

# INTEROPERABILIDADE DE ACERVOS DIGITAIS: INTEGRAÇÃO DO CPDOC COM A PLATAFORMA ROSSIO

UM ESTUDO DE CASO SOBRE INTEROPERABILIDADE E TECNOLOGIAS DIGITAIS NA GESTÃO DE  
ACERVOS

TATIANA LAGE<sup>1\*</sup>  
FUNDAÇÃO GETULIO VARGAS

**RESUMO:** O artigo aborda a implementação de práticas de interoperabilidade entre o acervo do CPDOC (Centro de Pesquisa e Documentação de História Contemporânea do Brasil) e o Rossio, uma infraestrutura digital dedicada à disponibilização de recursos culturais e históricos de Portugal. A pesquisa inclui o desenvolvimento de uma tabela de equivalência de metadados, utilizando o padrão Dublin Core, e a construção de um notebook em Python para a seleção e manipulação de recursos digitais do CPDOC. Os resultados incluem a integração dos recursos selecionados na plataforma Rossio, promovendo uma colaboração interinstitucional que amplia o acesso ao patrimônio cultural de ambos os países. O estudo discute os desafios enfrentados e as soluções propostas, destacando o papel da interoperabilidade e do uso de tecnologias abertas para a preservação e disseminação de acervos históricos.

**Palavras-chave:** Interoperabilidade, Dublin Core, CPDOC, Rossio, Patrimônio Cultural.

**ABSTRACT:** This article discusses the implementation of interoperability practices between the CPDOC (Center for Research and Documentation of Contemporary History of Brazil) archive and Rossio, a digital infrastructure dedicated to the dissemination of cultural and historical resources from Portugal. The research includes the development of a metadata equivalence table using the Dublin Core standard and the creation of a Python notebook for the selection and manipulation of CPDOC's digital resources. The results include the integration of selected resources into the Rossio platform, fostering an interinstitutional collaboration that enhances access to the cultural heritage of both countries. The study addresses the challenges faced and the solutions proposed, highlighting the role of interoperability and the use of open technologies for the preservation and dissemination of historical archives.

**Keywords:** Interoperability, Dublin Core, CPDOC, Rossio, Cultural Heritage.

---

<sup>1\*</sup> Tatiana Lage ([tatiana.lage28@gmail.com](mailto:tatiana.lage28@gmail.com)) é graduanda em Ciências Sociais pelo Centro de Documentação de História Contemporânea da Fundação Getulio Vargas (CPDOC FGV).

## **1. INTRODUÇÃO**

A interoperabilidade entre acervos culturais e históricos é uma questão central para a preservação e disseminação de informações em ambientes digitais. Com o avanço das tecnologias de dados abertos interligados, torna-se possível conectar diferentes repositórios de maneira eficiente, permitindo que pesquisadores, educadores e o público em geral acessem conteúdos diversos de forma integrada. Nesse contexto, o presente artigo tem como objetivo descrever o processo de interoperabilidade entre o acervo do CPDOC (Centro de Pesquisa e Documentação de História Contemporânea do Brasil) e a infraestrutura digital Rossio, dedicada à organização e disponibilização de recursos associados ao patrimônio cultural português. O projeto envolveu o desenvolvimento de uma tabela de equivalência de metadados entre o CPDOC e o padrão Dublin Core, amplamente utilizado em acervos digitais, além da criação de um notebook em Python para a seleção e manipulação de recursos provenientes do acervo do CPDOC relacionados a Portugal. A integração dos dados selecionados à plataforma Rossio resultou em uma importante colaboração interinstitucional, enriquecendo os acervos disponíveis e promovendo um acesso mais eficiente ao patrimônio cultural. Este artigo discute as etapas dessa integração, incluindo a construção da tabela de equivalência, os desafios técnicos encontrados e as soluções propostas para facilitar a interoperabilidade entre os acervos. Além disso, são analisados os impactos da adoção de padrões de metadados na organização e recuperação de informações culturais, destacando o papel do Dublin Core como ferramenta central para a padronização de acervos digitais.

## **2. CONTEXTUALIZAÇÃO TEÓRICA**

### **2.1 Centro de Pesquisa e Documentação de História Contemporânea do Brasil**

O Centro de Pesquisa e Documentação de História Contemporânea do Brasil (CPDOC), fundado em 1973, faz parte da Fundação Getúlio Vargas e é um dos principais acervos históricos do Brasil. O CPDOC foi criado com o objetivo de reunir, preservar e dar acesso a documentos que são fundamentais para a compreensão da história contemporânea brasileira, especialmente após a década de 1930. Seu acervo inclui manuscritos, entrevistas

orais, fotografias, materiais audiovisuais e iconográficos, que documentam a vida política, econômica e social do país<sup>2</sup>.

Um dos grandes diferenciais do CPDOC é o foco em arquivos pessoais, especialmente aqueles de figuras de destaque na política e nas ciências sociais brasileiras, como Getúlio Vargas e Oswaldo Aranha. A ideia de que arquivos privados pessoais podem ter grande valor para a pesquisa histórica foi um dos princípios que guiaram a fundação do CPDOC, reforçando o conceito de que "o fundo privado pessoal tem interesse para a pesquisa"<sup>3</sup>.

De acordo com dados apresentados por Souza et al. (2012), o acervo do CPDOC abrigava, à época, aproximadamente 200 fundos, totalizando cerca de 1,8 milhão de documentos em diferentes formatos, incluindo textos, imagens e vídeos<sup>4</sup>. Esses arquivos, doados ao CPDOC por personalidades públicas brasileiras, formam o mais importante acervo de arquivos pessoais de figuras públicas no país. O acervo também inclui um programa de História Oral, iniciado em 1975, com cerca de 1.000 entrevistas gravadas, correspondendo a mais de 5.000 horas de gravação, das quais metade estava disponível para consulta online<sup>5</sup>.

Outro destaque do CPDOC é o Dicionário Histórico Biográfico Brasileiro (DHBB), que começou a ser desenvolvido em 1974 e inclui verbetes sobre personagens e eventos relevantes para a história do Brasil contemporâneo. A versão mais recente do dicionário, lançada na web em 2010, conta com 7.553 verbetes, dos quais 6.584 são biográficos e 969 tratam de instituições, eventos e conceitos relacionados ao período pós-1930<sup>6</sup>.

Nos últimos anos, o CPDOC enfrentou desafios relacionados à digitalização de seu acervo, com o intuito de ampliar o acesso a seus documentos históricos. A digitalização de fotografias, iniciada nos anos 2000, foi uma das primeiras iniciativas bem-sucedidas, seguida pela digitalização de documentos textuais, como o fundo Getúlio Vargas<sup>7</sup>. Esses esforços permitiram que o CPDOC se posicionasse como pioneiro na digitalização de acervos históricos no Brasil, disponibilizando online uma vasta gama de documentos acessíveis a pesquisadores de todo o mundo<sup>8</sup>. A digitalização também trouxe novos desafios, especialmente no que diz respeito à preservação e ao armazenamento seguro dos documentos digitalizados. Como Spohr ressalta, "nossas preocupações com o gerenciamento de nosso

---

<sup>2</sup> Spohr, Martina. *O acervo histórico do CPDOC: novas perspectivas*. Revista do Arquivo Geral da Cidade do Rio de Janeiro, 2013, p. 269.

<sup>3</sup> Spohr, 2013, p. 270.

<sup>4</sup> Souza, Renato. Portal Semântico do CPDOC. IV Encontro de Bases de Dados Sobre Informações Arquivísticas, 2012, p. 2.

<sup>5</sup> Souza, 2012, p. 2.

<sup>6</sup> Souza, 2012, p. 3

<sup>7</sup> Spohr, 2013, p. 273

<sup>8</sup> Spohr, 2013, p. 274.

acervo e a salvaguarda de boa parte do acervo digitalizado tornaram-se verdadeiras políticas dentro do CPDOC<sup>9</sup>. Portanto, o acervo implementou um sistema de armazenamento digital robusto, além de ter adotado políticas de preservação preventiva para seus acervos físicos e digitais, buscando garantir o acesso contínuo e seguro ao patrimônio histórico que administra<sup>10</sup>.

## 2.2 INFRAESTRUTURA ROSSIO

O ROSSIO é uma infraestrutura digital inovadora, destinada a agregar, organizar e disponibilizar recursos digitais nas áreas de Ciências Sociais, Artes e Humanidades. Seu principal objetivo é promover o acesso livre ao patrimônio cultural português, facilitando a pesquisa e o uso de recursos digitais de diversas instituições de ensino superior e culturais portuguesas<sup>11</sup>.

Infraestrutura digital é definida como um conjunto complexo de dados, pessoas e serviços, que integra recursos diversos e os disponibiliza para uma ampla base de usuários por um longo prazo. Essas infraestruturas são essenciais para apoiar a pesquisa colaborativa e individual, além de facilitar o compartilhamento de resultados de forma acessível e reutilizável<sup>12</sup>. A infraestrutura ROSSIO, coordenada pela NOVA School of Social Sciences and Humanities, é composta por um consórcio de sete instituições portuguesas do patrimônio cultural e acadêmico<sup>13</sup>.

O ROSSIO se destaca também pela sua capacidade de coletar metadados de provedores de dados, utilizando o protocolo OAI-PMH (Open Archives Initiative Protocol for Metadata Harvesting). Esse protocolo é fundamental para a interoperabilidade entre bibliotecas e repositórios digitais, permitindo que os provedores de dados compartilhem seus metadados de maneira eficiente e acessível. Isso facilita a descoberta e o uso de informações digitais em larga escala<sup>14</sup>.

---

<sup>9</sup> Spohr, 2013, p. 276.

<sup>10</sup> Spohr, 2013, p. 277.

<sup>11</sup> Silva, G.M. et al. ROSSIO Infrastructure: A Digital Humanities Platform to Explore the Portuguese Cultural Heritage. *Information* 2022, 13, 50, p. 2.

<sup>12</sup> Silva et al., 2022, p. 2.

<sup>13</sup> Silva et al., 2022, p. 3.

<sup>14</sup> Silva et al., 2022, p. 5.

A infraestrutura ROSSIO adota o European Data Model (EDM), um modelo de dados que permite representar objetos digitais de diferentes domínios. Esse modelo padroniza a representação e o intercâmbio de metadados, garantindo consistência e uniformidade na forma como os dados são organizados e processados<sup>15</sup>. O ROSSIO utiliza um perfil de aplicação desse modelo conhecido como EDM-DRD, que foi desenvolvido para melhorar a capacidade de troca de dados e enriquecer os metadados recebidos dos provedores de dados<sup>16</sup>.

O ROSSIO oferece uma série de serviços inovadores, como um portal de descoberta, exposições digitais e um ambiente virtual de pesquisa (VRE). O portal de descoberta permite que usuários acessem mais de 5 milhões de recursos digitais catalogados, enquanto as exposições digitais e as coleções oferecem uma forma interativa e envolvente de explorar o patrimônio cultural português<sup>17</sup>. Além disso, a infraestrutura ROSSIO está alinhada com os princípios FAIR (Findable, Accessible, Interoperable, Reusable), que visam melhorar a descoberta, o acesso, a interoperabilidade e a reutilização de dados científicos. Esses princípios garantem que os dados sejam facilmente encontrados, acessíveis gratuitamente, estruturados para interoperabilidade e reutilizáveis com licenças apropriadas<sup>18</sup>.

### **2.3 INTEROPERABILIDADE ENTRE ACERVOS DIGITAIS**

A interoperabilidade desempenha um papel crucial na integração de acervos digitais, permitindo que diferentes sistemas conversem entre si de maneira eficiente e coerente. No contexto de bibliotecas, museus e arquivos, essa integração facilita a troca de dados entre repositórios, garantindo que o conteúdo compartilhado mantenha seu significado original. Conforme Carlos Henrique Marcondes explora em sua obra *Dados Abertos Interligados: Publicação, Recuperação e Integração de Acervos de Arquivos, Bibliotecas e Museus na Web*, a adoção de tecnologias de dados abertos vinculados e padrões semânticos revolucionou a forma como acervos digitais podem ser acessados e integrados<sup>19</sup>.

Segundo Marcondes, os sistemas de informação que utilizam essas tecnologias não apenas publicam dados, mas também "compreendem seus significados", promovendo uma interação mais eficiente e inteligente entre os usuários e as bases de dados<sup>20</sup>. Esse modelo, que

---

<sup>15</sup> Silva et al., 2022, p. 6.

<sup>16</sup> Silva et al., 2022, p. 7.

<sup>17</sup> Silva et al., 2022, p. 12.

<sup>18</sup> Silva et al., 2022, p. 14.

<sup>19</sup> Carlos Henrique Marcondes, *Dados Abertos Interligados: Publicação, Recuperação e Integração de Acervos de Arquivos, Bibliotecas e Museus na Web*, 2021, p. 45.

<sup>20</sup> Marcondes, 2021, p. 47.

se apoia na Web Semântica, permite que conteúdos de diferentes acervos sejam interligados de forma a criar uma rede de conhecimento mais rica e acessível. Uma das críticas centrais do autor aos catálogos digitais tradicionais é que, muitas vezes, eles operam como "silos" de informação, dificultando a conexão e o compartilhamento de dados entre diferentes repositórios<sup>21</sup>. Para superar essas limitações, Marcondes sugere que os acervos digitais sejam estruturados em formatos interoperáveis, utilizando padrões abertos que permitam sua integração por curadores originais e terceiros. Isso inclui o uso de identificadores persistentes, como URIs, e o estabelecimento de links semânticos entre catálogos e acervos, permitindo que informações dispersas se complementem e enriqueçam mutuamente<sup>22</sup>.

A interoperabilidade "não é apenas uma questão técnica, mas também uma questão de abertura e compartilhamento de conhecimento", o que implica que as instituições precisam adotar uma postura colaborativa para maximizar o potencial dessas tecnologias<sup>23</sup>. A padronização, nesse contexto, desempenha um papel fundamental, garantindo que diferentes sistemas sejam capazes de interpretar corretamente os metadados compartilhados, promovendo assim a integração eficiente de acervos<sup>24</sup>. No contexto da Web Semântica, os dados são descritos por meio de Resource Description Framework (RDF), que permite a criação de relações entre dados em diferentes sistemas, promovendo a interoperabilidade em uma escala global. Com o RDF, os conteúdos de diferentes acervos podem ser enriquecidos através de conexões semânticas, oferecendo uma experiência mais rica de recuperação de dados<sup>25</sup>. Para que isso aconteça de forma eficiente, é necessária a adoção de vocabulários padronizados, o que facilita a interoperabilidade semântica e garante que os dados possam ser compreendidos em diferentes contextos<sup>26</sup>.

A interoperabilidade, como descrita por Marcondes, é a chave para a criação de redes globais de dados que possam ser reutilizadas em múltiplos contextos, promovendo uma nova forma de colaboração e de gestão de acervos culturais e históricos<sup>27</sup>. No caso do CPDOC e do Rossio, a implementação de dados abertos interligados e a utilização de protocolos como o OAI-PMH facilitam a troca de metadados entre as instituições, promovendo um

---

<sup>21</sup> Marcondes, 2021, p. 52.

<sup>22</sup> Marcondes, 2021, p. 53.

<sup>23</sup> Marcondes, 2021, p. 55.

<sup>24</sup> Marcondes, 2021, p. 56.

<sup>25</sup> Marcondes, 2021, p. 58.

<sup>26</sup> Marcondes, 2021, p. 60.

<sup>27</sup> Marcondes, 2021, p. 61.

enriquecimento mútuo dos conteúdos e garantindo que os acervos possam ser acessados de maneira integrada<sup>28</sup>

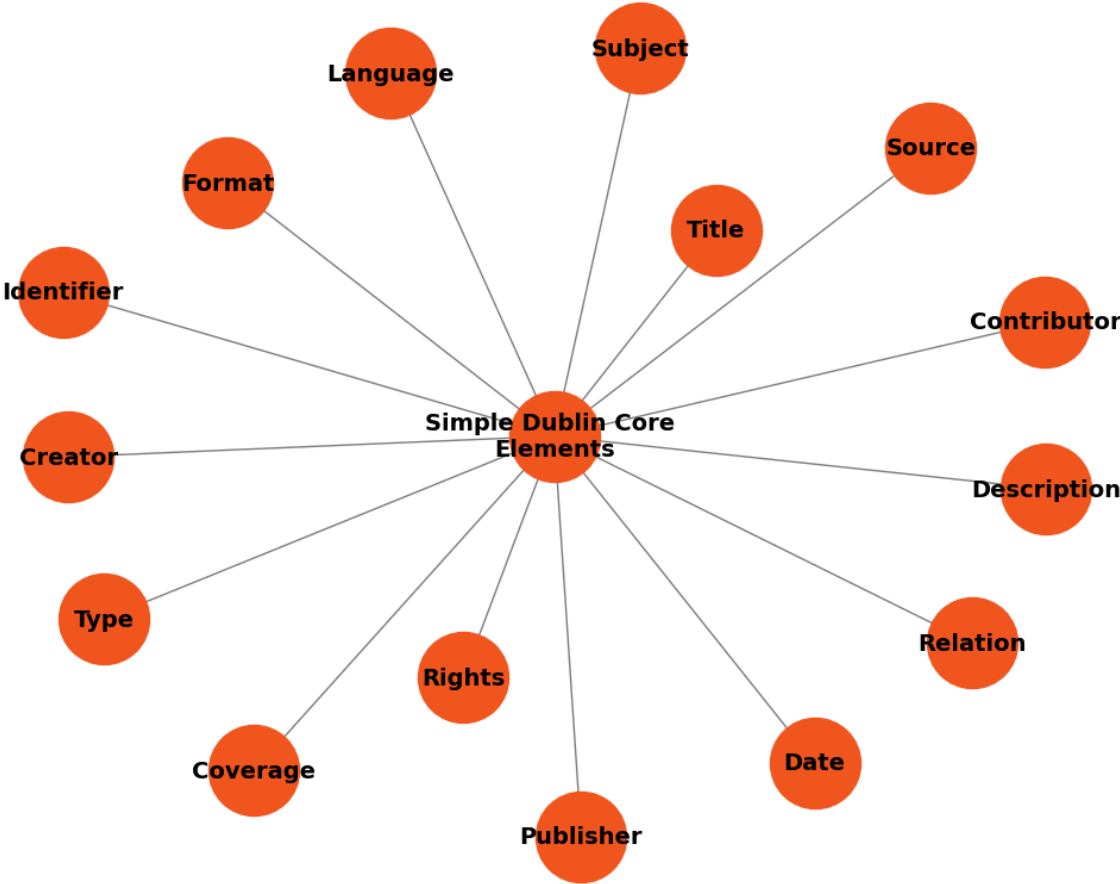
## **2.4 O PADRÃO DUBLIN CORE: ESTRUTURA E APLICAÇÕES**

O Dublin Core é um padrão internacional amplamente utilizado para a descrição de diversos tipos de recursos digitais, como documentos, imagens e acervos em bibliotecas, museus e arquivos. Seu principal objetivo é facilitar a descoberta, a organização e a recuperação de informações em ambientes digitais, oferecendo um conjunto de elementos de metadados padronizados que garantem a interoperabilidade entre diferentes sistemas<sup>29</sup>. Composto por 15 elementos principais, o Simple Dublin Core permite uma descrição padronizada e aplicável a uma ampla gama de recursos. Esses elementos incluem informações essenciais, como o título, o criador, o assunto e a descrição do recurso, entre outros, que facilitam a recuperação e o uso eficiente das informações. Para uma visão clara e organizada dos elementos, um mapa mental foi elaborado, apresentando uma síntese visual desses 15 componentes.

---

<sup>28</sup> Marcondes, 2021, p. 65.

<sup>29</sup> DCMI: Dublin Core Metadata Initiative. Disponível em: <https://www.dublincore.org/specifications/dublin-core/>. Acesso em:



30

Além do mapa mental, foram desenvolvidas duas tabelas descritivas que detalham os **nomes** e **descrições** de cada um dos elementos do **Simple Dublin Core**. Essas tabelas, divididas em duas partes (Tabela 1 e Tabela 2), servem como uma referência prática para a compreensão dos elementos e suas respectivas aplicações no contexto da interoperabilidade de acervos digitais (ver Tabelas 1 e 2).

<sup>30</sup> Figura 1 - Mapa mental dos elementos do Dublin Core. Fonte: Elaborado pela autora (2024).



Elementos Dublin Core pt.1	DESCRIÇÃO
Title (Título)	Representa o nome dado ao recurso.
Creator (Criador)	A pessoa ou entidade responsável pela criação do conteúdo do recurso.
Subject (Assunto)	Descreve o tema ou tópico principal do recurso.
Description (Descrição)	Fornecer uma visão geral ou resumo do conteúdo do recurso.
Publisher (Publicador)	A entidade responsável por disponibilizar o recurso.
Contributor (Contribuidor)	A entidade ou pessoa que contribuiu para o conteúdo do recurso, mas não é o criador principal.
Date (Data)	Refere-se à data em que o recurso foi criado ou disponibilizado.
Type (Tipo)	Indica a natureza ou gênero do conteúdo, como um artigo, imagem, áudio, etc.

31

Elementos Dublin Core pt.2	DESCRIÇÃO
Format (Formato)	Especifica o formato físico ou digital do recurso, como PDF, JPEG, MP3.
Identifier (Identificador)	Um identificador único para o recurso, como um URL, DOI, ou outro número de identificação.
Source (Fonte)	Indica o recurso original a partir do qual o conteúdo atual foi derivado, se aplicável.
Language (Idioma)	Define o idioma principal do conteúdo do recurso.
Relation (Relação)	Descreve a relação do recurso com outros recursos, como uma versão anterior, uma tradução, ou um conjunto de dados relacionado.
Coverage (Abrangência)	Define o escopo ou a abrangência espacial e temporal do conteúdo, como localizações geográficas ou períodos históricos.
Rights (Direitos)	Especifica os direitos de uso e acesso ao recurso, incluindo informações de copyright ou outras restrições.

32

Além do **Simple Dublin Core**, o padrão inclui o **Qualified Dublin Core**, que adiciona qualificadores para fornecer uma descrição mais detalhada e precisa dos recursos. Isso o torna adequado para uma variedade de contextos, como acervos acadêmicos e culturais, proporcionando uma organização mais robusta e flexível<sup>33</sup>.

O **Dublin Core** é amplamente reconhecido por organizações internacionais, como a **ISO 15836** e o **RFC 5013**, e pode ser utilizado em conjunto com outras tecnologias de

<sup>31</sup> Tabela 1 - Descrição dos elementos do Simple Dublin Core. Fonte: Adaptado de Dublin Core Metadata Initiative, traduzido pela autora (2024).  
<sup>32</sup> Tabela 2 - Descrição dos elementos do Simple Dublin Core (continuação). Fonte: Adaptado de Dublin Core Metadata Initiative, traduzido pela autora (2024).  
<sup>33</sup> Marcondes, 2021, p. 53.

metadados, como **XML** e **RDF** (Resource Description Framework). Essa integração facilita sua aplicação em sistemas de dados abertos interligados, promovendo o compartilhamento eficiente de informações e a interoperabilidade entre diferentes repositórios<sup>34</sup>.

A simplicidade e a flexibilidade do **Dublin Core** tornam-no uma solução eficiente para a organização de acervos digitais. Sua adoção permite a interoperabilidade entre sistemas distintos, facilitando a integração de acervos em uma rede global de informações. Por essas razões, o padrão é amplamente utilizado em instituições de memória, como bibliotecas e arquivos, que buscam formas eficazes de organizar e compartilhar seus acervos digitais<sup>35</sup>.

No contexto desta pesquisa, o estudo aprofundado do **Dublin Core** foi essencial para o desenvolvimento de uma **tabela de equivalência de metadados** entre o CPDOC e o padrão, promovendo maior interoperabilidade entre as instituições e facilitando a troca de informações entre bases de dados. Esse trabalho colaborativo foi fundamental para garantir a compatibilidade dos metadados do CPDOC com padrões amplamente aceitos e promover a integração eficiente com a infraestrutura digital do Rossio.

### 3. METODOLOGIA E IMPLEMENTAÇÃO

#### 3.1 DESENVOLVIMENTO DA TABELA DE CORRESPONDÊNCIA DE METADADOS

A primeira etapa do desenvolvimento da tabela de equivalência de metadados entre o CPDOC e o Dublin Core foi a criação de um template inicial. Esse template estabeleceu a correspondência entre os metadados utilizados nas diferentes tipologias de acervos do CPDOC — como manuscritos, fotografias e entrevistas — e os elementos do Dublin Core. O processo exigiu uma análise detalhada dos campos de metadados utilizados internamente no CPDOC e seu alinhamento com os elementos do Dublin Core, que fornecem uma estrutura padronizada e amplamente reconhecida para descrever recursos digitais.

Para as entrevistas, por exemplo, foi mapeada a função dos entrevistadores como os responsáveis pela criação. As datas iniciais e finais de gravação foram inseridas como o elemento "Date", e o formato da entrevista foi descrito com base na duração e tipo de mídia. As entrevistas foram identificadas digitalmente por URLs, e suas relações com projetos maiores foram documentadas. Além disso, os temas principais das entrevistas foram

---

<sup>34</sup> Marcondes, 2021, p. 58-60.

<sup>35</sup> Marcondes, 2021, p. 65

registrados por meio de descritores, assegurando que fossem recuperáveis a partir de palavras-chave associadas.

No caso dos manuscritos, o mapeamento enfrentou desafios específicos, como a ausência de alguns elementos fundamentais, como o criador e o título, que não estavam disponíveis nos metadados existentes no CPDOC. As datas de produção foram registradas de maneira mais flexível, utilizando períodos ou anos, e a descrição física do formato foi realizada com base no número de documentos e folhas. Como o CPDOC não possuía metadados relacionados à cobertura geográfica, essa informação foi omitida para os manuscritos.

Já para as fotografias, o fotógrafo, quando identificado, foi registrado como responsável pela criação. As datas de produção e os títulos foram registrados em conformidade com os padrões recomendados. Além disso, o formato físico das fotografias foi descrito com detalhes, e os temas principais foram documentados utilizando descritores, assim como nas entrevistas e manuscritos. As fotografias também foram identificadas digitalmente por URLs e associadas a projetos ou fundos maiores.

Um ponto importante de destaque é que em algumas situações, dois ou mais metadados do CPDOC precisaram ser mapeados para um único elemento do Dublin Core. Um exemplo disso foi o caso das entrevistas, nas quais dois metadados relacionados à data inicial e à data final precisaram ser agregados em um único campo "Date" no Dublin Core. Além disso, houve a ausência de elementos correspondentes em alguns casos, como no exemplo dos manuscritos, e a falta de informações sobre direitos autorais para as três tipologias de recursos (entrevistas, manuscritos e audiovisuais), o que impossibilitou o preenchimento do campo "Rights" no Dublin Core.

### **3.2 COLABORAÇÃO COM A EQUIPE DO ROSSIO**

Após a criação do template inicial de mapeamento, o projeto passou por um período de ajustes e refinamentos, de acordo com as especificidades dos metadados disponíveis no Rossio. A colaboração com a equipe técnica do Rossio foi essencial nesse processo. O objetivo principal dessa colaboração era garantir que as adaptações feitas nos metadados do CPDOC fossem totalmente compatíveis com a infraestrutura do Rossio, promovendo uma integração eficiente e sem perdas de informação.

Essa fase da interação entre as equipes do CPDOC e do Rossio envolveu reuniões técnicas para discutir detalhes sobre o formato de envio dos dados, as exigências de metadados e os procedimentos de validação para garantir a interoperabilidade entre as plataformas. Durante esse período, foi realizada uma seleção inicial de amostras com cinco entrevistas, dez materiais audiovisuais e dez manuscritos, todos com temas relacionados a Portugal. Esses dados foram então enviados para o Rossio para validação técnica. Com a resposta positiva da equipe do Rossio, foi possível dar continuidade à seleção completa de recursos do CPDOC relacionados a Portugal, que incluíram entrevistas, manuscritos e materiais audiovisuais. O resultado desse esforço foi a integração desses recursos à plataforma Rossio, onde passaram a fazer parte de um acervo digital compartilhado e acessível ao público.

### **3.3 DESENVOLVIMENTO DO NOTEBOOK EM PYTHON**

#### **3.3.1 Extração Inicial de Dados**

O desenvolvimento do notebook em Python teve como objetivo selecionar e manipular os recursos do acervo digital do CPDOC, relacionados a Portugal, a partir de um conjunto maior de dados previamente extraídos. A extração inicial das entrevistas, manuscritos e materiais audiovisuais foi realizada por Suemi Higuchi, utilizando consultas SQL diretamente na base de dados do CPDOC, obtendo todos os recursos disponíveis nessas categorias.

#### **3.3.2 Seleção de Recursos**

Após a extração inicial, o foco foi aplicar critérios específicos para selecionar os temas de interesse da plataforma Rossio, como "Portugal" e "Comunidade dos Países de Língua Portuguesa". Essa seleção foi realizada por meio do notebook, utilizando funções do pandas para filtrar os registros que continham os descritores temáticos correspondentes. A função `str.contains()` foi usada para identificar esses termos, garantindo a seleção precisa dos dados.

#### **3.3.3 Limpeza e Organização dos Dados**

Durante o processo de filtragem, o notebook também realizou a limpeza e organização dos dados. A verificação dos descritores e a padronização dos campos, como datas e formatos, foram automatizadas no notebook. Por exemplo, as datas dos documentos foram padronizadas

para o formato ISO (YYYY-MM-DD), assegurando consistência nos metadados, utilizando a função *pd.to\_datetime()* do pandas.

Durante a organização dos dados, o notebook também tratou da deduplicação de registros, que foi uma etapa essencial para garantir que cada recurso fosse representado apenas uma vez. Para isso, utilizei as funcionalidades nativas da biblioteca pandas para identificar e remover duplicatas ao comparar metadados como títulos e URLs. O processo foi realizado de forma controlada, garantindo que apenas uma instância de cada recurso fosse mantida no conjunto final de dados.

Para padronizar e adequar os dados ao formato exigido pela plataforma Rossio, foram necessárias algumas adaptações adicionais. Criei colunas específicas, como "NM\_TIPO\_UNIDADE\_DOCUMENTAL" e "PUBLISHER", para os três tipos de recursos. A coluna "NM\_TIPO\_UNIDADE\_DOCUMENTAL" foi preenchida com os valores "Entrevistas", "Audiovisual" ou "Textual", conforme o tipo documental de cada conjunto de dados. Já a coluna "PUBLISHER" foi padronizada com o valor "FGV CPDOC" para todos os recursos, indicando a instituição responsável pela disponibilização do acervo. No caso dos direitos autorais, apesar de sua relevância, não foi possível preencher o campo "Rights", já que essas informações não estavam disponíveis nos metadados fornecidos pelo CPDOC.

Outra adaptação necessária foi a combinação de colunas de localidade em uma única coluna para simplificar a visualização. Além disso, em alguns casos, cada entrevistador ou condição de acesso gerava uma nova linha para a mesma entrevista, o que resultava em duplicações. O notebook consolidou esses registros, reorganizando as informações de modo que cada entrevista fosse representada por uma única linha, com os entrevistadores e condições de acesso listados em uma mesma célula, separados por ponto e vírgula. Esse processo reduziu o número de linhas redundantes e permitiu uma melhor visualização e manipulação dos dados.

### 3.3.4 Geração do Arquivo Final

Finalmente, o notebook organizou os dados selecionados em arquivos **XLSX**, prontos para serem enviados à equipe do Rossio. Foram gerados três arquivos principais — um para cada tipologia (entrevistas, manuscritos e audiovisuais) —, contendo os dados já limpos e organizados.

### 3.4 VISUALIZAÇÃO DOS RECURSOS NO ROSSIO

Apesar de já termos enviado todos os recursos selecionados para o Rossio, até o momento, podemos visualizar apenas a amostra inicial de 25 recursos no portal. Essa amostra inclui entrevistas, manuscritos e materiais audiovisuais que abordam temas relacionados a Portugal. Para visualizar os recursos, é possível realizar uma pesquisa simples ou avançada no site do Rossio<sup>36</sup>. Entre os provedores de recursos listados, encontramos a Fundação Getúlio Vargas (FGV CPDOC), onde os materiais enviados estão disponíveis para consulta.

Um exemplo de recurso já disponível é a Carta de Alexandre Marcondes Filho a João Café Filho, onde o autor narra sua estadia em Portugal e menciona a recepção calorosa dos portugueses à visita presidencial de 1955. O documento possui a seguinte descrição: Título: Carta de Alexandre Marcondes Filho a João Café Filho: Estada em Portugal e Lembranças da Visita Presidencial de 1955; Descrição: Carta de Alexandre Marcondes Filho a João Café Filho narrando sua estada em Portugal e o entusiasmo com que os portugueses relembram a visita do Presidente em agosto de 1955. Paris. Descrição física: 1 doc, 2 f. Assunto: Brasil, Café Filho, Governo Café Filho (1954-1955), Viagens e visitas, Portugal Provedor: Fundação Getúlio Vargas Referência: AMF c 1955.10.00/1 Condições de reprodução: Direitos reservados para efeito de publicação, exposição e utilização comercial.

Além disso, o portal Rossio inclui um campo chamado "Recurso no Provedor", que redireciona o usuário para o link do recurso original no acervo digital do CPDOC, permitindo um acesso mais completo e contextualizado ao material. Esse exemplo demonstra como a integração inicial já está beneficiando o público ao oferecer acesso direto aos documentos do CPDOC no ambiente digital do Rossio, com links diretos para o acervo original.

## 4 DISCUSSÃO

A colaboração entre o **Rossio** e o **CPDOC** foi uma oportunidade exemplar de integração de acervos digitais, conectando os conhecimentos teóricos sobre interoperabilidade com uma aplicação prática que resultou em benefícios significativos para ambas as instituições. A disponibilização de recursos do acervo do CPDOC na plataforma Rossio expandiu a visibilidade e o acesso a documentos fundamentais para o estudo da história contemporânea do Brasil e de Portugal, enriquecendo a experiência de pesquisadores, educadores e do público em geral.

---

<sup>36</sup> Disponível em: <https://rossio.pt/front/home>. Acesso em: 15 de Agosto de 2024.

A integração de entrevistas, manuscritos e materiais audiovisuais contendo informações sobre Portugal trouxe uma importante dimensão intercultural à plataforma Rossio. Com essa colaboração, foi possível não apenas ampliar o volume e a diversidade de recursos disponíveis, mas também garantir que esses recursos fossem organizados e padronizados de acordo com práticas modernas de interoperabilidade, facilitando sua descoberta e uso em contextos de pesquisa mais amplos.

Além de enriquecer o acervo do Rossio, a parceria demonstrou a importância de se adotar padrões abertos e tecnologias interoperáveis para garantir que acervos culturais possam ser integrados e explorados de maneira eficiente. A padronização dos metadados utilizando o Dublin Core permitiu uma troca de informações mais fluida entre as plataformas, possibilitando que os recursos fossem apresentados de forma contextualizada, ao lado de outros documentos históricos e culturais relacionados. Essa abordagem cria um ecossistema de dados culturais interligados, onde as informações podem ser comparadas, complementadas e analisadas a partir de múltiplas perspectivas, promovendo um maior entendimento das conexões históricas entre Brasil e Portugal.

Essa integração também é um exemplo concreto de como colaborações interinstitucionais podem resultar na valorização do patrimônio cultural, ao mesmo tempo em que democratizam o acesso ao conhecimento. Ao garantir que documentos valiosos do acervo do CPDOC estivessem disponíveis para um público internacional através do Rossio, foi possível dar nova vida a recursos que antes eram acessíveis de forma limitada. Isso reforça a importância de iniciativas que promovam a disseminação aberta e global de documentos históricos, permitindo que as narrativas históricas sejam preservadas e continuamente reinterpretadas por diferentes gerações.

Por fim, a experiência colaborativa entre o CPDOC e o Rossio estabelece um modelo de interoperabilidade que pode servir de exemplo para outras instituições que buscam integrar seus acervos de forma eficiente. O uso de tecnologias de dados abertos interligados, como os adotados neste projeto, demonstra que, além de facilitar a gestão de acervos, essas soluções promovem novas formas de interação com o conhecimento e possibilitam uma exploração mais rica e significativa do patrimônio cultural. A parceria também destaca o papel central da interoperabilidade como uma ferramenta estratégica para a preservação e a valorização de documentos culturais e históricos, criando um impacto duradouro tanto para as instituições envolvidas quanto para a comunidade acadêmica e o público em geral.

## 5. CONCLUSÃO

A interoperabilidade entre o CPDOC e a plataforma Rossio demonstra a importância de práticas colaborativas e o uso de tecnologias digitais abertas para a preservação e disseminação do patrimônio cultural. O desenvolvimento de uma tabela de equivalência de metadados, utilizando o padrão Dublin Core, foi essencial para garantir a integração eficiente dos acervos, permitindo a harmonização de dados e o enriquecimento mútuo das plataformas. A adoção de práticas como o uso de identificadores persistentes e protocolos abertos, como o OAI-PMH, foi determinante para alcançar uma integração fluida e garantir que os recursos culturais pudessem ser acessados de forma contextualizada e significativa.

O processo de seleção e manipulação dos recursos relacionados a Portugal, realizado por meio do notebook desenvolvido em Python, mostrou-se uma solução eficaz para lidar com grandes volumes de dados e realizar a curadoria dos itens mais relevantes para a plataforma Rossio. A capacidade de automatizar a limpeza, normalização e deduplicação dos dados durante o processo de filtragem foi fundamental para manter a qualidade dos recursos selecionados e garantir a sua consistência com os padrões internacionais de metadados.

Além dos ganhos técnicos e operacionais, a parceria entre o CPDOC e o Rossio representa um exemplo de como a interoperabilidade pode não apenas ampliar o acesso ao conhecimento, mas também fortalecer o diálogo intercultural entre diferentes acervos e nações. A integração de recursos relacionados à história de Portugal, disponíveis no acervo do CPDOC, promove uma maior compreensão das conexões históricas entre Brasil e Portugal, oferecendo ao público uma experiência de pesquisa mais rica e profunda.

Por fim, a experiência deste projeto evidencia o potencial transformador das tecnologias de dados abertos e da interoperabilidade para a gestão e valorização de acervos culturais. A colaboração entre o CPDOC e o Rossio serve como um modelo de boas práticas para outras instituições que desejam integrar seus acervos digitais de maneira eficiente e promover o acesso global ao conhecimento. Continuar investindo em soluções abertas e colaborativas para a gestão de acervos será essencial para garantir que o patrimônio cultural e histórico continue a ser acessível e relevante para as futuras gerações.



## REFERÊNCIAS

DCMI: Dublin Core Metadata Initiative. Disponível em: <https://www.dublincore.org/specifications/dublin-core/>. Acesso em: 15 ago. 2024.

MARCONDES, Carlos Henrique. Dados Abertos Interligados: Publicação, Recuperação e Integração de Acervos de Arquivos, Bibliotecas e Museus na Web. 2021.

SILVA, G. M. et al. ROSSIO Infrastructure: A Digital Humanities Platform to Explore the Portuguese Cultural Heritage. *Information*, v. 13, n. 50, 2022.

SOUZA, Renato. Portal Semântico do CPDOC. IV Encontro de Bases de Dados Sobre Informações Arquivísticas, 2012.

SPOHR, Martina. O acervo histórico do CPDOC: novas perspectivas. *Revista do Arquivo Geral da Cidade do Rio de Janeiro*, 2013.

ROSSIO. Disponível em: <https://rossio.pt/front/home>. Acesso em: 15 ago. 2024.