesso, padrões de

representação, regras de armazenamento e backup, são requisitos atendidos dentro da estrutura tecnológica prevista.

Ainda no contexto do modelo de dados (figura 4), foi possível dar início ao desenvolvimento de um modelo semântico (figura 6) a partir das classes e associações apresentadas no modelo. O uso de representação em RDF com suas triplas sujeito-predicado-objeto guarda correlação com as classes e associações de uma representação para modelo de dados.

Vale destacar que é fundamental o perfil interdisciplinar para a equipe de investigação. Profissionais de informação na avaliação dos metadados e aspectos de representação semântica; profissionais de Ciência da Computação nas tarefas de avaliação da estrutura de Banco de Dados, extração dos dados, uso de ferramentas para transformação e complementação dos datasets, além da carga e disponibilização no CKAN nos formatos csv e JSON.

Adicionalmente, pode-se afirmar que a adoção do repositório CKAN reduziu a complexidade no processo de publicação. A existência de infraestrutura técnica, por meio

da equipe de dedicada, contribuiu para o sucesso na disponibilização de dados.

Quanto à possibilidade de reuso, o rol de Application Programming Interfaces (APIs) disponíveis (Ckan, [s.d]) deve impulsionar trabalhos com os datasets publicados.

O processo de finalização e avaliação dos resultados contará com a verificação do

## 6 CONSIDERAÇÕES FINAIS

Com a finalização do esforço de disponibilização dos conjuntos de dados se iniciou a preparação para a fase de pós-FAIRificação com as etapas que atenderão às atividades de Host e Access FAIRData com o uso do FairDataPoint (FDP) (Benhamed, et al., 2023). A jornada para implementação do FDP demandará esforço de customização que deverá ser avaliada em conjunto com os gestores do CKAN na Unirio.

As conclusões apontam para a viabilidade de compartilhar dados em plataformas distintas, mas que seguem padrões de interoperabilidade. Destaca-se ainda: a discussão sobre o uso de metadados para descrever os datasets, pois esses se referem a

## 7 REFERÊNCIAS

- Allemang, D., & Hendler, J. (2011). Semantic web for the working ontologist: effective modeling in RDFS and OWL. Elsevier.
- Baker, M. (2016). Is there a reproducibility crisis? *Nature*, 533, p. 452, 26 May 2016. <https://www.nature.com/news/1-500-scientists-lift-the-lid-on-reproducibility-1.19970>.
- Benhamed, O. M. et al. (2023). The FAIR Data Point: Interfaces and Tooling. *Data Intelligence 2023*; 5 (1): 184–201. doi: [https://doi.org/10.1162/dint\\_a\\_00161](https://doi.org/10.1162/dint_a_00161)
- Brasil. (2019, Dezembro 5). Infraestrutura nacional de dados abertos. <https://www.gov.br/governodigital/pt-br/dados-abertos/infraestrutura-nacional-de-dados-abertos>.
- nível de FAIRness dos datasets publicados. Nesse sentido, será essencial o uso de soluções como F-JUI Automated FAIR Data Assessment tool<sup>[5]</sup>, Self-Assessment Tool to Improve the FAIRness of Your Dataset – SATIFYD<sup>[6]</sup> e Ferramenta online proposta pela Data e FAIRDataBr, (Universidade Federal da Paraíba, 2022) para avaliar o nível de alinhamento com os princípios FAIR.
- conjuntos de dados que foram obtidos a partir de bancos de dados relacionais; e o uso do plano de gestão de dados no apoio a definição desta infraestrutura. O tempo e custo associados ao esforço de compartilhamento estão ligados à disponibilidade de plataformas tecnológicas.
- Ademais, com o desenvolvimento do protótipo já se vislumbrou a possibilidade de incrementar as conexões semânticas. Criar uma conexão com características da música pode ser uma evolução para possíveis integrações com outras estruturas semânticas, como, por exemplo, a Musicontology<sup>[7]</sup>, suas classes e propriedades, envolvendo artistas e demais músicos.
- Ckan (s.d.). API Guide. 2.10.1. <https://docs.ckan.org/en/2.10/api/index.html>
- Creative Commons (s.d.). Licenses. [https://creativecommons.org/licenses/?lang=pt\\_BR](https://creativecommons.org/licenses/?lang=pt_BR)
- Curry, R. G. & Aventurier, P. (2017). O paradigma da publicação de dados e suas diferentes abordagens. [Conference session]. Encontro Nacional de Pesquisa e Pós-graduação em Ciência da Informação, XVIII ENANCIB. <http://hdl.handle.net/20.500.11959/brapci/105144>.
- Fanelli, D. (2018). Opinion: Is science really facing a reproducibility crisis, and do we need it to? *Proceedings of the National*



- Academy of Sciences of the United States of America, 115(11), 2628-2631, mar. <https://www.pnas.org/doi/10.1073/pnas.1708272114>.
- Fenner, M., Crosas, M., Grethe, J.S., Kennedy D., Hermjakob H., Rocca-Serra P., Durand G., Berion R., Karcher S., Martone M. & Clark T. (2019). A data citation roadmap for scholarly data repositories. *Sci Data* 6(28). <https://doi.org/10.1038/s41597-019-0031-8>
- Gabriel Junior, R. F., Rocha, R. P., Caregnato, S. Vans, S., Pavão, C. G., Moura, A. P. M. de & Passos, P. C. S. J. (2022, agosto 9-11). Iniciativa para o desenvolvimento e implementação de repositórios de dados de pesquisa: RNP/Ibict/CNPq. [Conference session]. ENCONTRO DA REDE BRASILEIRA DE REPOSITÓRIOS DIGITAIS, Fiocruz-Icict e IBICT. <https://www.arca.fiocruz.br/handle/icict/54759>
- Henning, P. C., Ribeiro, C. J. S., Sales, L. F., Moreira, L. R, Santos & L. O. B. da S. (2018). Desmistificando os princípios FAIR: conceitos, métricas, tecnologias e aplicações inseridas no ecossistema dos dados FAIR. *Tendências da Pesquisa Brasileira em Ciência da Informação*, 11(1), 1-23. <https://revistas.ancib.org/index.php/tpbci/article/view/458>.
- Jacobsen, A. et al. (2020). A generic workflow for the Data FAIRification Process. *Data Intelligence*, Cambridge, 2(1-2), 56-65, 2020. [https://doi.org/10.1162/dint\\_a\\_00028](https://doi.org/10.1162/dint_a_00028).
- Jorge, V. A. & Albagli, S. (2018). Compartilhamento de dados de pesquisa em saúde: iniciativas do National Institutes of Health (NIH). *Reciis*, 12(4), out./dez. Disponível em: <https://www.reciis.icict.fiocruz.br/index.php/reciis/article/view/1499>.
- Kalinauskaitė, D. (2017). To be findable, accessible, interoperable and reusable: language data and technology infrastructure for supporting the FAIR data approach. [Conference session]. CEUR Workshop Proceedings, Kaunas. ICYRIME 2017: proceedings of the symposium for young researchers in informatics, mathematics and engineering. Aachen: CEUR-WS. <https://www.vdu.lt/cris/entities/publication/6138a749-b4c6-48b0-9020-b2b97ef1dd91/details>.
- Koche, J. C. (2015). O fluxograma da pesquisa científica. In J. C. KOCHÉ (ed.), *Fundamentos de metodologia científica: teoria da ciência e iniciação à pesquisa*. 34. Ed. (pp. 121-136). Petrópolis. Vozes.
- Novaes, F. C. P.; Tolare, J. B. & Moreira, W. (2023). Análise da representação da música em bibliotecas digitais. [Conference Session] ISKO-Brasil 2023. Organização e representação do conhecimento em diferentes contextos. PPGCI-UEL. (pp. 126-133). Londrina.
- Nüst, D., Granell, C., Hofer, B., Konkol, M., Ostemann, F. O, Sileryte, R. & Cerutti, V. (2018). Reproducible research and GIScience: an evaluation using AGILE conference papers. *PeerJ*, 6, e5072. <https://peerj.com/articles/5072/>.
- Open Knowledge Foundation (s.d). Open Database License. <https://opendefinition.org/licenses/odc-odbl/>
- Packer, A. L. & Santos, S. (2019). Ciência aberta e o novo modus operandi de comunicar pesquisa – parte I. <https://blog.scielo.org/blog/2019/08/01/ciencia-aberta-e-o-novo-modus-operandi-de-comunicar-pesquisa-parte-i/#.Y5Rn433MLIV>.
- Portugal. (2016). Ministério da Ciência, Tecnologia e Ensino Superior. *Ciência*

- aberta. <https://www.ciencia-aberta.pt/sobre-ciencia-aberta>
- Rautemberg, S., Souza, L. de, Dall’Agnol, J. M. H & Michelon, G. A. (2018). Guia prático para publicação de dados abertos conectados na web. Curitiba. Appris.
- Ribeiro, C. J. S., Sousa, C. S. de, Tertulino, C. I., Amaral, I. S. do, Santos, M. J. S. dos, Soares, Cóquero, S. de M. S., & Ulhoa, M. T de (2023). Knowledge Organization no Processo de FAIRificação de Datasets: Estruturando a Semântica e Interligando as Notícias do Banco de Dados de Periódicos Musicais Oitocentistas. \*[Conference Session]. [Conference Session] ISKO-Brasil 2023. Organização e representação do conhecimento em diferentes contextos. PPGCI-UEL. Londrina. doi <https://doi.org/10.29327/5261847>
- Ribeiro, C. J. S., & Almeida, R. F. de. (2011). Dados Abertos Governamentais (Open Government Data): instrumento para exercício de cidadania pela sociedade. [Conference Session]. Encontro Nacional de Pesquisa em Pós-Graduação em Ciência da Informação, Brasília, 12. <http://hdl.handle.net/20.500.11959/brapci/182610>.
- Ribeiro, C. J. S. & Velho, A. C. M. (2021). Princípios FAIR e a gestão de bases governamentais: análise do compartilhamento de dados de registros civis por meio da iniciativa GovData. In L. F. Sales, V. S. de O. Veiga, P. Henning, L. F. Sayão. (Eds.), Princípios FAIR aplicados à gestão de dados de pesquisa. (pp 69-83). Ibict. <https://ridi.ibict.br/handle/123456789/1182>
- Sayão, L. F. & Sales, L. F. (2015). Guia de gestão de dados de pesquisa para bibliotecários e pesquisadores. CNEN/IEN. <http://www.aben.com.br/Arquivos/420/420.pdf>.
- Severino, A. J. (2007). Metodologia do trabalho científico. Cortez.
- University of Twente. (2022). FAIR principles and the FAIRification process (2021-2B) [Web course]. University of Twente. The Netherlands.
- Veiga, V. S. de O., Dib, S. F., Henning, P. C., Nascimento, F. J. T., Gunaes, P. , Penedo, E., Guimarães, C. (2022). Plano de Gestão de Dados de Pesquisa - PGD: guia de elaboração. [Working Paper]. Fiocruz/Icict, <https://www.arca.fiocruz.br/handle/icict/54805>
- Wilkinson, M. D. et al. (2017). A design framework and exemplar metrics for FAIRness. [bioRxiv preprint] <https://www.biorxiv.org/content/biorxiv/early/2017/11/29/225490.full.pdf>.

## 8 NOTAS

[1] Realizado com o apoio do Conselho Nacional de Desenvolvimento Científico e Tecnológico – CNPq

[2] Ver: <http://dados.unirio.br/>

[3] Ver: <https://musicabrazilis.org.br/>

[4] Ver: <https://memoria.bn.br/hdb/periodico.aspx>

[5] Ver: <https://www.fairsfair.eu/f-uji-automated-fair-data-assessment-tool>

[6] Ver: SATIFYD. Disponível em: <https://satisfyd.dans.knaw.nl/>

[7] Ver: <http://musicontology.com/>