



UNIVERSIDADE DE BRASÍLIA
FACULDADE DE CIÊNCIA DA INFORMAÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA INFORMAÇÃO

JOYCE SIQUEIRA

BRASILIANA MUSEUS:
SERVIÇO DE BUSCA E RECUPERAÇÃO DA INFORMAÇÃO A PARTIR DA
AGREGAÇÃO DOS METADOS DOS MUSEUS VINCULADOS AO INSTITUTO
BRASILEIRO DE MUSEUS

BRASÍLIA - DF
2023

JOYCE SIQUEIRA

BRASILIANA MUSEUS:
SERVIÇO DE BUSCA E RECUPERAÇÃO DA INFORMAÇÃO A PARTIR DA
AGREGAÇÃO DOS METADOS DOS MUSEUS VINCULADOS AO INSTITUTO
BRASILEIRO DE MUSEUS

Tese submetida ao Programa de Pós-Graduação em Ciência da Informação (PPGCINF), da Faculdade de Ciência da Informação (FCI), da Universidade de Brasília (UnB) como requisito parcial para obtenção do título de Doutora em Ciência da Informação.

Área de concentração: Gestão, Organização e Comunicação da Informação e do Conhecimento

Linha de pesquisa: Gestão, Tecnologias e Organização da Informação e do Conhecimento

Orientador: Prof. Dr. Dalton Lopes Martins

BRASÍLIA – DF
2023

FOLHA DE APROVAÇÃO

Título: “ BRASILIANA MUSEUS: Serviço de Busca e Recuperação da Informação a partir da Agregação Automatizada dos Metadados dos Objetos Digitais dos Acervos Administrados pelo Instituto Brasileiro de Museus ”

Autor (a): Joyce Siqueira

Área de concentração: Gestão, Organização e Comunicação da Informação e do Conhecimento

Linha de pesquisa: Gestão, Tecnologias e Organização da Informação e do Conhecimento

Tese submetida à Comissão Examinadora designada pelo Colegiado do Programa de Pós-graduação em Ciência da Informação da Faculdade de Ciência da Informação da Universidade de Brasília como requisito parcial para obtenção do título de **DOCTOR** em Ciência da Informação.

Tese aprovada em: 24 de fevereiro 2023.

Presidente (UnB/PPGCINF): DALTON LOPES MARTINS

Membro Interno (UnB/PPGCINF): FERNANDO CESAR LIMA LEITE

Membro Externo (USP): JOSE EDUARDO SANTAREM SEGUNDO

Membro Externo (UFES): DANIELA LUCAS DA SILVA LEMOS

Suplente (UnB/PPGCINF): JOÃO DE MELO MARICATO



Documento assinado eletronicamente por **Dalton Lopes Martins, Coordenador(a) da Pós-Graduação da Faculdade de Ciência da Informação**, em 09/03/2023, às 15:45, conforme horário oficial de Brasília, com fundamento na Instrução da Reitoria 0003/2016 da Universidade de Brasília.



Documento assinado eletronicamente por **Jose Eduardo Santarem Segundo, Usuário Externo**, em 09/03/2023, às 19:33, conforme horário oficial de Brasília, com fundamento na Instrução da Reitoria 0003/2016 da Universidade de Brasília.



Documento assinado eletronicamente por **Daniela Lucas da Silva Lemos, Usuário Externo**, em 14/03/2023, às 18:37, conforme horário oficial de Brasília, com fundamento na Instrução da Reitoria 0003/2016 da Universidade de Brasília.



Documento assinado eletronicamente por **Fernando César Lima Leite, Professor(a) de Magistério Superior da Faculdade de Ciência da Informação**, em 15/03/2023, às 10:05, conforme horário oficial de Brasília, com fundamento na Instrução da Reitoria 0003/2016 da Universidade de Brasília.



A autenticidade deste documento pode ser conferida no site http://sei.unb.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0, informando o código verificador **9257295** e o código CRC **D970FAAA**.

Referência: Processo nº 23106.009844/2023-38

SEI nº 9257295

Ficha catalográfica elaborada automaticamente,
com os dados fornecidos pelo(a) autor(a)

SS618b Siqueira, Joyce
BRASILIANA MUSEUS: SERVIÇO DE BUSCA E RECUPERAÇÃO DA
INFORMAÇÃO A PARTIR DA AGREGAÇÃO DOS METADOS DOS MUSEUS
VINCULADOS AO INSTITUTO BRASILEIRO DE MUSEUS / Joyce
Siqueira; orientador Dalton Lopes Martins. -- Brasília, 2023.
172 p.

Tese(Doutorado em Ciência da Informação) -- Universidade
de Brasília, 2023.

1. Agregação. 2. Tainacan. 3. Elastic Stack. 4. Instituto
Brasileiro de Museus - Ibram. 5. Brasiliana Museus. I.
Lopes Martins, Dalton, orient. II. Título.

AGRADECIMENTOS

Agradeço a meus pais, Eliete Deise Siqueira e Hélio Siqueira, e a minha irmã Elisa Freeman, por sempre apoiarem minhas decisões; ao meu sobrinho, Henry Stephen Freeman, apenas para registrar que ele é muito fofo, e a minha sobrinha Sophie Eleanor Freeman que ainda irá nascer; ao meu orientador, Prof. Dalton Lopes Martins, que me apresentou a área da Ciência da Informação e me oportunizou contribuir com ela; à equipe do Laboratório de Inteligência de Redes, da Universidade de Brasília, em especial ao Vinícius Nunes Medeiros, com suas sempre valiosas contribuições. Agradeço aos professores Daniela Lucas da Silva Lemos, Fernando César Lima Leite e José Eduardo Santarém, que, com importantes considerações, trouxeram maior qualidade ao trabalho. E agradeço aos meus amigos, Angélica Lorraine de Moraes Roepke, Cleyton Rafael Gomes Silva, Paulo Henrique Gomes Silva e Raquel Oliveira Araújo, que, atuantes nos momentos descontraídos, trouxeram leveza ao processo.

RESUMO

BRASILIANA MUSEUS: SERVIÇO DE BUSCA E RECUPERAÇÃO DA INFORMAÇÃO A PARTIR DA AGREGAÇÃO DOS METADOS DOS MUSEUS VINCULADOS AO INSTITUTO BRASILEIRO DE MUSEUS

O Instituto Brasileiro de Museus (Ibram) administra diretamente 27 museus federais e, destes, 20 possuem seus acervos digitais, parciais ou totais, disponíveis na Web por meio do Tainacan, totalizando 22 coleções museais. Este cenário trouxe relevantes e excitantes possibilidades, mas junto a estas, trouxe desafios na busca e recuperação da informação. Assim, este estudo tem como problema de pesquisa a seguinte questão: quais são as etapas de tratamento informacional, suas características, os requisitos funcionais que devem desempenhar e como essas devem se relacionar para a elaboração de um modelo conceitual que resulte na construção de um serviço de busca e recuperação de informações agregadas dos repositórios digitais dos acervos museais do Ibram? Dessa forma, o objetivo geral desta pesquisa é elaborar um modelo conceitual completo de um novo serviço de busca e recuperação da informação a partir da agregação automatizada de metadados museais disponíveis em formato aberto nos repositórios digitais dos acervos do Ibram. Os estudos contemplaram uma pesquisa de natureza aplicada, que utiliza como técnicas de pesquisa: a pesquisa documental; a pesquisa bibliográfica e; a Revisão Sistemática de Literatura (RSL) e como método de pesquisa, o estudo de casos múltiplos. O levantamento realizado por meio das técnicas de pesquisa resultou, principalmente, na descrição de diversos agregadores de dados culturais construídos por diferentes países, buscando identificar o contexto no qual estão inseridos, os principais resultados, os processos e as tecnologias utilizadas, em prol de subsidiar as diretrizes para construção do modelo brasileiro. A partir dos resultados obtidos e das necessidades do Ibram, pode-se concluir que as tecnologias encontradas não atendiam as especificidades brasileiras em sua totalidade. Dessa forma, deu-se início a escolha das tecnologias, buscando por ferramentas modernas, gratuitas, com comunidade ativa, com alta performance, escaláveis, que permitissem integração com outras ferramentas e que também fossem relativamente simples de implantar. Optou-se então pelo uso da pilha Elastic Stack, que, composta por quatro ferramentas: Beats, Logstash, Elasticsearch e Kibana, visa obter dados de qualquer fonte, em qualquer formato, para pesquisar, analisar e visualizar tudo em tempo real. Assim, elas se integram ao Tainacan, compondo um processo de agregação formado por cinco etapas principais: i) mapeamento dos metadados dos museus de origem para o padrão do Ibram, disponibilizando dados no formato aberto; ii) coleta automatizada dos metadados dos objetos digitais das coleções; iii) agregação e transformação destes dados; iv) envio e armazenamento dos metadados no Tainacan, para criação do Portal Brasileira Museus; v) envio e armazenamento dos metadados para o Elasticsearch e; vi) criação do painel analítico, interativo, composto por elementos de visualização. O modelo conceitual criado foi prototipado e intitulado Brasileira Museus, revelando o serviço como uma solução eficiente, quer do ponto de vista da socialização e divulgação dos objetos digitais, quer na monitorização e acompanhamento da informação dos acervos por meio de painéis analíticos, além de representar uma solução tecnológica barata e viável de ser implementada.

Palavras-chave: Agregação. Tainacan. Elastic Stack. Instituto Brasileiro de Museus - Ibram. Brasileira Museus.

ABSTRACT

BRASILIANA MUSEUS: SERVICE FOR SEARCHING AND RETRIEVING INFORMATION FROM THE AGGREGATION OF THE METATA OF THE MUSEUMS LINKED TO THE BRAZILIAN INSTITUTE OF MUSEUMS

The Brazilian Institute of Museums (Ibram) directly manages 27 federal museums and, of these, 20 have their digital collections, partial or total, available on the Web through Tainacan, totaling 22 museum collections. This scenario brought relevant and exciting possibilities, but along with these, it brought challenges in the search and retrieval of information. Thus, this study has the following question as a research problem: what are the informational processing steps, their characteristics, the functional requirements they must perform and how these must be related to the elaboration of a conceptual model that results in the construction of a service of search and retrieval of aggregated information from the digital repositories of Ibram's museum collections? Thus, the general objective of this research is to elaborate a complete conceptual model of a new information search and retrieval service based on the automated aggregation of museum metadata available in open format in the digital repositories of Ibram's collections. The studies contemplated research of an applied nature, which uses the following research techniques: documentary research; bibliographic research and the Systematic Literature Review (SLR) and as a research method, the study of multiple cases. The survey carried out through research techniques resulted, mainly, in the description of several cultural data aggregators built by different countries, seeking to identify the context in which they are inserted, the main results, the processes and the technologies used, to subsidize the guidelines for building the Brazilian aggregation service.. Thus, the choice of technologies began, looking for modern, free tools, with an active community, with high performance, scalable, that allowed integration with other tools and that were also relatively simple to implement. So, It was opted to use the Elastic Stack, which consists of four tools: Beats, Logstash, Elasticsearch and Kibana, which aims to obtain data from any source, in any format, to search, analyze and visualize everything in real time. Thus, they are integrated into Tainacan, composing an aggregation process consisting of five main steps: i) mapping of metadata from the original museums to the Ibram standard, making data available in an open format; ii) automated collection of metadata of digital objects in the collections; iii) aggregation and transformation of these data; iv) sending and storing metadata in Tainacan, for the creation of the Portal Brasiliana Museum; v) sending and storing metadata to Elasticsearch and; vi) creation of the analytical, interactive panel, composed of visualization elements. The conceptual model created was prototyped and titled Brasiliana Museus, revealing the service as an efficient solution, both from the point of view of socializing and disseminating digital objects, and in monitoring and following up information on collections through analytical panels, in addition to representing a technological solution that is cheap and viable to implement.

Keywords: Aggregation. Tainacan. Elastic Stack. Brazilian Institute of Museums - Ibram. Brasiliana Museums.

LISTA DE FIGURAS

Figura 1. Exemplo de busca textual realizada por meio da palavra-chave 'cadeira' .	36
Figura 2. Resultado da busca textual do termo 'cadeira branca'	36
Figura 3. Busca textual por palavras-chaves utilizadas separadamente	37
Figura 4. Opções de filtros da ferramenta busca avançada do Tainacan	37
Figura 5. Resultado parcial da busca facetada realizada por meio do metadado 'estado de conservação'	38
Figura 6. Resultado parcial da busca realizada com os recursos busca textual e busca facetada	39
Figura 7. Workflow de agregação de dados da <i>American Art Collaborative</i>	63
Figura 8. Workflow de agregação de dados da Trove	64
Figura 9. Arquitetura da plataforma Supplejack	66
Figura 10. Workflow de agregação de dados da Europeana	67
Figura 11. Arquitetura do D-NET Software Toolkit	69
Figura 12. Workflow de agregação de dados do <i>Parthenos Aggregator</i>	70
Figura 13. Workflow de agregação de dados da Mexicana	71
Figura 14. Workflow de agregação de dados do Projeto UNLV's Linked Data	73
Figura 15. Arquitetura do sistema CARARE	84
Figura 16. Arquitetura do repositório MoRe	85
Figura 17. Workflow de ingestão de conteúdo do projeto EUScreen (2012)	86
Figura 18. Workflow de ingestão de conteúdo do projeto EUScreen (2013)	86
Figura 19. Fluxo de trabalho de agregação do projeto HOPE	87
Figura 20. Arquitetura da Federação de Bibliotecas Digitais Culturais da Natural Europe	89
Figura 21. Workflow de ingestão de dados do EFG	91
Figura 22. Arquitetura do <i>Open Cultural Digital Content Infrastructure</i>	92
Figura 23. Arquitetura do META-SHARE	93
Figura 24. Arquitetura do D-NET adaptado para uso no Projeto EAGLE	94
Figura 25. Workflow de ingestão de dados do DPLA	96
Figura 26. Visão geral da arquitetura do Projeto ECLAP	97
Figura 27. Infraestrutura do Projeto ARIADNE	99
Figura 28. Workflow para agregação de dados do software DAL	100

Figura 29. Etapas previstas pela Europeia e DPLA para agregação de dados, com foco na qualidade de dados	110
Figura 30. Arquitetura do serviço de busca e recuperação da informação Brasileira Museus.....	126
Figura 31. Interface de mapeamento dos metadados para INCBM, no Tainacan...	128
Figura 32. Trecho do arquivo JSON do Museu da Abolição	129
Figura 33. Estrutura do arquivo Logstash	132
Figura 34. Esquema da fila de submissão de itens no Tainacan	133
Figura 35. Interface da ferramenta Elasticsearch.....	134
Figura 36. Interface do Kibana para criação de um gráfico com o número de itens de cada museu.....	135
Figura 37. Workflow do Brasileira Museus: serviço de agregação de metadados museais do Ibram.....	136
Figura 38. Painel de Visualização Analítica dos Acervos Digitais Integrados do Ibram	140
Figura 39. Exemplo com o resultado da seleção do Museu Villa-Lobos no gráfico de museus.....	141
Figura 40. Exemplo de consulta com filtro pré-configurado no Kibana	142
Figura 41. Interface para uma consulta simples, por palavra-chave e a partir da adição de novos filtros.....	142
Figura 42. Portal Brasileira Museus (Acervo em Rede) - página principal	144
Figura 43. Resultado da busca por itens com a Classificação “05 interiores”, nos museus Regional do Caeté, Arte Religiosa e Tradicional e das Missões.....	145
Figura 44. Itens museais referenciados com a classificação ‘pinturas’	146
Figura 45. Busca simples e avançada.....	147

LISTA DE GRÁFICOS

Gráfico 1. Números referentes à execução da <i>string</i> de pesquisa da RSL	61
Gráfico 2. Ocorrência das etapas para agregação de dados – RSL	106
Gráfico 3. Proporção de ferramentas atualizadas e desatualizadas	108
Gráfico 4. Proporção de licenças gratuitas e proprietárias	109

LISTA DE QUADROS

Quadro 1. Museus federais vinculados ao Ibram, com seus respectivos Estados de localização, link dos acervos digitais e quantidade de itens públicos.....	27
Quadro 2. Descrição do caso Museu Casa de Hera	42
Quadro 3. Descrição do caso Museu Casa da Princesa	43
Quadro 4. Descrição do caso Museu Casa de Benjamin Constant.....	43
Quadro 5. Descrição do caso Museu Casa Histórica de Alcântara	44
Quadro 6. Descrição do Caso Museu da Abolição.....	44
Quadro 7. Descrição do caso Museu da Inconfidência	45
Quadro 8. Descrição do caso Museu das Bandeiras	46
Quadro 9. Descrição do caso Museu das Missões	46
Quadro 10. Descrição do caso Museu de Arqueologia de Itaipu	47
Quadro 11. Descrição do caso Museu de Arte Religiosa e Tradicional.....	47
Quadro 12. Descrição do caso Museu de Arte Sacra da Boa Morte	48
Quadro 13. Descrição do caso Museu do Diamante	48
Quadro 14. Descrição do caso Museu do Ouro	49
Quadro 15. Descrição do caso Museu Histórico Nacional	50
Quadro 16. Descrição do caso Museu Regional Casa dos Ottoni.....	50
Quadro 17. Descrição do caso Museu Regional do Caeté.....	51
Quadro 18. Descrição do caso Museu Regional São João Del Rey	52
Quadro 19. Descrição do caso Museu Solar Monjardim	52
Quadro 20. Descrição do caso Museu Victor Meirelles	53
Quadro 21. Descrição do caso Museu Villa-Lobos.....	53
Quadro 22. Relação de instituições, com seu respectivo país de origem e nome do projeto de agregação de dados.....	62
Quadro 23. Portais de busca integrada de agregadores internacionais	75
Quadro 24. Sumarização das etapas dos workflows de agregação de dados	76
Quadro 25. Sumarização das etapas dos workflows de agregação de dados utilizando a nomenclatura original das publicações	77
Quadro 26. Portais de busca integrada de agregadores nacionais.....	81
Quadro 27. Relação de projetos selecionados pela RSL, em ordem crescente por ano de publicação	83
Quadro 28. Relação de tecnologias utilizadas nos agregadores da RSL.....	101

Quadro 29. Relação de projetos, período de execução e portal agregador	103
Quadro 30. Sumarização das etapas dos workflows de agregação de dados utilizando a nomenclatura original das publicações - RSL	104
Quadro 31. Sumarização das etapas dos workflows de agregação de dados - RSL	105
Quadro 32. Resposta à pergunta 03: Quais tecnologias são utilizadas?	107
Quadro 33. Categorias de requisição de metadados	114
Quadro 34. Agregadores de dados internacionais	123
Quadro 35. Mapeamento de dados parcial do Museu da Inconfidência.....	128
Quadro 36. URL de um critério combinado de busca.....	147

LISTA DE SIGLAS E ABREVIações

AAC - *American Art Collaborative*

ACDM - *ARIADNE Catalogue Data Model*

API - *Application Programming Interface*

API RESTful - *Application Programming Interface - Representational State Transfer*

APP - *Application*

ATT - *Getty Art & Architecture Thesaurus*

CAInf - *Coordenação de Arquitetura da Informação Museal*

CARARE - *Connecting Archaeology and Architecture in Europeana*

CCO - *Cataloging Cultural Objects*

CDWA - *Categories for the Description of Works of Art*

CE - *Critério de Exclusão*

CHO - *Cultural Heritage Object*

CI - *Critério de Inclusão*

CRM - *Conceptual Reference Model*

CSV - *Comma Separated Values*

DAL - *Data Aggregation Lab*

DCRM - *Descriptive Cataloging of Rare Materials*

DHI - *Digital Humanities Infrastructures*

DPLA - *Digital Public Library of America*

DPLA MAP - *Digital Public Library of America Metadata Application Profile*

DQC - *Data Quality Committee*

EAGLE - *Europeana Network of Ancient Greek and Latin Epigraphy*

ECLAP - *European Collected Library of Artistic Performance Content Aggregator*

EDM - *Europeana Data Model*

EFG - *European Film Gateway*

ESE - *Europeana Semantic Elements*

FTP - *File Transfer Protocol*

FUNARJ - *Fundação Anita Mantuano de Artes do Estado do Rio de Janeiro*

GPLv3 – *General Public License version 3*

GUI - *Graphical User Interface*

HOPE - *Heritage of the People's Europe*

HTML - *HyperText Markup Language*

HTTP - *HyperText Transfer Protocol*
HTTPS - *HyperText Transfer Protocol Secure*
Ibict - Instituto Brasileiro de Informação em Ciência e Tecnologia
Ibram - Instituto Brasileiro de Museus
INBCM - Inventário Nacional de Bens Culturais Musealizados
Iphan - Instituto do Patrimônio Histórico e Artístico Nacional
ISO - *International Organization for Standardization*
JASIST - *Journal of the Association for Information Science and Technology*
JSON - *JavaScript Object Notation*
JRR - *Joint Resource Registry*
LCNAF - *Library of Congress Names*
LIDO - *Lightweight Information Describing Objects*
LOD - *Linked Open Data*
MAP - *Metadata Application Profile*
MARC - *Machine Readable Cataloging*
MMAT - *Multimedia Authoring Tool*
MoRe - *Monument Repository*
NDLTD - *Networked Digital Library of Theses and Dissertations*
NECE - *Natural Europe Cultural Environment*
NECHI - *Natural Europe Cultural Heritage Infrastructure*
NISO - *National Information Standards Organization*
OAI-PMH - *Open Archives Initiative Protocol for Metadata Harvesting*
PARTENOS - *Pooling Activities, Resources and Tools for Heritage E-research Networking, Optimization and Synergies*
PDF - *Portable Document Format*
PNM - Política Nacional de Museus
PPGCINF - Programa de Pós-Graduação em Ciência da Informação
RDA - *Resource Description and Access*
RDF - *Resource Description Framework*
REST - *Representational State Transfer*
RFID - *Radio Frequency Identification*
RI - *Recuperação da Informação*
RSL - Revisão Sistemática da Literatura

SFTP - *Secure File Transfer Protocol*
SIP - *Submission Information Packages*
SISGAM - Sistema de Gerenciamento de Acervos Museológicos
SPARQL - *SPARQL Protocol and RDF Query Language*
SQL - *Standard Query Language*
SRI - Sistema de Recuperação da Informação
TI - Tecnologia da Informação
UF - Unidade Federativa
UFG - Universidade Federal de Goiás
ULAN - *Union List of Artist Names*
UnB - Universidade de Brasília
UNLV - *University of Nevada*
URI - *Uniform Resource Identifier*
URL - *Uniform Resource Locator*
VIAF - *Virtual International Authority File*
WIDaT - Workshop de Informação, Dados e Tecnologia
WoS - *Web of Science*
WWW - *World Wide Web*
XML - *eXtensible Markup Language*

SUMÁRIO

1 INTRODUÇÃO	15
1.1 Problema de Pesquisa	17
1.2 Objetivos	17
1.2.1 Objetivo Geral	17
1.2.2 Objetivos Específicos	17
1.3 Justificativa	18
2 FUNDAMENTAÇÃO TEÓRICA	20
3 METODOLOGIA	26
3.1 Contextualização	26
3.2 Estudo de Casos Múltiplos	40
3.3 Pesquisa Bibliográfica	55
3.4 Pesquisa Documental	55
3.5 Revisão Sistemática de Literatura	57
3.6 Versões dos softwares	61
4 RESULTADOS	62
4.1. Revisão Bibliográfica	62
4.2 Revisão Sistemática de Literatura	82
4.3 Elementos para a Construção de uma Política de Qualidade de Dados para Agregar Acervos Culturais Digitais: Casos DPLA e Europeana	109
4.4 Brasileira Museus: Serviço de Busca e Recuperação da Informação Museal	122
4.5 Workflow do serviço de busca e recuperação da informação	135
4.6 Painel de Visualização Analítica dos Acervos Digitais Integrados	138
4.7 Portal Brasileira Museus: repositório final	143
5 DISCUSSÕES	149
6 PREMIAÇÕES E PUBLICAÇÕES	153
6.1 Premiações	153
6.2 Publicações	153
7. CONSIDERAÇÕES FINAIS	156
Referências	158

1 INTRODUÇÃO

As novas tecnologias estão propiciando a reinvenção do espaço museal, dando às instituições culturais a oportunidade de extrapolar a missão de custodiar, preservar e manter os acervos, permitindo-as aperfeiçoar a comunicação com o público em geral, disseminando seu conteúdo para que se constitua como recurso de memória, de história, de cultura, de educação e de cidadania (MARCONDES, 2016; MUCHACHO, 2005).

A democratização do acesso aos museus por meio dos acervos digitais potencializa a propagação do conhecimento, além de ampliar exponencialmente a visibilidade das instituições, anteriormente restritas à visita *in loco*, trazendo um novo fôlego para museus, arquivos e bibliotecas no século 21 (GOV.BR, 2023L; SANTAREM SEGUNDO; SILVA; MARTINS, 2019).

Assim, instituições nacionais e internacionais responsáveis pela guarda do patrimônio cultural têm investido na digitalização e difusão de seus acervos, resultando em inúmeros museus digitais distribuídos na Web, como exemplos: a Pinacoteca, no Brasil; o Museu do Louvre, na França; a Casa Anne Frank, na Holanda; o Museu Nacional de História Natural, nos Estados Unidos da América, dentre outros.

Este cenário, embora muito rico, trouxe desafios à busca e recuperação da informação, afinal, há uma vasta e infindável quantidade de objetos digitais disponíveis. Visando amenizar esses desafios, instituições nacionais e, principalmente, internacionais apresentam diferentes soluções que ofertam aos usuários uma interface única para busca e recuperação, por meio de uma plataforma centralizadora que direciona os acessos aos museus de origem.

Como exemplo, cita-se: no Brasil, a Brasileira Iconográfica (BRASILIANA ICONOGRÁFICA, 2017), a Brasileira Fotográfica (BRASILIANA FOTOGRÁFICA, 2023a) e a Rede Web de Museus (GOVERNO DO RIO DE JANEIRO, 2023b); nos Estados Unidos da América, a American Art Collaborative (AAC), a Digital *Public Library of America* (DPLA), a *World Digital Library* e a *Google Arts & Culture*; na Nova Zelândia, a DigitalNZ; na Europa, a Europeana; no México, a *Mexicana*; na Austrália, a TROVE (SIQUEIRA; MARTINS, 2019b, 2020, 2021c; SANTAREM SEGUNDO; SILVA; MARTINS, 2019). São diferentes realidades, mas que têm alterado a maneira de se coletar e se disseminar conteúdo digital.

Neste intermeio, o Instituto Brasileiro de Museus (Ibram), visando incentivar a criação de ferramentas que permitam o empoderamento social de todos aqueles que têm desejo de memória, e em vistas a fortalecer a sua capacidade técnica, tem firmado parcerias institucionais que, dentre outras coisas, estimulem a cultura digital nos museus brasileiros (MINISTÉRIO DO TURISMO, 2019).

Dessa forma, uma etapa importante para o fomento da cultura digital foi realizada por meio da parceria firmada em 2016, entre o Ibram e a Universidade Federal de Goiás (UFG), cujo destaque foi o desenvolvimento do software livre e gratuito Tainacan, um repositório digital responsável pela gestão de acervos digitais de forma compatível à realidade das instituições culturais e científicas nacionais.

Como resultado da parceria, que prossegue junto à Universidade de Brasília (UnB), atualmente há 20 museus na Web, totalizando mais de 17 mil objetos digitais disponíveis em formato aberto (MARTINS; MARTINS, 2021), resultando em uma nova frente de pesquisa, o desenvolvimento de um modelo conceitual do serviço de busca e recuperação da informação, visando coletar, agregar, transformar e armazenar os dados, de forma a atender às expectativas e realidades dos museus brasileiros sob gestão do Ibram.

O contexto torna nítida a necessidade de uma solução interdisciplinar que envolva diretamente as áreas da Ciência da Computação e da Ciência da Informação. Saracevic (1995) já discutia a relevância da interdisciplinaridade entre as duas áreas, na aplicação de computadores e computação em produtos e serviços, por exemplo, associados: à representação da informação, sua organização intelectual e ligações; à busca, à recuperação e à filtragem de informações; ao uso, à qualidade, ao valor e ao impacto de informação; à avaliação de sistemas de informação na perspectiva do usuário e uso e afins.

Claramente, para agregar dados é necessário todo um aparato tecnológico, mas também é essencial que haja acervos pautados em sólidos processos de organização. Para Harpring (2013), as informações sobre o patrimônio artístico e cultural oferecem desafios únicos na exibição e recuperação. Os fatos sobre objetos culturais e seus criadores nem sempre são conhecidos ou diretos e é enganoso e contrário aos princípios da erudição deixar de expressar essa incerteza. Ao mesmo tempo, a recuperação eficiente requer indexação de acordo com regras consistentes, bem definidas e terminologia controlada.

Às práticas de organização dos museus do Ibram aliar-se-ão às novas tecnologias para que a busca integrada apresente excelentes resultados, elaborando-se para isso o seguinte problema de pesquisa.

1.1 Problema de Pesquisa

Quais são as etapas de tratamento informacional, suas características, os requisitos funcionais que devem desempenhar e como essas devem se relacionar para a elaboração de um modelo conceitual que resulte na construção de um serviço de busca e recuperação de informações agregadas dos repositórios digitais dos acervos museais do Ibram?

1.2 Objetivos

1.2.1 Objetivo Geral

Elaborar o modelo conceitual de um novo serviço de busca e recuperação da informação a partir da agregação automatizada de metadados museais disponíveis em formato aberto nos repositórios digitais dos acervos do Ibram.

1.2.2 Objetivos Específicos

- Identificar e caracterizar as etapas utilizadas por serviços de agregação de objetos digitais culturais disponíveis.
- Definir quais são as etapas e selecionar as tecnologias necessárias à implantação de um serviço de busca e recuperação de informação, integrado, robusto e escalável, que se adeque às necessidades brasileiras, com ênfase nos acervos museais do Ibram.
- Determinar e avaliar as condições necessárias e suficientes para agregação dos acervos museais do Ibram.
- Implementar o protótipo do painel de visualização analítica, que, composto por gráficos, métricas, tabelas etc., oferece uma visão macro de todos os seus museus.
- Implementar o protótipo do Portal Brasileira Museus, um portal agregador na Web que permite a busca e recuperação de objetos digitais.

1.3 Justificativa

Agregar dados é tema de estudo em diversas áreas de atuação, como a Bioinformática, a Educação, a Geografia, a Medicina, a Música, a Cultura, entre outras. Por exemplo, à Bioinformática integram-se dados de expressão gênica da *Drosophila melanogaster*, conhecida como mosca da fruta, para selecionar genes e validar resultados experimentais (MILES et al., 2010); à Geografia integram-se dados de localizações geográficas para colaborar com a avaliação de pedidos de licenças para uso da água (ZIÉBELIM et al., 2017); à Medicina integram-se dados de formulários e sistemas para estudos epidemiológicos para determinar as causas de doenças, como obesidade, depressão, demência (KIRSTEN et al., 2017); à Música integram-se semanticamente informações de diferentes músicas para oferecer serviços de qualidade aos usuários (GÄNGLER, 2011).

Freire et al. (2017) realizaram um levantamento sobre as tecnologias Web para agregação de dados, trazendo importantes constatações. Primeiro, no domínio do patrimônio cultural a abordagem tecnológica majoritária é o protocolo *Open Archives Initiative Protocol for Metadata Harvesting - OAI-PMH* que mantém um bom grau de interoperabilidade, mas que com o avanço tecnológico deixou ter a motivação para sua adoção tão clara quanto costumava ser; segundo, destacam especificidades que dificultam o processo de agregação, tais como: a heterogeneidade de subdomínios, compostos por bibliotecas, arquivos e museus, que possuem suas próprias práticas e padrões, e os orçamentos limitados, que tornam mais lenta a velocidade e a extensão da inovação e adoção de novas tecnologias; terceiro, os metadados são transferidos para o agregador, mas continuam a evoluir no provedor de dados, portanto, o agregador precisa atualizar periodicamente sua cópia dos dados.

Outro fator crucial, também destacado por Freire et al. (2017), é a prática de acordar o uso de um único modelo de dados de referência, que geralmente tem a mesma semântica dos dados originais e que atendem às necessidades de informação dos serviços prestados pelo agregador.

Assim, ao longo dos anos, várias pesquisas foram desenvolvidas em prol de solucionar, satisfatoriamente, o problema da agregação de dados, visto não existir uma solução simples, mas, pelo contrário, muitos desafios de engenharia de software e modelagem de dados (AVAZPOUR; GRUNDY; ZHU, 2019). Além disso, há uma

infinidade de opções de técnicas e de tecnologias que podem ser selecionadas com base em diferentes limitações, por exemplo, de infraestrutura e/ou orçamentárias, diferentes objetivos ou apenas com diferentes contextos, dificultando soluções mais genéricas.

O cenário internacional está à frente com muitas soluções avançadas e de sucesso, de tal forma que estudos apresentam o retorno financeiro que os agregadores internacionais trouxeram. Duas importantes instituições, a British Library e a Europeia, realizaram estudos, ambos publicados em 2013, os quais indicaram que os serviços on-line da British Library estão avaliados em 19.5 milhões de libras por ano, gerando benefícios a estudantes, pesquisadores e cidadãos do Reino Unido. A Europeia concluiu que o impacto econômico para o continente europeu seria de aproximadamente 78.8 milhões de euros em um cenário de base, 59.6 milhões de euros em um cenário pessimista e 97.7 milhões em um cenário otimista (MARTINS; SILVA; CARMO, 2018).

As iniciativas internacionais, embora mais consolidadas e mais bem estruturadas, não disponibilizam documentação suficiente e são claramente muito específicas à realidade informacional de cada contexto no qual estão inseridas, sendo praticamente impossível reutilizá-las no Brasil (SIQUEIRA; MARTINS, 2021c).

Concomitante a isso, no Brasil as iniciativas são limitadas, isoladas e necessitam de apoio técnico e de modelos de governança escaláveis para o território brasileiro. Além disso, identifica-se a falta de incentivo político e econômico para ofertar soluções nessa direção (MARTINS et al. (2022); SANTARÉM SEGUNDO; SILVA; MARTINS, 2019).

A análise de todos os casos apresenta uma infinidade de técnicas, de tecnologias, de motivações, de finalidades, de níveis de investimento financeiro e de aportes teóricos e, por isso, cada uma delas contribui a seu modo, na construção de um novo agregador de dados, independente da área de atuação. Nesta pesquisa o enfoque é a agregação de metadados culturais voltado às especificidades do Ibram.

2 FUNDAMENTAÇÃO TEÓRICA

Os museus têm cada vez mais usufruído dos avanços tecnológicos a partir de uma infinidade de aplicações que, a depender da maturidade tecnológica da instituição, vão desde uma simples *home page*, com informações relevantes à visita presencial, até sofisticadas exposições virtuais com uso de realidade aumentada, como exemplo (LAPA, 2011).

No entanto, quando a inovação tem relação com os acervos, além dos imprescindíveis recursos tecnológicos, estes devem estar pautados em sólidos processos de organização, ou seja, organizados de forma a criar capacidades, impondo intencionalmente ordem e estrutura (GLUSHKO, 2013).

Para Svenonius (2000) há pontos de vista adotados por três movimentos filosóficos que são relevantes para a base intelectual da organização da informação: i) filosofia de sistemas, cuja teoria geral dos sistemas é a base, e deriva da prática da análise de sistemas, ou seja, a análise de um objeto de forma a vê-lo como um sistema cujas várias partes são integradas com a finalidade de alcançar determinados objetivos; ii) filosofia da ciência, cuja metodologia científica tem sido um foco central, com a filosofia dominante da ciência como positivismo lógico, expresso pelo princípio da verificabilidade e; iii) filosofia da linguagem, que dominou duas filosofias do século XX, o mencionado positivismo lógico e o princípio linguístico, para o qual o significado de uma palavra é seu uso e esse uso é governado por regras muito parecidas com as regras que governam os movimentos nos jogos de linguagem.

Glushko (2013) categoriza quatro tipos de coisas passíveis de organização: as 'coisas físicas', como exemplo, as pinturas, as esculturas e outros artefatos museais; as 'informações sobre as coisas físicas', como os catálogos de fichas de biblioteca; as 'coisas digitais', como exemplo, os periódicos, os livros digitais, os programas de computador e os bancos de dados; e as 'informações sobre as coisas digitais', como os catálogos de bibliotecas digitais, os portais da web e os sites de agregação que organizam links para outros recursos digitais.

Complementando, Buckland (1991) considera que a materialidade das representações informacionais representa um meio de referenciar concretamente a informação, já que ao considerá-lo como coisa, torna possíveis seus processos descritivos, comunicativos e representativos em um sistema de informação.

Em se tratando de acervos digitais, há duas categorias: os digitalizados, que possuem uma base física que passa pelo processo de digitalização e os natos digitais, que nascem no formato digital. Em ambos os casos, para que seja possível recuperar a informação, é necessário organizá-los juntamente a informações que permitam sua identificação (IBRAM, 2020).

Para Glushko (2013), quando as coisas e as informações sobre elas estão em formato físico, é fácil ver que a primeira é um recurso primário e a segunda um recurso substituto ou associado que o descreve ou se relaciona com ele. Quando se trata de organizar informações sobre coisas digitais, o contraste é muito menos claro. Por exemplo, quando se pesquisa um livro usando um mecanismo de pesquisa, primeiro obtém-se a descrição do catálogo do livro e, muitas vezes, o próprio livro está a apenas a um clique de distância.

Além disso, Glushko (2013) conceitua 'Sistema Organizador' como sendo uma coleção de recursos organizados e as interações que eles suportam. Em se tratando de sistema organizador na web, mesmo que ela como um todo não exiba o arranjo intencional centralizado de recursos, podemos vê-la como consistindo em milhões de Sistemas Organizadores, cada um incorporando um arranjo intencional separado. Além disso, na maioria das vezes interagimos com a web indiretamente usando um mecanismo de busca, que atende à definição de Sistema Organizador por conta de seus algoritmos de indexação e recuperação.

A Recuperação da Informação (RI) se concentra em ajudar usuários a encontrar informações de seu interesse, com alta eficácia. Quando se une à indexação, busca-se também a maior eficiência, afinal, mesmo com um modelo de RI e interface eficazes, os usuários ainda ficarão frustrados se tiverem de esperar dois minutos pelos resultados de uma consulta (Baeza-Yates; Ribeiro-Neto, 2013).

A indexação automática, via algoritmos, ocorre por meio da criação de índices nos bancos de dados construídos a partir do texto para acelerar as buscas (BAEZA-YATES; RIBEIRO-NETO, 2013). Em outro contexto, há o processo de avaliação de informações e designação de termos de indexação usando um vocabulário controlado, que é realizado por trabalho humano, por catalogadores experientes que consideram as implicações de recuperação ao atribuir termos de indexação (HARPRING, 2013).

Ainda sobre Sistema Organizador, Glushko (2013) relata que uma importante contribuição é o conceito de 'recurso' que vai além das definições de 'coisas',

‘documentos’ e ‘informação’, ao mesmo tempo em que reconhece que o que está sendo organizado é apenas uma das dimensões que precisam ser consideradas.

Para Glushko (2013) o ‘recurso’ tem um senso comum de qualquer coisa de valor que possa apoiar a atividade orientada a objetivos. Tem significado especializado na arquitetura da Internet, tratando como um recurso primário qualquer coisa que possa ser identificada, sendo uma importante generalização do conceito porque permite que serviços baseados na web, feeds de dados, objetos com tags *Radio Frequency Identification* (RFID), sensores ou outros dispositivos inteligentes, ou agentes computacionais façam parte de Sistemas Organizadores.

Para Svenonius (2000), um sistema organizador, para ser eficaz, deve se basear em uma base intelectual que consiste em várias partes. São elas:

- uma ideologia, formulada em termos de propósitos e princípios que orientam seu desenho;
- a formalização dos processos envolvidos, tais como as conceituações linguísticas e os modelos entidade-atributo-relacionamento;
- o conhecimento adquirido por meio da pesquisa, particularmente aquele expresso na forma de generalizações de alto nível sobre o design e uso de sistemas organizadores e;
- na medida em que uma disciplina é definida por seus focos de pesquisa, os principais problemas que precisam ser resolvidos de forma inteligente a partir do avanço da Ciência da Informação.

Para o design do sistema de organização, Glushko (2013) apresenta seis grupos de perguntas: O que está sendo organizado? Por que está sendo organizado? Quanto está sendo organizado? Quando está sendo organizado? Como ou por quem, ou por quais processos computacionais estão sendo organizados? Onde está sendo organizado?

O quão bem essas decisões se aglutinam em um sistema organizador depende dos requisitos e objetivos dos usuários humanos e computacionais e da compreensão das restrições e compensações que qualquer conjunto de requisitos e objetivos impõe. Podemos aumentar a visão categórica dos Sistemas Organizadores pensando neles como existindo em um espaço de design multifacetado ou multidimensional no qual podemos considerar muitos tipos de coleções ao mesmo tempo (GLUSHKO, 2013).

Existem quatro atividades que ocorrem naturalmente em todo sistema organizador e quão explícitos eles são depende do escopo, da amplitude ou variedade dos recursos, da escala, do número de recursos que o Sistema Organizador abrange (GLUSHKO, 2013). São eles:

- Selecionar: determinar o escopo do Sistema Organizador especificando quais recursos devem ser incluídos.
- Organizar: especificar os princípios ou regras que serão seguidas para organizar os recursos.
- Projetar interações baseadas em recursos: projetar e implementar as ações, funções ou serviços que fazem uso dos recursos.
- Manter: gerenciar e adaptar os recursos e a organização impostos para apoiar as interações.

Essas atividades estão profundamente enraizadas nos currículos acadêmicos e nas práticas profissionais, com termos específicos de domínio para seus métodos e resultados. Museus costumam explicitar seus princípios de seleção nas políticas de desenvolvimento de coleções. Adicionar uma coleção a um museu é chamado de acessão. Documentar o conteúdo de museus para organizá-los é chamado de catalogação. As interações primárias para os usuários do museu estão em visualizar ou visitar a coleção. As atividades de manutenção são geralmente descritas como preservação ou curadoria (GLUSHKO, 2013).

A National Information Standards Organization – NISO (2004, tradução nossa) conceitua interoperabilidade como a possibilidade de diversos sistemas trocarem dados entre si com perda mínima de conteúdo e funcionalidade, de modo a garantir que pessoas, organizações e sistemas de informação (como os repositórios digitais) interajam satisfatoriamente.

Embora o conceito proposto pela NISO (2004) seja amplamente aceito, Santarém Segundo, Silva e Martins (2019) destacam que há um conjunto de definições que, de maneira geral, se relacionam ao domínio em que se apresentam. No mesmo estudo, os autores selecionam a definição de Innocenti, Vullo e Ross (2010), como sendo mais específica, no âmbito dos acervos digitais.

Innocenti, Vullo e Ross (2010, tradução nossa) entendem a interoperabilidade como a capacidade dos sistemas e das organizações de trabalharem em conjunto. É reconhecida como um passo fundamental na mudança de acervos digitais isolados

(os referidos autores tratam especificamente sobre as bibliotecas) para um espaço comum que permitirá aos usuários navegar através de diferentes acervos dentro de um único ambiente integrado.

Para Santarém Segundo, Silva e Martins (2019) a interoperabilidade é um aspecto tecnológico fundamental para a evolução e desenvolvimento dos mais variados tipos de acervos digitais, permitindo integração de conteúdo.

Em busca da interoperabilidade, instituições investem na adoção de padrões e práticas recomendadas para a produção de metadados, sendo, portanto, uma tentativa de se obter um vocabulário comum e consistente para descrever uma variedade de estruturas de dados capazes de satisfazer a várias comunidades no âmbito dos Sistemas de Recuperação da Informação (SRIs).

Os SRIs são a interface entre uma coleção de documentos e uma população de usuários, e se constituem, principalmente, por dois subsistemas interdependentes, quais sejam, o de entrada e o de saída (LANCASTER, 1986; 2004; ABADAL; CODINA, 2005; BAEZA-YATES; RIBEIRO-NETO, 2011). O subsistema de entrada refere-se ao processo de seleção de documentos e à descrição (ou catalogação) de suas características a partir de elementos e registros de metadados, por meio dos quais as bases de dados documentais são constituídas. Já o subsistema de saída refere-se à interface do usuário com o sistema, partindo da necessidade de informação do usuário, passando pela estratégia de busca e navegação, de modo a propiciar a recuperação de itens relevantes e, finalmente, o acesso ao documento (objeto digital) e/ou seus registros metadados.

Nesse sentido, os metadados são estruturados seguindo padrões para tratamento nos dados (GILLILAND, 2016), incluindo: i) estrutura de dados: conjunto de elementos de metadados ou esquemas de categorias que formam um registro de informação; ii) valores dos dados: linguagens documentárias, vocabulários controlados, arquivos de autoridade e ontologias de domínio usados para preencher os dados nos elementos de metadados; iii) conteúdo dos dados: regras e códigos de catalogação que orientam em formatações, sintaxes e relacionamentos para os valores de dados usados para preencher os elementos de metadados e; iv) comunicação de dados: padrões de metadados expressados em uma linguagem de representação legível para a máquina.

Desse modo, instrumentos de representação de relacionamentos semânticos e conceituais como vocabulários controlados (linguagens documentárias) podem ser usados como sistemas de organização do conhecimento (HJORLAND, 2007; LEMOS; SOUZA, 2020) para descrever melhor os dados, visando normalização, qualidade e intercâmbio de suas descrições e, com isso, melhorar o alcance da interoperabilidade entre diferentes esquemas de metadados e aplicações.

Um vocabulário controlado é um arranjo organizado de palavras e frases usadas para categorização, indexação e recuperação de informações. É relevante para indexação porque sem ele os catalogadores não usarão consistentemente o mesmo termo para se referir à mesma pessoa, lugar ou coisa. No processo de recuperação, vários usuários finais podem usar diferentes sinônimos ou termos mais genéricos para se referir a um determinado conceito. Por geralmente não serem especialistas precisam ser orientados porque podem não conhecer o termo correto (HARPRING, 2013).

Vocabulários controlados podem ser exemplificados por esquemas de classificação, listas de cabeçalhos de assuntos, tesouros, taxonomias e ontologias. A formulação de pontos de acesso controlados é geralmente governada por regras oriundas de uma política interna institucional ligada à qualidade de dados, o que permite a criação padronizada de metadados ao nomear, de forma consistente, os pontos de acesso aos documentos numa interface de um sistema de recuperação da informação, tal como um repositório (MARTINS et al., 2021).

Assim, as comunidades de arte e patrimônio cultural têm feito uso cada vez mais de vocabulários e outros padrões, pois buscam fornecer acesso a informações que antes eram mantidas em arquivos de papel ou isoladas em sistemas locais. Inspirados pelo poder dos bancos de dados online e da World Wide Web (WWW), os profissionais das várias comunidades de arte e patrimônio cultural agora veem o valor da troca eficiente de informações entre si (HARPRING, 2013).

3 METODOLOGIA

O estudo visa a construção de um modelo conceitual para busca e recuperação da informação a partir da agregação automatizada dos metadados dos museus digitais do Ibram. Dessa forma, se trata de uma pesquisa de natureza aplicada, de cunho descritivo, com abordagem qualitativa, que utiliza as técnicas de pesquisa: pesquisa documental, pesquisa bibliográfica e Revisão Sistemática de Literatura (RSL). Como método de pesquisa utiliza o estudo de casos múltiplos, tornando essencial conhecer mais a fundo o Ibram e o contexto da pesquisa.

3.1 Contextualização

O Ibram foi criado pelo presidente da República, Luiz Inácio Lula da Silva, em janeiro de 2009, com a assinatura da Lei n. 11.906, visando suceder ao Instituto do Patrimônio Histórico e Artístico Nacional (Iphan), nos direitos, deveres e obrigações relacionados aos museus federais (GOV.BR, 2023I, MINISTÉRIO DO TURISMO, 2019). No período, a autarquia foi vinculada ao Ministério da Cultura, porém, atualmente, 2022, está vinculada ao Ministério do Turismo.

O Instituto continua responsável pela Política Nacional de Museus (PNM) e pela melhoria dos serviços do setor, tais como: aumento de visitação e arrecadação; fomento de políticas de aquisição e preservação de acervos, e criação de ações integradas entre os museus brasileiros, além de administrar, diretamente, 30 museus federais (GOV.BR, 2023I; MINISTÉRIO DO TURISMO, 2019).

Por fim, o Ibram é responsável por: orientar, fomentar e garantir o cumprimento da função social dos Museus e o direito à memória, por meio da preservação dos acervos, ações de fortalecimento das identidades, saberes locais, iniciativas de memória e museologia social; contribuir para a formação de cidadãos; combater estatísticas de criminalidade e evasão escolar; e possibilitar o acesso à cultura a uma parte da população que não teria, ao menos em regra, condições de usufruí-la (MINISTÉRIO DO TURISMO, 2019, p. 7).

Os museus federais estão distribuídos em nove estados brasileiros, ocupando uma área de mais de 931mil m², compostos por 99 edificações, incluindo jardins históricos, espaços verdes, áreas de proteção ambiental e matas nativas (GOV.BR, 2023k).

Por serem lugares de transformação social e de desenvolvimento, fazem com que também esteja entre as prioridades do Ibram, incentivar a criação de ferramentas que permitam o empoderamento social de todos aqueles que têm desejo de memória. Por isso, com vistas a fortalecer a sua capacidade técnica, tem firmado parcerias institucionais, que dentre outras coisas, estimulem a cultura digital nos museus brasileiros (MINISTÉRIO DO TURISMO, 2019).

Dessa forma, uma etapa importante para o fomento da cultura digital foi realizada por meio da parceria firmada em 2016, entre o Ibram e a UFG, cujo destaque foi o desenvolvimento do software Tainacan, para gestão dos acervos digitais. Em 2019, o projeto foi transferido para a UnB, que prossegue a implementação do Tainacan e demais produtos de pesquisa por meio do Laboratório de Inteligência de Redes, vinculado ao curso de Ciência da Informação da Instituição.

O Ibram possui 30 museus vinculados, porém, em termos de pesquisa consideraremos apenas 27, ou seja, não consideraremos os subsidiários do Museu da República (Palácio Rio Negro), do Museu Imperial (Casa Geyer e Casa de Cláudio de Souza) e do Museu Raimundo Ottoni de Castro Maya (Museu Chácara do Céu e Museu do Açude). Destes, 20 (74%) possuem repositório digital com ao menos parte dos seus acervos no Tainacan, totalizando mais de 17 mil itens, conforme Quadro 1, que apresenta a relação de Museus com seus respectivos Estados de localização, link do acervo digital no Tainacan e quantidade de itens públicos disponíveis.

Quadro 1. Museus federais vinculados ao Ibram, com seus respectivos Estados de localização, link dos acervos digitais e quantidade de itens públicos

N.	Museu	Estado	Acervo Digital	N. de Itens
1	Museu Casa da Hera	RJ	https://museucasadahera.acervos.museus.gov.br	1.287
2	Museu Casa da Princesa	GO	http://museusibramgoias.acervos.museus.gov.br/museu-casa-da-princesa	799
3	Museu Casa de Benjamin Constant	RJ	http://museucasabenjaminconstant.acervos.museus.gov.br	983
4	Museu Casa Histórica de Alcântara	MA	http://museucasahistoricadealcantara.acervos.museus.gov.br	631
5	Museu Castro Maya	RJ	Não possui	0
6	Museu da Abolição	PE	https://museudaabolicao.acervos.museus.gov.br	301

7	Museu da Inconfidência	MG	http://museudainconfidencia.acervos.museus.gov.br	4.625
8	Museu da República	RJ	Não possui	0
9	Museu das Bandeiras	GO	http://museusibramgoias.acervos.museus.gov.br	441
10	Museu das Missões	RS	http://museudasmissoes.acervos.museus.gov.br	90
11	Museu de Arqueologia de Itaipu	RJ	http://museudearqueologiadeitaipu.museus.gov.br	1.040
12	Museu de Arte Religiosa e Tradicional	RJ	https://museudeartereligiosaetradicional.acervos.museus.gov.br	132
13	Museu de Arte Sacra da Boa Morte	GO	http://museusibramgoias.acervos.museus.gov.br/museu-casa-da-boa-morte	790
14	Museu de Arte Sacra de Paraty	RJ	Não possui	0
15	Museu do Diamante	MG	https://museudodiamante.acervos.museus.gov.br	895
16	Museu do Ouro – Casa Borba Gato	MG	https://museudoouro.acervos.museus.gov.br	136
17	Museu Forte Defensor Perpétuo	RJ	Não possui	0
18	Museu Histórico Nacional	RJ	http://mhn.acervos.museus.gov.br	2.062
19	Museu Imperial	RJ	Não possui	0
20	Museu Lasar Segall	SP	Não possui	0
21	Museu Nacional de Belas Artes	RJ	Não possui	0
22	Museu Regional Casa dos Ottoni	MG	https://museuregionalcasadosottoni.acervos.museus.gov.br	463
23	Museu Regional de Caeté	MG	https://museuregionaldecaete.acervos.museus.gov.br	243
24	Museu Regional de São João Del-Rei	MG	http://museuregionaldesaojoao delrei.acervos.museus.gov.br	328
25	Museu Solar Monjardim	ES	https://museusolarmonjardim.acervos.museus.gov.br	77
26	Museu Victor Meirelles	SC	http://museuvictormeirelles.acervos.museus.gov.br	237
27	Museu Villa-Lobos	RJ	http://museuvillalobos.acervos.museus.gov.br	1.812
TOTAL				17.370

Fonte: elaborado pelos autores (2023)

O Ibram disponibiliza sucinta descrição de cada museu (GOV.BR, 2023k), dos quais destacamos apenas os que possuem acervo digital no Tainacan.

O Museu Casa da Hera está localizado na Casa da Hera, antiga residência da família de Joaquim José Teixeira Leite, grande advogado e comissário de café, no século XIX. Seu acervo é composto por mobiliário, quadros e objetos de uso doméstico, além de uma biblioteca, uma coleção de trajes franceses e um piano Henri Herz. Em 1952, a construção foi tombada como patrimônio nacional (GOV.BR, 2023).

O Museu Casa da Princesa funciona na Casa Setecentista, antiga moradia senhorial do século XVIII. O acervo é formado por documentos históricos, fotografias e objetos que mostram formas do viver goiano dos séculos XVIII ao XX (GOV.BR, 2022).

O Museu Casa de Benjamin Constant foi a residência de Benjamin Constant, figura histórica da República brasileira. Possui uma exposição permanente de objetos, obras de arte e mobiliário que recriam o modo de vida do final do Século XIX e início do Século XX. Em 1891 foi adquirido pelo patrimônio público (GOV.BR, 2023a).

O Museu Casa Histórica de Alcântara teve sua abertura ao público em 2004 e tem por missão remontar aos tempos do Brasil Imperial, expondo a opulência dos hábitos e costumes do século XIX, da aristocracia rural da cidade de Alcântara. Seu acervo é composto por mobiliário, indumentárias, acessórios, louças, iconografias, cerâmicas e azulejos (GOV.BR, 2023b).

O Museu da Abolição, Centro de Referência da Cultura Afro-Brasileira, foi criado em 1957 e inaugurado em 1983. Está localizado na antiga sede do Engenho Madalena e residência do conselheiro abolicionista João Alfredo. Fechado em 1990 e reaberto em 1996, no Dia do Patrimônio Cultural, seu acervo dispõe de peças do cotidiano de senhores e escravos, desde objetos ligados ao sincretismo religioso até aqueles utilizados no tráfico negreiro (GOV.BR, 2021a)

O Museu da Inconfidência está localizado na antiga Casa de Câmara e Cadeia de Ouro Preto. Inaugurado em 1944, visa a preservar, pesquisar e divulgar objetos e documentos relacionados à Inconfidência Mineira. O acervo reúne peças históricas e artísticas. Além do Panteão dos Inconfidentes, guarda lápides com os restos mortais de 16 inconfidentes, incluindo o poeta Tomás Antônio Gonzaga (GOV.BR, 2023c).

O Museu das Bandeiras ocupa a antiga Casa de Câmara e Cadeia da Província de Goyaz. Criado em 1949, reúne o conjunto de documentos que representa uma das fontes de informação mais importantes sobre a administração pública da região Centro-Oeste durante o período colonial, imperial e republicano (GOV.BR, 2022a).

O Museu das Missões foi projetado em 1938 e está situado junto aos remanescentes do Sítio Histórico São Miguel Arcanjo, ambos reconhecidos como Patrimônio Cultural da Humanidade pela Unesco. Reúne a maior coleção pública de imagens missionárias em madeira policromada dos séculos XVII e XVIII do Mercosul (GOV.BR, 2023d).

O Museu de Arqueologia de Itaipu está sediado nos remanescentes do Recolhimento de Santa Teresa. O acervo é composto por artefatos produzidos pelos povos que viveram no litoral fluminense antes de 1500. São artefatos líticos e ósseos, concreções, matéria corante, ocre, restos ósseos humanos e remanescentes de fauna, além de blocos testemunhos do Sambaqui de Camboinhas (GOV.BR, 2023e).

O Museu de Arte Religiosa e Tradicional está localizado no Antigo Convento de Nossa Senhora dos Anjos, que oferece uma exposição permanente de arte sacra dos séculos XVII e XVIII. O prédio é tombado pelo Patrimônio Histórico e Artístico Nacional (GOV.BR, 2023f).

O Museu de Arte Sacra da Boa Morte está sediado na Igreja de Nossa Senhora da Boa Morte e tem seu acervo constituído por objetos litúrgicos, prataria e obras do escultor e pintor goiano José Joaquim da Veiga Valle, do século XIX (GOV.BR, 2022b).

O Museu do Diamante foi criado em 1954 e atende o visitante que busca conhecer os contextos sobre a vocação da cidade de Diamantina como polo minerador, em especial do ouro e do diamante. Seu acervo é composto por peças de valor histórico e artístico dos séculos XVIII e XIX (GOV.BR, 2021b).

O Museu do Ouro – Casa Borba Gato está instalado na antiga Casa de Intendência e Fundição do Ouro da Vila Real de Nossa Senhora da Conceição do Sabará. Foi criado em 1945 e inaugurado em 1946. O acervo é constituído por mobiliário, armaria, porcelanas, imaginária religiosa e objetos ligados à prática da mineração, dos séculos XVIII e XIX (GOV.BR, 2022c).

O Museu Histórico Nacional é um dos maiores do país e ocupa um conjunto arquitetônico de relevância histórica para o Rio de Janeiro. O acervo é composto por

pinturas, esculturas, armaria, viaturas, porcelanas, prataria e a maior coleção de moedas antigas da América Latina, além de uma biblioteca especializada em história do Brasil (GOV.BR, 2023g).

O Museu Regional Casa dos Ottoni ocupa uma construção do século XVIII e sua origem histórica está ligada aos Ottoni, descendentes do bandeirante Fernão Dias Paes Leme. Criado em 1949, o acervo é formado, principalmente, por imagens de arte católica (GOV.BR, 2023h).

O Museu Regional de Caeté está instalado em uma casa construída em fins do século XVIII e guarda o acervo composto por mobiliário e objetos de época, além de peças de arte e arte sacra populares. Atua também na preservação da riqueza cultural, dos saberes e fazeres do povo de Caeté e seus distritos (GOV.BR, 2022d).

O Museu Regional de São João Del-Rei é a antiga casa do Comendador João Antônio da Silva Mourão, na época do Segundo Império. O prédio foi tombado em 1946 e aberto à visita em 1963. Seu acervo é constituído, em sua maior parte, de peças de mobiliário, objetos de arte sacra e imaginária oitocentista procedentes de várias cidades de Minas Gerais (GOV.BR, 2023i).

O Museu Solar Monjardim está instalado em um casarão colonial, construído na década de 1780. Antiga sede da Fazenda Jucutuquara, o local abrigou importantes personagens da história regional e nacional. Com acervo histórico eclético, revela aspectos da vida cotidiana no século XIX, por meio de objetos, práticas sociais e manifestações artísticas da época (GOV.BR, 2021c).

O Museu Victor Meirelles está localizado na casa natal do pintor Victor Meirelles, construída no final do século XVIII. O imóvel tombado pelo Patrimônio Histórico e Artístico Nacional preserva a memória do pintor catarinense, um dos mais importantes artistas brasileiros do século XIX (GOV.BR, 2022e).

O Museu Villa-Lobos está instalado em um casarão tombado do século XIX. Reúne objetos e documentos referentes à vida e à obra do compositor e maestro Heitor Villa-Lobos. O acervo é composto por partituras, correspondências, recortes de jornais, discos, filmes, livros, condecorações, instrumentos musicais e objetos de uso pessoal (GOV.BR, 2023j).

Para catalogação de acervos, existem padrões de documentação internacionais bastante utilizados, como por exemplo *Categories for the Description*

of Works of Art (CDWA¹), um conjunto de diretrizes para a descrição de arte, arquitetura e outras obras culturais, o *Cataloging Cultural Objects* (CCO²), um manual para descrever, documentar e catalogar artefatos culturais, o CDWA Lite³, um esquema *eXtensible Markup Language* (XML) para descrever os registros principais para obras de arte e cultura material com base no CDWA e CCO, o *Lightweight Information Describing Objects* (LIDO⁴), um esquema que combina CDWA Lite e museumdat e que pode ser usado para enviar informações sobre todos os tipos de objetos do patrimônio cultural, o VRA Core⁵, um padrão de dados para a descrição de obras de cultura visual e o CIDOC *Conceptual Reference Model* (CRM⁶), uma ontologia de referência no campo do patrimônio cultural.

No entanto, todos são, em diferentes níveis, complexos para uso. Dessa forma, o Ibram optou por definir regras específicas para catalogação de seus acervos, criando o Inventário Nacional de Bens Culturais Musealizados (INBCM) - um instrumento da PNM, em consonância com o Decreto nº 8.124/2013, o qual regulamenta dispositivos da Lei nº 11.904/2009 e da Lei nº 11.906/2009 - que se trata de “um instrumento de inserção periódica de dados sobre os bens culturais musealizados [...], para fins de identificação, acautelamento e preservação” (BRASIL, 2021).

O INBCM tem como objetivos: identificar os bens culturais que integram o acervo dos museus brasileiros para fins de acautelamento, preservação e difusão das informações; contribuir para a proteção e preservação dos bens culturais musealizados; contribuir para a difusão dos museus brasileiros; manter atualizados os dados referentes aos bens culturais musealizados inventariados; fornecer padrões e procedimentos para identificar os bens culturais musealizados; e disponibilizar informações sobre os bens culturais que compõem o Inventário Nacional (MINISTÉRIO DO TURISMO, 2019, p. 11).

O INBCM está descrito na Resolução Normativa do Ibram n.6, de 2021, e possui 15 metadados: número de registro, outros números, situação, denominação, título, autor, classificação, resumo descritivo, dimensões (considera medidas bidimensionais, como altura x largura; tridimensionais, como altura x largura x

¹ https://www.getty.edu/research/publications/electronic_publications/cdwa/. Acesso em 11 jan. 2023

² <https://vraweb.org/resources/cataloging-cultural-objects/>. Acesso em 11 jan. 2023

³ https://www.getty.edu/research/publications/electronic_publications/cdwa/cdwalite.pdf. Acesso em 11 jan. 2023

⁴ <http://www.lido-schema.org/schema/v0.9/lido-v0.9-specification.pdf>. Acesso em 11 jan. 2023

⁵ <https://www.loc.gov/standards/vracore/>. Acesso em 11 jan. 2023

⁶ <https://www.cidoc-crm.org/>. Acesso em 11 jan. 2023

profundidade; circulares, como diâmetro x espessura e peso), material/técnica, estado de conservação, local de produção, data de produção, condições de reprodução e mídias relacionadas. O documento apresenta a descrição semântica de cada elemento, além da informação da obrigatoriedade ou não de cada metadado (BRASIL, 2021).

Entretanto, o INBCM, a priori, não pode ser considerado um guia de catalogação, tampouco um padrão de metadados, pois não apresenta orientações formais acerca dos preenchimentos dos elementos de metadados, em seus aspectos de valor e de conteúdo, não sugerindo, inclusive, vocabulários controlados para elementos de metadados específicos.

Os museus vinculados ao Ibram possuem a prática de uso de linguagens documentárias para o preenchimento de valores de dados destinados a entidades associadas à classificação de temas, assuntos ou contextos de uso (MARTINS et al., 2021), a saber: o Tesouro de Objetos do Patrimônio Cultural nos Museus Brasileiros (FERREZ, 2016) e o Thesaurus para Acervos Museológicos (FERREZ, 1987). Porém, o INBCM não faz menção a qualquer orientação acerca de qual versão do tesouro usar, podendo acarretar significativas diferenças terminológicas no processo de indexação em âmbito geral dos museus, o que pode causar inconvenientes de ordem semântica no momento da agregação de itens de acervos referentes a uma determinada temática.

A descrição de inúmeros itens dos museus geralmente é realizada de maneira independente, ressaltando diferentes características a serem descritas e diferentes terminologias para descrevê-las, culminando em situações problemáticas para os sistemas de recuperação da informação (LANCASTER, 1986; 2004; ABADAL; CODINA, 2005; BAEZA-YATES; RIBEIRO-NETO, 2011) como, por exemplo: i) busca feita por palavras isoladas e descontextualizadas, o que dificulta maior visibilidade do acervo sob a ótica dos usuários e, conseqüentemente, dos mecanismos de busca; ii) ambigüidade conceitual (de qual conceito precisamente está se falando?) e; iii) pouca relevância para o recurso recuperado.

Conforme supracitado, o software utilizado pelo Ibram é o Tainacan: software gestor de acervos digitais, uma ferramenta flexível, voltada à difusão e à interoperabilidade entre acervos digitais culturais brasileiros, permitindo catalogar, organizar, armazenar e compartilhar informações dos acervos com facilidade,

mantendo todos os requisitos de uma plataforma profissional para repositórios. Se adapta às necessidades do usuário, oferecendo uma série de recursos customizáveis, como a criação de coleções, metadados, itens, filtros e outros (GOV.BR, 2021d; TAINACAN.ORG, 2023).

É um software livre⁷, com licença *General Public License version 3* (GPLv3), que permite o seu uso comercial, a sua modificação, a sua distribuição, o uso de patentes e o seu uso privado, que vislumbrou sua manutenção a longo prazo, escolhendo como base a ecologia do WordPress⁸, amplamente difundida e com uma comunidade de desenvolvedores e usuários ativa, além de oferecer a facilidade para que profissionais da área da cultura, sem formação específica em Tecnologia da Informação (TI), pudessem instalar, customizar e utilizar de forma autônoma (MARTINS; MARTINS, 2020).

Por ser um plugin e um tema do WordPress, é necessário ter uma instalação do WordPress ativa e, assim, usufruir do painel administrativo que permite a gestão e a publicação de seu acervo, a criação de perfis de usuários, bem como páginas web para a comunicação dos acervos (TAINACAN.ORG, 2023). O download do Tainacan está disponível na biblioteca de plugins do WordPress⁹.

O Tainacan é uma ferramenta robusta, que disponibiliza um grande rol de funcionalidades, mas para compreensão destas, alguns conceitos básicos são importantes.

O repositório é composto por coleções, responsáveis por agrupar itens que compartilham um mesmo conjunto de metadados. Tais itens, por sua vez, trata-se do verdadeiro conteúdo do repositório: a 'pintura', o 'filme', o 'livro', compostos pelos metadados, mas também por documentos, que são a informação principal do item, ou seja, ao objeto a que os metadados se referem (TAINACAN, 2022).

Em relação aos metadados, cada coleção contém um conjunto de metadados, conforme supracitado, que podem ser livremente definidos e configurados pelo usuário, o qual pode escolher os nomes, os tipos (texto simples, área de texto, data, numérico, taxonomias etc.¹⁰), a obrigatoriedade ou não de preenchimento, a

⁷ <https://github.com/orgs/tainacan/repositories>. Acesso em 11 jan. 2023

⁸ Projeto de código aberto para criar sites, blogs ou aplicativos, <https://br.wordpress.org/>. Acesso em 11 jan. 2023

⁹ <https://br.wordpress.org/plugins/tainacan/>. Acesso em 11 jan. 2023

¹⁰ <https://tainacan.github.io/tainacan-wiki/#/pt-br/metadados?id=tipos-de-metadados>. Acesso em 11 jan. 2023

possibilidade de receber valores únicos ou múltiplos, a definição se são privados ou públicos, dentre outros (TAINACAN, 2022).


Ressalta-se o metadado do tipo taxonomia, pois este permite a criação e utilização de vocabulários controlados para classificação de itens. Este metadado torna o resultado do uso de filtros, para busca com navegação facetada, ainda mais eficiente (TAINACAN, 2022).

Outro ponto importante são as ferramentas de busca e recuperação da informação do Tainacan. A saber, para Kuhlthau (1991) “uma busca de informação é um processo de construção que envolve toda a experiência da pessoa, sentimentos, bem como pensamentos e ações” (p. 362, tradução nossa). Para Baeza-Yates; Ribeiro-Neto (2013), a Recuperação da Informação (RI) é uma área da Ciência da Computação que diz respeito à representação, ao armazenamento, à organização e ao acesso a itens, visando recuperar todos os documentos que são relevantes à necessidade de informação do usuário e, ao mesmo tempo, recuperar o menor número possível de documentos irrelevantes.

Dessa forma, são apresentadas ferramentas de busca e recuperação do Tainacan: as buscas textuais, avançadas, facetadas, dentro de arquivos PDF e a junção destas para maior refinamento das pesquisas.


A busca textual é realizada por meio de palavras-chaves que são procuradas no conteúdo de qualquer metadado do item. A Figura 1 ilustra o resultado da pesquisa realizada a partir da palavra-chave ‘cadeira’. No exemplo, recuperam-se três itens a partir de metadados distintos: dentro do texto, em ‘resumo descritivo’, assim como em ‘denominação’ e ‘título’.

Figura 1. Exemplo de busca textual realizada por meio da palavra-chave 'cadeira'




RESUMO DESCRITIVO

Sobre fundo escuro, figura masculina sentada em **cadeira** dourada estofada com tecido vermelho, voltado para a esquerda, rosto à frente. Cabelos curtos pretos, olhos claros e barba passa-piolho grisalha. Traja paletó, colete e gravata em laço na cor preta com camisa branca de colarinho alto.



DENOMINAÇÃO

Cadeira



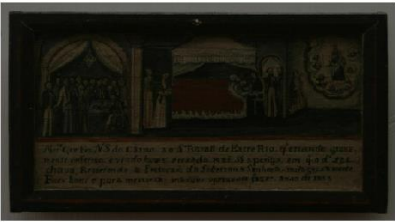
TÍTULO

Cadeira

Fonte: elaborada pela autora (2023)

Se for inserido um conjunto de palavras, o buscador retornará o item a partir do texto completo, como por exemplo a busca textual por 'cadeira branca', conforme ilustra a Figura 2.

Figura 2. Resultado da busca textual do termo 'cadeira branca'



Metadados

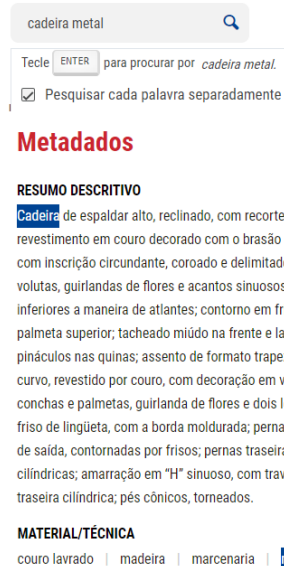
RESUMO DESCRITIVO

Ex-voto confeccionado em tábuas de madeira, de formato retangular, com cena pintada a têmpera na parte frontal e em sentido horizontal. Pintura dividida em duas partes, tendo na parte superior três cenas diferenciadas. Na parte superior, cena central com representação do Barão de Entre Rios feita através de figura masculina, de meia-idade, deitada sobre leito com dossel branco e cabeceira em tom verde, trajando camisola branca de mangas compridas, tendo o corpo coberto por colcha vermelha com barra em bordado branco. Ao lado do enfermo, figura masculina, representando o padre, sentado sobre **cadeira branca** segurando o braço esquerdo do enfermo junto à Bíblia. À direita, sob portal em arco, duas figuras masculinas, jovens, de pé, em trajes civis.

Fonte: elaborada pela autora (2023)

Na versão mais atual do Tainacan (0.19.3), a busca textual conta com um novo recurso, o de buscar por palavras-chave separadamente. Por exemplo, busca-se por ‘cadeira’ e ‘metal’. A Figura 3 ilustra o resultado.

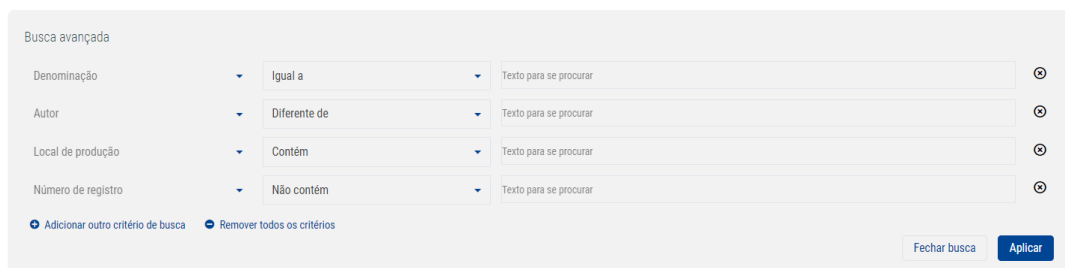
Figura 3. Busca textual por palavras-chaves utilizadas separadamente



Fonte: elaborada pela autora (2023)

A busca avançada traz maior controle ao usuário que possui um número maior de informações sobre o item, pois permite configurações mais específicas. Pode-se definir diferentes combinações utilizando os metadados junto a filtros, tais como: “igual a”, “diferente de”, “contém” e “não contém”. Assim como na busca textual, várias palavras no campo “texto para se procurar” levam à busca do termo completo. Nesse campo não há a possibilidade de buscar as palavras separadamente. A Figura 4 apresenta algumas opções disponíveis.

Figura 4. Opções de filtros da ferramenta busca avançada do Tainacan



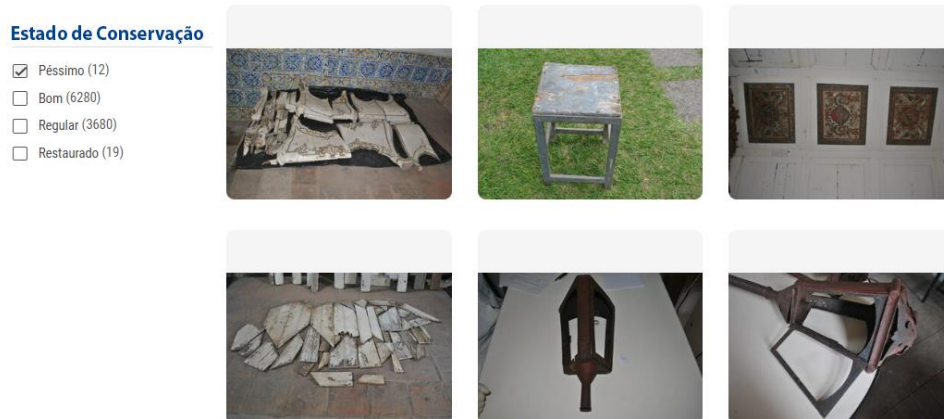
Fonte: elaborada pela autora (2023)

A busca por navegação facetada é possível, pois o Tainacan permite configurar qualquer metadado como uma faceta, possibilitando a criação de filtros pré-definidos.

Classificação facetada é o sistema que agrupa termos estruturados, nas bases da análise de um assunto, para a identificação de suas facetas, isto é, dos diferentes aspectos nele contidos. A análise facetada coordena conceitos, significando que um assunto, por mais completo que seja, pode ser representado (BARBOSA, 1972, p.74).

O tipo ideal de metadado para criação de facetas é o tipo ‘taxonomia’, o qual recebe um vocabulário controlado, sem ambiguidades, trazendo resultados positivos à RI. A Figura 5 mostra busca facetada realizada a partir do metadado ‘estado de conservação’.

Figura 5. Resultado parcial da busca facetada realizada por meio do metadado ‘estado de conservação’



Fonte: elaborada pela autora (2023)

A partir da versão 0.12, em uma abordagem experimental, há disponível a procura dentro de arquivos *Portable Document Format* (PDF), permitindo a busca textual dentro dos documentos no formato “PDF” (TAINACAN, 2022a).

É possível e recomendado que as formas de busca sejam utilizadas de maneira combinada. Por exemplo, a busca textual junto à busca facetada, conforme exemplo ilustrado na Figura 6. Nela se buscou pelo item ‘cadeira’, por meio da busca textual; porém, por meio da busca faceta, podemos reduzir a busca às peças que estão em bom estado de conservação, localizadas no Estado de Minas Gerais.

Figura 6. Resultado parcial da busca realizada com os recursos busca textual e busca facetada



Fonte: elaborada pela autora (2023)

Além disso, para interoperabilidade, o Tainacan implementa uma API RESTful - *Application Programming Interface - Representational State Transfer*¹¹, de leitura e escrita, que permite que outras aplicações interajam com seu repositório. Oferece, por padrão, um ponto de saída dos dados, ou seja, *Uniform Resource Locator* (URL) alternativos para visualização dos dados em diversos formatos, sendo eles: *JavaScript Object Notation* (JSON); *HyperText Markup Language* (HTML) e arquivos *Comma Separated Values* (CSV) (TAINACAN.ORG, 2023a; TAINACAN.ORG, 2022).

A partir das URLs pode-se criar *Application Programming Interface* – APIs que consomem os dados no formato aberto, gerando uma infinidade de novas aplicações.

Dados abertos são dados que podem ser livremente utilizados, reutilizados e redistribuídos por qualquer pessoa – sujeitos, no máximo, à exigência de citar a fonte original e compartilhar com as mesmas licenças em que as informações foram inicialmente apresentadas (ISOTANI; BITTENCOURT, 2015, p. 42).

Para Ofoeda, Boateng e Effah (2019) as APIs visam ao intercâmbio entre dois ou mais programas e são essenciais para melhorar a produtividade do desenvolvimento de software, principalmente por suportarem a sua reutilização, fornecendo-lhe novas funcionalidades, reduzindo o esforço e o tempo que os programadores gastam no seu desenvolvimento. Além disso, Ofoeda, Boateng e Effah (2019) apresentam uma definição menos técnica para APIs, no entanto esclarecedora:

¹¹ REST é um tipo de arquitetura de *software*, que indica um conjunto de restrições que devem ser seguidas no desenvolvimento de uma aplicação na internet.

as APIs fornecem um terreno comum para os softwares conversarem entre si, permitindo que diferentes deles troquem informações e serviços dentro das organizações, criando-lhe valor.

O software continua em franco desenvolvimento, atualmente pelo Laboratório de Inteligência de Redes da Universidade de Brasília, com apoio da UFG, do Instituto Brasileiro de Informação em Ciência e Tecnologia (Ibict) e do Ibram (TAINACAN.ORG, 2023).

Tornou-se uma consolidada ferramenta, atualmente com mais de 18 mil downloads realizados¹², utilizada por diferentes instituições, nacionais e internacionais, como exemplo: instituições de ensino públicas (MARTINS; MARTINS, 2020) e privadas; museus estaduais e municipais; instituições privadas e do terceiro setor; além de diversas temáticas de acervos, tais como: arquitetura, artesanato, documental, Covid-19, entre outros (TAINACAN, 2023; GOV.BR, 2022b).

Dada suas características de software livre e do crescimento de instalações e de usuários, criou-se uma comunidade que mantém um fórum de discussão¹³, descrito como um espaço para sanar dúvidas, para aprender e para se discutir sobre o Tainacan. O espaço possui três categorias principais, a de suporte, a de desenvolvimento e a de divulgação, que tem seus tópicos discutidos pelos administradores, moderadores e comunidade.

A compreensão sobre o Ibram e sobre o Tainacan colaboram no entendimento do estudo de caso proposto.

3.2 Estudo de Casos Múltiplos

O estudo de caso é um método de pesquisa que permite aos investigadores focar em um caso específico a fim de reter uma perspectiva holística do mundo real, permitindo, a partir de uma investigação empírica, aprofundar o conhecimento sobre um fenômeno contemporâneo e, assim, oferecer subsídios para novas investigações (YIN, 2015).

O método permite tanto o estudo de um único caso como o de casos múltiplos, visando a um conjunto único de conclusões. A lógica de utilização dos multicaseos diz respeito à replicação e não amostragem, ou seja, não permite generalização dos

¹² <https://br.wordpress.org/plugins/tainacan/advanced/>. Acesso em 09 jan. 2023

¹³ <https://tainacan.discourse.group/>. Acesso em 11 jan. 2023

resultados para toda a população, mas sim a possibilidade de previsão de resultados similares ou a possibilidade de produzir resultados contrários por razões previsíveis, de modo semelhante ao método de experimentos (YIN, 2015).

Assim, este estudo de casos múltiplos segue as diretrizes sugeridas por Yin (2015) o qual propõe uma estratégia detalhada para construção de um protocolo de aplicação que visa à maior confiabilidade da pesquisa.

3.2.1. Protocolo de Estudo de Caso

Segundo Yin (2015), um protocolo de estudo de caso deve ter quatro seções, são elas:

3.2.1.1 Visão Geral

As questões anteriores ao estudo de caso foram levantadas a partir das técnicas de pesquisa: pesquisa documental; pesquisa bibliográfica e RSL. O detalhamento da metodologia empregada em cada técnica, assim como seus resultados, estão apresentados nas seções 3.3 Pesquisa Bibliográfica, 3.4 Pesquisa Documental, 3.5 Revisão Sistemática de Literatura e 4 RESULTADOS.

O protocolo de realização de cada técnica de pesquisa foi idealizado pela pesquisadora com o acompanhamento do Professor Orientador e, posteriormente, validado por um grupo de estudo interdisciplinar composto por mais quatro integrantes do Laboratório de Inteligência de Redes, da UnB.

3.2.1.2 Procedimentos de coleta de dados

O estudo de caso será realizado com os 20 museus digitais do Ibram que, por meio do software Tainacan, disponibilizam os metadados públicos em formato de dados abertos. A relação de museus está apresentada no Quadro 1. Museus federais vinculados ao Ibram, com seus respectivos Estados de localização, link dos acervos digitais e quantidade de itens públicos.

No entanto, como etapa prévia à coleta destes dados, é necessário, por uma única vez, que todos os museus mapeiem seus metadados para os elementos descritivos propostos no INBCM. Para isso, cada museu deve atualizar sua instalação

Tainacan para versão 0.15 ou superior, instalar o plugin de mapeamento de dados para o INBCM e realizar o mapeamento.

Para esta etapa o contato é realizado via Coordenação de Arquitetura da Informação Museal (CAInf), do Ibram, que apoia o uso dos acervos pelos museus no Tainacan, e pode realizar todos os passos por meio de uma solicitação via e-mail.

Cada museu passa a disponibilizar uma URL para acesso à API JSON, contendo todos os dados abertos e mapeados para o INBCM. Assim, a descrição de cada caso está disponível nos Quadro 2 a Quadro 21.

Quadro 2. Descrição do caso Museu Casa de Hera

Museu Casa de Hera	
Versão do Tainacan: 0.19.3	
Coleção Acervo Museológico	
Link	http://museucasadahera.acervos.museus.gov.br/acervo-museologico
API JSON	http://museucasadahera.acervos.museus.gov.br/wp-json/tainacan/v2/collection/7/items/?perpage=50&order=DESC&orderby=date&exposer=json-flat&mapper=inbcm-ibram&paged=1
Metadados	Número Tainacan; Revisar número Tainacan; Nº de tomo; Denominação; Título; Autor/Fabricante; Resumo descritivo; Material; Técnica; Altura Total (item + acessórios); Altura (item + acessórios); Largura (item + acessórios); Largura (item + acessórios); Comprimento (item + acessórios); Diâmetro (item + acessórios); Profundidade (item + acessórios); Circunferência (item + acessórios); Classificação; Época; Data de produção; Local de produção; Categoria; Espécie; Inscrições; Partes integrantes; Etiqueta; Fontes e Condições de reprodução
Metadados mapeados	outros-numeros; denominacao; titulo; autor; resumo-descritivo; material-tecnica; classificacao; data-producao; local-producao e condicoes-reproducao
Coleção Indumentária	
Link	http://museucasadahera.acervos.museus.gov.br/indumentaria
API JSON	http://museucasadahera.acervos.museus.gov.br/wp-json/tainacan/v2/collection/42232/items/?perpage=50&order=DESC&orderby=date&exposer=json-flat&mapper=inbcm-ibram&paged=1
Metadados	Nº de tomo; Número Tainacan; Revisar número Tainacan; Denominação; Título; Autor/Fabricante; Resumo descritivo; Material; Técnica; Classificação; Época; Data de produção; Altura Frente; Altura Costas; Circunferência Cintura; Circunferência Barra; Circunferência Punho Direito; Circunferência Punho Esquerdo;

	Comprimento Manga Direita; Comprimento Manga Esquerda; Pescoço; Ombro; Altura Salto Direito; Altura Salto Esquerdo; Largura Salto Direito; Largura Salto Esquerdo; Circunferência Cabeça; Largura da Copa/Aba; Partes integrantes; Local de produção; Categoria; Espécie; Inscrições; Etiqueta; Fontes; Condições de reprodução; Número Tainacan e Número Tainacan
Metadados mapeados	outros-numeros; denominacao; titulo; autor; resumo-descritivo; material-tecnica; classificacao; data-producao; local-producao e condicoes-reproducao

Fonte: Acervo Digital do Museu Casa de Hera (2023)

Quadro 3. Descrição do caso Museu Casa da Princesa

Museu Casa da Princesa	
Versão do Tainacan: 0.19.3	
Coleção Museu Casa da Princesa	
Link	http://museusibramgoias.acervos.museus.gov.br/museu-casa-da-princesa
API JSON	https://museusibramgoias.acervos.museus.gov.br/wp-json/tainacan/v2/collection/1227/items/?perpage=96&order=ASC&orderby=meta_value&metakey=81804&exposer=json-flat&mapper=inbcm-ibram&paged=1
Metadados	Número de registro; Outros números; Situação; Denominação; Título; Autor; Classificação; Resumo descritivo; Altura; Largura; Comprimento; Profundidade; Diâmetro; Peso Material; Técnica; Marcas/Inscrições; Local de produção; Data de produção; Localização; Modo de aquisição; Data de aquisição; Procedência; Fonte da informação; Histórico/Observação e Código Thesaurus
Metadados mapeados	numero-registro ; outros-numeros; situacao; denominacao; titulo; autor; classificacao; resumo-descritivo; dimensoes-altura; dimensoes-largura; dimensoes-profundidade-comprimento; dimensoes-diametro; dimensoes-peso; material-tecnica; local-producao e data-producao

Fonte: Acervo Digital do Museu Casa da Princesa (2023)

Quadro 4. Descrição do caso Museu Casa de Benjamin Constant

Museu Casa de Benjamin Constant	
Versão do Tainacan: 0.19.3	
Coleção Acervo Museológico	
Link	http://museucasabenjaminconstant.acervos.museus.gov.br/acervo-museologico

API JSON	http://museucasabenjaminconstant.acervos.museus.gov.br/wp-json/tainacan/v2/collection/5347/items/?perpage=96&order=DESC&orderby=date&exposer=json-flat&mapper=inbcm-ibram&paged=1
Metadados	Denominação; Número de registro; Coleção; Título; Classificação; Data de produção; Autor; Fabricante; Material; Técnica; Local de produção; Marcas/Inscrições; Altura (cm); Largura (cm); Comprimento (cm); Diâmetro (cm); Profundidade (cm); Espessura (cm); Resumo descritivo; Características técnicas; Características estilísticas; Características iconográficas/ornamentais; Dados históricos e Condições de reprodução
Metadados mapeados	denominacao; numero-registro; titulo; classificacao; data-producao; autor; material-tecnica; local-producao; dimensoes-altura; dimensoes-largura; dimensoes-profundidade-comprimento; dimensoes-diametro; dimensoes-espessura; resumo-descritivo e condicoes-reproducao

Fonte: Acervo Digital do Museu Casa de Benjamin Constant (2023)

Quadro 5. Descrição do caso Museu Casa Histórica de Alcântara

Museu Casa Histórica de Alcântara	
Versão do Tainacan: 0.19.3	
Coleção Acervo Museológico	
Link	http://museucasahistoricadealcantara.acervos.museus.gov.br/objetos
API JSON	https://museucasahistoricadealcantara.acervos.museus.gov.br/wp-json/tainacan/v2/collection/14284/items/?perpage=96&order=ASC&orderby=date&exposer=json-flat&mapper=inbcm-ibram&paged=1
Metadados	Número de registro; Denominação; Título; Autor; Fábrica/Oficina; Classificação; Altura (cm); Resumo descritivo; Largura (cm); Comprimento/Profundidade (cm); Peso (g); Material/Técnica; Técnicas; Marcas e inscrições; Local de produção; Século; Data de produção; Condições de reprodução; Histórico; Usos; Função e Bibliografia
Metadados mapeados	numero-registro; denominacao; titulo; autor; classificacao; dimensoes-altura; resumo-descritivo; dimensoes-largura; dimensoes-profundidade-comprimento; dimensoes-peso; material-tecnica; local-producao; data-producao e condicoes-reproducao

Fonte: Acervo Digital do Museu Casa Histórica de Alcântara (2023)

Quadro 6. Descrição do Caso Museu da Abolição

Museu da Abolição

Versão do Tainacan: 0.19.3	
Coleção Acervo Museológico	
Link	https://museudaabolicao.acervos.museus.gov.br/acervo_museologico/
API JSON	https://museudaabolicao.acervos.museus.gov.br/wp-json/tainacan/v2/collection/7/items/?perpage=96&order=DESC&orderby=date&exposer=json-flat&mapper=inbcm-ibram&paged=1
Metadados	Número de registro; Denominação; Título; Autor; Classificação; Resumo descritivo; Altura; Peso; Largura/Diâmetro; Profundidade/Espessura; Material; Técnica; Local de produção; Data de produção; Condições de reprodução; Mídias relacionadas; Coleção; Faz conjunto com; Assunto; Tipologia; Moldura, base, estojo; Inscrições; Histórico do objeto; mídias relacionadas e condições de reprodução
Metadados mapeados	numero-registro ; denominação; titulo; autor; classificação; resumo-descritivo; dimensoes-altura; dimensoes-peso; dimensoes-largura; dimensoes-espessura; material-tecnica; local-producao; data-producao; condicoes-reproducao e midias-relacionadas

Fonte: Acervo Digital do Museu da Abolição (2023)

Quadro 7. Descrição do caso Museu da Inconfidência

Museu da Inconfidência	
Versão do Tainacan: 0.19.3	
Coleção Acervo Museológico	
Link	http://museudainconfidencia.acervos.museus.gov.br/acervo-museologico
API JSON	https://museudainconfidencia.acervos.museus.gov.br/wp-json/tainacan/v2/collection/9/items/?perpage=96&order=ASC&orderby=date&exposer=json-flat&mapper=inbcm-ibram&paged=1
Metadados	Número de registro; Denominação; Título; Autor; Fabricante; Marcas/Inscrições; Classificação; Estilo; Altura (cm); Largura (cm); Profundidade (cm); Comprimento (cm); Espessura (cm); Diâmetro (cm); Peso (Kg); Material; Técnica; Temas; Estado de conservação; Local de produção; Complemento - Local de produção; Data de produção e Fotografia
Metadados mapeados	numero-registro; denominacao; titulo; autor; classificacao; dimensoes-altura; dimensoes-largura; dimensoes-profundidade-comprimento; dimensoes-espessura; dimensoes-diametro; dimensoes-peso; material-tecnica; estado-conservacao; local-producao e data-producao

Fonte: Acervo Digital do Museu da Inconfidência (2023)

Quadro 8. Descrição do caso Museu das Bandeiras

Museu das Bandeiras	
Versão do Tainacan: 0.19.3	
Coleção Museus das Bandeiras	
Link	http://museusibramgoias.acervos.museus.gov.br/museu-das-bandeiras
API JSON	https://museusibramgoias.acervos.museus.gov.br/wp-json/tainacan/v2/collection/1093/items/?perpage=96&order=ASC&orderby=meta_value&metakey=81804&exposer=json-flat&mapper=inbcm-ibram&paged=1
Metadados	Número de registro; Outros números; Situação; Denominação; Título; Autor; Classificação; Resumo descritivo; Altura; Largura; Comprimento; Profundidade; Diâmetro; Peso; Material; Técnica; Marcas/Inscrições; Local de produção; Data de produção; Localização; Modo de aquisição; Data de aquisição; Procedência; Fonte de aquisição; Tema; Exposições; Histórico/Observação; Acervo; Categoria; Coleção; Código Thesaurus e Destaque no acervo
Metadados mapeados	numero-registro; outros-numeros; situacao; denominacao; titulo; autor; classificacao; resumo-descritivo; dimensoes-altura; dimensoes-largura; dimensoes-profundidade-comprimento; dimensoes-diametro; dimensoes-peso; material-tecnica; local-producao e data-producao

Fonte: Acervo Digital do Museu das Bandeiras (2023)

Quadro 9. Descrição do caso Museu das Missões

Museu das Missões	
Versão do Tainacan: 0.19.3	
Coleção Acervo Museológico	
Link	http://museudasmissoes.acervos.museus.gov.br/acervo-museologico
API JSON	https://museudasmissoes.acervos.museus.gov.br/wp-json/tainacan/v2/collection/6409/items/?perpage=96&order=ASC&orderby=date&exposer=json-flat&mapper=inbcm-ibram&paged=1
Metadados	Título para o público; Número de registro; Título em inglês; Denominação; Título; Autor; Coleção; Classificação; Material/Técnica; Marcas/Inscrições; Local de produção; Data de produção; Altura (cm); Largura (cm); Diâmetro (cm); Profundidade (cm); Peso (Kg); Temas; Estilo; Movimento; Escola/Grupo Cultural; Exposições; Resumo descritivo; Procedência; Desdobramento

	número; Referências bibliográficas/arquivísticas; Domínio público e Condições de reprodução
Metadados mapeados	título; numero-registro; autor; classificacao; material-tecnica; local-producao; data-producao; dimensoes-altura; dimensoes-largura; dimensoes-diametro; dimensoes-profundidade-comprimento; dimensoes-peso; resumo-descritivo e condicoes-reproducao;

Fonte: Acervo Digital do Museu das Missões (2023)

Quadro 10. Descrição do caso Museu de Arqueologia de Itaipu

Museu de Arqueologia de Itaipu	
Versão do Tainacan: 0.19.3	
Coleção MAI	
Link	http://museudearqueologiadeitaipu.museus.gov.br/museu-itaipu
API JSON	http://museudearqueologiadeitaipu.museus.gov.br/wp-json/tainacan/v2/collection/94553/items/?perpage=96&order=DESC&orderby=date&exposer=json-flat&mapper=inbcm-ibram&paged=1
Metadados	Número de registro; Outros números; Denominação; Data da entrada no acervo; Procedência; Modo de aquisição; Doador; Material/Técnica; Classificação; Coleção; Autor; Resumo descritivo; Nº de partes; Estado de conservação; Datação; Localização atual; Data da compilação da ficha; Largura – Mesial (cm); Largura – Proximal (cm); Largura - Zona distal (cm); Largura - Zona mesial (cm); Largura - Zona proximal (cm); Largura – Distal (cm); Diâmetro/Comprimento da base (cm); Diâmetro/Medida do ombro (cm); Diâmetro/Comprimento da boca (cm); Altura (cm); Largura (cm); Espessura (cm); Diâmetro (cm); Comprimento (cm); Peso (g); Matéria prima; Processos curatoriais; Observações; Equipe MAI; Data MAI; Localização; Histórico; Projeto de Catalogação; Data do Projeto de Catalogação; Situação e Condições de reprodução
Metadados mapeados	numero-registro; outros-numeros; denominacao; material-tecnica; classificacao; autor; resumo-descritivo; estado-conservacao; data-producao; dimensoes-largura; dimensoes-diametro; dimensoes-altura; dimensoes-espessura; dimensoes-profundidade-comprimento; dimensoes-peso e situação

Fonte: Acervo Digital do Museu de Arqueologia de Itaipu (2023)

Quadro 11. Descrição do caso Museu de Arte Religiosa e Tradicional

Museu de Arte Religiosa e Tradicional	
Versão do Tainacan: 0.19.3	
Coleção Acervo Museológico	

Link	https://museudeartereligiosaetradicional.acervos.museus.gov.br/ace-rvo-museologico
API JSON	https://museudeartereligiosaetradicional.acervos.museus.gov.br/wp-json/tainacan/v2/collection/878/items/?perpage=96&order=DESC&orderby=date&exposer=json-flat&mapper=inbcm-ibram&paged=1
Metadados	Número de registro; Denominação; Título; Autor; Ano de produção; Século de produção; Dimensões (cm); Material; Técnica; Classificação; Resumo descritivo; Local de produção; Condições de reprodução e Local de produção
Metadados mapeados	numero-registro; titulo; autor; data-producao; dimensoes; material-tecnica; classificacao; resumo-descritivo; condicoes-reproducao e local-producao

Fonte: Acervo Digital do Museu de Arte Religiosa e Tradicional (2023)

Quadro 12. Descrição do caso Museu de Arte Sacra da Boa Morte

Museu de Arte Sacra da Boa Morte	
Versão do Tainacan: 0.19.3	
Coleção Museu de Arte Sacra da Boa Morte	
Link	http://museusibramgoias.acervos.museus.gov.br/museu-casa-da-boa-morte
API JSON	https://museusibramgoias.acervos.museus.gov.br/wp-json/tainacan/v2/collection/3387/items/?perpage=96&order=ASC&orderby=meta_value&metakey=699&exposer=json-flat&mapper=inbcm-ibram&paged=1
Metadados	Número de registro; Outros números; Situação; Título; Denominação; Autor; Classificação; Resumo descritivo; Altura; Largura; Comprimento; Profundidade; Diâmetro; Peso; Material; Técnica; Marcas/Inscrições; Local de produção; Data de produção; Localização; Modo de aquisição; Data de aquisição; Procedência e Histórico/Observação
Metadados mapeados	numero-registro; outros-numeros; situacao; denominacao; autor; classificacao; resumo-descritivo; dimensoes-altura; dimensoes-largura; dimensoes-profundidade-comprimento; dimensoes-diametro; dimensoes-peso; material-tecnica; local-producao e data-producao

Fonte: Acervo Digital do Museu de Arte Sacra da Boa Morte (2023)

Quadro 13. Descrição do caso Museu do Diamante

Museu do Diamante
Versão do Tainacan: 0.19.3

Coleção Acervo Museológico	
Link	https://museudodiamante.acervos.museus.gov.br/acervo-museologico
API JSON	https://museudodiamante.acervos.museus.gov.br/wp-json/tainacan/v2/collection/11603/items/?perpage=96&order=ASC&orderby=date&exposer=json-flat&mapper=inbcm-ibram&paged=1
Metadados	Título; Número de registro; Denominação; Classificação; Local de produção; Data de produção; Acervo; Procedência; Autor; Material/Técnica; Temas; Altura (cm); Largura (cm); Situação; Comprimento (cm); Profundidade (cm); Diâmetro (cm); Circunferência (cm); Espessura (cm); Peso (kg); Resumo descritivo; Data da fotografia; Marcas/Inscrições; Dados históricos; Referências bibliográficas/arquivísticas; Condições de reprodução e Mídias relacionadas
Metadados mapeados	titulo; numero-registro; denominacao; classificacao; local-producao; data-producao; autor; material-tecnica; dimensoes-altura; dimensoes-largura; situacao; dimensoes-profundidade-comprimento; dimensoes-diametro; dimensoes-espessura; dimensoes-peso; resumo-descritivo; condicoes-reproducao e midias-relacionadas

Fonte: Acervo Digital do Museu do Diamante (2023)

Quadro 14. Descrição do caso Museu do Ouro

Museu do Ouro	
Versão do Tainacan: 0.19.3	
Coleção Acervo Museológico	
Link	https://museudoouro.acervos.museus.gov.br/acervo
API JSON	https://museudoouro.acervos.museus.gov.br/wp-json/tainacan/v2/collection/3583/items/?perpage=96&order=ASC&orderby=date&exposer=json-flat&mapper=inbcm-ibram&paged=1
Metadados	número de registro; Denominação; Título; Classificação; Local de produção; Data de produção; Autor; Conjunto com nº; Material/Técnica; Palavras-chave; Fotógrafo; Data da fotografia; Esfera de proteção legal; Tipo de tombamento; Marcas/Inscrições; Dimensões; Resumo descritivo; Características estilísticas; Características iconográficas/ornamentais; Dados históricos; Referências bibliográficas/arquivísticas; Condições de reprodução; Domínio público e Mídias relacionadas
Metadados mapeados	numero-registro; denominacao; titulo; classificacao; local-producao; data-producao; autor; material-tecnica; dimensoes; resumo-descritivo; condicoes-reproducao e midias-relacionadas

Fonte: Acervo Digital do Museu do Ouro (2023)

Quadro 15. Descrição do caso Museu Histórico Nacional

Museu Histórico Nacional	
Versão do Tainacan: 0.19.3	
Coleção Acervo Museológico	
Link	http://mhn.acervos.museus.gov.br/reserva-tecnica
API JSON	https://mhn.acervos.museus.gov.br/wp-json/tainacan/v2/collection/24/items/?perpage=96&order=ASC&orderby=date&exposer=json-flat&mapper=inbcm-ibram&paged=1
Metadados	Número de registro; Outros números; Denominação; Resumo descritivo; Faz conjunto ou par com; Título; Autor; Técnica; Material; Forma de aquisição; Fonte de aquisição; Referência de aquisição; Local de produção; Data de Produção; Classe; Estado de conservação; Termos de Indexação; Altura (cm); Largura (cm); Comprimento (cm); Diâmetro (cm); Peso (g); Condições de reprodução; Observações; Exposição; Referências expográficas; Referências bibliográficas; Autoria das fotos; Exposições; Destaque do carrossel da página principal; Termos de indexação - revisar e transferir os termos para o "Termos de indexação"; Sobre Aleijadinho Temp e Condições de reprodução
Metadados mapeados	numero-registro; outros-numeros; denominacao; resumo-descritivo; titulo; autor; material-tecnica; local-producao; data-producao; classificacao; estado-conservacao; dimensoes-altura; dimensoes-largura; dimensoes-profundidade-comprimento; dimensoes-diametro; dimensoes-peso e condicoes-reproducao
Coleção Moedas de Ouro	
Link	https://mhn.acervos.museus.gov.br/moedas-de-ouro
API JSON	https://mhn.acervos.museus.gov.br/wp-json/tainacan/v2/collection/286327/items/?perpage=96&order=ASC&orderby=date&exposer=json-flat&mapper=inbcm-ibram&paged=1
Metadados	Resumo descritivo; Denominação; Outros títulos; Número de registro; Classe; Título; País; Local de produção; Data de produção; Governante; Material; Técnica; Estado de conservação; Diâmetro (mm); Peso (g); Exposição moedas de ouro; Referências; Exposição; Modo de aquisição e Condições de reprodução
Metadados mapeados	resumo-descritivo; denominacao; numero-registro; classificacao; titulo; local-producao; data-producao; material-tecnica; estado-conservacao; dimensoes-diametro e dimensoes-peso

Fonte: Acervo Digital do Museu Histórico Nacional (2023)

Quadro 16. Descrição do caso Museu Regional Casa dos Ottoni

Museu Regional Casa dos Ottoni	
Versão do Tainacan: 0.19.3	
Coleção Acervo Museológico	
Link	https://museuregionalcasadosottoni.acervos.museus.gov.br/acervo-museologico
API JSON	https://museuregionalcasadosottoni.acervos.museus.gov.br/wp-json/tainacan/v2/collection/31492/items/?perpage=96&order=ASC&orderby=date&exposer=json-flat&mapper=inbcm-ibram&paged=1
Metadados	Número de registro; Título para o público; Denominação; Autor; Material; Técnica; Origem; Ano de produção; data de produção; Classificação; Altura (cm); Comprimento (cm); Largura (cm); Espessura (cm); Profundidade (cm); Diâmetro (cm); Circunferência (cm); Resumo descritivo; Condições de reprodução; Características técnicas; Características estilísticas; Características iconográficas/ornamentais; Licenças (Creative Commons); Dados históricos; Referências bibliográficas/arquivísticas; Temas; Conjunto com nº; Marcas/inscrições/legendas e Época
Metadados mapeados	numero-registro; denominacao; autor; material-tecnica; local-producao; data-producao; classificacao; dimensoes-altura; dimensoes-profundidade-comprimento; dimensoes-largura; dimensoes-espessura; dimensoes-diametro; resumo-descritivo e condicoes-reproducao

Fonte: Acervo Digital do Museu Regional Casa dos Ottoni (2023)

Quadro 17. Descrição do caso Museu Regional do Caeté

Museu Regional do Caeté	
Versão do Tainacan: 0.19.3	
Coleção Acervo Museológico	
Link	https://museuregionaldecaete.acervos.museus.gov.br/acervo/
API JSON	https://museuregionaldecaete.acervos.museus.gov.br/wp-json/tainacan/v2/collection/848/items/?perpage=96&order=DESC&orderby=date&exposer=json-flat&mapper=inbcm-ibram&paged=1
Metadados	Número de registro; Denominação; Título; Autor; Classificação; Resumo descritivo; Altura (cm); Largura (cm); Comprimento (cm); Profundidade (cm); Diâmetro (cm); Circunferência (cm); Peso (g); Material/técnica; Local de produção; Data de produção; Marcas/inscrições; Conjunto com; Itens relacionados; Fotografia; Data da foto; Características técnicas; Características estilísticas; Características iconográficas/ornamentais; Dados históricos e Referências bibliográficas/arquivísticas
Metadados mapeados	numero-registro; denominacao; titulo; autor; classificacao; resumo-descritivo; dimensoes-altura; dimensoes-largura; dimensoes-

	profundidade-comprimento; dimensoes-diametro; dimensoes-peso; material-tecnica; local-producao e data-producao
--	--

Fonte: Acervo Digital do Museu Regional do Caeté (2023)

Quadro 18. Descrição do caso Museu Regional São João Del Rey

Museu Regional São João Del Rey	
Versão do Tainacan: 0.19.3	
Coleção Acervo Museológico	
Link	http://museuregionaldesaojoaodelrei.acervos.museus.gov.br/acervo_museologico
API JSON	https://museuregionaldesaojoaodelrei.acervos.museus.gov.br/wp-json/tainacan/v2/collection/11451/items/?perpage=96&order=ASC&orderby=date&exposer=json-flat&mapper=inbcm-ibram&paged=1
Metadados	Número de registro; Denominação; Título; Autor; Classificação; Resumo Descritivo; Altura (cm); Largura (cm); Comprimento (cm); Profundidade (cm); Espessura (cm); Diâmetro (cm); Peso (Kg); Material; Técnica; Marcas/Inscrições; Local de produção; Data de produção; Temas; Conjunto com n°; Dados históricos; Características Iconográficas/Ornamentais; Características Estilísticas; Características técnicas; Referências Bibliográficas/Arquivísticas e Documentação fotográfica
Metadados mapeados	numero-registro; denominacao; titulo; autor; classificacao; resumo-descritivo; dimensoes-altura; dimensoes-largura; dimensoes-profundidade-comprimento; dimensoes-espessura; dimensoes-diametro; dimensoes-peso; material-tecnica; local-producao e data-producao

Fonte: Acervo Digital do Museu Regional São João Del Rey (2023)

Quadro 19. Descrição do caso Museu Solar Monjardim

Museu Solar Monjardim	
Versão do Tainacan: 0.19.3	
Coleção Acervo Museológico	
Link	https://museusolarmonjardim.acervos.museus.gov.br/acervo
API JSON	https://museusolarmonjardim.acervos.museus.gov.br/wp-json/tainacan/v2/collection/109360/items/?perpage=96&order=ASC&orderby=date&exposer=json-flat&mapper=inbcm-ibram&paged=1
Metadados	número de registro; Denominação; Título; Autoria; Resumo descritivo; Data de produção; Local de produção; Classificação;

	Material/técnica; Altura (cm); Altura da cruz (cm); Altura do Cristo (cm); Altura da cabeça (cm); Outras medidas de altura (cm); Lagura (cm); Largura da cruz (cm); Largura do Cristo (cm); Largura da cabeça (cm); Outras medidas de largura (cm); Profundidade (cm); Profundidade da cruz (cm); Profundidade do Cristo (cm); Profundidade da cabeça (cm); Outras medidas de profundidade (cm); Características estilísticas; Características técnicas; Características iconográficas/ornamentais; Marcas/Inscrições/Legendas; Dados históricos; Referências bibliográficas/arquivísticas e Projeto
Metadados mapeados	numero-registro; denominacao; titulo; autor; resumo-descritivo; local-producao; classificacao; material-tecnica; dimensoes-altura; dimensoes-largura e dimensoes-profundidade-comprimento

Fonte: Acervo Digital do Museu Solar Monjardim (2023)

Quadro 20. Descrição do caso Museu Victor Meirelles

Museu Victor Meirelles	
Versão do Tainacan: 0.19.3	
Coleção do Acervo do Museu Victor Meirelles	
Link	http://museuvictormeirelles.acervos.museus.gov.br/mvm-acervo
API JSON	https://museuvictormeirelles.acervos.museus.gov.br/wp-json/tainacan/v2/collection/18004/items/?perpage=96&order=ASC&orderby=date&exposer=json-flat&mapper=inbcm-ibram&paged=1
Metadados	Classificação; Coleção; Denominação; Título; Informações sobre o autor; País de produção; Estado de produção; Cidade de produção; Data de produção/datação; Material/Técnica; Dimensões; Duração; Marcas/Inscrições; Estado de conservação; Notas sobre o estado de conservação; Procedência; Exposições; Bibliografia; Resumo descritivo; Descrição de conteúdo; Estilos/Temas; Exposições; Domínio público; Mídias relacionadas e Condições de reprodução
Metadados mapeados	classificacao; denominacao; titulo; autor; local-producao; data-producao; material-tecnica; dimensoes; estado-conservacao; resumo-descritivo; midias-relacionadas e condicoes-reproducao

Fonte: Acervo Digital do Museu Victor Meirelles (2023)

Quadro 21. Descrição do caso Museu Villa-Lobos

Museu Villa-Lobos	
Versão do Tainacan: 0.19.3	
Coleção Fotografias	
Link	http://museuvillalobos.acervos.museus.gov.br/fotografias

API JSON	https://museuvillalobos.acervos.museus.gov.br/wp-json/tainacan/v2/collection/1570/items/?perpage=96&order=ASC&orderby=date&exposer=json-flat&mapper=inbcm-ibram&paged=1
Metadados	Número de registro; Denominação; Título; Autor; Classificação; Resumo descritivo; Dimensões; Material/Técnica; Data de produção; Local de produção; Estado de conservação; Procedência; Modo de aquisição e Condições de reprodução
Metadados mapeados	numero-registro; denominacao; titulo; autor; classificacao; resumo-descritivo; dimensoes; material-tecnica; data-producao; local-producao; estado-conservacao e condicoes-reproducao

Fonte: Acervo Digital do Museu Villa-Lobos (2023)

Apenas os metadados públicos mapeados são coletados e utilizados na agregação, independente se vazios ou preenchidos. Os demais são desconsiderados, sendo possível serem visualizados apenas nos museus de origem.

3.2.1.3 Questões de estudo de caso

Em posse dos dados de todos os museus, a questão a ser analisada é se o que se tem é suficiente para tornar possível a agregação dos dados, sem interesse nas causas de como os repositórios foram montados, ou seja, há condições suficientes para produzir uma experiência de agregação? E essas condições são as mesmas para todos os museus, ou seja, é possível adicionar novos museus ao serviço de busca e recuperação, seguindo o mesmo workflow?

3.2.1.4 Guia para o relatório de estudo de caso

O relatório contará com a narrativa sobre as práticas identificadas visando à concretização do serviço de busca e recuperação da informação a partir dos casos estudados. O relatório demonstra as condições como se apresentam, sem que haja interferência no modo de trabalho dos museus, respeitando sua diversidade.

3.3 Pesquisa Bibliográfica

Pesquisa de caráter descritivo, de natureza qualitativa, realizada em bases de dados acadêmicos e na literatura cinzenta, com o intuito de encontrar workflows de agregação de dados culturais.

Foram utilizadas três fontes distintas de pesquisa: o Google, para localizar workflows na literatura cinzenta; a EBSCOhost, pela produção científica internacional, e a BRAPCI, por ser específica da área de Ciência da Informação no Brasil, tornando a pesquisa mais ampla.

As buscas foram realizadas a partir das palavras: “*pipeline*”, “*architecture*”, “*aggregation*”, “*metadata ingest*”, “*metadata aggregation*”, “*aggregative data infrastructures*” e suas versões em português. Além de projetos de agregação de dados culturais conhecidos de diferentes regiões do mundo, como cita Navarrete (2016): “*Europeana*”, “*Mexicana*”, “*Digital Public Library of America*”, “*Trove*” e “*DigitalNZ*”.

As pesquisas resultaram em oito projetos de agregação, realizados por sete instituições. O resultado foi publicado em 2019, no III WIDaT - Workshop de Informação, Dados e Tecnologia (SIQUEIRA; MARTINS, 2019b) e uma versão estendida, em 2020, pela Revista Ciência da Informação do Instituto Brasileiro de Informação em Ciência e Tecnologia - Ibict (SIQUEIRA; MARTINS, 2020).

3.4 Pesquisa Documental

Pesquisa de natureza qualitativa, de cunho descritivo, envolvendo a DPLA e a Europeana, realizando análise comparativa de suas políticas de qualidade de dados, com intuito de acessar fontes documentais a partir do sítio web das duas instituições, incluindo manuais, tutoriais, vídeos explicativos, materiais didáticos, entre outros conteúdos que fossem voltados explicitamente para apoiar as instituições desejosas de se tornarem provedores de dados para esses agregadores.

A saber, a *Digital Public Library of America.Inc* é uma instituição sem fins lucrativos que visa a ampliar o valor das bibliotecas e organizações culturais como fontes mais confiáveis de conhecimento compartilhado dos norte-americanos. Mantém o portal de busca integrada *Digital Public Library of America* – DPLA¹⁴ (DPLA,

¹⁴ <https://dp.la/>. Acesso em 11 jan. 2023

2023g), no qual são disponibilizados milhões de materiais de bibliotecas, arquivos, museus e outras instituições culturais de todo o país em uma única experiência de descoberta (DPLA, 2023a). A *Europeana Foundation* é uma organização independente e sem fins lucrativos que opera a plataforma *Europeana Collections*¹⁵ (EUROPEANA, 2023b), a qual contém milhares de itens oriundos de arquivos, bibliotecas e museus europeus, fornecendo acesso a livros, a músicas, a obras de arte, dentre outros tópicos, com ferramentas sofisticadas de pesquisa (EUROPEANA, 2023a; EUROPEANA PRO, 2023b).

Para efeitos da presente pesquisa, política de qualidade de dados é um conjunto de diretrizes técnicas que norteiam os interessados em agregar seus dados (DPLA, 2023c; SCHOLZ, 2019). De forma geral, a política, quando seguida, visa fornecer uma melhor experiência de usuário, criando coleções completas e coerentes (DPLA, 2023c). Além disso, possui caráter formativo, visto que os parceiros são encorajados a não apenas enviar um mínimo de metadados e qualidade de conteúdo, mas a buscar metadados ricos e a mais alta qualidade de dados possível (SCHOLZ, 2019).

Assim sendo, o objetivo é propor elementos preponderantes para uma política de qualidade de dados que auxilie procedimentos de agregação de acervos em rede, buscando: i) identificar e selecionar documentos estratégicos oriundos das instituições investigadas; e ii) analisar e comparar elementos de política de qualidade dos dados mapeados nesta investigação.

As técnicas envolvidas incluem a definição de categorias para apoiar análise de conteúdo dos materiais selecionados. Segundo Bardin (2016), as categorias são estabelecidas com base nos seguintes critérios: semântica (temas); sintaxe (verbos, adjetivos, pronomes); léxicos (agrupando sinônimos ou antônimos, ou de acordo com o significado de palavras); e expressão (agrupando desvios de escrita ou linguísticos).

O critério usado foi o semântico, para o qual se alinhou temas ligados aos campos da informação e da tecnologia. Nesse viés, as categorias de análise (tanto individual quanto comparativa) foram definidas e são descritas como se segue.

- Provedores de dados: instituições culturais individuais ou agregadoras que enviam dados para agregação.

¹⁵ <https://www.europeana.eu/>. Acesso em 11 jan. 2023

- Solicitação de participação: meios disponibilizados pelas instituições agregadoras para iniciar o processo de agregação.
- Modelo de metadados: especificação formal do modelo de dados, no qual estão listadas as classes e propriedades dos metadados.
- Contrato de troca de dados: documento que rege a parceria entre instituição agregadora e provedor de dados.
- Licenças de uso: rol de licenças de uso pré-selecionadas pelas instituições agregadoras.
- Uso dos metadados: definição dos elementos mínimos de metadados para agregação.
- Critérios técnicos para qualidade de dados: definição de critérios de qualidade para o conteúdo do metadado e/ou mídia digital.
- Cronograma de coleta: especificação de como são estruturadas as coletas que visam atualizar os agregadores.
- Validação e publicação de dados: protótipo dos dados agregados, que devem ser validados pelo provedor de dados para publicação.

A pesquisa *“Elements for Constructing a Data Quality Policy to Aggregate Digital Cultural Collections: Cases of the Digital Public Library of America and Europeana Foundation”* foi publicada em 2021, pelo *“Institute for Computer Sciences, Social Informatics and Telecommunications Engineering”* (SIQUEIRA et al, 2021).

3.5 Revisão Sistemática de Literatura

A RSL é uma forma de estudo secundário que revisa estudos primários relacionados a questões de pesquisa, utilizando uma metodologia bem definida, a fim de identificar, analisar e interpretar as evidências de forma imparcial e replicável (KITCHENHAM; CHARTERS, 2007). Esta RSL segue as diretrizes propostas por Kitchenham (2004), que define e descreve três etapas: Planejamento, Execução e Resultados.

3.5.1 Planejamento

O planejamento contempla a identificação da necessidade de revisão, a elaboração das questões de pesquisa, a escolha das bases de dados, bem como a criação da *string* de busca e a elaboração dos critérios de inclusão e exclusão.

3.5.1.1 Identificação da necessidade de revisão

A revisão visa fornecer subsídios técnicos e conceituais para pesquisadores envolvidos no esforço brasileiro de construção de uma ferramenta de agregação de dados de objetos digitais na área de cultura. Espera-se que os resultados apresentem como esse desafio é enfrentado por outras iniciativas e facilite a produção de uma solução local.

3.5.1.2 Questões de pesquisa

O RSL visa responder a cinco questões:

- Q01. Quais são os projetos?
- Q02. Quais são as etapas planejadas?
 - Contemplando um breve histórico do projeto; o *workflow* de agregação e o detalhamento das etapas previstas no *workflow*.
- Q03. Quais tecnologias são utilizadas?
- Q04. As etapas são executadas de forma manual, automática ou semiautomática?
- Q05. Quais realizam busca semântica?

3.5.1.3 Base de dados

Foram selecionadas três importantes bases de dados: a *Networked Digital Library of Theses and Dissertations* (NDLTD)¹⁶, uma organização internacional para a divulgação de teses e dissertações e a *Scopus*¹⁷ e a *Web of Science* (WoS)¹⁸, por sua reconhecida quantidade e variedade de publicações e áreas do conhecimento.

Destacamos que, embora a WoS e a Scopus sejam as maiores bases multidisciplinares do mundo, existem algumas limitações, como centralizar as áreas de Ciências Exatas e Biológicas, com abrangência restrita para as Ciências Humanas

¹⁶ <http://www.ndltd.org/>. Acesso em 11 jan. 2023

¹⁷ <https://www.scopus.com/>. Acesso em 11 jan. 2023

¹⁸ <https://www.webofknowledge.com/>. Acesso em 11 jan. 2023

e Sociais, com foco em publicações feitas por países como Estados Unidos, Inglaterra e Holanda, com menor representação para os países da América Latina e África; e priorizando publicações em língua inglesa (LARIVIÈRE e SUGIMOTO, 2018; MARTÍN-MARTÍN et. al, 2018).

3.5.1.4 *String* de busca

A definição da *string* de busca trouxe um desafio adicional, pois o assunto é relativamente novo, além de ser bastante polissêmico devido ao seu alto grau de interdisciplinaridade. Portanto, não há consenso sobre qual(is) termo(s) melhor o(s) representa(m).

Assim, num primeiro momento, exploramos o tema “serviços de descoberta em escala na web” (SIQUEIRA e MARTINS, 2019b); na sequência, entendendo que o termo não atendia à busca por agregadores de dados, passamos a identificar, na literatura cinzenta, grandes projetos para agregação de dados (SIQUEIRA e MARTINS, 2019a, 2020).

Ambos os estudos se mostraram essenciais para compreendermos e ampliarmos os termos de busca, resultando na *string* que segue devidamente formatada de acordo com as especificidades do buscador de cada base de dados. No NDLTD, realizamos a busca em todos os campos disponíveis, enquanto no Scopus e WoS realizamos a busca nos campos título, resumo e palavras-chave.

- NDLTD: ((*aggregat* OR integration*) AND (*infrastructure OR workflow OR flux OR flow*) AND (*metadata*) NOT (“*real time*” OR *real-time* OR *realtime* OR “*big data*”))
- Scopus: *TITLE-ABS-KEY* ((*aggregat* OR integration*) AND (*infrastructure OR workflow OR flux OR flow*) AND (*metadata*) AND NOT (“*real time*” OR *real-time* OR *realtime* OR “*big data*”))
- WoS: *ts =* ((*aggregat* OR integration*) AND (*infrastructure OR workflow OR flux OR flow*) AND (*metadata*) NO (“*real time*” OR *real-time* OR *realtime* OR “*big data*”))

3.5.1.5 Critérios de seleção

Os critérios de seleção foram divididos em critérios de inclusão (CI) e critérios de exclusão (CE), conforme segue.

3.5.1.5.1 Critérios de inclusão

- CI 01. Línguas portuguesa, inglesa ou espanhola.
- CI 02. Período de 2010 a 2020.
- CI 03. Soluções para a agregação de dados com foco no patrimônio cultural.
- CI 04. Fluxo de trabalho de agregação.

3.5.1.5.2 Critérios de exclusão

- CE 01. Publicações duplicadas.
- CE 02. Publicações inacessíveis.
- CE 03. Publicações fora do escopo.
 - CE 03.1. Relacionado ao processamento de dados em tempo real.
 - CE 03.2. Relacionado à computação ubíqua.

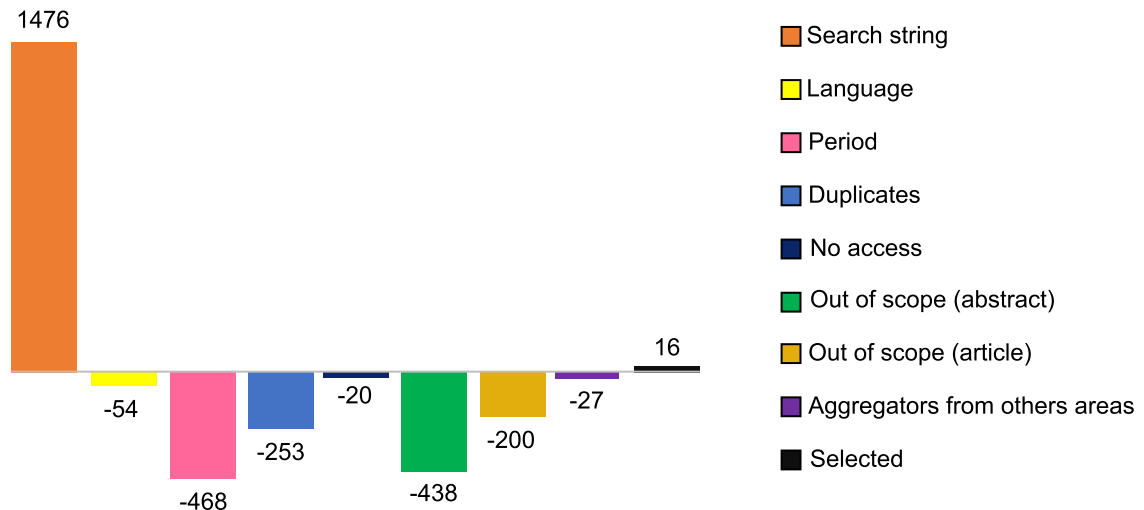
Para explicar alguns dos critérios escolhidos, a CI 04 visava a publicações que apresentassem o fluxo de trabalho, pois esse é passível de replicação. Assim, ele estaria mais propenso a oferecer respostas para as questões de pesquisa. Na CE02, excluimos as publicações indisponíveis nas bases de dados às quais temos acesso, considerando também aquelas pagas pela Universidade de Brasília. No CE03 excluimos publicações sem relação com o tema proposto. Também foram desconsideradas pesquisas que apresentavam soluções voltadas ao processamento de dados em tempo real (CE03.1), já que para o patrimônio cultural a inclusão de dados não requer esse tipo de abordagem, e também foram excluídas as relacionadas à computação ubíqua (CE03.2) que, embora tratem de agregação de dados, têm finalidades específicas, além do nosso interesse.

3.5.2 Execução

A execução da *string* de busca foi realizada em 15 de maio de 2020, retornando um total de 1.476 publicações, sendo 72 no ND LTD, 939 no Scopus e 465 no WoS. Os critérios de seleção foram aplicados da seguinte forma: os critérios de inclusão CI 01 e CI 02 foram aplicados com os filtros das próprias ferramentas de busca. Foram realizados os critérios de inclusão CI 03, CI 04 e os critérios de exclusão CE 03, CE 03.1 e CE 03.2 em duas etapas, sendo a primeira a leitura dos resumos e, em caso

de dúvida, a leitura do documento completo. O Gráfico 1 mostra os números relativos à execução da *string* e a redução do número de publicações a cada critério aplicado

Gráfico 1. Números referentes à execução da *string* de pesquisa da RSL



Fonte: Siqueira; Martins (2021c, p. 4)

Os 16 documentos selecionados subsidiam as respostas às questões de pesquisa. O resultado dessa etapa da pesquisa foi publicado em 2021, na revista *Journal of the Association for Information Science and Technology* (JASIST) (SIQUEIRA; MARTINS, 2021b).

3.6 Versões dos softwares

Para o correto funcionamento da solução, torna-se necessário o uso de versões mínimas dos softwares:

- Tainacan dos museus de origem: versão 0.15 ou superior com o plugin de mapeamento de metadados para o INBCN instalado.
- Elastic Stack: versão 7.9 (não foram testadas versões superiores)
- Para o Tainacan do agregador sugere-se a versão mais atual do software.

4 RESULTADOS

Os resultados apresentam o detalhamento das ações executadas para alcançar o objetivo geral da pesquisa.

4.1. Revisão Bibliográfica

Foram elencados oito projetos de agregação de dados, listados no Quadro 22. Em alguns projetos não foi encontrado o workflow, mas sim sua arquitetura, que, por mostrar etapas, foi considerada neste estudo.

Quadro 22. Relação de instituições, com seu respectivo país de origem e nome do projeto de agregação de dados

N.	Instituição	País	Projeto
01	American Art Collaborative	EUA	AAC
02	Biblioteca Nacional	Austrália	Trove
03	Biblioteca Nacional	Nova Zelândia	DigitalNZ
04	Fundação Europeia	União Europeia	Europeana
05	Instituto de Ciência e Tecnologia da Informação	Itália	D-NET Software
06	Instituto de Ciência e Tecnologia da Informação	Itália	Parthenos Aggregator
07	Secretaria de Cultura	México	Repositório Mexicana
08	Universidade de Nevada	EUA	UNLV's Linked Data Project

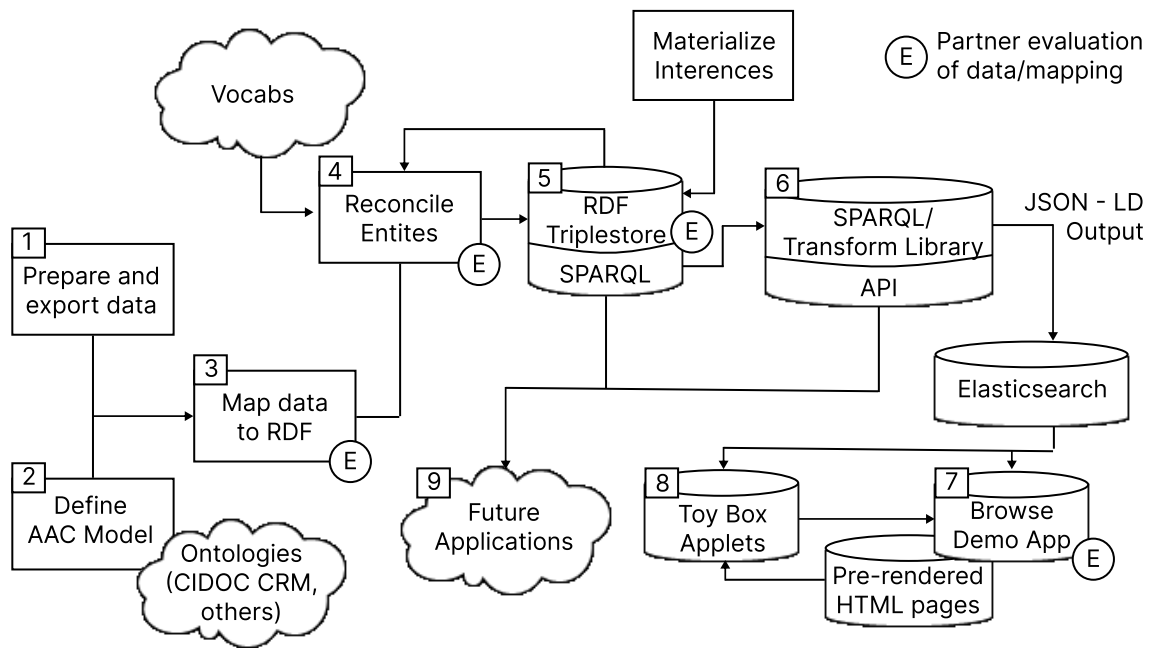
Fonte: Siqueira; Martins (2020, p. 100)

A seguir os workflows e suas descrições.

4.1.1 American Art Collaborative

A *American Art Collaborative* (AAC) é um consórcio de 14 instituições de arte que visam a investigar e construir uma massa crítica de *Linked Open Data* (LOD). Para Fink (2018), LOD se trata de um método para publicar dados estruturados na web de forma que as informações sejam interconectadas e, assim, tornadas amplamente úteis. A Figura 7 apresenta o workflow proposto pela AAC.

Figura 7. Workflow de agregação de dados da *American Art Collaborative*



Fonte: Fink (2018, p. 32), adaptada.

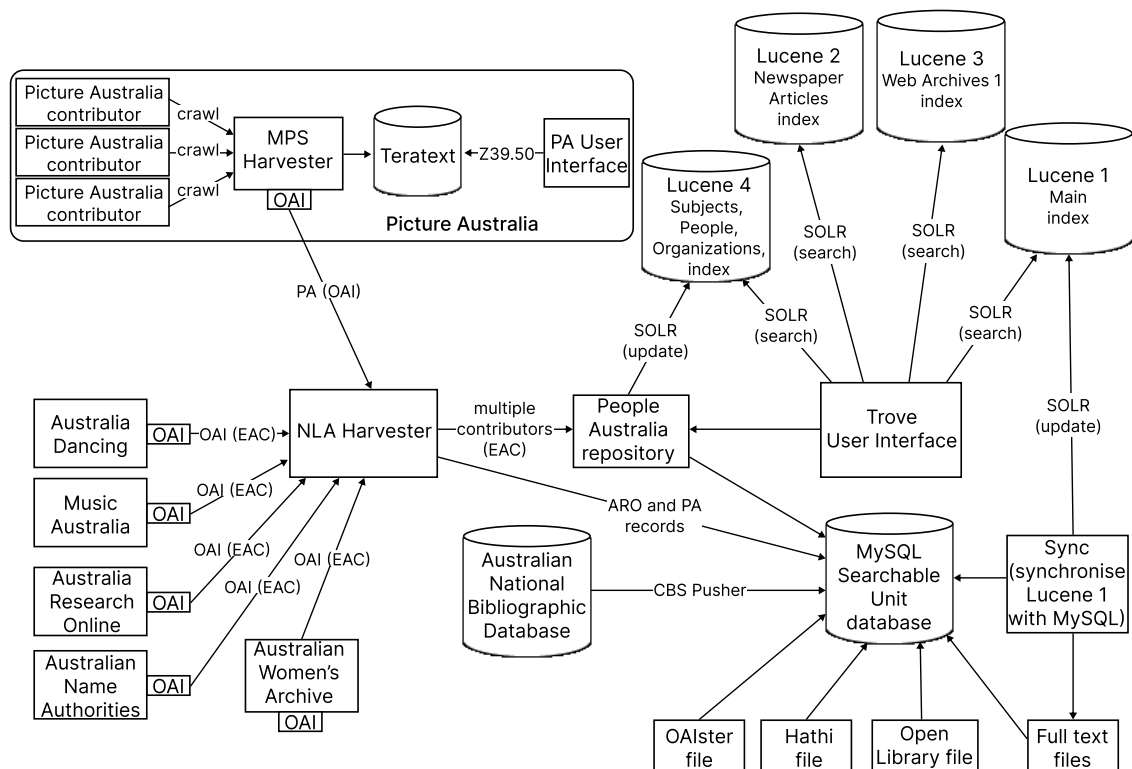
O workflow prevê nove etapas. A Etapa 1. “*Prepare and export data*” visa a fornecer dados principais e adicionais, considerados úteis pelos parceiros, que, na prática, exportam dados brutos de seus sistemas de origem e os armazenam em um repositório GitHub compartilhado. A etapa 2. “*Define AAC Model*” trata-se de um conjunto geral de orientações sobre ontologias que podem ser adotadas e reutilizadas, tendo por objetivo constituir um modelo conceitual único para agregar os dados das diferentes instituições. A etapa 3. “*Map data to Resource Description Framework (RDF)*” visa a mapear os dados das instituições parceiras para um modelo de destino. Assim, usando o modelo de destino e a ferramenta de integração de dados Karma, os dados de cada parceiro são convertidos em RDF. A etapa 4. “*Reconcilie Entities*” visa a mapear entidades individuais para IDs comuns, considerando, sempre que possível, vocabulários públicos padronizados. A etapa 5. “*RDF Triple Store - SPARQL*” visa a armazenar e fornecer acesso aos dados vinculados enriquecidos pela reconciliação de entidades e permite consultas *SPARQL Protocol and RDF Query Language - SPARQL*. A etapa 6. “*SPARQL/Transform library – API*” visa a compilar um conjunto

de consultas claras, reutilizáveis e focadas em entidades que navegam no gráfico de dados vinculados para fornecer documentos JSON-LD a desenvolvedores e humanistas digitais. A etapa 7. “*Browse Demo App*” visa a apresentar uma ilustração inicial suficientemente rica de dados vinculados para os usuários. A etapa 8. “*Toy Box Applets*” visa a ampliar as possibilidades e ideias do AAC, por meio de um aplicativo de navegação. A etapa 9. “*Future Applications*” visa a continuar a explorar casos de uso e aplicativos para dados vinculados por meio de contribuições de dados de parceiros (FINK, 2018).

4.1.2 Trove

A Biblioteca Nacional da Austrália desenvolveu o Trove, que objetiva disponibilizar recursos culturais relacionados à Austrália. O Trove oferece um mecanismo de busca integrada em bibliotecas, museus, arquivos e outras organizações de pesquisa, além de um conjunto de serviços (TROVE, 2023). A Figura 8 apresenta o workflow proposto.

Figura 8. Workflow de agregação de dados da Trove



Fonte: National Library of Australia (2010), adaptada.

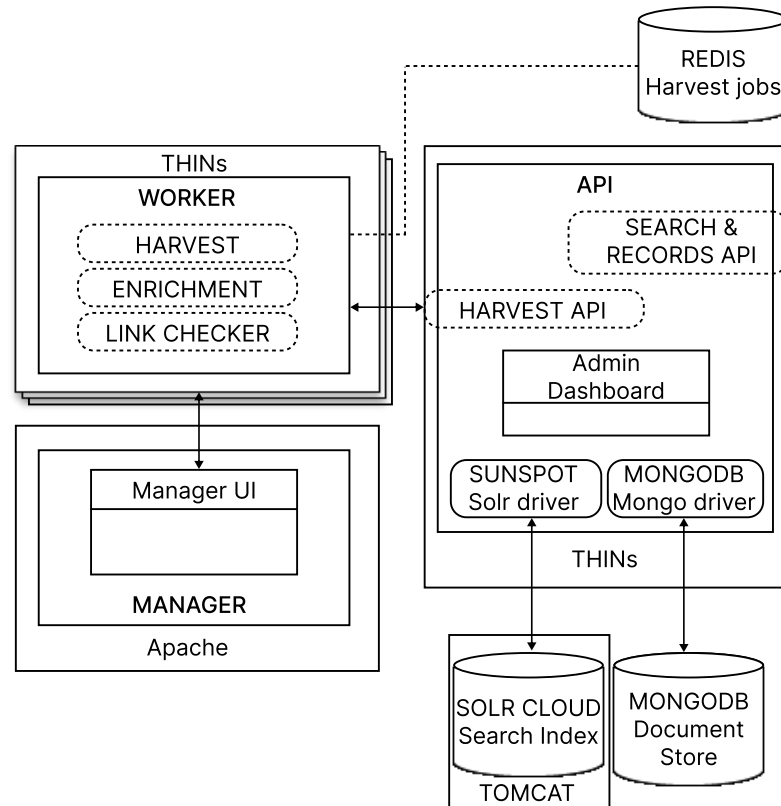
O Trove mantém um site com informações relevantes do projeto, no entanto, ainda que apresente o workflow de agregação, não traz, nas fontes pesquisadas, documentação que explique cada etapa do workflow. Contudo, na análise realizada pela pesquisa, percebe-se um grande uso do protocolo OAI-PMH para coleta de dados dos provedores, bem como a criação de diferentes índices para indexação ágil e recuperação da informação utilizando a tecnologia Apache Lucene¹⁹, que é um software voltado para busca e indexação de documentos de alta escalabilidade e aplicado em projetos que exigem processamento de dados massivos. A arquitetura se vale de diversas camadas, mas devido à falta de documentação explícita, somente se pode inferir como as camadas se relacionam, sem condições de uma análise crítica da solução para eventual replicação.

4.1.3 DigitalNZ

A Biblioteca Nacional da Nova Zelândia junto à Rede do povo Aotearoa Kaharoa desenvolveu, no início de 2006, o DigitalNZ, que utiliza o software Supplejack para agregação de dados (DIGITAL NEW ZEALAND, 2023). A Figura 9 apresenta a arquitetura da plataforma Supplejack.

¹⁹ <https://lucene.apache.org/>. Acesso em 11 jan. 2023

Figura 9. Arquitetura da plataforma Supplejack



Fonte: Supplejack (2023), adaptada.

A arquitetura é composta por: “*Manager*” que apresenta uma interface para o usuário controlar as atividades do software; “*Worker*” que realiza as atividades de coleta, enriquecimento e verificação de links; “*API*”, um *wrapper* público para pesquisar o repositório de índice e metadados e “*Common*” que são ajudantes compartilhados entre o “*Worker*” e o “*Manager*” (SUPPLEJACK, 2023).

Como apresentado na Figura 3, o Supplejack depende da integração com um índice de pesquisa, o padrão é Solr e um repositório de metadados que utiliza o MongoDB. O Apache Solr é uma tecnologia voltada à pesquisa e à indexação de documentos massivos, e o MongoDB, um banco de dados NoSQL²⁰, também utilizado em projetos contemporâneos que envolvem novas arquiteturas para processamento

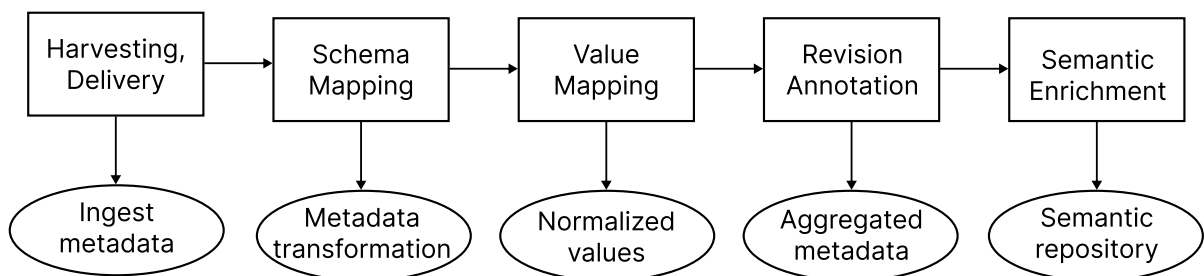
²⁰ NoSql é um termo genérico que representa os bancos de dados não relacionais. Trata-se de bancos de dados que fornecem mecanismos para armazenamento e recuperação de dados, mas que são modelados de formas diferentes das relações tabulares usadas nos bancos de dados relacionais.

de dados massivos baseados em informação semiestruturada ou mesmo desestruturada.

4.1.4 Europeana

A Fundação Europeia desenvolveu a Europeana, que reuniu mais de 55 milhões de objetos digitais das coleções on-line de mais de 3.500 galerias, bibliotecas, museus, coleções audiovisuais e arquivos de toda a Europa (KOLLIA et al., 2012). A Figura 10 apresenta seu workflow de agregação.

Figura 10. Workflow de agregação de dados da Europeana



Fonte: Kollia et al. (2012, p. 70), adaptada.

A primeira etapa, “*Harvesting, Delivery*”, refere-se à coleta de metadados de provedores de conteúdo, por meio de protocolos de entrega, como o OAI-PMH, o *Hypertext Transfer Protocol* (HTTP) e o *File Transfer Protocol* (FTP). A segunda etapa, “*Schema Mapping*”, alinha os metadados coletados a um modelo de referência comum. Nesta etapa, uma interface gráfica colabora com o mapeamento, por meio de uma linguagem de mapeamento compreensível por máquinas. A terceira etapa, “*Value Mapping*”, foca no alinhamento e na transformação dos termos constantes nos metadados coletados para arquivo de autoridade ou fonte externa, ou seja, permite a normalização de datas, localizações, países, idiomas, dentre outros. A quarta etapa, “*Revision, Annotation*”, permite adicionar anotações para atribuir metadados não disponíveis no contexto original, e, por último, a quinta etapa, “*Semantic Enrichment*”, foca na transformação dos dados em um modelo semântico, extração e identificação de recursos e implantação em RDF (KOLLIA et al., 2012).

O artigo não detalha os recursos tecnológicos utilizados, não deixando claro quais ferramentas e tecnologias são utilizadas em cada etapa. Novamente, há

dificuldade de se encontrar evidências que facilitem a replicação ou mesmo a adaptação de soluções em outros contextos.

A Europeia, por agregar inúmeros agregadores menores, tem soluções implantadas em seis projetos encontrados na RSL (SIQUEIRA; MARTINS, 2021b), descrita no item 5.2 Revisão Sistemática de Literatura - RSL.

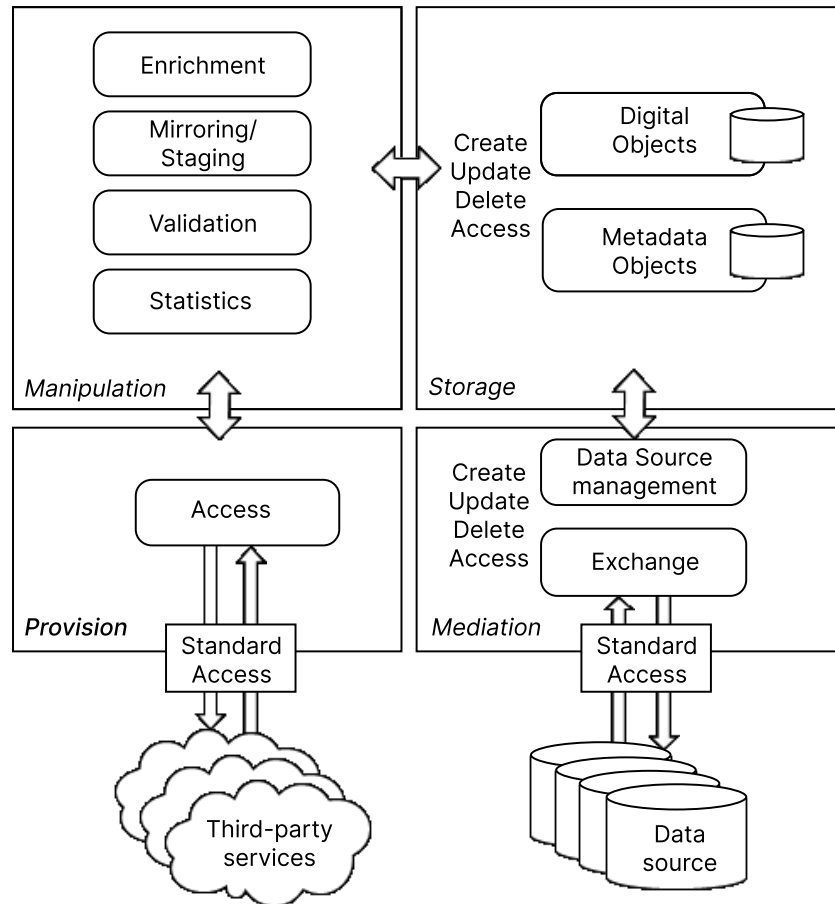
4.1.5 D-NET

O *Istituto di Scienza e Tecnologie dell'Informazione* desenvolveu o D-NET, uma estrutura orientada a serviços, de uso geral, na qual os designers podem construir infraestruturas agregadas autônomas, robustas, escaláveis e personalizadas, de maneira econômica.

Oferece serviços de gerenciamento de dados capazes de fornecer acesso a diferentes tipos de fontes de dados externas, armazenar e processar objetos de informações de qualquer modelo de dados, convertê-los em formatos comuns e expor objetos de informações a aplicativos de terceiros por meio de vários acessos padrão (MANGHI et al., 2014).

Neste estudo, categorizamos D-NET e o Supplejack em um mesmo nicho, por serem ambos softwares que têm por objetivo fornecer um serviço completo para agregação de dados. A Figura 11 apresenta a arquitetura do software.

Figura 11. Arquitetura do D-NET Software Toolkit



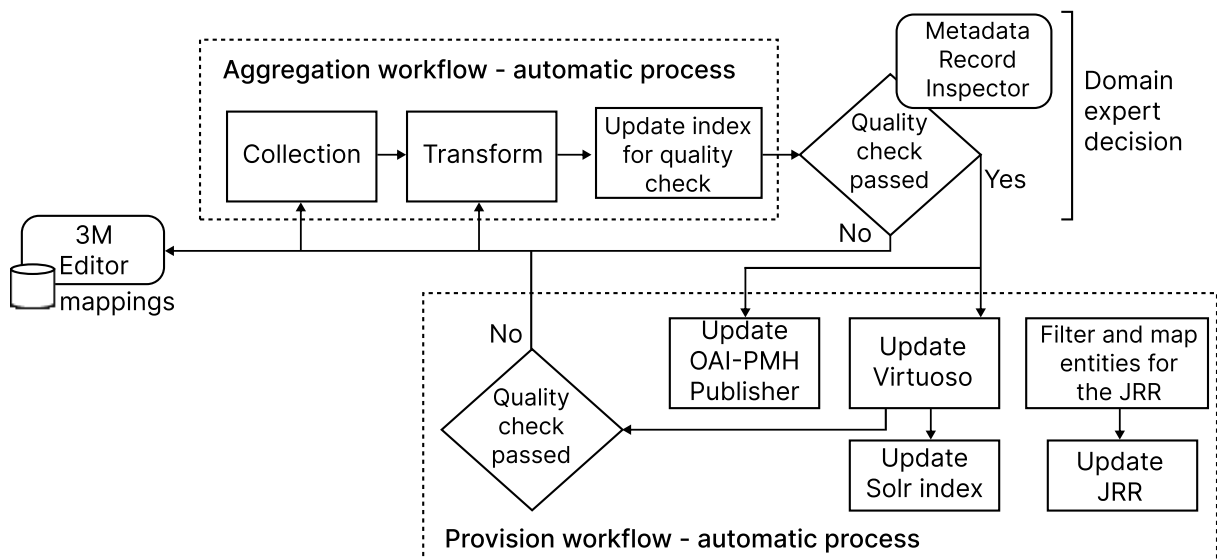
Fonte: Manghi et al. (2014, p. 327), adaptada.

O D-NET subdivide sua arquitetura em quatro camadas principais: *Manipulation*, *Storage*, *Provision* e *Mediation*. Os serviços de manipulação, “*Manipulation*”, foram projetados para executar o enriquecimento, a validação, o espelhamento e a geração de estatísticas. Os serviços de armazenamento, “*Storage*”, propiciam o armazenamento de objetos, agrupando tecnologias conhecidas, de código-fonte aberto (índices de texto completo, bancos de dados relacionais, repositórios de documentos etc). Os serviços de fornecimento de dados, “*Provision*”, fazem interface com aplicativos externos, como, por exemplo, portais para uso dos usuários finais ou serviços de terceiros. Os serviços de mediação, “*Mediation*”, visam a buscar dados de fontes externas e importá-los para a infraestrutura agregada, como, por exemplo, objetos em conformidade com um determinado recurso de modelo de dados (BARDI; MANGHI e ZOPPI, 2012).

4.1.6 Parthenos Aggregator

As infraestruturas de humanidades digitais (*Digital Humanities Infrastructures* - DHIs) apoiam pesquisadores no campo das ciências humanas, oferecendo-lhes um ambiente digital no qual podem encontrar e usar ferramentas e dados de pesquisa para conduzir suas atividades. Há um número crescente de DHIs e, para integrá-los, a Comissão Europeia lançou o projeto *Pooling Activities, Resources and Tools for Heritage E-research Networking, Optimization and Synergies* - PARTHENOS (FROSINI et al., 2018). A Figura 12 apresenta o workflow proposto.

Figura 12. Workflow de agregação de dados do *Parthenos Aggregator*



Fonte: Frosini et al. (2018, p. 40), adaptada.

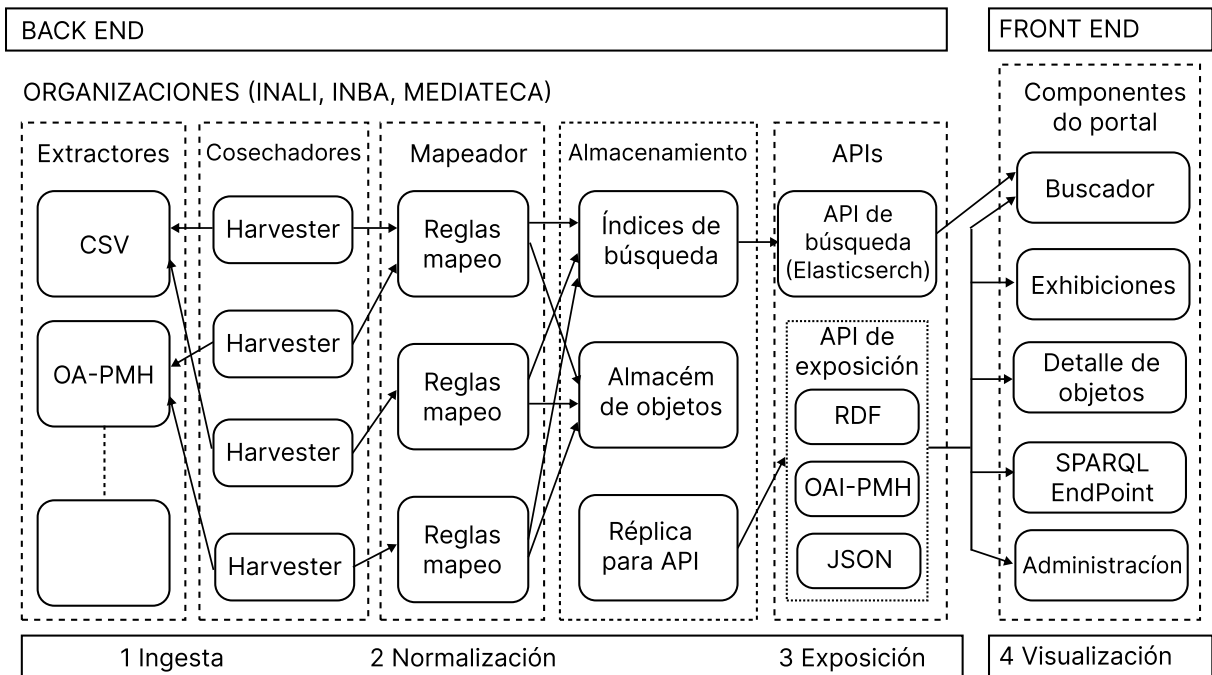
O workflow se divide em dois fluxos de trabalho: “*aggregation workflow*” e “*provision workflow*”. No fluxo de agregação (*aggregation workflow*), a etapa “*Collection*” visa a lidar com a coleta de metadados por meio de diferentes protocolos de acesso: OAI-PMH, FTP, *Secure File Transfer Protocol* (SFTP), HTTP, *Representational State Transfer* - RESTful. A etapa “*Transformation*” visa a mapear os metadados a uma ontologia única. A “*Metadata Cleaner*” trata-se do serviço que harmoniza valores em registros de metadados com base em um conjunto de tesouros. Após esta etapa, inicia-se o processo de inspeção, na etapa “*Metadata Record Inspector*”, na qual uma *Graphical User Interface* (GUI), da Web integrada ao D-NET (MANGHI et al., 2014), fornece dados aos curadores com uma visão geral das

informações, possibilitando pesquisas e navegação entre os registros para verificar a correção da fase de transformação (por exemplo, metadados sem mapeamento, erros ou inconsistências semânticas) e a fase de limpeza. Uma vez positivamente verificado, os registros podem ser exportados publicamente na “*OAI-PMH publisher service*”. O serviço oferece Interfaces OAI-PMH para aplicativos de terceiros que desejam acessar metadados. “*Index service*” orienta a alimentação de Índices Solr e é responsável por transformar os registros de metadados agregados em documentos Solr. O JRR (*Joint Resource Registry*) é o Registro Conjunto de Recursos, um componente PARTHENOS concebido para apoiar a gestão dos recursos integrados (FROSINI et al., 2018).

4.1.7 Mexicana

A Secretaria de Cultura do México desenvolveu a Mexicana, um Repositório do Patrimônio Cultural do México, livre e aberto, que tem o objetivo principal de difundir e vincular os acervos do patrimônio cultural do México (MÉXICO, 2018). Seu workflow está apresentado na Figura 13.

Figura 13. Workflow de agregação de dados da Mexicana



Fonte: México (2018), adaptada

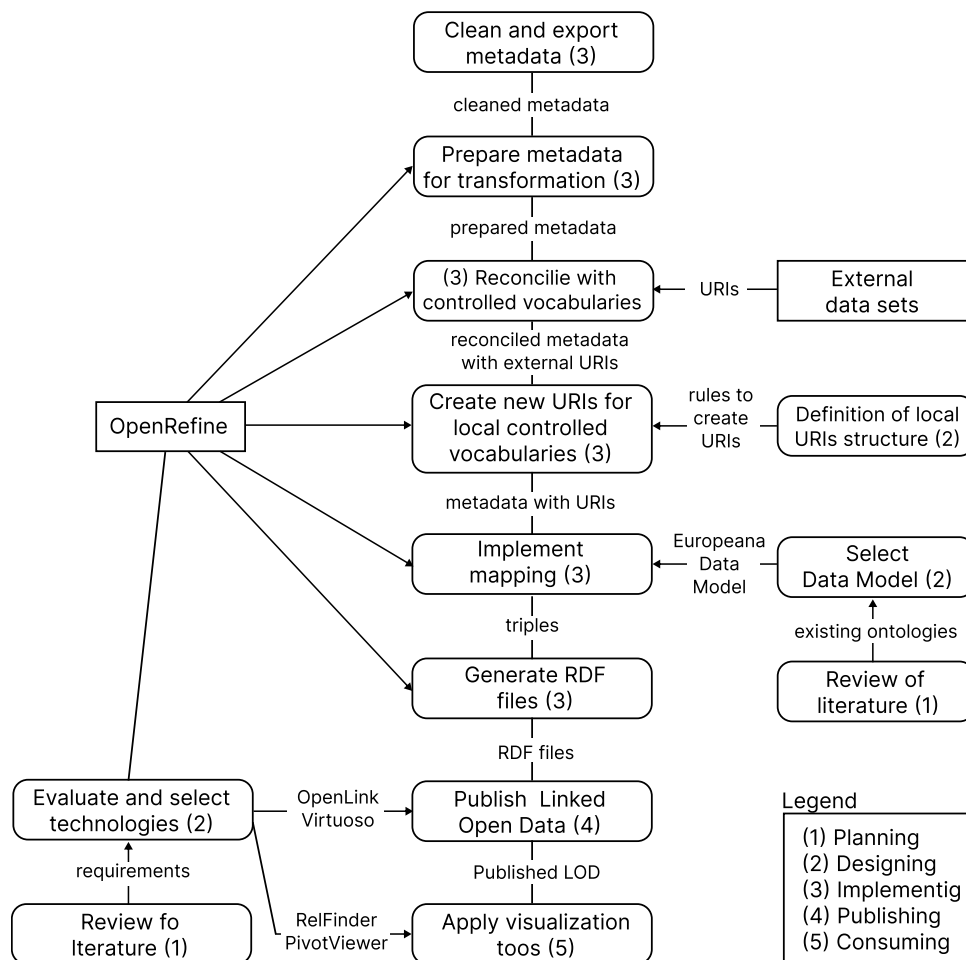
A Secretaria de Cultura do México desenvolveu documento explicativo sobre o projeto Mexicana, no qual consta o workflow apresentado, que é dividido em *Back End* e *Front End*. No *Back End*, a Etapa 1. “*Extractores*” trata dos componentes de código responsáveis pela cópia de dados locais ou remotos, realizando uma reestruturação mínima dos dados. Nesta etapa, considera-se a criação de extratores para diferentes formatos e a criação de interfaces para facilitar a extensão da funcionalidade. A Etapa 2. “*Cosechadores*” trata de componente de código configurável que permite gerenciar os extratores e o mapeador dos dados (próxima etapa), de acordo com regras estabelecidas, tais como: execução sob demanda e por periodicidade. A Etapa 3. “*Mapeador*” permite configurar e executar regras de mapeamento definidas entre os dados gerados pelos extratores e o esquema de dados unificados do sistema. A Etapa 4. “*Almacenamiento*”, como o próprio nome diz, visa a armazenar os dados coletados. Nessa etapa, no workflow, há dois itens além do armazenamento: o “*Índices de búsqueda*” no qual o componente explora os serviços do ElasticSearch para gerar índices de pesquisa de metadados de objetos incorporados ao armazenamento do sistema e a “*Réplica para API*”, que replica os objetos digitais para fornecer seu acesso rápido através do APIs de exposição. A Etapa 5. “*API*” reforça o uso servidor ElasticSearch para indexação e recuperação de metadados, assim como a “*API de Búsqueda*” trata sobre formatos para exibição de objetos digitais e seus metadados (MÉXICO, 2018).

O *Front end*, etapa voltada à visualização dos dados, é composto por: “*Buscador*”, que trata sobre pesquisa desagregada de objetos digitais inseridos no sistema através da exploração da API; “*Exhibiciones*”, que permite a configuração de coleções de objetos digitais agrupados por tema ou evento específico; “*Detalle de objetos*”, que permite visualizar o arquivo com os detalhes dos objetos digitais inseridos no sistema e no metadados associados; o “*EndPoint SPARQL*”, que permite a execução de consultas SPARQL para o modelo de dados do sistema unificado através da estrutura semântica definida pelo Modelo de Dados e, por fim, “*Administración*”, que permite a administração de usuário e fluxo de trabalho para a configuração e execução das coletas para fontes de dados (MÉXICO, 2018).

4.1.8 UNLV'S Linked Data Project

A Universidade de Nevada (UNLV), por meio da equipe do departamento de Coleções Digitais das Bibliotecas da Universidade, reuniu esforços para encontrar maneiras de tornar mais eficiente a descoberta e o uso das informações, iniciando, assim, estudos para adoção do LOD, culminando no desenvolvimento do Projeto UNLV's Linked Data (SOUTHWICK, 2015). A Figura 14 apresenta o workflow.

Figura 14. Workflow de agregação de dados do Projeto UNLV's Linked Data



Fonte: Southwick (2015, p. 13), adaptada.

Southwick (2015) afirma que optou pela adoção de tecnologias com código aberto, sem qualquer adaptação ou desenvolvimento, ou seja, “no estado em que se encontram”.

O projeto foi dividido em cinco etapas: “*Planning*”, “*Designing*”, “*Implementing*”, “*Publishing LOD*” e “*Consuming LOD*”. A etapa “*Planning*” é composta

por duas revisões de literatura e retrata o período de estudo necessário ao desenvolvimento do protótipo (SOUTHWICK, 2015).

A segunda etapa, *“Designing”*, se desdobra em três atividades: *“Evaluate e select technologies”*, na qual a equipe do projeto avaliou várias tecnologias e selecionou seis para aplicação no protótipo. A autora destaca que, embora as tecnologias selecionadas tenham funcionado bem, não significa que são as únicas ou as melhores; *“Select data model”*, que diz respeito à seleção da ontologia utilizada, a partir da investigação de modelos de dados utilizados por outras instituições; e *“Definition of local URIs structure”*, fase na qual optou-se por criar URIs apenas para “coisas” que ainda não receberam URIs de outros provedores de dados (SOUTHWICK, 2015).

A terceira fase, *“Implementing”*, é composta por seis etapas: *“Clean and export metadata”*, na qual o CONTENTdm é utilizado como sistema de gerenciamento de conteúdo e este exporta metadados em formato de planilha delimitada por tabulação, que pode ser importada para o OpenRefine, cuja limpeza por ele executada consiste em cumprir rigorosamente os termos usados nas coleções que adotam um determinado vocabulário controlado e criar vocabulários controlados locais consistentes; *“Prepare metadata for transformation”*, fase de preparação dos metadados para gerar LOD. Para tal, utilizaram-se funções do OpenRefine, como: remover espaços em branco, separar tipos diferentes de dados e separar valores agrupados em um único campo; *“Reconcilie with controlled vocabularies”* que trata da reconciliação feita usando a extensão *“OpenRefine Resource Description Framework”*, do OpenRefine e; *“Generate RDF files”* o qual trata da geração de arquivos RDF que serão utilizados nas próximas etapas. As etapas *“Create new URIs for local controlled vocabularies”* e *“Implement mapping”* são a efetivação das etapas *“Definition of local URIs structure”* e *“Select data model”*, da fase *“Designing”* (SOUTHWICK, 2015).

A quarta fase, a *“Publishing LOD”*, é composta por uma única atividade, *“Publish Linked Open Data”*, na qual o arquivo RDF é publicado à comunidade. A última fase, *“Consuming LOD”*, também é composta por uma única atividade, *“Apply visualization tools”*, na qual foram realizados três experimentos com ferramentas de visualização para arquivos RDF: com Pivot Viewer, útil para visualizar imagens de maneira dinâmica, pois é baseado em consultas SPARQL; com RelFinder, que visa a

encontrar relacionamentos entre as “coisas” e também com ReIFinder, mas com o conceito de relacionamento expandido, considerando relacionamentos que duas “coisas” tinham uma ou mais “coisas” em comum (SOUTHWICK, 2015).

Os portais de busca integrada estão disponíveis para acesso e o Quadro 23 apresenta o link para cada uma delas, com exceção do D-NET que é um software para agregação.

Quadro 23. Portais de busca integrada de agregadores internacionais

Projeto	Sites
AAC	https://americanart.si.edu/search
Trove	https://trove.nla.gov.au/
DigitalNZ	https://digitalnz.org/
Europeana	https://www.europeana.eu/pt/collections
D-NET Software	-
Parthenos Aggregator	http://www.parthenos-project.eu/portal
Repositório Mexicana	https://mexicana.cultura.gob.mx/
UNLV's Linked Data Project	https://www.library.unlv.edu/linked-data

Fonte: Siqueira; Martins (2020, p. 110)

Considerando a análise dos workflows (SIQUEIRA; MARTINS, 2019a, 2020) elencaram-se oito fases para agregação, sendo elas: extração, estruturação, transformação, reconciliação, armazenamento, exposição, publicação e novas aplicações. De forma sintética, estas etapas significam:

- Extrair: extrair os dados em sua forma bruta, que podem estar, por exemplo, em PDF, em planilhas eletrônicas, documentos de texto, XML, em bancos de dados relacionais, dentre outras opções.
- Estruturar: definir um padrão de metadados, selecionar vocabulários controlados pré-existentes e ontologias para aplicação nos dados.
- Transformar: realizar a normalização, limpeza e correção sintática dos dados.
- Reconciliar: enriquecer os metadados por meio de outros dados existentes na web.
- Armazenar: selecionar onde os dados coletados serão armazenados.
- Publicar: disponibilizar a interface única de busca integrada.

- Expor: disponibilizar os dados agregados por meio de API, que os exponham em formato RDF, OAI-PMH ou JSON.
- Novas aplicações: permitir, a partir dos arquivos disponibilizados na etapa 'Expor', que novas aplicações possam ser criadas.

O Quadro 24 mostra quais etapas foram encontradas e quais projetos contemplam quais etapas, considerando a presença (X) ou não (-) de cada uma delas, para visualização geral.

Quadro 24. Sumarização das etapas dos workflows de agregação de dados

Projeto/Etapas	Extrair	Estruturar	Transformar	Reconciliar	Armacenar	Publicar	Expor	Novas aplicações
AAC	X	X	-	X	X	X	X	X
Trove ¹	X	-	-	-	X	X	-	-
DigitalNZ	X	-	-	-	X	X	-	-
Europeana ²	X	X	-	X	X ²	X ²	-	-
D-NET Software	X	X	X	X	X	X	X	X
Parthenos Aggregator	X	X	X	X	X	X	X	X
Repositório Mexicana	X	X	-	X	X	X	X	-
UNLV's Linked Data Project	X	X	X	X	X	X	X	X
¹ Dados observados somente a partir da visualização dos workflows. ² Itens não explícitos no workflow, mas identificados na documentação.								

Fonte: Siqueira; Martins (2020, p. 111)

O Quadro 25 apresenta a nomenclatura original das etapas na literatura revisada, classificando-as dentro das etapas elencadas neste estudo. Este quadro também nos ajuda a visualizar quais são os nomes mais usados para nomear cada etapa, para colaborar com pesquisas futuras.

Quadro 25. Sumarização das etapas dos workflows de agregação de dados utilizando a nomenclatura original das publicações

Projeto/ Etapas	Extrair	Estruturar	Transformar	Reconciliar	Armazenar	Publicar	Expor	Novas aplicações
AAC	<i>Prepare and export</i>	<i>Define AAC Model</i>	-	<i>Reconcilie Entities</i>	<i>RDF Triple Store - SPARQL</i>	<i>Browse Demo APP e Toy Box Applets</i>	<i>SPARQL/ Transform Library API</i>	<i>Future Applications</i>
Trove ¹	<i>NLA Harvest</i>	-	-	-	<i>MySQL Seachable Unit Database</i>	<i>Trove User Interface</i>	-	-
DigitalNZ	<i>Manager e Common</i>	-	-	-	<i>API</i>	<i>API</i>	-	-
Europeana	<i>Harvesting Delivery</i>	<i>Schema Mapping</i>	-	<i>Value Mapping</i>	<i>X ²</i>	<i>X ²</i>	-	-
D-NET Software	<i>Mediation</i>	<i>Manipulation</i>	<i>Manipulation</i>	<i>Manipulation</i>	<i>Storage</i>	<i>Provision</i>	<i>Provision</i>	<i>Provision</i>
Parthenos Aggregator	<i>Collection</i>	<i>Transformation</i>	<i>Metadata Record Inspector</i>	<i>Metadata Cleaner</i>	<i>Index Service</i>	<i>X ²</i>	<i>OAI-PMH publisher service</i>	<i>OAI-PMH publisher service</i>
Repositório Mexicana	<i>Extractores e Cosechadores</i>	<i>Mapeador</i>	-	-	<i>Almacenamiento</i>	<i>Buscador e Exhibiciones</i>	<i>API de Búsqueda</i>	-
UNLV's Linked Data Project	<i>Clean and export</i>	<i>Implement mapping</i>	<i>Prepare metadata for transformation</i>	<i>Reconcilie with controlled vocabularies e Create new URIs for local controlled vocabularies</i>	<i>Publish LOD</i>	<i>Publish LOD</i>	<i>Apply visualization tools</i>	<i>Publish LOD</i>
¹ Dados observados somente a partir da visualização do workflow								
² Item não explícito no workflow								

Fonte: Siqueira; Martins (2020, p. 113)

A documentação na qual os workflows estão inseridos apresenta alguns dados extras. Além disso, percebe-se pouca preocupação com a qualidade dos dados inseridos, ou seja, com os dados coletados na etapa de extração, havendo pouca menção a etapas tradicionais de projetos de análise de dados, envolvendo limpeza, tratamento e normalização de dados.

Além das etapas, as publicações apresentam algumas ferramentas de software utilizadas para execução do workflow. De forma geral, os workflows são genéricos demais e não apresentam o fluxo real de processos necessários, contrariando, assim, um dos princípios básicos de um workflow, que é a possibilidade de ser replicado. Além disso, percebe-se a necessidade de um conhecimento técnico avançado e extremamente especializado para compreensão de todas as etapas.

Após as revisões supracitadas (SIQUEIRA; MARTINS, 2019a, 2020), com objetivo de encontrar agregadores brasileiros, uma nova busca foi realizada, unicamente na literatura cinzenta, apontando três agregadores: o Brasiliana Fotográfico, o Brasiliana Iconográfico e o Rede Web Museus, que são descritos a seguir.

4.1.9 Brasiliana Fotográfica

O Brasiliana Fotográfica²¹ foi lançado em 2015, fruto da parceria entre a Biblioteca Nacional Digital e o Instituto Moreira Salles. Trata-se de um espaço que objetiva dar visibilidade e fomentar debates, abordando os acervos como fonte primária e como patrimônio digital a ser preservado (BRASILIANA FOTOGRAFICA, 2023b; MARCONDES; BETTENCOURT, 2019).

Atualmente integra mais de 9.000 itens de documentos fotográficos referentes ao Brasil, de onze instituições: o Arquivo Geral da Cidade do Rio de Janeiro, o Arquivo Nacional, a Diretoria do Patrimônio Histórico e Documentação da Marinha, a Casa de Oswaldo Cruz/Fiocruz, a Fundação da Biblioteca Nacional, a Fundação Joaquim Nabuco, o Instituto Moreira Salles, o *Leibniz-Institut fuer Laenderkunde*, o Museu Aeroespacial, o Museu da República e o Museu Histórico Nacional (BRASILIANA FOTOGRAFICA, 2022b, 2022c).

Desde o princípio do projeto, os idealizadores tinham a intenção de expandir seu uso a outras instituições, nacionais e internacionais, públicas ou privadas. Dessa forma, o que inicialmente contava com apenas duas instituições, a Biblioteca Nacional Digital e o Instituto Moreira Salles, no momento abarca onze, conforme supracitado (BRASILIANA FOTOGRAFICA, 2022b).

Em relação às tecnologias, o Brasiliana Fotográfica utiliza o WordPress, um software para criação de sites, blogs ou aplicativos (WORDPRESS, 2023), em conjunto com o DSpace²², um software para criação de repositórios digitais abertos (DSPACE, 2023), ambos de código aberto e gratuitos (DSPACE, 2023; WORDPRESS, 2023).

Sobre a coleta de dados, elas podem ser realizadas de três maneiras distintas: diretamente no DSpace, por meio de login e senha de acessos; por meio da

²¹ <https://brasilianafotografica.bn.gov.br/>. Acesso em 11 jan. 2023

²² <https://duraspace.org/dspace/>. Acesso em 11 jan. 2023

exportação dos dados nos formatos *Machine Readable Cataloging* - MARC/XML, ISO2709 ou planilha Excel e, por último; coleta de dados por meio do protocolo OAI-PMH (MARCONDES; BETTENCOURT, 2019).

Em relação aos modelos de dados, o padrão utilizado é o Dublin Core²³ visando garantir compatibilidade, interoperabilidade e vida longa ao projeto (HISTÓRIA CIÊNCIAS SAÚDE MANGUINHOS, 2015). No portal, há documentos técnicos que orientam os provedores de dados sobre a captura, o formato e o tratamento dos arquivos e sobre licença de uso. Não há disponível o workflow ou a descrição de como é realizada a agregação.

4.1.10 Brasileira Iconográfico

Lançado em 2017, o portal Brasileira Iconográfico²⁴ visa a reunir fontes iconográficas (desenhos, aquarelas, pinturas, gravuras e impressos) datadas do século XVI até início do século XX, dispersas por coleções públicas e privadas, nacionais e internacionais. O portal prevê que o espaço seja também destinado ao debate e à reflexão sobre o tema (MARCONDES; BETTENCOURT, 2019; BRASILIANA ICONOGRAFICA, 2017).

Atualmente o portal Brasileira Iconográfico integra mais de 2.500 itens iconográficos, de quatro importantes instituições brasileiras: a Fundação Biblioteca Nacional, o Instituto Moreira Salles, a Pinacoteca de São Paulo e o Instituto Itaú Cultural. O portal possui o propósito de se expandir para outras instituições, porém ainda se mantém com os provedores de dados iniciais. Prevê-se que, no futuro, o portal possa ser disponibilizado em outros idiomas, para abrigar contribuições vindas de diferentes partes do mundo (BRASILIANA ICONOGRAFICA, 2017).

Sobre as tecnologias, o portal foi desenvolvido a partir da integração do repositório DSpace com a plataforma Shiro²⁵ - um software de código aberto, sob a licença *Creative Commons Attribution-NonCommercial 4.0 International License* - que se trata de facilitador para gestão e potencialização da visualização de acervos que, com arquitetura baseada nos padrões da web 3.0, amplia as possibilidades de recuperação da informação e visualização do conteúdo (PLANO B, 2023a, 2023b).

²³ <https://www.dublincore.org/>. Acesso em 11 jan. 2023

²⁴ <https://www.brasilianaiconografica.art.br/>. Acesso em 11 jan. 2023

²⁵ <https://plano-b.com.br/shiro>. Acesso em 11 jan. 2023

A inclusão de dados pelos parceiros é feita de forma similar à adotada pela Brasileira Fotográfica, e utiliza o Dublin Core como padrão de dados (MARCONDES; BETTENCOURT, 2019).

O site não apresenta demais informações sobre a agregação dos dados, trazendo informações somente sobre autoria e recorte cronológico. Não há disponível o workflow ou a descrição de como é realizada a agregação.

4.1.11 Rede Web de Museus

A Rede Web de Museus do estado do Rio de Janeiro²⁶ é um projeto desenvolvido desde 2014, que visa a estabelecer uma política estadual integrada e colaborativa para os museus do Rio de Janeiro, facilitando o compartilhamento e o gerenciamento de informações. Tem como principal objetivo oferecer uma infraestrutura gerencial, tecnológica e de padronização que permita dinamizar a gestão cooperativa sobre os acervos. Para isso, a Rede dispõe de um conjunto de instrumentos normativos, metodológicos, tecnológicos e gerenciais e do portal unificado de consulta pública para disponibilizar aos seus colaboradores (GOVERNO DO RIO DE JANEIRO, 2023b).

Atualmente a Rede Web de Museus possui 35 instituições participantes, com um total de mais de 60.000 registros, que são: nove unidades da Fundação Anita Mantuano de Artes do Estado do Rio de Janeiro (FUNARJ), dois setores da Fundação Theatro Municipal do Estado do Rio de Janeiro, três palácios da Casa Civil, três museus do Município do Rio de Janeiro, dois museus geridos pelo Município de Niterói, quatro museus da esfera federal e onze museus privados (GOVERNO DO RIO DE JANEIRO, 2023a).

Como tecnologia, utiliza o SISGAM - Sistema de Gerenciamento de Acervos Museológicos, principal produto gerado pelo Projeto Rede de Museus, que é uma plataforma de gestão e registro de acervos responsável pela interligação das unidades museológicas por meio de um sistema comum, utilizando normas e padrões que ampliam as possibilidades de acesso aos seus conteúdos e imagens (GOVERNO DO RIO DE JANEIRO, 2023a; DE SOUZA; MACHADO, 2019).

O SISGAM não só possibilita a disponibilização online dos acervos, mas também promove dinamização nos procedimentos de documentação e catalogação.

²⁶ <http://www.museusdoestado.rj.gov.br/sisgam/>. Acesso em 11 jan. 2023

Ele realiza um controle eficaz para as instituições, envolvendo a entrada do objeto no museu, sua pesquisa, conservação, circulação, segurança e a sua disponibilização na web. Para viabilizar a exportação de dados gerados, há três possibilidades de formatos: XML, Excel e Dublin Core (DE SOUZA; MACHADO, 2019).

O site apresenta tutoriais sobre o SISGAM, porém, embora haja a informação de que os arquivos serão atualizados, no momento, os links encontram-se inacessíveis. Não há disponível o workflow ou a descrição de como é realizada a agregação.

Os portais de busca integrada estão disponíveis para acesso e o Quadro 26 apresenta o link para cada uma delas.

Quadro 26. Portais de busca integrada de agregadores nacionais

Projeto	Sites
Brasileana Fotográfica	https://brasilianafotografica.bn.gov.br/
Brasileana Iconográfica	https://www.brasilianaiconografica.art.br/
Rede Web de Museus	http://www.museusdoestado.rj.gov.br/sisgam/

Fonte: elaborado pelos autores (2022)

Em relação aos agregadores nacionais, analisando-os constatamos que não há informação suficiente para compreender, de maneira mais profunda, como a agregação ocorre em cada projeto. Embora todos permitam a entrada de novas instituições, não há disponível uma política clara de como o processo deve se dar. O Brasileira Fotográfica e o Iconográfica utilizam o repositório DSpace e o protocolo OAI-PMH. Sobre o DSpace, em relato feito pela Biblioteca Nacional Digital, a tentativa da adoção do DSpace não rendeu o resultado esperado, pois apesar de possuir código aberto, havia pouquíssima mão de obra capaz de executar as customizações necessárias e a migração segura dos dados de um sistema para o outro (MARTINS, 2018).

O fato de ambos utilizarem o OAI-PMH corrobora Freire et al. (2017), que aponta o protocolo como o mais usado para coleta de dados no ambiente cultural. Em estudo realizado por Siqueira e Martins (2021a), que analisou as instituições vinculadas ao Ministério da Cultura, hoje Ministério do Turismo, também concluiu que o padrão de metadados mais encontrado foi o Dublin Core e o protocolo de coleta de dados, o OAI-PMH.

É importante destacar que ser mais utilizado não é garantia de resultados totalmente satisfatórios, há questões referentes a alto tempo de resposta, resultados duplicados, impossibilidade de refinar os resultados, um ponto único de falha e pontos de estrangulamento, pouca escalabilidade, replicação de dados e manipulação de versões de serviços dos fornecedores, além de requerer dos usuários certa experiência para identificar os dados pertinentes (BRIGHAM et al. 2016; PAVÃO; CAREGNATO, 2015).

Os provedores de dados devem utilizar algumas ferramentas em comum, como XML e o Dublin Core ao trocar os dados. Assim, eles destacam que para um resultado plenamente satisfatório, padrões mais específicos, como o MARC, poderiam ser utilizados, porém a diversidade dos itens culturais torna difícil a correspondência entre seus padrões e, por isso, o uso de um padrão mais abrangente como o Dublin Core foi necessário, mesmo existindo a perda de especificidade e de valores de metadados (CUNHA; ALVES, 2019; FREIRE et al, 2017).

Com o intuito de conhecer o atual estado da arte do tema, identificar novas soluções e apoiar as decisões relativas à implementação do projeto de agregação do Ibram foi realizada uma revisão sistemática de literatura.

4.2 Revisão Sistemática de Literatura

Na RSL foram elencadas cinco questões de pesquisa, descritas na seção de Metodologia, cujas respostas estão a seguir.

4.2.1 Quais são os projetos?

Foram localizadas 16 publicações que trouxeram informações sobre 12 diferentes projetos, que estão listados no Quadro 27.

Quadro 27. Relação de projetos selecionados pela RSL, em ordem crescente por ano de publicação

No.	Projects
1	Connecting Archaeology and Architecture in Europeana Project – CARARE Project (Gavrilis, Dallas, & Angelis, 2013; Papatheodorou et al., 2011)
2	EUscreen Project (Oomen & Tzouvaras, 2012; Oomen, Tzouvaras, & Hyypää, 2013)
3	Heritage of the People's Europe Project – HOPE Project (Bardi, Manghi, & Zoppi, 2012, 2014)
4	Natural Europe Project (Makris et al., 2013)
5	European Film Gateway Project – EFG Project (Artini et al., 2013)
6	Open Cultural Digital Content Infrastructure (Stathopoulou, Stathopoulos, Georgiadis, Houssos, & Sachini, 2014)
7	Meta-Share (Piperidis et al., 2014)
8	Europeana network of Ancient Greek and Latin Epigraphy Project – EAGLE Project (Mannocci, Casarosa, Manghi, & Zoppi, 2014)
9	Digital Public Library of America – DPLA (Matienzo & Rudersdorf, 2014)
10	European Collected Library of Artistic Performance Content Aggregator – ECLAP (Bellini, Bruno, Nesi, & Paolucci, 2015)
11	ARIADNE Project (Ronzino, Felicetti, & Giorgio, 2016)
12	Data Aggregation Lab Software Tool – DAL (Freire, 2019)

Fonte: Siqueira; Martins (2021c, p. 5)

A próxima questão, a Q02, traz o detalhamento de cada projeto, apresentado em ordem crescente do seu ano de publicação, contemplando um breve histórico, o *workflow* de agregação e a descrição das etapas previstas no *workflow*.

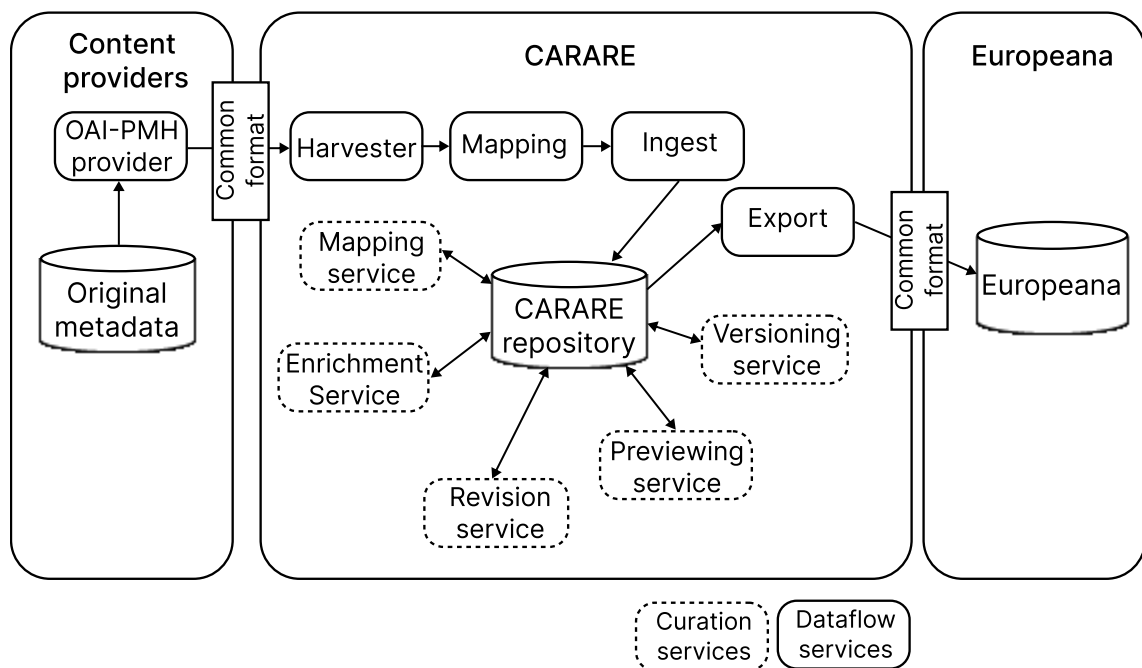
4.2.2 Quais são as etapas planejadas?

4.2.2.1 Projeto CARARE

A rede de melhores práticas CARARE (*Connecting Archaeology and Architecture in Europeana*) foi criada para aumentar a quantidade e a qualidade dos conteúdos digitais do patrimônio arqueológico e arquitetônico da Europeana. A arquitetura CARARE, mostrada na Figura 15, apresenta um repositório central

chamado *Monument Repository* (MoRe), que fornece serviços, tais como normalização de informações geográficas, monitoramento de integridade de metadados, enriquecimento semântico etc., antes de serem mapeados para o *Europeana Data Model* (EDM), e fornecidos à Europeana (PAPATHEODOROU et al., 2011).

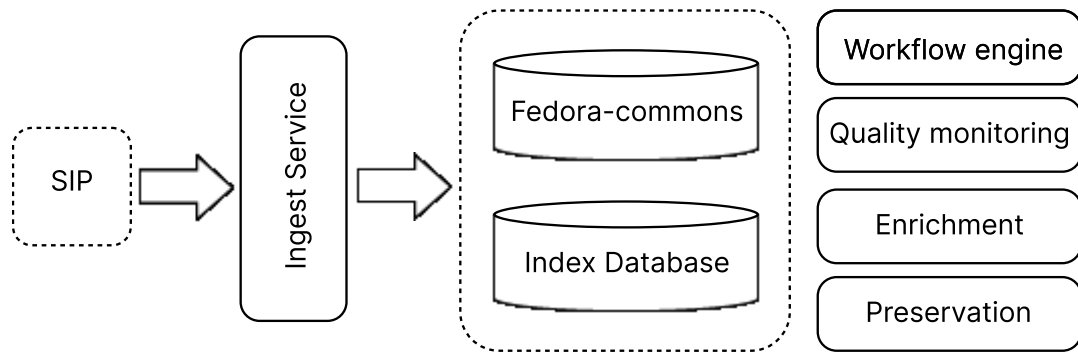
Figura 15. Arquitetura do sistema CARARE



Fonte: Papatheodorou et al. (2011, pág. 420), adaptada.

Os autores não fornecem detalhes sobre cada etapa, mostrando apenas o diagrama de fluxo de trabalho para análise. No entanto, Gavrilis et al. (2013) fornecem mais informações sobre o projeto, especialmente sobre o repositório MoRe, conforme ilustrado na Figura 16.

Figura 16. Arquitetura do repositório MoRe



Fonte: Gavrilis et al. (2013, pág. 3), adaptada.

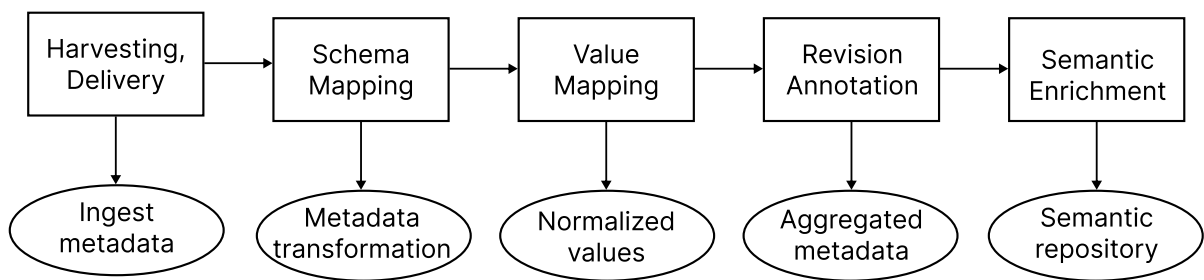
Os *Submission Information Packages* (SIP) são fontes de informações que na camada de *Ingest Service* são manipulados, seguindo especificações precisas quanto ao seu conteúdo e estrutura. Os proprietários de conteúdo inicialmente mapeiam metadados internos em registros compatíveis com o esquema CARARE, usando uma ferramenta de transformação, como MINT ou Repox. O MoRe, por meio de seu próprio mecanismo de indexação, define quais partes dos metadados serão indexadas e a estrutura do banco de dados *Standard Query Language* (SQL), ao qual serão indexados. A arquitetura central, *Fedora-Commons* e *Index Database*, consiste em uma camada de serviços que recebe pacotes de informações, realiza o pré-processamento e armazenamento de dados resultantes em uma instalação do Fedora. Os índices desses fluxos de dados são armazenados em um banco de dados MySQL e mantidos em sincronia com o Fedora. Na camada *Workflow engine* são realizados serviços simples de limpeza e enriquecimento. Serviços complexos, como adicionar relacionamentos, remover registros duplicados e assim por diante, são executados posteriormente; a camada *Quality Monitoring* garante que os proprietários dos dados possam controlar efetivamente o status dos objetos de metadados ingeridos; a camada de Curadoria permite o monitoramento e objetos de informação e realiza tarefas de avaliação, incluindo enriquecimento (*Enrichment*), visando a melhorar a qualidade do conteúdo, incluindo limpeza de elementos e atributos, eliminação de duplicação, preenchimento de atributos, adição de relacionamentos e transformação espacial. Na camada de preservação (*Preservation*) a cada ação que gera novo conteúdo ou, de alguma forma, o modifica, uma nova versão de fluxo de

dados é produzida e armazenada no Fedora junto com o log de eventos PREMIS (GAVRILIS et al., 2013).

4.2.2.2 Projeto EUScreen

O projeto EUScreen oferece acesso a milhares de itens do patrimônio audiovisual, reunindo clipes de eventos sociais, culturais, políticos e econômicos. Permite ainda explorar programas televisivos centrados na experiência cotidiana e funciona como agregador de domínios para a Europeia (OOMEN, TZOUVARAS, 2012). O fluxo de trabalho consiste em cinco fases, conforme mostrado na Figura 17.

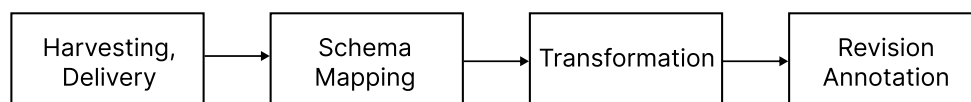
Figura 17. Workflow de ingestão de conteúdo do projeto EUScreen (2012)



Fonte: Oomen; Tzouvaras (2012, p. 2), adaptada.

O artigo não detalha as etapas, porém, em trabalho mais recente, Oomen, Tzouvaras e Hyypää (2013) apresentam um fluxo semelhante, mostrado na Figura 18.

Figura 18. Workflow de ingestão de conteúdo do projeto EUScreen (2013)



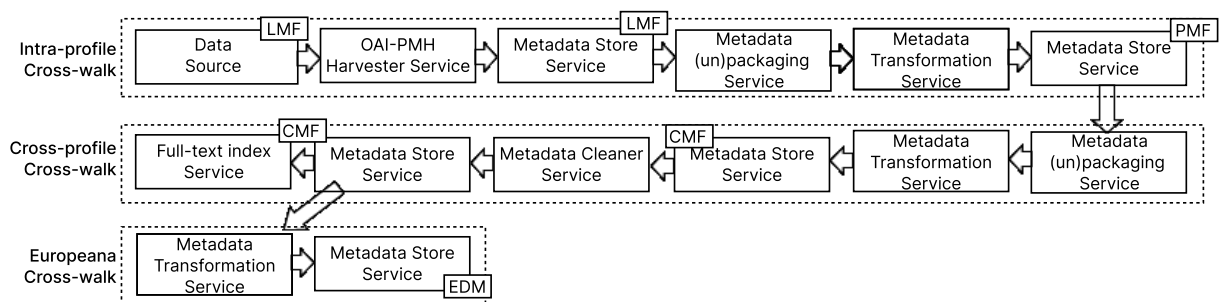
Fonte: Oomen; Tzouvaras; Hyypää (2013, p. 481), adaptada.

O fluxo apresenta a ingestão de metadados, sua transformação em um esquema de referência comum, na etapa de enriquecimento e, finalmente, na publicação como dados vinculados. Os autores não detalham as etapas, mas incluem a ferramenta usada para ingestão, transformação e enriquecimento, a *Mint Platform*. A publicação traz informações sobre o *front end*, que consiste em APIs “sem servidor”, nas quais um sistema proxy trata da comunicação com os serviços de *back-end*. Por fim, foi realizada a vinculação interna e externa ao conteúdo do EUScreen e o repositório resultante tornou-se acessível por meio de um *endpoint* de consulta SPARQL (OOMEN; TZOUVARAS; HYYPPÄÄ, 2013).

4.2.2.3 Projeto HOPE

O Projeto HOPE (*Heritage of the People's Europe*) utiliza o D-NET Software Toolkit (MANGHI et al., 2014) apresentado como um candidato ideal para a criação de infraestruturas de agregação de dados sustentáveis, extensíveis, escaláveis e dinâmicas para o patrimônio cultural. A Figura 19 apresenta o fluxo de trabalho de agregação do HOPE usando D-NET (BARDI; MANGHI; ZOPPI, 2012).

Figura 19. Fluxo de trabalho de agregação do projeto HOPE



Fonte: Bardi; Manghi; Zoppi (2012, p. 248), adaptada.

O uso de *crosswalk* (ou mapeamentos) resolve a heterogeneidade estrutural e semântica dos registros de metadados. Como mostra a Figura 19, existem três fases: a *intra-profile*, na qual ocorre o mapeamento de registros de metadados de todas as fontes de dados do mesmo perfil para registros de metadados em conformidade com um determinado modelo de dados padrão; o *cross-profile*, no qual é fornecido mapeamento estável desses registros de metadados para registros do

modelo de dados HOPE e o *cross-walk Europeana*, este último destinado a mapear para o EDM e armazenar dados (BARDI; MANGHI; ZOPPI, 2012).

Os serviços (*Harvester Service*) podem coletar registros via OAI-PMH, FTP, SFTP, HTTP e Hyper Text Transfer Protocol Secure (HTTPS). Uma instância do *Transformation service* é criada para cada fonte de dados para transformar registros de entrada com base em mapeamentos definidos com a ajuda do provedor de conteúdo. Depois que os registros são transformados no formato de metadados de perfil, eles são processados por outra instância do *Transformer Service* configurada com o cruzamento do perfil de metadados para o formato de metadados comum do HOPE. Nesse ponto, o fluxo vai para o *Cleaner service*, que aplica a transformação semântica de valores. Os registros limpos são entregues ao índice para uso do portal e transformados em EDM para serem coletados pelo OAI-PMH pela Europeana (BARDI; MANGHI; ZOPPI, 2012).

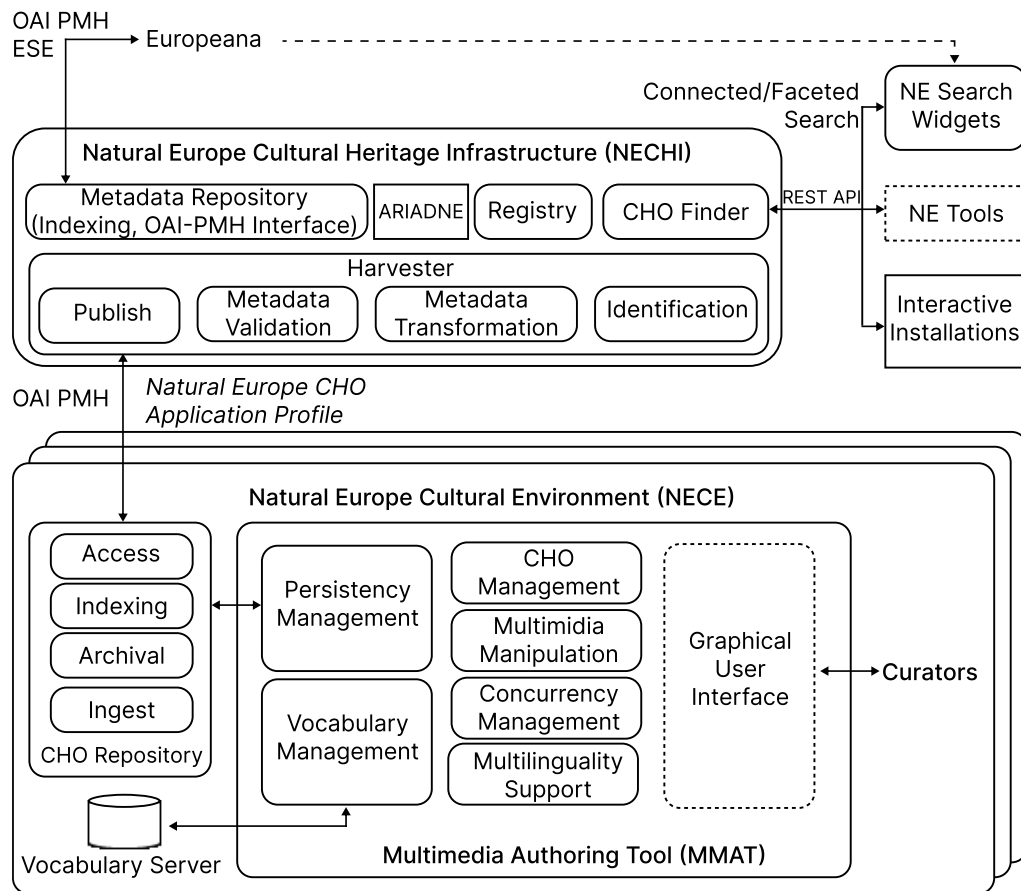
As informações são armazenadas no *Metadata Store Service* e podem ser acessadas por meio do *Full-Text Index Service*. Foram implementados os serviços de curadoria, que fornecem ferramentas para encontrar erros de mapeamento, inconsistências semânticas e enriquecimento de dados, mas ele não faz parte do fluxo de trabalho (BARDI; MANGHI; ZOPPI, 2012).

4.2.2.4 Projeto *Natural Europe*

O projeto *Natural Europe* oferece uma solução coordenada que visa a melhorar a disponibilidade e a relevância dos conteúdos culturais ambientais, num contexto multilíngue e multicultural. O conteúdo é coletado de seis Museus de História Natural da Europa (MAKRIS et al., 2013).

A Figura 20 apresenta a estrutura da Federação de Bibliotecas Culturais Digitais da Europa Natural, juntamente com suas ferramentas e serviços, que consistem em: *Natural Europe Cultural Environment* (NECE), ou seja, a infraestrutura e o conjunto de ferramentas implantados em cada Museu, e a *Natural Europe Cultural Heritage Infrastructure* (NECHI), interconectando as bibliotecas digitais dos museus, expondo ainda mais seus registros de metadados ao Europeana (MAKRIS et al., 2013).

Figura 20. Arquitetura da Federação de Bibliotecas Digitais Culturais da Natural Europe



Fonte: Makris et al. (2013, pág. 3), adaptada.

O NECE consiste na *Multimedia Authoring Tool* (MMAT), um sistema multilíngue de gerenciamento baseado na web que facilita o enriquecimento dos metadados dos objetos do patrimônio cultural (*Cultural Heritage Object*) - CHO: *CHO Management*, responsável por criar, recuperar, atualizar e excluir CHOs, registros/coleções e usuários; *Multimedia Manipulation*, que gerencia todas as funcionalidades de arquivos multimídia no sistema, incluindo geração de miniaturas e extração de metadados de arquivos de mídia, que são utilizados para a criação e enriquecimento de registros dos CHO; *Concurrency Management*, que fornece a funcionalidade básica para acesso simultâneo aos dados no repositório. Ele garante que não haja problemas de consistência quando vários usuários tentarem acessar o mesmo recurso; *Vocabulary Management*, que permite o acesso a termos taxonômicos, vocabulário e arquivos de autoridade, os quais estão guardados no

Vocabulary Server, Persistency Management, que gerencia o upload ou recuperação de pacotes de informações para o *CHO Repository*, responsável pela ingestão, manutenção e disseminação de conteúdo e metadados e adota o Modelo de Referência OAIS; *Graphical User Interface*, responsável pela interação do usuário, apresentação das informações e comunicação com o servidor (MAKRIS et al., 2013).

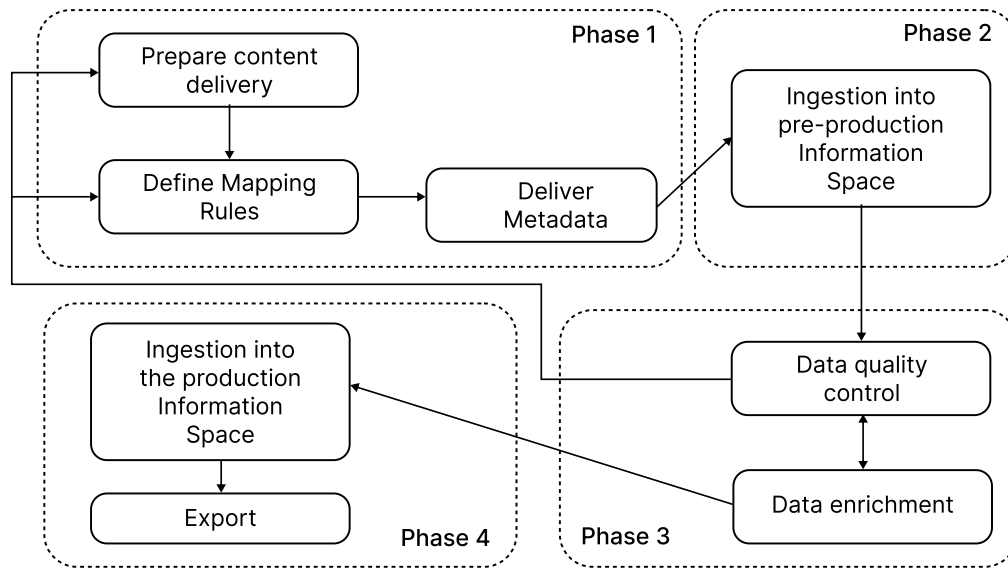
Natural Europe CHO Application Profile é um superconjunto do formato de metadados da *Europeana Semantic Elements* (ESE) e foi desenvolvido através de um processo interativo envolvendo especialistas do domínio dos museus e parceiros técnicos do projeto (MAKRIS et al., 2013).

O NECHI foi baseado em tecnologias e serviços ARIADNE (RONZINO et al., 2016). Ele utiliza o OAI-PMH para recolher os metadados dos destinos e publicá-los num repositório central através do serviço de publicação que integra os serviços: *Publish*, que publica os metadados recolhidos num repositório central; *Metadata Validation*, que fornece validação sintática e semântica de instâncias de metadados em relação a perfis de aplicativos predefinidos; *Metadata Transformation*, que converte metadados do formato *Natural Europe CHO Application Profile* para o formato especificado pela Europeana; *Identification*, que fornece identificadores digitais persistentes para recursos na infraestrutura ARIADNE e; por fim, o *Search Widgets* que suporta uma pesquisa simples, facetada ou conectada (MAKRIS et al., 2013).

4.2.2.5 Projeto EFG

Arquivos de filmes contendo coleções de material digital foram criados em muitos países europeus. Assim, o Projeto EFG (*European Film Gateway*) projetou um modelo de dados comum para informações de filmes, no qual os modelos de dados de arquivo podem ser mapeados de forma otimizada. O EFG possui uma infraestrutura de dados baseada no software D-NET (MANGHI et al., 2014), que foi ampliado com ferramentas avançadas para curadoria de dados (ARTINI et al., 2013). A Figura 21 apresenta o fluxo de trabalho.

Figura 21. Workflow de ingestão de dados do EFG



Fonte: Artini et al. (2013, pág. 36), adaptada.

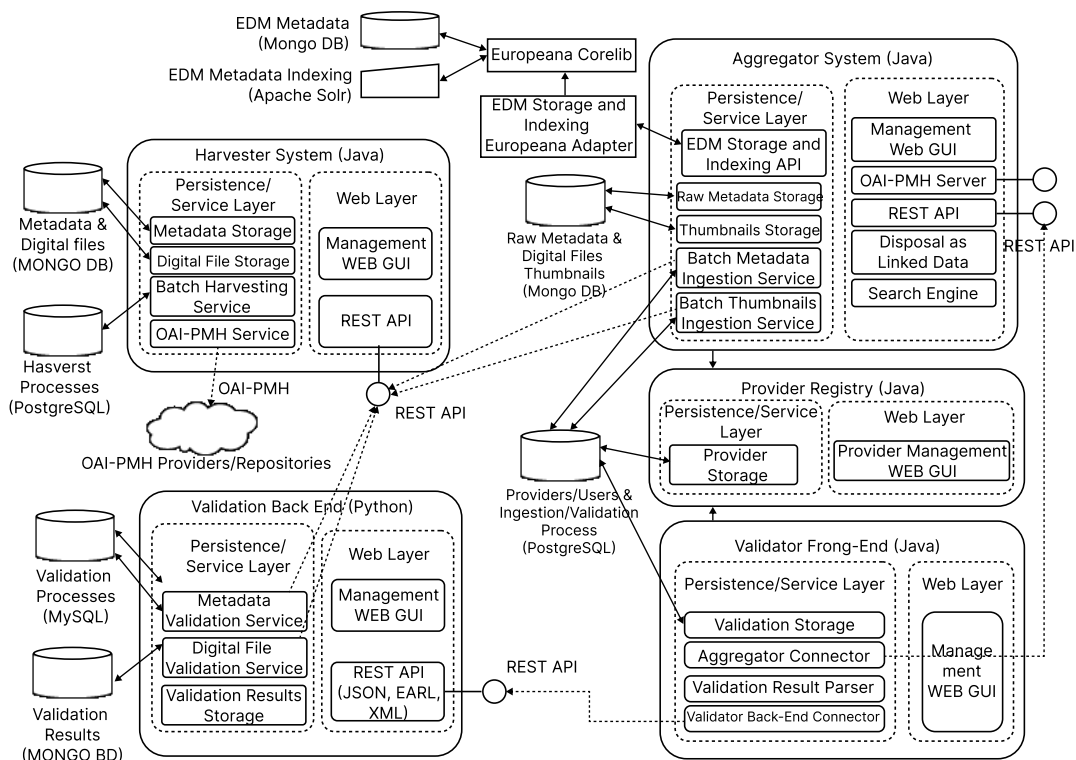
Na Fase 1, *Metadata Mapping Definition*, especialistas em arquivos analisam os metadados que fornecem para determinar como essas informações podem ser mapeados estrutural e semanticamente para o esquema de metadados *EFG metadata schema*. As regras de mapeamento são entregues aos administradores de infraestrutura, que as codificam na forma de scripts D-NET. Na Fase 2, *Metadata Transformation and Cleaning*, os metadados são coletados via protocolos OAI-PMH ou FTP para serem processados pelos scripts de mapeamento produzidos na Fase 1 e gerar os registros EFG correspondentes. Os registros resultantes não estão imediatamente disponíveis para acesso, mas são armazenados em um espaço de informação de “pré-produção”, onde a Fase 3 pode ocorrer. Na Fase 3, *Metadata Quality Control and Enrichment*, os registros podem ser validados e inspecionados para identificar erros de mapeamento, erros de digitação e duplicatas. Especificamente, a ferramenta *Content Checker* pode ser usada para verificar se o mapeamento estrutural foi executado corretamente. A *Vocabulary Checker Tool* notifica os provedores de dados sobre registros EFG que ainda não estão em conformidade com vocabulários comuns e o *Authority File Manager*, uma ferramenta para mesclar registros duplicados e desambiguar informações. Na Fase 4, *Metadata Publishing*, os registros aprovados na Fase 3 ficam visíveis no portal do EFG

e podem ser exportados para fornecedores externos, como a Europeana (ARTINI et al., 2013).

4.2.2.6 Open Cultural Digital Content Infrastructure

Stathopoulou et al. (2014) apresentam o desenvolvimento de uma infraestrutura para agregar registros de metadados e arquivos digitais produzidos centralmente e validar automaticamente sua conformidade com especificações de interoperabilidade e qualidade. A Figura 22 apresenta a arquitetura do projeto.

Figura 22. Arquitetura do *Open Cultural Digital Content Infrastructure*



Fonte: Stathopoulou et al. (2014, pág. 286), adaptada.

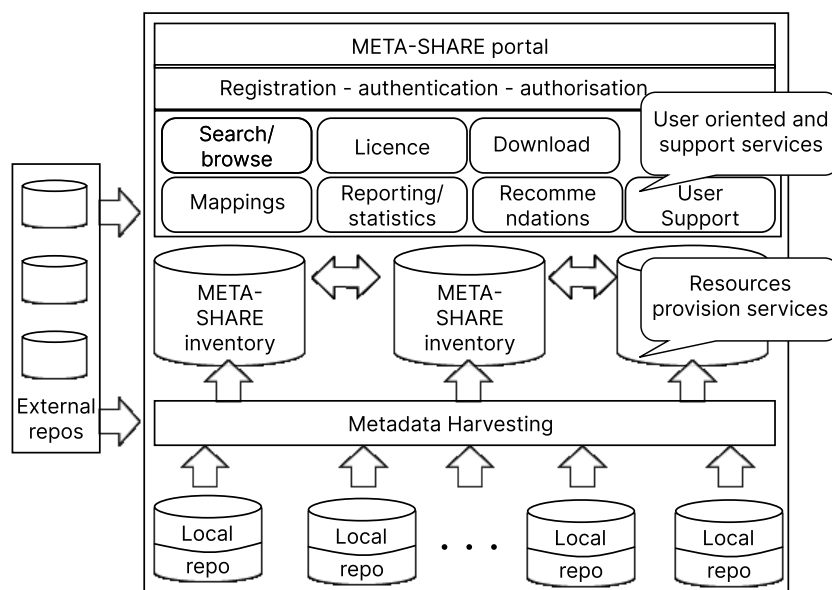
Embora apresente uma arquitetura detalhada na Figura 22, o artigo não explica todos os itens, enfatizando as camadas de validação: *Validação Back End* e *Front-End Validator*. Em geral, informam que o sistema é construído com uma camada autônoma de coleta de dados, que coleta e armazena metadados e arquivos digitais. A API REST foi implementada para permitir que componentes de software externos acionem e gerenciem as coleções. Em relação à camada *Validation Back*

End, o *Open Cultural Digital Content Infrastructure* permite especificar regras de validação complexas em qualquer nível (repositório, metadados, arquivos digitais), realizar validações e registrar resultados detalhados, enquanto o *Front-End Validator* fornece uma interface de usuário, que poderá realizar procedimentos de validação, agregar relatórios e conectar-se com os provedores de repositório e procedimentos administrativos para validações (STATHOPOULOU et al., 2014).

4.2.2.7 META-SHARE

Piperidis et al. (2014) apresentam o META-SHARE, uma infraestrutura formada por uma rede de repositórios que armazenam recursos de linguagem (dados, ferramentas e serviços de processamento) documentados com metadados de alta qualidade, agregados em inventários centrais, permitindo busca e acesso uniformes. A Figura 23 apresenta sua arquitetura.

Figura 23. Arquitetura do META-SHARE



Fonte: Piperidis et al. (2014, pág. 1533), adaptada.

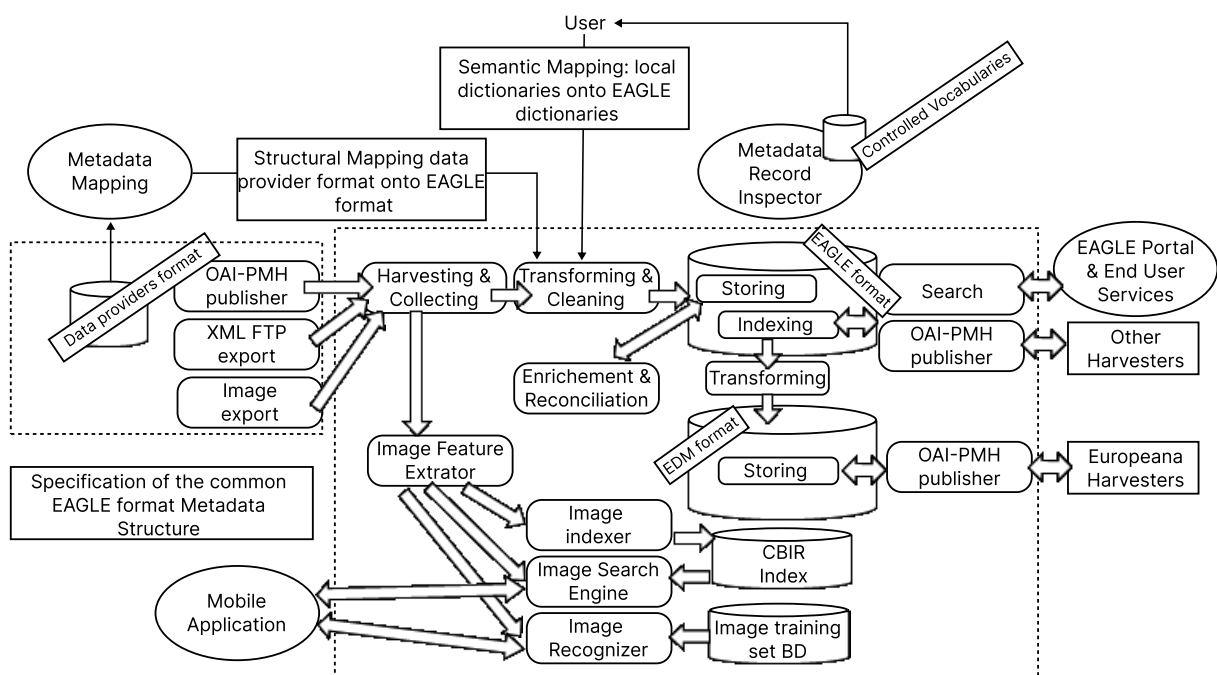
Os autores não detalham todas as etapas, mas explicam que os registros de metadados são coletados e armazenados nos servidores centrais do META-SHARE, que compartilham metadados, criam, hospedam e mantêm um inventário central, incluindo descrições baseadas em metadados de todos os recursos disponíveis na

rede distribuída. Os provedores de recursos podem criar, armazenar e editar descrições de recursos usando um editor implementado pelo META-SHARE, bem como carregar recursos reais diretamente ou fornecer um link para um armazenamento existente. Eles também podem obter estatísticas sobre o número de visualizações e downloads de seus recursos, bem como a origem dos espectadores. Os data centers ou outras organizações com catálogos de recursos existentes podem obter suporte para mapear seus metadados no esquema do modelo de infraestrutura. Os consumidores de recursos podem se registrar, criar um perfil de usuário (PIPERIDIS et al., 2014).

4.2.2.8 Projeto EAGLE

O projeto EAGLE (*Europeana Network of Ancient Greek and Latin Epigraphy*) visa fornecer à Europeana uma coleção abrangente de fontes históricas da região do Mediterrâneo e fornecer um portal amigável para acessar a mesma coleção, construída com o material de cerca de 15 arquivos epigráficos diferentes. A infraestrutura de agregação de dados EAGLE é alimentada pelo software D-NET (MANNOCCI et al., 2014). A Figura 24 apresenta o fluxo de trabalho do EAGLE.

Figura 24. Arquitetura do D-NET adaptado para uso no Projeto EAGLE



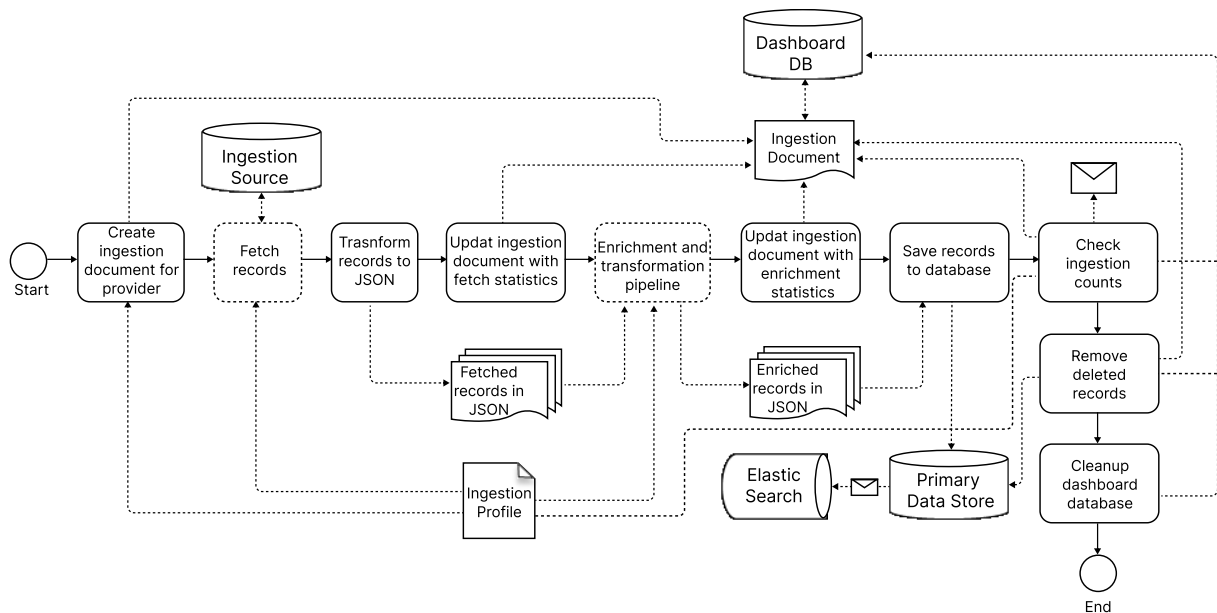
Fonte: Mannocci et al. (2014, pág. 293), adaptada.

O processo baseado no software D-NET (MANGHI et al., 2014) consiste em quatro fases. A definição de mapeamento de metadados, que em cooperação com especialistas de domínio fornece registros, regras estruturais e semânticas para mapear os registros para esquema de metadados EAGLE e que são posteriormente codificados na forma de scripts D-NET; transformação e limpeza de metadados, em que os registros de metadados são coletados e processados para gerar os “objetos EAGLE”, criando assim o espaço de pré-produção; controle de qualidade de metadados, em que os registros são inspecionados e validados para identificar erros de mapeamento e outros, como digitação; provisionamento de metadados, no qual os registros que passam na Fase 3 são movidos para o Espaço de Informação de Produção, onde são indexados e ficam disponíveis para ingestão na Europeia ou para consulta e navegação no portal EAGLE (MANNOCCI et al., 2014).

4.2.2.9 DPLA

A DPLA - *Digital Public Library of America*, após seu primeiro ano de agregação de objetos de bibliotecas, arquivos e museus norte-americanos, realizou uma revisão das tecnologias e processos em que o trabalho ocorre, bem como os modelos de dados utilizados para agregar os vários padrões de metadados dos Hubs (provedores de conteúdo) e a natureza da colaboração entre DPLA e Hubs, resultando na publicação (MATIENZO; RUDERSDORF, 2014). Entre os tópicos discutidos, um deles é a apresentação do workflow de ingestão na DPLA, ilustrado na Figura 25.

Figura 25. Workflow de ingestão de dados do DPLA



Fonte: Matienzo; Rudersdorf (2014, p. 3), adaptada.

O fluxo de trabalho de ingestão é chamado por um script que lê o perfil de ingestão de uma determinada origem e cria um documento de ingestão no banco de dados do painel, que contém temporariamente uma representação de cada registro para permitir que a equipe identifique quais partes foram alteradas. Após a criação do documento de ingestão, a equipe invoca a tarefa de pesquisa, que obtém os metadados a serem ingeridos da origem definida no perfil. Os metadados são transformados de seu formato nativo, geralmente XML, para uma expressão JSON e mantidos em disco em um local temporário. Após a conclusão do processo de busca, o documento de ingestão é atualizado para conter a localização dos dados transformados em JSON (MATIENZO; RUDERSDORF, 2014).

A equipe de ingestão invoca as tarefas de transformação e enriquecimento. Essas tarefas mapeiam e transformam os metadados serializados em JSON para o DPLA MAP (*Digital Public Library of America Metadata Application Profile*) e normalizam, aprimoram os metadados. O MAP DPLA é uma extensão do EDM e, como o EDM, incorpora ou faz referência a uma variedade de padrões e modelos (MATIENZO; RUDERSDORF, 2014).

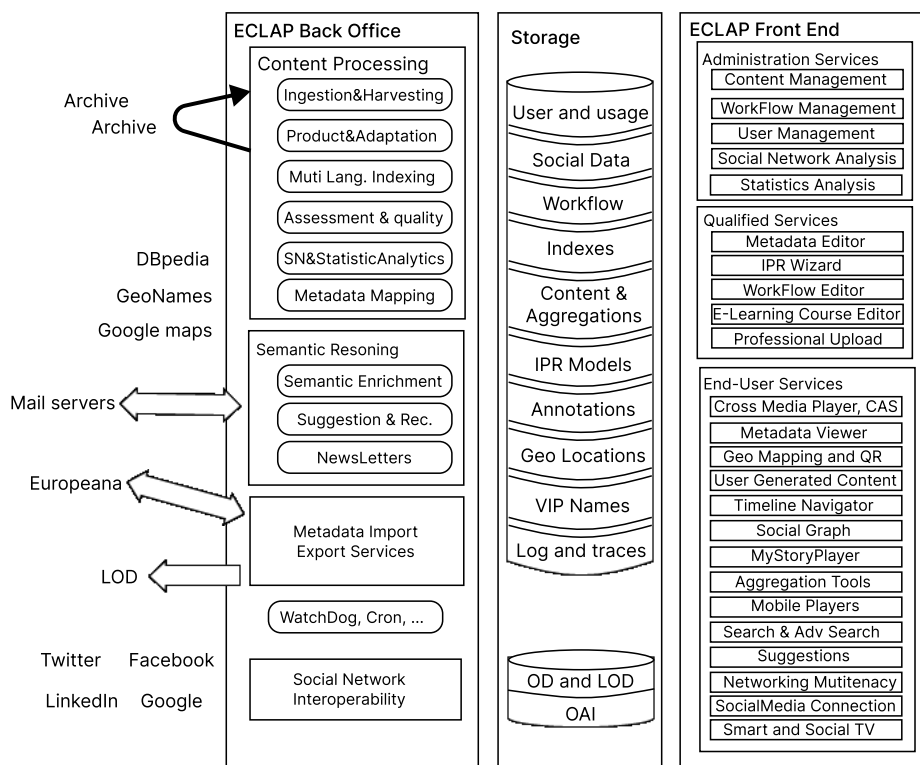
Depois de concluídos, os registros são mantidos temporariamente em disco como uma serialização JSON-LD MAP e o documento de ingestão é atualizado com

informações sobre os processos de transformação e enriquecimento, incluindo a localização dos registros e a extensão de quaisquer falhas de processo. A equipe de ingestão armazena os dados e identifica o número de registros novos, atualizados ou excluídos e alerta automaticamente a equipe quando os valores estão acima de um determinado limite. Por fim, a equipe de ingestão realiza duas tarefas finais: remover registros excluídos e limpar o banco de dados do painel (MATIENZO; RUDERSDORF, 2014).

4.2.2.10 Agregador de conteúdo ECLAP

A solução ECLAP (*European Collected Library of Artistic Performance Content Agregator*) coleta e agrega dados de instituições culturais voltadas para as artes cênicas (BELLINI et al., 2015). A Figura 26 apresenta uma visão geral da arquitetura do ECLAP.

Figura 26. Visão geral da arquitetura do Projeto ECLAP



Fonte: Bellini et al. (2015, pág. 8), adaptada.

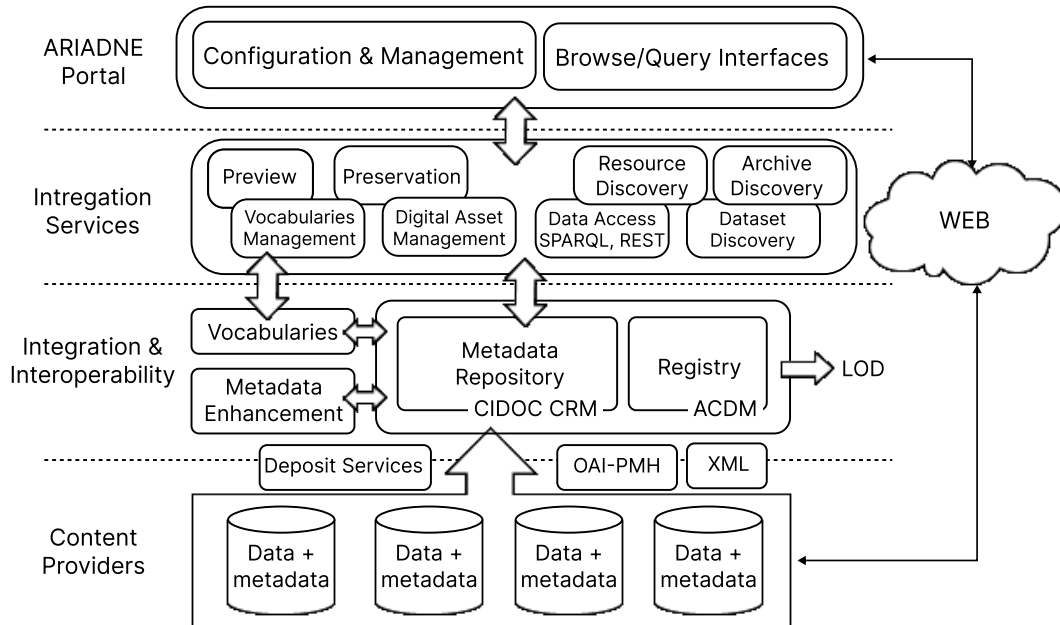
A Figura 26 retrata o *Back Office* da ECLAP, que consiste em quatro etapas. A primeira, Processamento de Conteúdo, inclui as atividades de: Ingestão e Coleta, que faz a ingestão de metadados de diversos provedores de conteúdo por meio de processos de coleta e aquisição; Produção e adaptação, que consiste em converter arquivos digitais nos diversos formatos necessários para reproduzi-los em diferentes dispositivos e serviços e extrair miniaturas para o front-end; Indexação multilíngue, que executa a indexação multilíngue de conteúdo de mídia cruzada para fornecer serviço de pesquisa de front-end; Avaliação da qualidade, que consiste em uma avaliação geral da qualidade do conteúdo ingerido, como identificar conteúdos que não possuam o nível mínimo de integridade e consistência dos metadados em relação ao nível de aceitação imposto; Mapeamento de metadados, em que os conjuntos de metadados ingeridos são fornecidos em diferentes esquemas e devem ser mapeados de acordo com o modelo de ingestão semântica ECLAP. A segunda etapa, Raciocínio Semântico, inclui o Enriquecimento Semântico, que visa a enriquecer os descritores de conteúdo por meio do processamento de metadados com algoritmos de processamento de linguagem natural para extrair nomes de pessoas, nomes geográficos e datas, a fim de encontrar sinônimos, desambiguação de nomes, nomes geográficos e datas, conectando nomes de pessoas com DBpedia, entre outros. Sugestões, Rec e Newsletters, consistem em computar semelhanças entre usuários e conteúdo, para produzir sugestões e recomendações. O terceiro, *Metadata Import Export Service*, inclui processos e serviços para produzir metadados no EDM, mapeando uma parte do modelo semântico ECLAP. Uma vez produzido, os descritores de metadados são exportados através de um OAI-PMH para a Europeia. Além disso, o LOD do modelo semântico ECLAP também está disponível. A última etapa, Interoperabilidade de Redes Sociais, inclui atividades para a integração da infraestrutura da ECLAP com redes de terceiros e serviços sociais (BELLINI et al., 2015).

4.2.2.11 Projeto ARIADNE

O projeto ARIADNE visa proporcionar aos investigadores acesso integrado e garantir a interoperabilidade semântica de conjuntos de dados arqueológicos distribuídos por toda a Europa. As descrições de conteúdo foram codificadas usando o Modelo de Dados de Catálogo ARIADNE (*ARIADNE Catalogue Data Model* -

ACDM). Para abordar a complexidade da integração de dados arqueológicos, a ontologia CIDOC-CRM foi escolhida como a espinha dorsal do ACDM (RONZINO et al., 2016). A Figura 27 mostra seu fluxo de trabalho.

Figura 27. Infraestrutura do Projeto ARIADNE



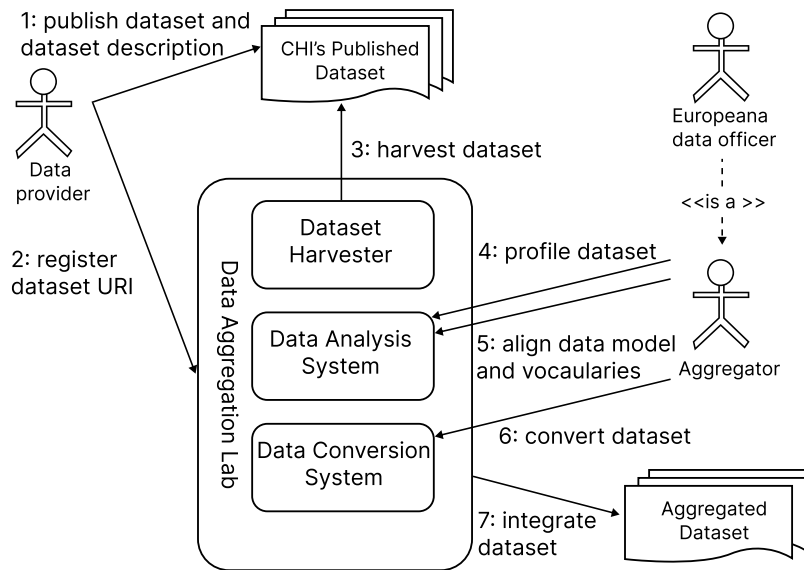
Fonte: Ronzino et al. (2016, pág. 47), adaptada.

Os autores não detalham as etapas, mas informam que a agregação de vários bancos de dados foi transformada mapeando seus esquemas individuais para o Modelo de Referência ARIADNE. O processo de mapeamento foi suportado pelo *X3ML Mapping Framework*, ferramenta que garante a integridade dos dados iniciais e preserva o significado original (RONZINO et al., 2016).

4.2.2.12 Data Aggregation Lab

Freire (2019) descreve o software *Data Aggregation Lab* (DAL), que implementa o fluxo de trabalho de agregação de metadados do patrimônio cultural, baseado na solução de agregação atual da Europeia. A Figura 28 apresenta o fluxo de trabalho de agregação proposto.

Figura 28. Workflow para agregação de dados do software DAL



Fonte: Freire (2019, p. 421), adaptada.

O fluxo de trabalho começa depois que os metadados são publicados por um provedor de dados e termina com a preparação de uma exportação de conjunto de dados para fornecê-los aos sistemas de ingestão e integração. Segundo os autores, o número de etapas do fluxo de trabalho pode variar um pouco dependendo da tecnologia em uso, mas, em geral, é composto por sete etapas principais (FREIRE, 2019).

Na etapa *"Publish dataset and dataset description"*, o provedor de dados cria um recurso RDF que descreve o conjunto de dados. Contém informações descritivas sobre o conteúdo do patrimônio cultural e detalhes técnicos sobre como o conjunto de dados pode ser coletado; em *"Register dataset URI"*, o provedor notifica o agregador da disponibilidade de seus conjuntos de dados para agregação, fornecendo o URI do recurso RDF, criado na primeira etapa; em *"Harvest dataset"*, o conjunto de dados é coletado pelo DAL, aplicando o mecanismo de coleta especificado na primeira etapa; em *"Profile dataset"*, o conjunto de dados de perfil é perfilado pelo agregador, dando suporte às ferramentas de análise de dados. Essas ferramentas analisam aspectos importantes para integrar os dados no conjunto agregado central. Ele permite que o agregador detecte deficiências no conjunto de dados e avalie o esforço e as tarefas específicas necessárias para as etapas a seguir; em *"Align data model and vocabularies"*, o agregador alinha o modelo de dados e os vocabulários usados

no conjunto de dados do provedor de dados ao EDM; em “*Convert dataset*”, o data set é convertido para EDM e o resultado é verificado; finalmente, em “*Integrate dataset*”, o DAL interage com a ingestão regular e integração de dados dos sistemas agregadores e fornece o conjunto de dados convertido (FREIRE, 2019).

A próxima questão, a Q03, lista quais foram as tecnologias utilizadas no processo de agregação dos projetos.

4.2.3 Quais tecnologias são utilizadas?

As tecnologias, sua categoria e descrição são apresentadas no Quadro 28.

Quadro 28. Relação de tecnologias utilizadas nos agregadores da RSL

Technology	License	Distribution	Link	Current
Data aggregator				
Ariadne	Not found	Not found	https://ariadne-infrastructure.eu/	Yes
CHO Repository	Apache License	Apache License	https://github.com/psu-libraries/cho	Yes
D-NET	GNU Affero GPL	GNU Affero GPL	http://www.d-net.research-infrastructures.eu/	No
META-SHARE	BSD License	BSD License	https://github.com/metashare/META-SHARE	No
MINT	GNU Affero GPL	GNU Affero GPL	https://github.com/mint-ntua/Mint	No
MoRe	Not found	Not found	https://locloud.eu/resources/more-repository/	No
Repos	Not found	Not found	https://doi.org/10.1007/978-3-642-04346-8_65	No
Application for smartphone				
Flagship Mobile	Proprietary	Free	https://www.eagle-network.eu/resources/flagship-mobile-app/	Yes
Consultation language				
SPARQL Query Language	W3C Software License	W3C Software License	https://www.w3.org/TR/rdf-sparql-query/	Yes
Mapping				
X3ML Mapping Repository	Apache License	Apache License	http://ceur-ws.org/Vol-1656/paper1.pdf	No
Fedora	Apache License	Apache License	https://duraspace.org/fedora/	Yes
Cloud hosting service				
GRNET VIMA	Not applicable	Not applicable	https://grnet.gr/services/cloud-services/vima/	Yes
NoSQL Database Management System				
Apache CouchDB	Apache License	Apache License	https://couchdb.apache.org/	Yes
Apache Solr	Apache License	Apache License	https://lucene.apache.org/solr/	Yes
Elastic Search	Apache License	Apache License	https://www.elastic.co/pt/what-is/elasticsearch	Yes
Mongo DB	Server Side Public License	Server Side Public License	https://www.mongodb.com/	Yes
Relational Database Manager System				
MySQL	GNU Affero GPL	GNU Affero GPL	https://www.mysql.com/	Yes
PostgreSQL	PostgreSQL License	PostgreSQL License	https://www.postgresql.org/	Yes
System for data enrichment				
Multimedia Authoring	Not found	Not found	http://natural-europe.tuc.gr/music/mmat	No

Fonte: Siqueira; Martins (2021c, p. 15)

A Q04 aborda a maneira como as etapas do processo de agregação acontecem, se de forma manual, automática ou semiautomática.

4.2.4 Como as etapas são executadas?

Os projetos CARARE, Hope, EFG, Eagle, ARIADNE e ECLAP foram considerados semiautomáticos e o DPLA, manual. Não há projetos que realizem todas as etapas automaticamente. Nos projetos DAL, EUscreen, Natural Europe, META-SHARE e Open Cultural Digital Content, não foram encontradas evidências que nos levassem a esta resposta. Destaca-se que esta informação não é encontrada diretamente nas publicações, sendo inferida pelos pesquisadores. A Q05 aborda quais projetos realizam a busca semântica.

4.2.5 Quais apresentam busca semântica?

Os projetos CARARE, DAL, DPLA, ECLAP, EFG, EUscreen, Hope e Open Cultural Digital Content realizam busca semântica, enquanto o Meta-Share não. Os demais projetos não trouxeram essa informação.

As respostas às questões de pesquisa nos levam a algumas reflexões que são apresentadas nesta seção. Em relação a Q01, buscou-se identificar o local de origem dos projetos, se ainda estão em execução e se os portais agregadores estão disponíveis²⁷. O Quadro 29 contém esses dados, retirados de publicações e/ou buscas na Internet.

²⁷ A RSL visa traçar o histórico de publicação de um período específico; entretanto, buscamos identificar quais dos projetos ainda estão em uso ou foram atualizados, porém, ressaltamos que a falta de artigos atualizados não nos permite afirmar se houve mudanças nas tecnologias ou fluxos de trabalho, sendo apenas uma forma de apresentar o panorama atual dos projetos.

Quadro 29. Relação de projetos, período de execução e portal agregador

Europe		
Project (execution)	Successor (execution)	Aggregator portal
CARARE (2010–2013)	CARARE pro (no information)	https://www.carare.eu/explore/
EUScreen (2009–2012)	EUScreenXL (2013–2016)	http://www.euscreen.eu/
HOPE (2010–2013)	–	https://socialhistoryportal.org/
Natural Europe (2010–2013)	–	–
EFG (2008–2011)	EFG1914 (2012–2014)	https://www.europeanfilmgateway.eu/
Open Cultural Digital (no information)	–	–
Meta-Share (no information)	–	http://www.meta-share.org/
EAGLE (no information)	–	https://www.eagle-network.eu/
ECLAP (no information)	–	http://www.eclap.eu
Ariadne (2013–2017)	Ariadne Plus (2019–now)	http://portal.ariadne-infrastructure.eu/
DAL (no information)	–	–
North of America		
Project (execution)	Successor (execution)	Aggregator portal
DPLA (no information)	–	https://dp.la/

Fonte: Siqueira; Martins (2021c, p. 16)

Observamos a predominância de iniciativas originadas no continente europeu, 11 em 12, sendo os provedores de dados majoritários para a Europeia, que trabalha com mais de 50 milhões de itens digitalizados, como livros, músicas, obras de arte etc., oferecendo ferramentas de pesquisa e filtro. A exceção é o DPLA, que se originou nos Estados Unidos. Não encontramos os portais agregadores dos projetos *Natural Europe* e *Open Cultural Digital*.

O objetivo da Q2 foi identificar as etapas necessárias para a agregação de dados. Para tanto, sistematizamos as etapas do Quadro 30 para melhor compreensão de todo o processo. Ressaltamos a dificuldade na elaboração do Quadro 30, pois, embora as publicações apresentem o fluxo de trabalho de agregação, as etapas não são, em sua totalidade, evidentes e/ou descritas. Selecionamos apenas aqueles que apresentavam mais de uma ocorrência, ou seja, mais de um agregador os citou.

Quadro 30. Sumarização das etapas dos workflows de agregação de dados utilizando a nomenclatura original das publicações

Steps, projects	Harvesting	Ingestion	Mapping	Indexing	Storing	Monitoring	Enriching	Displaying	Publishing LOD
CARARE	Harvester	Ingest	Mapping	Indexing	Fedora-commons e index database	Quality monitoring	Curation	Export	-
EUScreen	Harvesting, delivery	-	Schema mapping	-	-	-	Transformation, revision, and annotation and semantic mapping	-	Contains this step
HOPE	Harvester service	-	Transformation and (un)packaging	Full-text index	Store	Content checker and record tagging service	Cleaner service	-	-
Natural Europe	Harvester	Publish	Transformation	-	-	-	Validation	-	-
EFG	Transformation and clean	Deliver metadata	Mapping definition and transformation and clean	-	Transformation and clean	Quality control and enrichment	Quality control and enrichment	Metadata publish	-
Open cultural	Contains this step	-	Contains this step	Contains this step	Contains this step	-	-	Contains this step	-
Meta-Share	Contains this step	-	Contains this step	-	Contains this step	-	-	-	-
EAGLE	Transformation and clean	-	Mapping definition and transformation and clean	Production information space	Transformation and clean	Quality control	Enrichment and reconciliation	OAI-PMH Publisher	-
DPLA	Contains this step	-	Contains this step	-	Contains this step	-	Contains this step	-	-
ECLAP	Ingestion and harvesting	Ingestion and harvesting	Mapping and import export	Multilingual indexing	Storage	Qualified and administrator	Semantic enrichment	Contains this step	Contains this step
ARIADNE	Contains this step	Contains this step	Contains this step	-	Contains this step	-	Enhancement	-	Contains this step
DAL	Harvest dataset	-	Align data model and vocabular; convert dataset and integrate dataset	-	-	-	-	-	Contains this step

Fonte: Siqueira; Martins (2021c, p. 17)

Alguns fluxos de trabalho não apresentam todas as etapas descritas no artigo, portanto, quando citados, mas não apresentados no fluxo de trabalho, incluímos a etapa na Quadro 30 com um “*Contains this step*”.

Para melhor compreensão do Quadro 30, elaboramos as descrições de cada categoria. O Quadro 31 os apresenta.

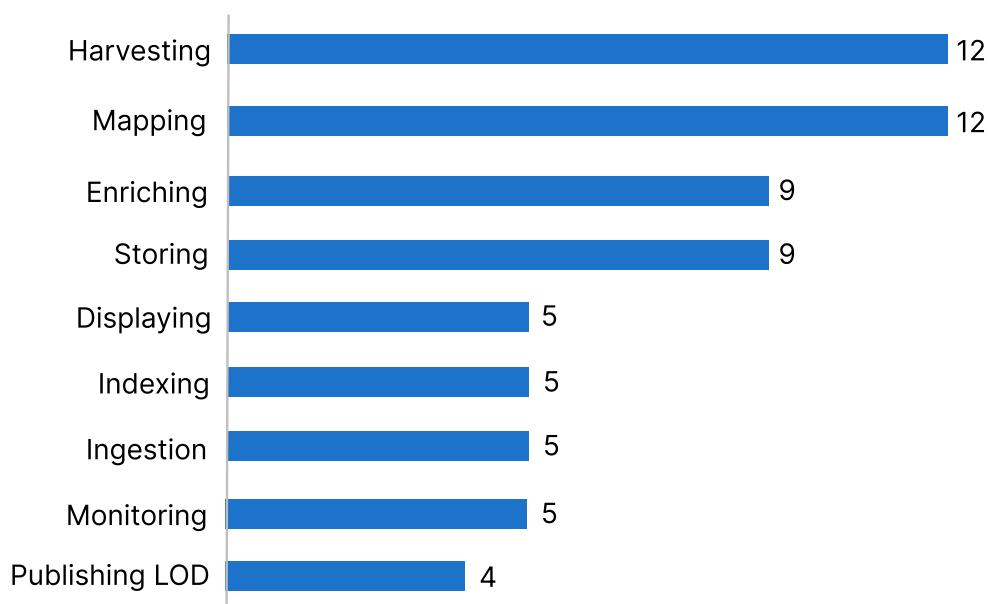
Quadro 31. Sumarização das etapas dos workflows de agregação de dados - RSL

No.	Step	Description
1	Harvesting	Extraction of data in its raw form. In general, in OAI-PMH format
2	Ingestion	Receive the data extracted in the collection
3	Mapping	Mapping metadata schemas, performing the transformation from the source scheme to the target scheme
4	Indexing	Metadata indexing service
5	Storing	Store the collected data
6	Monitoring	Interface for monitoring results for possible human interference
7	Enriching	Enrich the metadata through other existing data on the web
8	Displaying	Make the aggregated data available through API, which displays the data in RDF, OAI-PMH or JSON format
9	Publishing LOD	Publish on Linked Open Data

Fonte: Siqueira; Martins (2021c, p. 18)

Apenas duas etapas se repetem em todas as soluções: a *Harvesting*, pois é uma etapa obrigatória, e *Mapping*, essencial pela diversidade de objetos culturais disponíveis. No entanto, consideramos que as etapas de ingestão e armazenamento também devem ser incluídas em todas as soluções, embora não tenham sido declaradas explicitamente. O Gráfico 2 mostra o número de ocorrências de cada etapa.

Gráfico 2. Ocorrência das etapas para agregação de dados – RSL



Fonte: Siqueira; Martins (2021c, p. 18)

Algumas etapas não foram incluídas no Quadro 30, por serem únicas nos agregadores, são elas: no CARARE, existem as etapas de “limpeza”, que são serviços simples de limpeza e preenchimento de dados e “preservação”, que visa a manter os metadados dos registros fornecidos aos o repositório, possibilitando sua revisão, versionamento e validação; no ECLAP, há compatibilidade com redes sociais e algoritmos para sugestões de objetos digitais do perfil do usuário; no EUscen, existe o “Mapeamento de Valor”, mas a etapa não é descrita e fornece um endpoint SPARQL para consultas; na Natural Europe, há a etapa de “Identificação”, que fornece identificadores digitais persistentes para recursos na infraestrutura e os “Widges de pesquisa”, que suportam uma pesquisa simples, facetada ou conectada. No ARIADNE, existem etapas no fluxo de trabalho que não estão descritas no texto.

As soluções são personalizadas e, aparentemente, pouco escaláveis, não sendo possível responder a cenários e contextos diferentes dos produzidos, exceto o projeto DAL, o único que possui uma solução mais genérica. De forma indireta, o D-NET aparece em três dos projetos e pode ser considerada uma solução mais genérica. Em geral, há diferenças significativas no fluxo de dados, mas não nas etapas.

Em relação a Q03, nem todos os artigos indicaram quais tecnologias são utilizadas. A partir dos artigos que o fizeram, descobrimos que havia tecnologias diferentes em uso e não uma mais usada. Para complementar a resposta, pesquisamos as licenças de uso e distribuição e oferecemos o link para acessar a ferramenta e verificamos se ela está atualizada ou não. Para efeitos desta pesquisa, foram consideradas atuais as com atualizações ocorridas nos últimos dois anos. Apresentamos o resultado no Quadro 32.

Quadro 32. Resposta à pergunta 03: Quais tecnologias são utilizadas?

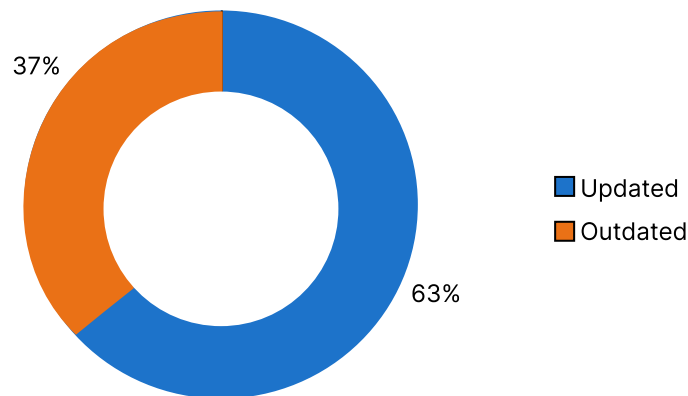
Technology	License	Distribution	Link	Current
Data aggregator				
Ariadne	Not found	Not found	https://ariadne-infrastructure.eu/	Yes
CHO Repository	Apache License	Apache License	https://github.com/psu-libraries/cho	Yes
D-NET	GNU Affero GPL	GNU Affero GPL	http://www.d-net.research-infrastructures.eu/	No
META-SHARE	BSD License	BSD License	https://github.com/metashare/META-SHARE	No
MINT	GNU Affero GPL	GNU Affero GPL	https://github.com/mint-ntua/Mint	No
MoRe	Not found	Not found	https://locloud.eu/resources/more-repository/	No
Repos	Not found	Not found	https://doi.org/10.1007/978-3-642-04346-8_65	No
Application for smartphone				
Flagship Mobile	Proprietary	Free	https://www.eagle-network.eu/resources/flagship-mobile-app/	Yes
Consultation language				
SPARQL Query Language	W3C Software License	W3C Software License	https://www.w3.org/TR/rdf-sparql-query/	Yes
Mapping				
X3ML Mapping	Apache License	Apache License	http://ceur-ws.org/Vol-1656/paper1.pdf	No
Repository				
Fedora	Apache License	Apache License	https://duraspace.org/fedora/	Yes
Cloud hosting service				
GRNET VIMA	Not applicable	Not applicable	https://grnet.gr/services/cloud-services/vima/	Yes
NoSQL Database Management System				
Apache CouchBD	Apache License	Apache License	https://couchdb.apache.org/	Yes
Apache Solr	Apache License	Apache License	https://lucene.apache.org/solr/	Yes
Elastic Search	Apache License	Apache License	https://www.elastic.co/pt/what-is/elasticsearch	Yes
Mongo DB	Server Side Public License	Server Side Public License	https://www.mongodb.com/	Yes
Relational Database Manager System				
MySQL	GNU Affero GPL	GNU Affero GPL	https://www.mysql.com/	Yes
PostgreSQL	PostgreSQL License	PostgreSQL License	https://www.postgresql.org/	Yes
System for data enrichment				
Multimedia Authoring	Not found	Not found	http://natural-europe.tuc.gr/music/mmat	No

Fonte: Siqueira; Martins (2021c, p. 19)

Não ficou claro como essas tecnologias são usadas ou como elas são integradas. Concomitantemente, a documentação mostrou-se bastante carente de detalhes e com longas discussões sobre o assunto.

Identificamos a tecnologia D-NET como mais frequente; porém, no repositório GitHub observamos que a última atualização foi realizada há 4 anos, o que dificulta a adoção em novos projetos. Quanto às atualizações, o Gráfico 3 mostra a proporção de ferramentas atualizadas em relação às desatualizadas. E podemos observar no Quadro 32 que as ferramentas desenvolvidas especificamente para agregação estão em sua maioria desatualizadas e que as ferramentas com uso mais amplo, com comunidade ativa, estão atualizadas.

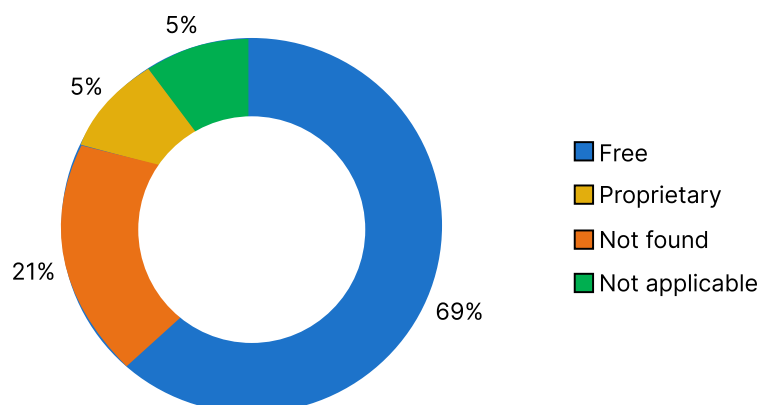
Gráfico 3. Proporção de ferramentas atualizadas e desatualizadas



Fonte: Siqueira; Martins (2021c, p. 20)

Outro ponto importante em relação às ferramentas é sua licença de uso e distribuição, pois isso impacta diretamente nos custos do projeto e na sua viabilidade. O Gráfico 4 mostra as licenças de uso de cada ferramenta.

Gráfico 4. Proporção de licenças gratuitas e proprietárias



Fonte: Siqueira; Martins (2021c, p. 20)

Em relação à execução das etapas serem manuais, semiautomáticas ou automáticas (Q4), a informação não é explícita e não é um tema amplamente discutido nas publicações, embora sua relevância seja inegável, principalmente considerando a escalabilidade. Os projetos CARARE, Hope, EFG, Eagle e ECLAP são apresentados como semiautomáticos, no entanto, isso foi verificado a partir da presença da etapa “*Monitoring*”, por exemplo. Portanto, devemos considerar esses resultados como inconclusivos nesta pesquisa.

4.3 Elementos para a Construção de uma Política de Qualidade de Dados para Agregar Acervos Culturais Digitais: Casos DPLA e Europeana

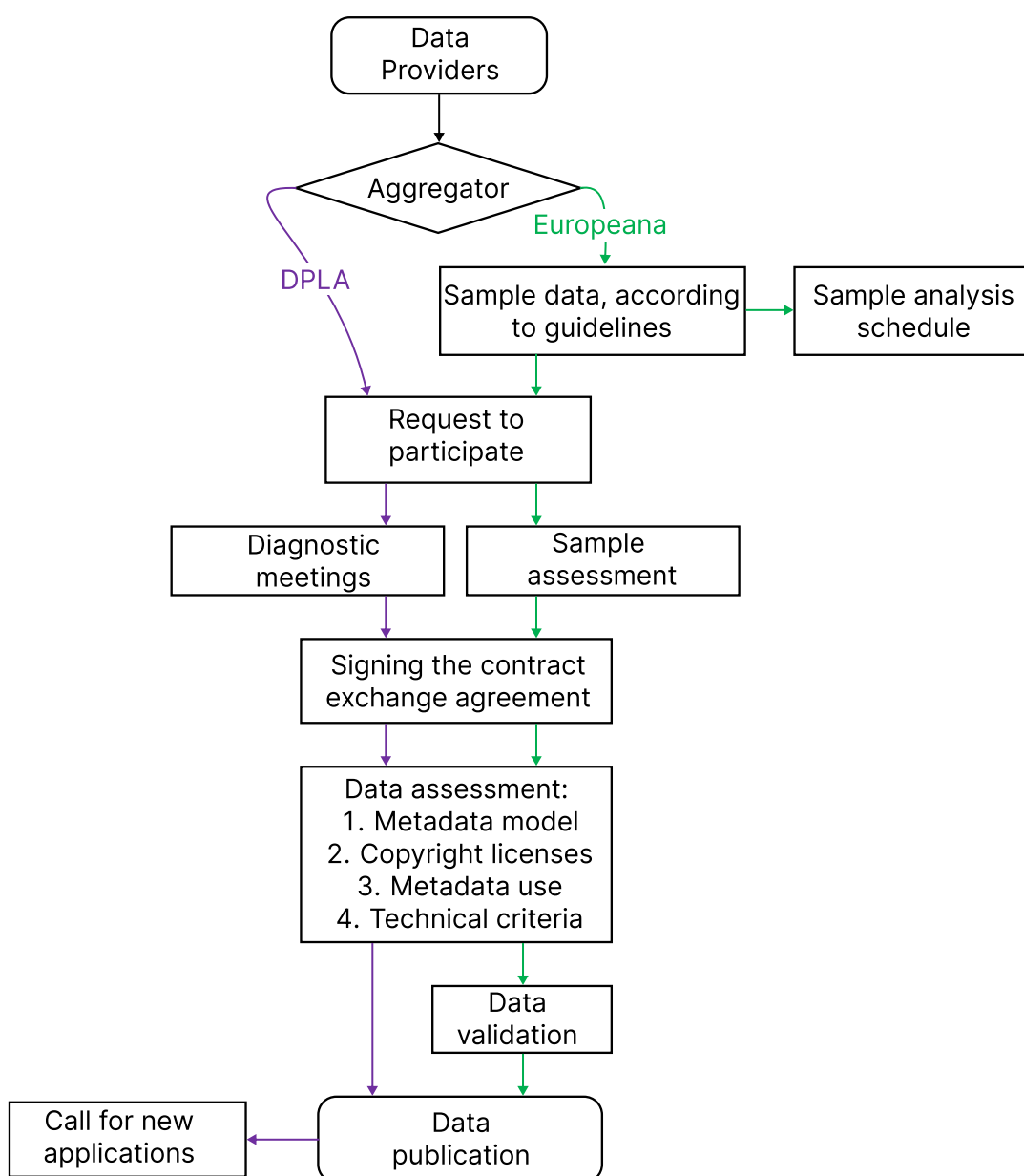
Identificados os workflows de agregação, buscou-se compreender o fluxo de execução dos processos de agregação de dados de duas importantes instituições, a *Digital Public Library of America.Inc* e a *Europeana Foundation*, apresentando, principalmente, como seus parceiros fornecem os dados e uma análise dos elementos para maior qualidade deles.

As instituições foram selecionadas por serem paradigmáticas para o campo, por motivos geopolíticos, uma vez que a *Digital Public Library of America.Inc* atende instituições dos Estados Unidos da América e a *Europeana Foundation*, as de toda a Europa. Considerou-se a *Europeana Foundation* por envolver mais de um país e diversos idiomas de origem e a *Digital Public Library of America.Inc* apenas um país e um idioma; a relevância da *Europeana Foundation* enquanto serviço de agregação e; a diferença significativa nos processos, sendo o da *Digital Public Library of*

America.Inc muito mais simples e manual que o da *Europeana Foundation*. Além disso, em análise exploratória inicial, observou-se que a *Digital Public Library of America.Inc* poderia fornecer um modelo mais próximo da realidade brasileira.

Como resultado, obteve-se o total de nove etapas, que são ilustradas na Figura 29, demonstrando a perspectiva de cada instituição, considerando o acontecimento das ações.

Figura 29. Etapas previstas pela Europeana e DPLA para agregação de dados, com foco na qualidade de dados



Fonte: Siqueira et al. (2021, p. 111)

Na sequência, as etapas são descritas: provedores de dados; solicitação de participação; modelo de metadados; contrato de troca de dados; licenças de uso; cronograma de coleta; uso dos metadados; critérios técnicos para qualidade de dados e; validação e publicação de dados.

4.3.1 Provedores de dados

Provedores de dados são instituições que disponibilizam seus acervos para agregação. Há dois tipos: provedores individuais, que enviam seus dados diretamente à instituição agregadora; e os agregadores, que agregam dados de provedores individuais, normalmente agrupados por um nicho específico, por exemplo, por região ou temática, e enviam os acervos para a instituição agregadora. Tanto a DPLA quanto a Europeia aceitam receber dados dos dois tipos de provedores, considerando critérios particulares a cada uma delas.

Em números, a DPLA possui, atualmente, 43 provedores de dados, sendo destes, 29 agregadores e aproximadamente 41 milhões de itens oriundos de instituições culturais dos Estados Unidos. A Europeia, por sua vez, possui 38 provedores de dados, todos agregadores, e mais de 50 milhões de itens digitalizados sobre o acervo do patrimônio cultural da Europa. Ambas disponibilizam a lista completa de provedores de dados em seus sites, DPLA²⁸ (DPLA, 2023f) e a Europeia²⁹ (EUROPEANA PRO, 2023a).

4.3.2 Solicitação de participação

Para se tornar um provedor de dados, a DPLA solicita o envio de um e-mail que dará início a uma série de conversas obrigatórias com os membros da equipe. Na sequência, para agregadores, é solicitado o preenchimento de um formulário que visa a fornecer um diagnóstico do parceiro de dados. Após a aceitação, os novos parceiros serão notificados e convidados a revisar as próximas etapas, a saber: o contrato para

²⁸ <https://pro.dp.la/hubs/our-hubs>. Acesso em 11 jan. 2023

²⁹ <https://pro.europeana.eu/page/aggregators>. Acesso em 11 jan. 2023

troca de dados, o acordo de associação e o formulário de metadados, tecnologia e informações de conteúdo (DPLA, 2023b).

A Europeana solicita o envio de um e-mail para sua equipe, que solicitará uma amostra dos dados em formato compactado, contendo registros em XML, individuais, ou via OAI-PMH. Os dados devem estar devidamente mapeados de acordo com o modelo de dados EDM e serão verificados e validados em conformidade com todos os aspectos do *Europeana Publishing Guide* (SCHOLZ, 2022), automaticamente, usando a ferramenta de processamento Metis, disponível no GitHub (GITHUB, 2023)

Ambas as instituições seguem um modelo de metadados, apresentado na próxima seção. No entanto, destaca-se que para ser um provedor de dados da Europeana é preciso conhecer previamente sua documentação, incluindo seu modelo de metadados, diferente da DPLA, que realiza o processo junto ao provedor de dados após o aceite, identificando nos metadados de origem quais são compatíveis com o modelo de dados, o *Metadata Application Profile* (MAP), via planilha no Google Drive (DPLA, 2023e).

4.3.3 Modelo de Metadados

Cada instituição possui seu próprio modelo de metadados, o que é essencial para que haja um padrão para agregação. A *Europeana Foundation* desenvolveu o EDM, que é um modelo teórico de dados que permite que estes sejam apresentados de acordo com as práticas dos vários domínios que contribuem com dados para a *Europeana Collection* (CLAYPHAN; CHARLES; ISAAC, 2016).

A DPLA desenvolveu o MAP, que é a base de como os dados são estruturados e validados e orienta como estes são armazenados, serializados e disponibilizados. O MAP foi projetado com base no EDM (GUEGUEN et al., 2017).

4.3.4 Contrato de troca de dados

Quando aceitos, os provedores de dados precisam firmar, oficialmente, a parceria com a instituição agregadora. Parte relevante desse processo é a assinatura de um contrato de troca de dados, que estabelece as regras que norteiam a parceria.

Os documentos são apresentados na íntegra por ambas as instituições. A DPLA utiliza o *Data Exchange Agreement* (DPLA, 2017a) e a Europeana, o *Europeana Data Exchange Agreement* (EUROPEANA FOUNDATION, 2023).

Dentre todo o conteúdo descrito nestes contratos, um ponto em destaque é a necessidade de o provedor de dados garantir que seus metadados e mídias agregadas possuam licença de uso.

4.3.5 Licenças de uso

Conforme supracitado, as licenças são parte fundamental do processo de agregação de dados, pois expressam como o usuário final poderá acessar e reutilizar o objeto digital. A DPLA e a Europeana possuem comportamento semelhante em relação às licenças. No que diz respeito aos metadados, ambas exigem que estes sejam disponibilizados sob a licença *CC0 1.0 Universal Public Domain Dedication*³⁰. Em relação às mídias, a DPLA e a Europeana elencam a seus parceiros 12 declarações de direitos padronizados e recomendações para o uso (DPLA, 2017b; EUROPEANA PRO, 2023c).

A partir de breve análise das licenças, constata-se que essas visam a permitir a reutilização dos metadados museais de forma irrestrita, mas para as mídias, há gradações em relação ao acesso e reuso.

4.3.6 Cronograma de coleta

DPLA não apresenta um cronograma bem definido para inclusão de novos provedores, porém mantém um calendário on-line (DPLA, 2022) que apresenta as datas de realização de novas coletas para atualização dos dados.

A Europeana informa que a primeira análise dos dados é realizada por ordem de chegada. No entanto, coleções temáticas, metadados e conteúdo de qualidade, demandas do usuário e prioridades de negócios, definirão quais coleções receberão mais atenção. Os dados serão processados e publicados continuamente ao longo do ano.

4.3.7 Uso dos Metadados

Os esquemas de metadados utilizados pelas instituições agregadoras possuem uma diversidade de campos e, neste estudo, destacamos, no Quadro 33, as categorias: obrigatório; obrigatório sempre que houver; obrigatório, a partir de um grupo predefinido de metadados e; recomendáveis.

³⁰ <https://creativecommons.org/publicdomain/zero/1.0/>. Acesso em 11 jan. 2023.

Quadro 33. Categorias de requisição de metadados

No.	Metadata	DPLA	Europeana
1	Subject	SourceResource.subject ^d	dc:subject ^c
2	Location	SourceResource.spatial ^d	dcterms:spatial ^c
3	Type or genre	–	dc:type ^c
4	Time	–	dcterms:temporal ^c
5	Date created	SourceResource.date ^d	–
6	Copyright license	SourceResource.rights ^a	edm:rights e sua URI
7	Unique identifier	–	edm:ProvidedCHO, edm:WebResource, ore:Aggregation ^a
8	Thumbnail link	Aggregation.object ^b	–
9	Link to the original record	Aggregation.isShownAt ^a	edm:isShownBy ou edm:isShownAt ^a
10	Metadata populated	–	edm:ugc = true ^b
11	Name of the institution	Aggregation.dataProvider ^a	edm:dataProvider ^a
12	Name of the aggregator	Aggregation.provider ^b	edm:provider ^b
13	Textual objects	–	dc:language ^b
14	Type of resource	SourceResource.type ^b	edm:type ^b
15	Collection title	Collection.title ^b	–
16	Title	SourceResource.title ^a	dc:title ^c
17	Description	–	dc:description ^c

^aRequired

^bRequired when present

^cRequired, based on a predefined metadata group

^dRecommended

Fonte: Siqueira et al. (2020, p. 114)

As descrições dos metadados podem ser encontrados nos documentos sobre os modelos de metadados, MAP (GUEGUEN et al., 2017) e EDM (CLAYPHAN; CHARLES; ISAAC, 2016).

A DPLA lista 11 elementos de metadados, sendo destes, quatro obrigatórios, quatro obrigatórios sempre que houver e, três recomendáveis. A Europeana lista 14 elementos, sendo seis obrigatórios e quatro obrigatórios sempre que houver. A diferença entre os valores se dá pela Europeana organizar a requisição de metadados por grupos, solicitando que apenas um deles seja obrigatório, como é o caso de dc:title e dc:description.

Definidos os metadados, a próxima seção diz respeito à qualidade do conteúdo dos metadados e das mídias.

4.3.8 Critérios técnicos para qualidade de dados

Além dos metadados mínimos, é de extrema importância que haja qualidade no conteúdo disponibilizado, incluindo aqui, as mídias digitais. Assim, para orientar os provedores de dados quanto à qualidade dos próprios metadados e mídias e, a partir desta informação, quando necessário, melhorá-los - tanto a DPLA quanto a Europeia - criaram-se diretrizes técnicas que orientam as instituições interessadas em agregar seus dados.

A DPLA elenca diretrizes em torno de camadas de recomendação para cada propriedade do MAP. A Europeia, por sua vez, apresenta um esquema mais complexo, envolvendo dois conjuntos de camadas, um referente à qualidade de metadados e outro referente à qualidade de mídias.

A DPLA, por meio do documento *DPLA Metadata Quality Guidelines* (DPLA, 2023d) apresenta três camadas: “Requisitos mínimos de qualidade”; “Recomendações para melhor qualidade” e; “Recomendações para a máxima qualidade”.

As diretrizes da camada “Requisitos mínimos de qualidade” são consideradas necessárias à agregação e normalmente estão relacionadas ao uso correto da propriedade do metadado ou a problemas de granularidade. Cada propriedade pode ter ou não requisitos mínimos, mas quando houver, a diretriz deve ser seguida. Um exemplo, o metadado *title*, obrigatório, não possui requisito mínimo de qualidade, no entanto, o metadado, *alternative title*, opcional, tem como requisito mínimo não ser uma cópia do título principal, embora títulos traduzidos sejam aceitáveis.

Na próxima camada, as “Recomendações para melhor qualidade” consistem em um conjunto de diretrizes que visam a melhorar os metadados dentro do contexto de agregação, criando melhor consistência geral por meio de ações como o uso de autoridades ou padrões de conteúdo, por exemplo, o metadado *creator*, opcional, tem como recomendação o uso de listas de autoridade de nomes, como LCNAF (*Library of Congress Names*³¹), VIAF (*The Virtual International Authority File*³²), ULAN (*Union List of Artist Names*³³) etc; um segundo exemplo, o metadado *extent*, opcional, tem

³¹ <https://id.loc.gov/authorities/names.html>. Acesso em 11 jan. 2023

³² <http://viaf.org/>. Acesso em 11 jan. 2023

³³ <https://www.getty.edu/research/tools/vocabularies/ulan/index.html>. Acesso em 11 jan. 2023

como recomendação o uso de padrões de conteúdo como CCO³⁴, RDA (*Resource Description and Access*³⁵), DCRM (*Descriptive Cataloging of Rare Materials*³⁶).

Por último, as “Recomendações para a máxima qualidade” apresentam diretrizes relacionadas ao uso de URIs, ou seja, um identificador único, que possibilita o uso do *LOD*, ou dados abertos ligados. O documento observa que os requisitos relacionados a dados ligados estão nessa camada superior, pois entende-se que ainda é um conceito novo no âmbito das bibliotecas, arquivos e museus (DPLA, 2023b).

No caso da Europeana, os requisitos e recomendações para qualidade de dados são apresentados em dois documentos principais, o *Europeana Publishing Guide* (SCHOLZ, 2022), que detalha as diretrizes para agregação de dados de maneira mais técnica, e o *Europeana Publishing Framework* (EUROPEANA FOUNDATION, 2019) que os apresenta, porém com o objetivo principal de informar aos provedores de dados todas as vantagens de se oferecer metadados e mídias de qualidade.

A Europeana oferece diretrizes para os metadados e para as mídias. A forma como os metadados podem ser disponibilizados pelos provedores de dados é classificada em camadas ‘A’, ‘B’ e ‘C’, cuja ordem representa o nível de reuso, sendo o menor representada pela letra ‘A’ e o maior, pela letra ‘C’. A camada ‘A’ representa a *Europeana Collections* como um motor de busca básico; a ‘B’, como uma plataforma de exploração e a ‘C’, como uma plataforma de conhecimento.

Quando a qualidade dos metadados corresponder à camada A, a instituição provedora de dados pode utilizar a plataforma *Europeana Collections* como um motor de busca básico, ou seja, os usuários podem encontrar objetos específicos se tiverem as informações precisas para encontrá-lo, podendo utilizar os filtros para a busca. Para que seja atingido esse requisito, pelo menos 25% dos elementos de metadados que foram fornecidos e são considerados relevantes precisam de minimamente um rótulo de idioma. Além disso, os metadados devem incluir pelo menos um dos elementos ativadores extraídos dos cenários de descoberta e uso.

Os cenários de descoberta e uso foram desenvolvidos pelo Comitê de Qualidade de Dados (*Data Quality Committee - DQC*) e refletem as necessidades dos

³⁴ <https://vraweb.org/resources/cataloging-cultural-objects/>. Acesso em 11 jan. 2023

³⁵ <https://www.rdatoolkit.org/about>. Acesso em 11 jan. 2023

³⁶ <http://rbms.info/dcrm/dcrmmss/>. Acesso em 11 jan. 2023

usuários ao acesso à informação, listando os problemas e propondo ações. Estes cenários visam a demonstrar situações específicas aos provedores de dados para motivá-los a fornecer metadados cada vez mais completos, com inclusão de elementos ativadores (elementos de metadados opcionais), que podem suportar cenários de uso específicos, visando à maior recuperação de informações e melhoria dos serviços nas coleções (CHARLES; ISAAC; HILL, 2015).

Caso os metadados sejam disponibilizados em concordância com a camada B, a instituição pode utilizar a *Europeana Collections* como plataforma de exploração, ou seja, os usuários poderão encontrar os objetos buscando tanto por informações específicas quanto gerais. Dessa forma, as coleções poderão ser contextualizadas e encontradas em coleções temáticas; a descoberta pode ser aprimorada por meio da possibilidade de busca multilíngue, além de serem apresentadas no recurso de páginas de entidades da plataforma. Para tal, pelo menos 50% dos elementos de metadados relevantes fornecidos precisam ter pelo menos um rótulo de idioma. Os metadados também devem incluir o mínimo de três elementos ativadores que abrangem pelo menos dois cenários distintos de descoberta e uso, e conter uma classe contextual com todos os elementos mínimos necessários ou um link para o vocabulário LOD.

Classes contextuais visam a dar significado aos objetos, facilitando a recuperação. O EDM acomoda quatro classes contextuais, edm:Agent, edm:Place, edm:TimeSpan e skos:Concept, que podem capturar entidades distintas relacionadas aos objetos do acervo do patrimônio cultural. Recomenda-se o uso de referências para vocabulários multilíngues e LOD, como *Getty Art & Architecture Thesaurus (AAT)*, Wikidata ou Geonames (CHARLES; ISAAC; HILL, 2015).

Se os dados correspondem à camada de qualidade C, o provedor de dados pode usar a *Europeana Collections* como uma plataforma de conhecimento. Isso implicaria em o usuário encontrar com maior precisão o objeto que busca sem ter informações específicas sobre ele, pois os metadados contextuais seriam fornecidos e se abriria uma rede de conhecimentos vinculados, permitindo a descoberta das coleções por meio das relações entre elas e suas identidades. Dessa forma, os metadados podem ser utilizados nos projetos parceiros como a *Historiana*³⁷ (projeto

³⁷ <https://historiana.eu/>. Acesso em 11 jan. 2023

para o setor educacional) e CLARIN³⁸ (*European Research Infrastructure for Language Resources and Technology* - projeto de infraestrutura para pesquisa) e demais projetos que envolvam as indústrias criativas. Além disso, essas informações contextualizadas e enriquecidas podem ser utilizadas em aulas, estudos e aplicações, o que traria um maior alcance e exposição das coleções dentro e além da plataforma *Europeana Collections*. Para que isso seja possível, pelo menos 75% dos elementos de metadados considerados relevantes precisam ter pelo menos um rótulo de idioma. Os metadados devem incluir pelo menos três elementos habilitadores distintos, cobrindo pelo menos dois cenários distintos de descoberta e uso, e duas classes contextuais distintas com todos os elementos mínimos necessários ou links para os vocabulários de LOD.

Com base nas informações levantadas pode-se perceber que ambos os agregadores, Europeana e DPLA, apresentam as primeiras camadas como requisitos mínimos de entrada na rede, mas também orientam as instituições para que de forma progressiva elas possam tratar seus dados de modo a oferecê-los com melhor qualidade, e assim explorar de forma mais eficiente os recursos de busca e descoberta de informação fornecidos pelas plataformas.

A forma como as mídias são disponibilizados pelos provedores de dados é classificada por números, '1' para *Europeana Collections* como motor de pesquisa; '2' como vitrine; '3' como uma plataforma para o reuso não comercial e '4' como uma plataforma de livre reuso.

As camadas 1, 2, 3 e 4 se referem à forma como as mídias podem ser disponibilizadas pelos provedores de dados. É importante ressaltar que a numeração representa uma gradação em relação à possibilidade de reuso das mídias disponibilizadas, ou seja, o número 1 representa o menor uso possível e o número 4 o maior uso possível.

Se os dados relativos às mídias atingirem a camada de qualidade 1, ela permite que o provedor de dados use a plataforma *Europeana Collections* como um motor de pesquisa por onde os usuários podem navegar e encontrar objetos relativos à sua busca. Dessa forma, o objeto digital é visualizado por meio do site da instituição de origem e não no ambiente da *Europeana Collections*.

³⁸ <https://www.clarin.eu/>. Acesso em 11 jan. 2023

Se a qualidade dos dados do provedor atingir a camada 2, essa permite que a plataforma *Europeana Collections* seja usada como vitrine para os acervos dos provedores de dados, em que os usuários visualizam versões de boa qualidade de suas coleções. Dessa forma, os objetos também podem ser apresentados em alguma coleção temática, havendo assim uma maior exposição das coleções.

A camada 3 de qualidade de dados permite que o provedor use a *Europeana Collections* como uma plataforma distribuidora para o reuso não comercial, ou seja, os usuários podem ver versões de qualidade dos objetos diretamente na plataforma e incorporar as coleções em projetos existentes como a Historiana ou CLARIN. Isso seria possível por meio das declarações de direitos de uso que permitem algum reuso. Dessa forma, as coleções podem ser utilizadas em sites, aplicativos e outros serviços sem fins comerciais.

A camada 4 permite que o provedor de dados utilize a *Europeana Collections* como uma plataforma de livre reuso, dessa forma, os dados referentes às coleções podem ser incorporados em projetos e parcerias com as indústrias criativas; podem ser compartilhados com a hashtag *#OpenCollections* nas mídias sociais; podem ser usados em plataformas abertas como as plataformas da Fundação Wikimedia e também podem ser utilizados em *hackathons*. Para que isso seja possível, a coleção deve ser disponibilizada sob declarações de direitos que permitam a reutilização livre e gratuita, dessa forma as coleções podem ser usadas em sites, aplicativos, serviços e produtos comerciais e não comerciais.

Os requisitos específicos para cada tipo de mídia (imagem, áudio, vídeo e 3D) estão descritos detalhadamente no documento *Europeana Publishing Guide* (SCHOLZ, 2022).

4.3.9 Validação e Publicação de dados

Ao final do processo, a Europeana disponibiliza aos provedores de dados um link para a visualização prévia ao provedor de dados que confirmará se a publicação pode avançar ou se serão necessárias alterações (SCHOLZ, 2022). A DPLA não cita esse tipo de serviço.

De acordo com os dados obtidos, podemos observar que os processos de agregação da DPLA são mais simplificados e manuais se comparados aos processos complexos da Europeana. Um exemplo é a forma como a Europeana solicita que os

dados de amostra sejam encaminhados, já seguindo a documentação disponível e, de início, realiza a validação dos dados, automaticamente, via *software* Metis (GITHUB, 2023). Por outro lado, a DPLA, por exemplo, mapeia os metadados por meio uma planilha eletrônica, de forma manual, após a aceitação da instituição como provedora de dados.

No caso da Europeia, o nível de requisito é muito rígido e as etapas, complexas, o que é positivo se considerarmos os efeitos pedagógicos para o campo, visto que nos apresenta caminhos que devem ser percorridos, mas, por outro lado, gera dificuldade extra aos provedores de dados interessados. O processo parece demandar certo nível de maturidade técnica e de gestão por parte dos provedores interessados na publicização de seus dados na rede.

A DPLA com seu modelo de dados MAP foca numa interface comum para descrever as mídias, refletindo informação sobre o conteúdo, o que difere da Europeia com seu modelo EDM, cuja distinção ocorre na descrição subjacente ao conteúdo (ou na descrição do conhecimento sobre o recurso) por meio de tecnologias semânticas, especialmente com uso de vocabulários semânticos disponíveis em ambiente LOD, com vistas à publicação formal e explícita de conhecimento na Web.

Logo, provedores de dados que se beneficiam desse tipo de serviço de agregação como o do *Europeana Collections*, conseguem ampliar possibilidades de pesquisa para seus usuários em redes de conhecimentos ligadas semanticamente na Web de dados. Isso inclui laboratórios de pesquisa multimídia, indústrias criativas, setores culturais e de educação, mídias sociais - a exemplo da Wikidata -, apenas para citar alguns exemplos, o que possibilitaria maior profundidade na exploração da informação dentro e para além da plataforma de agregação.

A Europeia atua em um cenário que presume maior maturidade organizacional nos processos de gestão da informação de seus parceiros. Vale destacar que a instituição se vale essencialmente de agregadores nacionais nos respectivos países europeus para coletar dados, além de alguns projetos temáticos específicos, como na área do cinema, por exemplo. Dessa forma, ela atua a partir de parâmetros de qualidade e estrutura mais rígida dos dados para iniciar o processo. Já na DPLA a situação se mostra bastante diferente. Ela atua diretamente na possibilidade de apoiar uma instituição a mapear os seus dados para o modelo requerido, facilitando para que a instituição tenha um suporte técnico e apoio

institucional para como lidar com seus dados e em como fornecer dados de maneira adequada para serem agregados. Esse efeito pedagógico e de suporte às instituições é de especial importância para a presente pesquisa.

Percebe-se que ambas as instituições buscam fornecer recomendações relacionadas à qualidade dos metadados dentro do contexto de agregação, além de um *framework* geral de metadados para aplicações em mídias, cujos conteúdos sejam processáveis por máquina. Entretanto, observam-se diferenças pontuais nas propostas dos *frameworks* envolvidos no que diz respeito à modelagem de informação multimídia. Apesar de a visão sobre metadados ser similar tanto para a DPLA quanto para a Europeia, algumas diferenças em relação a propósito semântico são evidentes. A Europeia possui explicitamente a intenção de colaborar e se tornar um provedor de dados para as redes semânticas de dados abertos ligados. A DPLA não parece possuir esse objetivo de forma explícita.

Também é importante destacar a observação que a Europeia aponta para usos possíveis dos dados agregados em cenários de maior complexidade, desenvolvendo e estabelecendo parcerias com projetos temáticos de reuso de dados, assim como relação com amplas bases de conhecimento, como a Wikidata. Sem dúvida, pode-se inferir que essa maior complexidade nos procedimentos de tratamento e agregação de dados pode se reverter em importantes benefícios em dados de maior qualidade e com maior potencial de reuso em aplicações e serviços futuros.

Em relação aos requisitos de exigência dos metadados mínimos necessários para agregação em suas diferentes categorias de obrigatoriedade, a DPLA também se apresentou como um conjunto menor de dados, solicitando apenas 11 campos, e a Europeia, 14. Uma das diferenças fundamentais que se observou foi o requisito da linguagem, bastante destacado em vários pontos pela Europeia e com menor ênfase pela DPLA. Sem dúvida, o fato de a Europeia agregar diferentes países e diferentes línguas oficiais torna esse fator de maior complexidade, gerando demandas que vão se traduzir em elementos de seu modelo de dados.

Com base na revisão de literatura realizada, pode-se constatar a alta relevância do mapeamento de dados para o sucesso do projeto de agregação do Ibram. Na pesquisa bibliográfica, conforme Quadro 24. Sumarização das etapas dos workflows de agregação, pudemos identificar que apenas dois agregadores não se

referiram ao mapeamento. Na pesquisa bibliográfica sobre os agregadores brasileiros, pudemos identificar que eles optaram pelo Dublin Core. Na RSL, o Gráfico 2. Ocorrência das etapas para agregação de dados, nos mostra que em todos os agregadores encontrados, há a realização do mapeamento de dados. No estudo sobre a DPLA e Europeana, podemos identificar o uso dos modelos MAP e EDM, utilizados para mapeamento dos dados.

A análise dos diferentes workflows de agregação de dados nos permitiu compreender o atual estado da arte do tema. Dessa forma, foi possível identificar que não há consenso em relação à quantidade de etapas, às nomenclaturas ou às tecnologias utilizadas, demonstrando o estágio inicial em que o tema se encontra ou revelando que as soluções são altamente customizadas, exigindo soluções locais para problemas específicos.

É importante reconhecer que esse é um tema ainda novo para a Ciência da Informação e que esforços de pesquisa e desenvolvimento devem ser feitos para que se compreendam as possíveis aplicações dessas tecnologias, dado que elas não apenas são novas técnicas, mas representam novas formas de se pensar nos dados e em um ecossistema completo de serviços analíticos.

Assim, desenvolveu-se um novo serviço, apresentado a seguir

4.4 Brasileira Museus: Serviço de Busca e Recuperação da Informação Museal

O Brasileira Museus é um serviço de busca e recuperação da informação desenvolvido especialmente para o Ibram, composto por dois produtos: o painel de visualização analítica e o portal agregador.

O painel de visualização oferece recursos para uma gestão da documentação museológica mais eficaz, pois apresenta, de forma unificada e interativa, uma visão macro de todos os seus museus, por meio de gráficos, métricas, tabelas e mapas. O portal agregador é uma plataforma centralizadora que permite a busca e recuperação de objetos digitais e direciona os acessos aos museus de origem.

As etapas para construção do serviço estão apresentadas a seguir.

4.4.1 Etapas e tecnologias

Foram identificados 20 diferentes projetos internacionais para a agregação de dados culturais, muitos dos quais tratam da *Europeana Foundation*, ou seja, diferentes

projetos que utilizam processos e tecnologias semelhantes. Nem todos os artigos indicaram quais tecnologias usavam, mas a partir dos que o fizeram, constatou-se que há várias tecnologias diferentes em uso e não uma mais usada. Verificou-se também se elas estavam atualizadas ou não (para efeito desta pesquisa, consideramos as atualizações ocorridas nos últimos dois anos). O Quadro 34 mostra os principais agregadores encontrados.

Quadro 34. Agregadores de dados internacionais

Data aggregator	Link	Current
Ariadne	https://ariadne-infrastructure.eu/	Yes
CHO repository	https://github.com/psu-libraries/cho	No
D-NET	https://github.com/dnet-team/dnet-basic-aggregator	No
META-SHARE	https://github.com/metashare/META-SHARE	No
MINT	https://github.com/mint-ntua/Mint	No
MoRe	http://more.dcu.gr/	No
Reposx	https://doi.org/10.1007/978-3-642-04346-8_65	No
Supplejack	https://digitalnz.org/developers/supplejack	Yes

Fonte: Siqueira; Martins; Medeiros (2022, p. 86)

Identificamos a D-NET com a tecnologia mais frequente; porém, no GitHub observamos que a última atualização do repositório foi realizada há mais de quatro anos, o que torna difícil a adoção em novos projetos.

No Brasil, embora existam iniciativas que visam a integrar acervos culturais, estes são limitados, isolados e requerem suporte técnico e modelos de governança escaláveis para o território brasileiro. Constatou-se que a simples replicação de soluções internacionais não seria possível, porque são diversificadas e atendem às especificidades de cada localidade; não há documentação suficiente disponível para tentar implementar algo e não há consenso em melhores soluções e/ou tecnologias. Ou seja, mesmo que fosse possível replicar, provavelmente não seria efetivo no contexto brasileiro, pois são realidades diferentes, considerando infraestrutura, investimento, mão de obra especializada, entre outros.

Estas deveriam ser modernas, gratuitas, com comunidade ativa, com alta performance, escaláveis, que permitissem integração com outras ferramentas e que também fossem relativamente simples de implantar.

Após alguns testes com ferramentas como Supplejack³⁹ e Vufind⁴⁰ sem resultados positivos, a pilha Elastic Stack⁴¹ demonstrou atender os requisitos mínimos, trazendo excelentes resultados, tendo sido escolhida para desenvolver o agregador.

Ressalta-se que a tecnologia atende aos critérios mínimos buscados e não se pretende defini-la como a melhor, mas sim como a mais adequada ao contexto da pesquisa.

A pilha Elastic Stack, também conhecida como ELK Stack, é composta por quatro ferramentas: a Beats - Filebeat⁴², a Logstash⁴³, a Elasticsearch⁴⁴ e a Kibana⁴⁵, visando a obter dados de maneira confiável e segura, a partir de qualquer fonte, em qualquer formato, para pesquisar, analisar e visualizar tudo em tempo real, simplificando a ingestão, a visualização e os relatórios de dados (ELASTIC, 2023a, tradução nossa).

Beats é uma plataforma para agentes de dados de finalidade única, enviar dados de centenas ou milhares de computadores para o Logstash ou o Elasticsearch. São sete agentes de dados, dentre os quais utilizamos o Filebeat, que oferece uma maneira leve de encaminhar e centralizar logs e arquivos, independentemente do dispositivo coletado. Ele utiliza um protocolo para suportar a demanda de volumes maiores de dados, ou seja, se o Logstash está ocupado processando dados, ele avisa o Filebeat para desacelerar sua leitura (ELASTIC, 2023b, tradução nossa).

Logstash é um mecanismo de coleta de dados em tempo real que pode unificar dinamicamente os dados de fontes díspares e normalizar os dados em destinos previamente definidos. Um dos produtos principais do Elastic Stack processa os dados independentemente do formato e da complexidade do lado do servidor e permite fazer a ingestão de dados de várias fontes simultaneamente, enriquecê-los e transformá-los antes de serem indexados no Elasticsearch (ELASTIC, 2023f, tradução nossa).

O Logstash oferece uma rica coleção com mais de 200 plugins de entrada, filtro, codec e saída. Os plugins estão disponíveis como pacotes independentes chamados *gems* e hospedados em [RubyGems.org](https://www.rubygems.org), que podem ser combinados e orquestrados

³⁹ <https://digitalnz.github.io/supplejack/>. Acesso em 11 jan. 2023

⁴⁰ <https://vufind.org/vufind/>. Acesso em 20 jan. 2023

⁴¹ <https://www.elastic.co/elastic-stack/>. Acesso em 11 jan. 2023

⁴² <https://www.elastic.co/beats/filebeat>. Acesso em 11 jan. 2023

⁴³ <https://www.elastic.co/logstash/>. Acesso em 11 jan. 2023

⁴⁴ <https://www.elastic.co/elasticsearch/>. Acesso em 11 jan. 2023

⁴⁵ <https://www.elastic.co/kibana/>. Acesso em 11 jan. 2023

em diferentes entradas, filtros e saídas. Caso nenhum plugin atenda à demanda, novos plugins podem ser facilmente desenvolvidos a partir da API para desenvolvimento de plugins e um gerador de plugins⁴⁶ (ELASTIC, 2023f, 2023i, tradução nossa).

Elasticsearch é um software que provê uma interface RESTful de pesquisa e análise de dados. É um mecanismo de busca e análise de dados distribuído para todos os tipos de dados. Sua velocidade, escalabilidade, capacidade de indexar muitos tipos de conteúdo e seu amplo conjunto de recursos significam que ele pode ser usado para inúmeros casos de uso. Os dados recebidos são indexados e os usuários podem executar consultas complexas com base em seus dados e usar agregações para recuperar resumos complexos dos dados (ELASTIC, 2023h, tradução nossa).

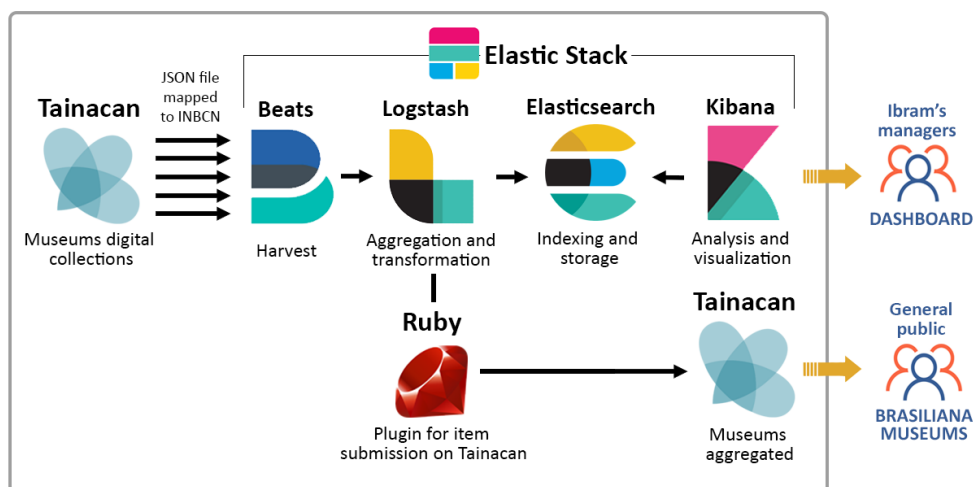
É possível criar vários índices (*index*) do Elasticsearch, que são um conjunto de documentos que estão relacionados entre si, armazenados como documentos JSON. O Elasticsearch usa uma estrutura de dados chamada *índice invertido*, que é projetada para permitir buscas de texto completo muito rápidas (ELASTIC, 2023h, tradução nossa).

Por fim, o Kibana é uma interface de usuário para visualização de dados do Elasticsearch e navegação no Elastic Stack. É uma ferramenta de visualização e gerenciamento de dados que fornece histogramas em tempo real, gráficos e mapas. Inclui aplicações avançadas como o Canvas, que permite aos usuários criar infográficos dinâmicos personalizados com base em seus dados, e o Elastic Maps para visualizar dados geoespaciais (ELASTIC, 2023c).

O conjunto de ferramentas permitiu o desenvolvimento de dois fluxos de trabalho diferentes, cada um para sua própria finalidade. O caminho 1 refere-se à criação do painel analítico; e o caminho 2, à criação do Portal Brasileira Museus. A Figura 30 apresenta a sua arquitetura completa.

⁴⁶ <https://www.elastic.co/guide/en/logstash/5.5/plugin-generator.html>. Acesso em 13 jan. 2023

Figura 30. Arquitetura do serviço de busca e recuperação da informação Brasileira Museus



Fonte: elaborada pelos autores (2022)

Dessa forma, para o serviço de busca e recuperação, Portal e Painel Analítico, o fluxo de trabalho é composto por seis etapas, a saber:

1. mapeamento dos metadados dos museus de origem para o padrão do Ibram, disponibilizando dados no formato aberto, no Tainacan;
2. coleta automatizada dos metadados dos objetos digitais das coleções, por meio do Beats - Filebeat;
3. agregação e transformação dos dados, por meio do Logstash;
4. envio e armazenamento dos metadados no Tainacan, para criação do Portal Brasileira Museus, ainda por meio do Logstash, porém, a partir do desenvolvimento de um novo plugin escrito com a linguagem de programação Ruby;
5. envio e armazenamento dos metadados integrados para o Elasticsearch;
6. criação do painel analítico, interativo, composto por elementos de visualização, com Kibana.

Cada etapa é descrita a seguir.

4.4.2 Tainacan: mapeando os museus de origem

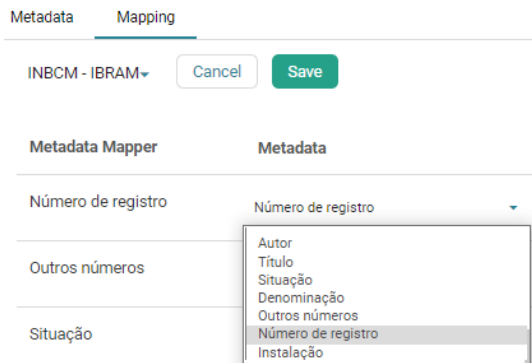
Para que a agregação seja realizada é essencial que cada museu disponibilize seus dados em formato aberto. No entanto, também é essencial que todos os museus tenham compatibilidade nos dados fornecidos, ou seja, é preciso assegurar que os dados dialoguem entre si, que todos possuam o mesmo padrão de metadados, resultando em uma mesma abordagem sintática e semântica na agregação. Assim, o requisito mínimo definido para promover a agregação de dados é a disponibilidade de dados abertos, no formato JSON, mapeado para o INBCM.

Por padrão, o Tainacan permite o mapeamento dos dados para o padrão Dublin Core. Tal procedimento é denominado na literatura como *crosswalk*: os elementos de um esquema de metadados são associados a outro, viabilizando a interoperabilidade e permitindo que coleções heterogêneas possam ser pesquisadas simultaneamente (CHAN; ZENG, 2006).

A partir dessa necessidade, a equipe do Tainacan desenvolveu uma nova funcionalidade, o plugin intitulado “*Tainacan Mappers IBRAM INBCM*”, que por ser específico para o Ibram, não compõe o pacote padrão do Tainacan, devendo ser instalado para uso. Dessa forma, a partir da versão 0.15 do Tainacan é possível adicionar o plugin desenvolvido especificamente para o Ibram, que mapeia os metadados para este padrão (SIQUEIRA; MARTINS; MEDEIROS, 2022).

O plugin oferece uma interface simplificada para mapeamento de metadados de uma coleção, Figura 31, em grau de equivalência um a um, ou seja, um metadados do museu de origem pode ser mapeado para um metadado do INBCM. Em elementos onde não há equivalência exata, os metadados são desconsiderados. Em relação ao mapeamento semântico, ele é realizado por pessoal especializado (SIQUEIRA; MARTINS; MEDEIROS, 2022).

Figura 31. Interface de mapeamento dos metadados para INCBM, no Tainacan



Fonte: Siqueira; Martins; Medeiros (2022, p. 189)

No Tainacan o usuário é livre para modelar quaisquer metadados que desejar e determinar se eles são públicos ou privados. Dada essa característica, embora o INBCM descreva quais metadados são obrigatórios ou opcionais e não especifique quais devem ser mantidos privados ou públicos, o mapeador considera todos os metadados existentes, desprezando as validações de metadados privados e obrigatórios. Porém, só disponibiliza no arquivo final, no formato JSON, os metadados mapeados e públicos.

Para exemplificar, o Quadro 35 apresenta parte do mapeamento do Museu da Inconfidência, que possui, ao todo, 50 metadados, sendo 23 privados e 27 públicos.

Quadro 35. Mapeamento de dados parcial do Museu da Inconfidência

INBCM Metadados	Museu da Inconfidência		Mapeado?
	Metadados	Status	
N. de registro	N. de registro	público	Sim
Outros números	Outros números	privado	Sim
Diâmetro	Não há	público	Não
Não há	Editor	privado	Não

Fonte: Siqueira; Martins; Medeiros (2022, p. 189)

Dessa forma, embora seja possível mapear dados privados, apenas os públicos serão disponibilizados, não havendo necessidade de o usuário se preocupar com essas questões.

A Figura 32 apresenta um exemplo de arquivo JSON, do Museu da Abolição, no qual constam os dados de paginação, o código da coleção e os dados mapeados de um item. Apenas os dados do metadado “Número de registro” está aberto, para exemplificação.

Figura 32. Trecho do arquivo JSON do Museu da Abolição

```

▼ pagination:
  per_page: 50
  total_page: 7
  total_items: 301
  current_page: 1
  next_page: 2
  next_page_link: "/tainacan/v2/collection/7/items?page=5"
▼ items:
  ▼ 0:
    id: 5668
    data:
      ▼ numero-registro:
        label: "Número de registro"
        value: "2016.199"
      denominacao: {...}
      titulo: {...}
      autor: {...}
      classificacao: {...}
      resumo-descritivo: {...}
      dimensoes-altura: {...}
      dimensoes-peso: {...}
      dimensoes-largura: {...}
      dimensoes-espessura: {...}
      material-tecnica: {...}
      local-producao: {...}
      data-producao: {...}
      condicoes-reproducao: {...}
      midias-relacionadas: {...}
    url: "https://museudaabolicao.acervos.museus.g
    document:
      type: "attachment"
      value: "https://museudaabolicao.acervos.museus.g
    thumbnail:
      ▼ 0: "https://museudaabolicao.acervos.museus.g
      1: 2560
      2: 1665
      3: false
    creation_date: "março 8, 2021"
    modification_date: "abril 26, 2021"

```

Fonte: API JSON do Museu da Abolição (2022) ⁴⁷

Como mostra a Figura 32, além dos metadados mapeados, o arquivo JSON conta com: as informações de paginação (no Tainacan, a lista de itens é separada em páginas); o número de identificação do acervo (código único gerado pelo Tainacan para identificar cada coleção/acervo); o URL do item no museu de origem e o URL da imagem de miniatura do item.

⁴⁷ <https://museudaabolicao.acervos.museus.gov.br/wp-json/tainacan/v2/collection/7/items/?perpage=50&order=DESC&orderby=date&exposer=json-flat&mapper=inbcm-ibram&paged=1>. Acesso em 11 jan. 2023

A partir deste arquivo, o Filebeat coletará todos os dados, como detalhado a seguir.

4.4.3 Beats - Filebeat: coleta automatizada dos itens de museus

O Filebeat trabalha a partir de dois arquivos de configuração. O primeiro diz respeito à configuração necessária para coleta de cada acervo, de cada museu. Esse arquivo contém basicamente o URL da API JSON que será coletada, a configuração sobre a paginação dos itens e a inclusão de quatro novos dados, que caracterizam cada museu no agregador, são eles: Museu, Coleção, Cidade e Unidade Federativa - UF. Os itens são incluídos manualmente no código do Filebeat e no Portal Brasileira Museus tornam-se taxonomias importantes a busca e a recuperação dos dados. Dessa forma, para os 20 museus agregados, há 22 arquivos configurados, pois o Museu Casa de Hera e o Museu Histórico Nacional possuem dois acervos/coleções cada.

O segundo trata-se do arquivo de configuração, que orquestra o acesso a arquivos individuais e envia seus resultados para o Logstash. Estes arquivos estão disponíveis no GitHub ⁴⁸ ⁴⁹.

4.4.4 Logstash: agregação de coleções e transformação de dados

Após a coleta das fontes de dados realizada pelo Filebeat, o Logstash as recebe, agrega-as e realiza transformações nos dados, quando necessário.

Em suma, o Logstash funciona da seguinte forma: há dois elementos obrigatórios, o de entrada (*input*) e o de saída (*output*), além de um elemento opcional, o filtro (*filter*). Os elementos de entrada consomem dados de uma fonte, os elementos de filtro modificam os dados conforme as especificações desejadas e os elementos de saída gravam os dados em um destino (ELASTIC, 2023e).

Para solução do Ibram foram necessários três filtros padrão: o *prune*, o *mutate* e o *fingerprint*. O filtro *prune* remove campos de eventos com base em listas brancas ou listas negras. Isso pode ser útil para um filtro JSON que cria campos com nomes desconhecidos, dos quais se deseja manter apenas um subconjunto deles (ELASTIC,

⁴⁸ https://github.com/tainacan/tainacan-elk/tree/museus_federais/filebeat/inputs.d. Acesso em 11 jan. 2023

⁴⁹ https://github.com/tainacan/tainacan-elk/blob/museus_federais/filebeat/filebeat.yml. Acesso em 11 jan. 2023

2023d). O filtro realiza a limpeza dos dados recebidos, entregando os dados relevantes, configurados no Logstash, removendo quaisquer outros enviados pelo Filebeat.

O filtro *mutate* permite que que modificações gerais sejam executadas nos dados. No Ibram, foram usados os seguintes subfiltros de *mutate*: o *capitalize*⁵⁰, para que o campo “classificação”, tenha sempre a primeira letra em maiúsculo; o *lowercase*⁵¹, para que o campo “URL”, por exemplo, seja composto somente por letras minúsculas; o *remove_field*⁵², para retirar o campo *payload*, coletado pelo Filebeat; o *rename*⁵³, para dar os nomes corretos a cada campo; o *strip*⁵⁴, para retirar os possíveis espaços em branco antes e depois de todos os dados e; o *split*⁵⁵, para separar dados, como por exemplo, no metadado “material-tecnica”.

Outro filtro utilizado, essencial para a solução de agregação, é o *fingerprint*, ou filtro de impressão digital, que visa a criar hashes consistentes a partir de um ou mais campos, armazenando o resultado em um novo campo. Segundo Elastic (2023g), este filtro pode ser utilizado para criar IDs de documentos consistentes quando os eventos são inseridos em Elasticsearch. Essa abordagem significa que os documentos existentes podem ser atualizados em vez novamente criados.

Para o código único foi criado uma hash utilizando dois campos: a instalação, ou seja, o nome do museu, e o código de identificação do item, ambos coletados no museus de origem. Estes campos foram escolhidos por não poderem ser modificados pelo museu de origem, garantindo que o item sempre tenha o mesmo código único.

Além dos filtros padrão, o Logstash permite o desenvolvimento de novos filtros. Assim, a partir dos dados coletados e tratados, o intuito é que cada item seja enviado a uma nova instalação Tainacan, que será o Portal Brasileira Museus. O filtro intitulado *tainacan_submission* foi elaborado em conjunto com Vinícius Nunes

⁵⁰ <https://www.elastic.co/guide/en/logstash/current/plugins-filters-mutate.html#plugins-filters-mutate-capitalize>. Acesso em 11 jan. 2023

⁵¹ <https://www.elastic.co/guide/en/logstash/current/plugins-filters-mutate.html#plugins-filters-mutate-lowercase>. Acesso em 11 jan. 2023

⁵² https://www.elastic.co/guide/en/logstash/current/plugins-filters-mutate.html#plugins-filters-mutate-remove_field. Acesso em 11 jan. 2023

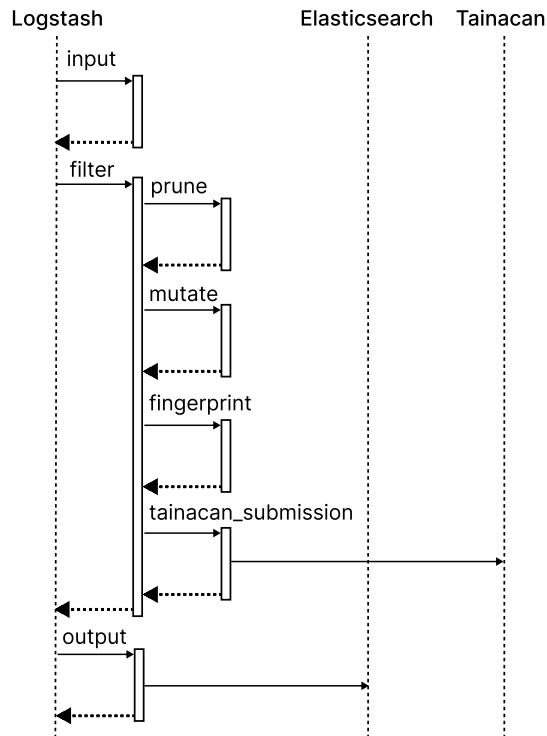
⁵³ <https://www.elastic.co/guide/en/logstash/current/plugins-filters-mutate.html#plugins-filters-mutate-rename>. Acesso em 11 jan. 2023

⁵⁴ <https://www.elastic.co/guide/en/logstash/current/plugins-filters-mutate.html#plugins-filters-mutate-strip>. Acesso em 11 jan. 2023

⁵⁵ <https://www.elastic.co/guide/en/logstash/current/plugins-filters-mutate.html#plugins-filters-mutate-split>. Acesso em 11 jan. 2023

Medeiros, da equipe de desenvolvimento do Tainacan. A Figura 33 mostra a estrutura base do arquivo Logstash.

Figura 33. Estrutura do arquivo Logstash

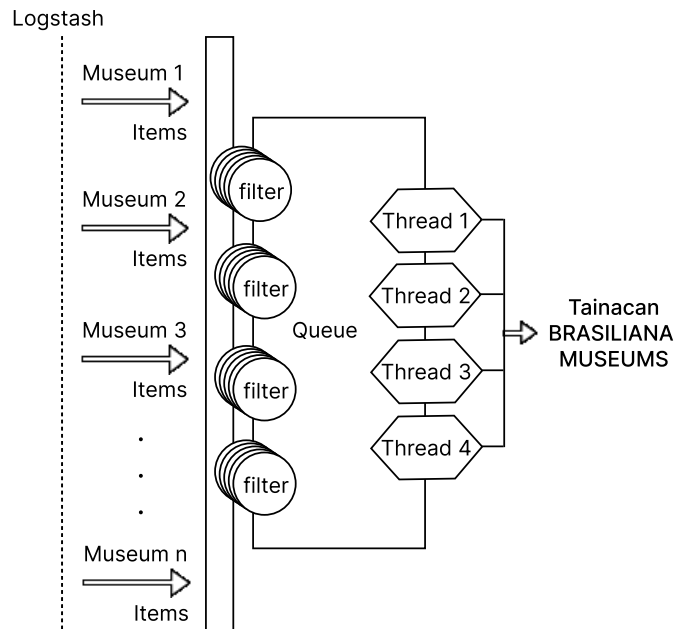


Fonte: Siqueira; Martins; Medeiros (2022, p. 91)

Para o melhor desempenho das submissões no Tainacan, que começa a receber um alto fluxo de inserções simultâneas vindas do Logstash, uma fila de processos, que organiza e encaminha os dados para a API do Tainacan, foi implementado. Todos os itens processados são incluídos no final do fila, enquanto um conjunto de threads (quatro, por padrão) são responsáveis por processar a saída dos itens dessa fila, efetuando o envio dos itens no Tainacan. Após o processamento de todos os itens da fila, eles estarão disponíveis para acesso na instalação do Tainacan. O código Ruby está disponível no GitHub⁵⁶. A Figura 34 apresenta o esquema da fila de submissão.

⁵⁶ https://github.com/tainacan/tainacan-elk/tree/museus_federais/logstash/plugins/src/filters/logstash-filter-tainacan_submission. Acesso em 11 jan. 2023

Figura 34. Esquema da fila de submissão de itens no Tainacan



Fonte: Siqueira; Martins; Medeiros (2022, p. 93)

No plug-in de saída, os dados são enviados para o armazenamento do Elasticsearch, para o qual foram elaboradas duas alternativas. Na primeira, caso seja identificado erro no item submetido ao Tainacan, este é enviado para um *index* de erros, no Elasticsearch, chamado *tainacan_integracao_erros*. Um exemplo de erro: um item com taxonomia inexistente no Tesouro proposto pelo Ibram. O segundo caminho, caso o item esteja correto, este, que foi corretamente armazenado no Tainacan, é enviado para o *index* oficial da agregação, o *tainacan_integracao*, que será utilizado pelo Kibana, para criação do Painel Analítico.

4.4.5 Elasticsearch: armazenamento de dados

O Elasticsearch é uma ferramenta robusta com diferentes possibilidades, dentre elas, várias operações de pesquisa, recuperação de dados e compartilhamento. A Figura 35 mostra a interface da ferramenta.

Figura 35. Interface da ferramenta Elasticsearch



Fonte: elaborada pela autora (2023)

Com os dados armazenados no Elasticsearch, é possível, por meio do Kibana, criar painéis com os dados.

4.4.6 Kibana: painel analítico

A ferramenta Kibana acessa os dados integrados armazenados no Elasticsearch, possibilitando a criação de painéis interativos, compostos por elementos de visualização, como gráficos, títulos, métricas e mapas. Os painéis, que podem ser facilmente criados e/ou editados, oferecem aos gestores diferentes visualizações das coleções, em uma única interface totalmente interativa que reage a cada nova busca.

Por estar diretamente relacionado ao processo de agregação de dados, as atualizações realizadas nos museus de origem também são realizadas nos painéis. Para exemplificar, apresentamos, na Figura 36, a interface de criação da visualização *Horizontal Bar*, ou seja, um gráfico de barras horizontais que demonstra a quantidade de itens por museu.

Figura 36. Interface do Kibana para criação de um gráfico com o número de itens de cada museu



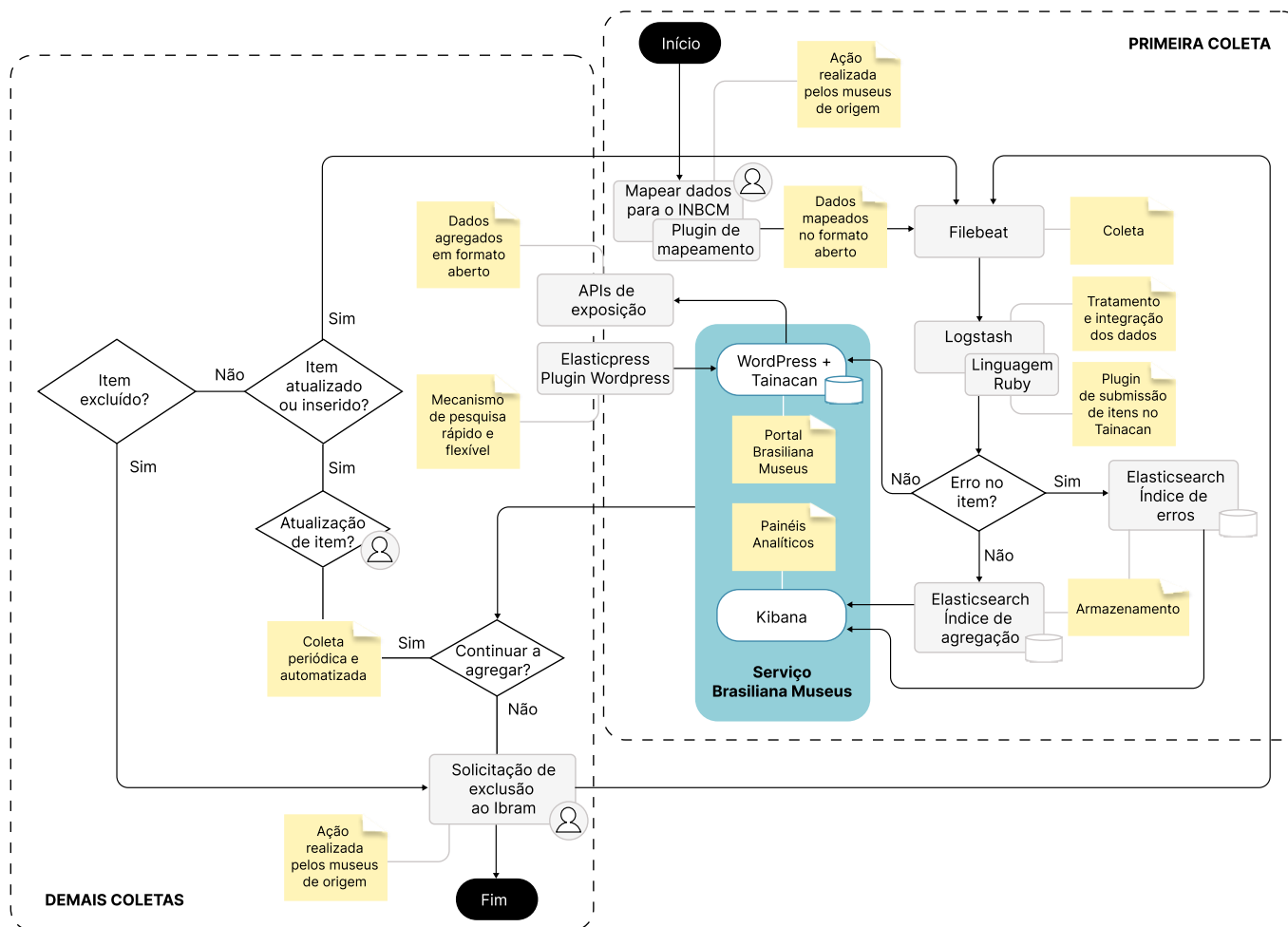
Fonte: SIQUEIRA; MARTINS (2021)

Dessa forma, um painel é composto por um conjunto de visualizações, que podem ser facilmente criadas, editadas e estarem em constante melhoria para atender as necessidades do Ibram no decorrer do tempo.

4.5 Workflow do serviço de busca e recuperação da informação

A partir das tecnologias e das etapas foi possível desenvolver o workflow de agregação de metadados culturais, que apresenta todos as etapas, as tecnologias e os requisitos mínimos para agregação. A Figura 37 apresenta o workflow.

Figura 37. Workflow do Brasiliana Museus: serviço de busca e recuperação da informação do Ibram



Fonte: elaborado pelos autores (2023)

O workflow do serviço Brasileira Museus visa a ilustrar graficamente seus principais fluxos, considerando processos, tecnologias e atores, demonstrando a viabilidade e simplicidade da solução.

Conforme supracitado, o primeiro passo diz respeito à intenção do museu em disponibilizar seus dados abertos mapeados para o INBCM. O mapeamento deve ser realizado pelos profissionais dos museus que, conhecendo os dados, podem realizar o processo corretamente. Os dados coletados são agregados, transformados e enviados para o Tainacan e para o Elasticsearch, que se divide em dois index, o primeiro receberá os dados corretamente submetidos no Tainacan e o segundo, os com erros de submissão.

Ambos os index são usados para a criação dos painéis de visualização analítica. Já os dados submetidos ao Tainacan compõem o Portal Brasileira Museus, junto às páginas Web desenvolvidas com o WordPress. Além disso, foi adicionado ao Tainacan de integração o plugin do WordPress Elasticpress⁵⁷, para agilizar as consultas.

O Tainacan também oferece, por default, APIs de saída de dados em diferentes formatos e, dessa forma, os dados agregados também poderão ser utilizados por outras aplicações. Neste ponto, destacamos uma limitação relevante, consequência do uso do INBCM e da prática de catalogação dos museus: não há disponível a licenças de uso para cada item, o que dificulta sobremaneira a utilização destes em diversas atividades e áreas de atuação. Por exemplo, é possível vincular as imagens em aplicações com fins lucrativos? A Europeana, em sua documentação, apresenta fortemente a importância da definição correta de licenças de uso para a reutilização dos dados (metadados, imagens, arquivos). Dessa forma, mesmo que o reúso da informação seja tecnologicamente viável, a questão ainda demanda de estudo e ações do Ibram.

A Figura 37 foi dividida em dois momentos, o da primeira coleta, quando os dados são coletados pela primeira vez, e da segunda coleta em diante, quando os dados precisam ser atualizados ou excluídos da base.

O serviço foi avaliado a partir de uma técnica de Engenharia de Software, chamado teste funcional, no qual foram testados sete casos de teste, são eles: i) inserir novo item; ii) inserir novo item com termo de tesouro inexistente; iii) atualizar

⁵⁷ <https://br.wordpress.org/plugins/elasticpress/>. Acesso em 21 jan. 2023

item com alteração no conteúdo de um metadado; iv) atualizar o nome da coleção; v) alterar a URL da instalação; vi) excluir um item e; vii) excluir um item, a partir da alteração do status de “público” para “privado”.

Os resultados apresentam melhorias importantes, que foram implementadas da solução apresentada, mas também demonstrou uma limitação do projeto, que é a exclusão de dados de origem. Dessa forma, a inclusão e atualização de itens nas bases são coletadas e atualizadas de forma automatizada, porém, a exclusão de itens, requer mais estudos. Assim, as exclusões devem ser solicitadas para atualização nas bases de dados do serviço de agregação.

Destaca-se a flexibilidade da solução. Embora tenha sido pensado a partir do uso do software Tainacan com os elementos descritivos INBCM, caso a instituição opte por outros repositórios, DSpace⁵⁸, Omeka⁵⁹ etc., ou outros formatos de dados, desde que estes permitam o consumo dos seus dados no formato JSON, seriam necessárias poucas alterações no código-fonte. O Filebeat permite a coleta de outros tipos de dados, mas, para tal, seriam necessárias mudanças mais significativas no código-fonte, possíveis, porém, significativas.

Dessa forma, não necessariamente os dados precisam ser extraídos do Tainacan, mas de qualquer repositório que disponha de uma API. É essencial que os dados conversem entre si e eles podem ser mapeados para qualquer outro padrão. Por exemplo, será um facilitador se os dados de origem estiverem todos utilizando Dublin Core, tornando o mapeamento menos essencial.

Com a coleta feita, pode-se manter o Tainacan como repositório final, sem necessidade de mudança de código, assim como pode-se adaptá-lo à submissão em novo um software, o que, de fato, demandaria recursos de profissionais capacitados para execução.

4.6 Painel de Visualização Analítica dos Acervos Digitais Integrados

Durante o processo de prototipação da busca integrada, identificou-se a possibilidade de se oferecer painéis de visualização dos acervos digitais integrados do Ibram.

Segundo Sarikaya et al (2019):

⁵⁸ <https://dspace.lyrasis.org/>. Acesso em 21 jan. 2023

⁵⁹ <https://omeka.org/>. Acesso em 21 jan. 2023

Dados onipresentes e tecnologias de visualização disponíveis ao público ampliaram a adoção do painel para novos domínios. Consequentemente, o conceito de painel evoluiu de telas de relatório para incluir interfaces interativas, com múltiplas visualizações e propósitos, incluindo comunicação, aprendizado e motivação, além das noções clássicas de monitoramento e suporte à decisão (p. 683, tradução nossa).

Os painéis são extremamente importantes em um mundo baseado em dados. Empresas, organizações sem fins lucrativos e grupos comunitários dependem de painéis todos os dias para realizar seu trabalho. Para muitos, os painéis podem ser o primeiro (ou único) encontro com todos os dados. Além disso, o uso do painel ultrapassou os limites organizacionais, incluindo os indivíduos e o público em geral (SARIKAYA *et al*, 2019).

Os painéis são ferramentas para a visualização da informação, especificamente, nesta pesquisa, a Visualização Analítica, que procura focar no pensamento analítico por meio de interfaces gráficas, impulsionada pelo crescimento das bases dados e dos processos de gestão de conhecimento, bem como do processo de diminuição de incertezas para tomada de decisão e análise do comportamento cognitivo (FERREIRA, 2012).

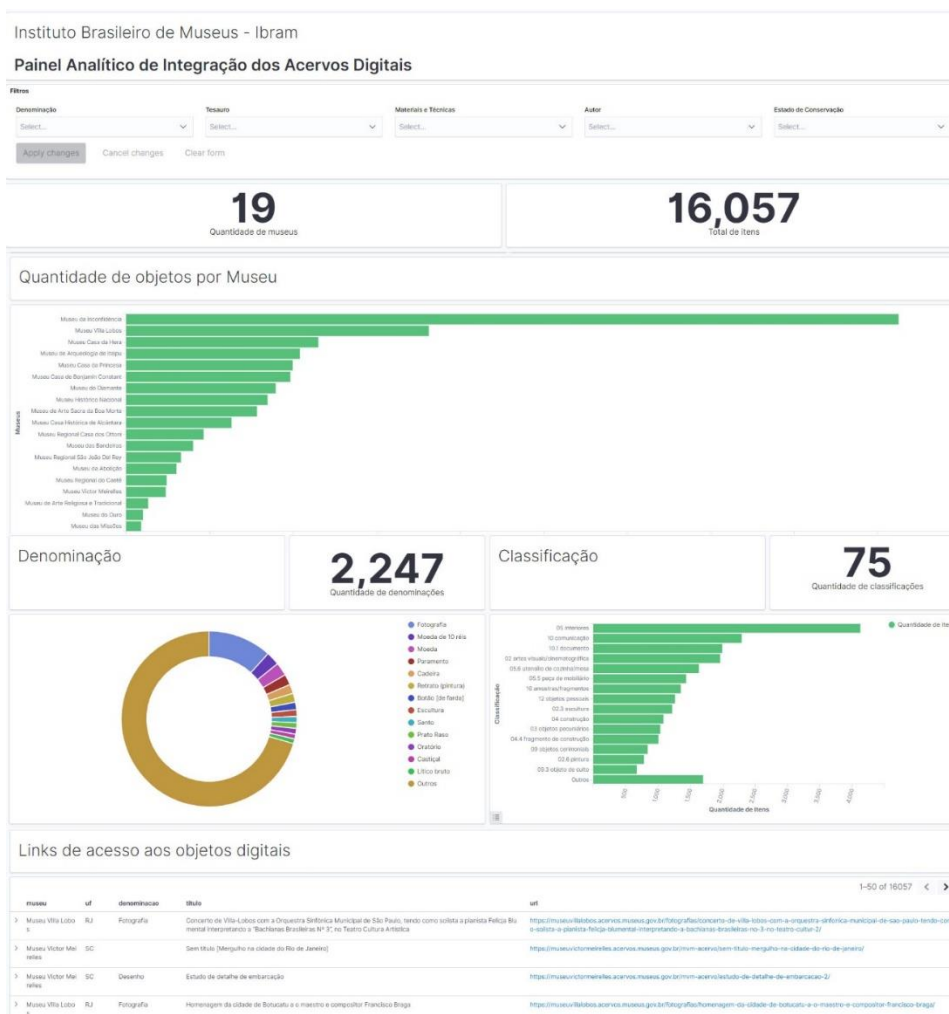
Além disso, dados integrados podem demonstrar inconsistências sintáticas ou semânticas, possibilitando aos museus melhorarem, na base, suas informações a partir de boas práticas de catalogação (GILLILAND, 2016), incluindo criação e uso de metadados, de vocabulários controlados (ou linguagens documentárias) e de regras de catalogação, com as quais se determinam como elaborar a descrição de um recurso de informação e os pontos de acesso, tornando-se práticas essenciais na padronização e na descrição, viabilizando, portanto, interoperabilidade e agregação de acervos digitais. Nesse viés, instituições de memória cumprem a função de fornecer uma melhor experiência de busca para o usuário, potencializando um caráter formativo de sua equipe interna, tendo em vista que os museus são encorajados a disponibilizar conteúdo de qualidade na rede.

A partir da comparação de diferentes instituições, torna-se possível encontrar padrões e conexões que interessam (FERREIRA, 2012). Os indicadores podem garantir um melhor planejamento estratégico, maior produtividade do trabalho, agilidade no acesso aos dados, melhorando a eficiência da gestão, gerando efeitos nas dimensões administrativas, culturais e educacionais.

O painel também pode, a critério do Ibram, ficar disponível ao público em geral, permitindo que pesquisadores, por exemplo, possam enriquecer sua fonte de dados para pesquisa; que os usuários possam ter acesso aos *links* dos objetos digitais por meio de simples pesquisas; que os museus possam monitorar o seu acervo em comparação aos demais. As possibilidades de análises e públicos-alvo são inúmeras, pelo fato de o painel abrir um leque de novas possibilidades, até então, inimagináveis se considerarmos cada museu de forma independente.

Dessa forma, um painel é composto por um conjunto de visualizações. No caso do Ibram, por nove títulos, cinco filtros, oito métricas numéricas, seis gráficos, uma tabela, um quadro de seleção, podendo, a qualquer momento, adicionar novas visualizações, a depender das necessidades da Instituição. A Figura 38 apresenta parte do painel.

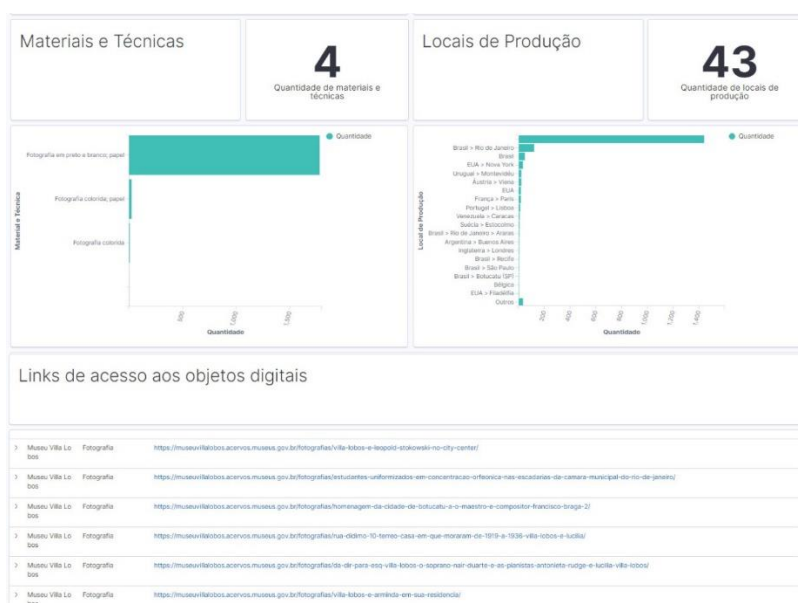
Figura 38. Painel de Visualização Analítica dos Acervos Digitais Integrados do Ibram



Fonte: elaborada pelos autores (2022)

Por fim, demonstramos algumas formas de se realizar as buscas no painel de maneira rápida e eficiente. A primeira, é possível interagir com os gráficos. Por exemplo, ao clicar em Museu Villa-Lobos, no gráfico de museus, todo painel se altera, mostrando apenas os números referentes a este museu, conforme demonstra a Figura 39.

Figura 39. Exemplo com o resultado da seleção do Museu Villa-Lobos no gráfico de museus



Fonte: Siqueira; Martins (2021, p. 11)

Outra maneira de realizar a seleção dos dados, são os filtros pré-configurados, que permitem a obtenção de resultados que combinem diferentes seleções. A Figura 40 apresenta o exemplo de busca por itens cujas denominações selecionadas são oratório ou cadeira, em materiais e técnicas foi selecionada a madeira e que esteja em bom estado de conservação.

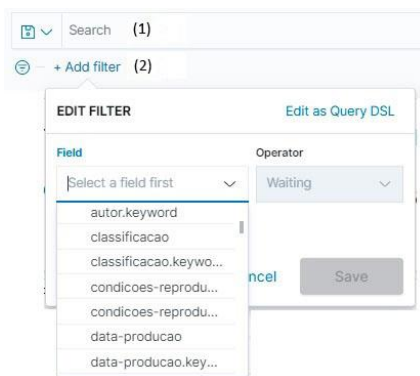
Figura 40. Exemplo de consulta com filtro pré-configurado no Kibana



Fonte: Siqueira; Martins (2021, p. 11)

O Kibana também permite realizar buscas que independem das visualizações pré-definidas, adicionando e combinando diferentes preferências (2), ou uma busca simples, por palavra-chave (1), conforme Figura 41.

Figura 41. Interface para uma consulta simples, por palavra-chave e a partir da adição de novos filtros



Fonte: Siqueira; Martins (2021, p. 12)

O painel criado utilizando o software Kibana abre diversas possibilidades, e permite aos gestores um controle de todos os acervos em uma única interface. Por ser utilizado junto à Elastic Search, ao FileBeat e ao Logstash, as atualizações realizadas nos museus de origem serão automaticamente atualizadas. Além disso, o

Kibana permite a criação de diversos outros gráficos, tabelas, a partir das necessidades apresentadas pelo Ibram, tornando o painel ainda mais completo.

Para visualizar o resultado final e realizar as consultas desejadas, disponibilizamos o acesso para o painel de visualização analítica dos acervos integrados dos museus vinculados ao Ibram⁶⁰.

4.7 Portal Brasileira Museus: repositório final

A quinta etapa contempla o objetivo específico: implementar o protótipo do Portal Brasileira Museus, um portal agregador de objetos digitais museais brasileiro, que permite a busca e a recuperação dos acervos em uma interface única. Por ser um portal que utiliza o *software* Tainacan, também pode usufruir de todos os recursos de criação de páginas do WordPress. Dessa forma, além do repositório digital, conta com a página principal (

Figura 42); uma página sobre o agregador, com informações gerais; uma página para cada museu - como exemplo cita-se o Museu da Bandeira¹, em Goiás, que possui foto, um breve histórico, endereço, telefone, link do museu; uma página de curadorias, com algumas coleções previamente selecionadas e, por fim, uma página de contatos.

O portal Brasileira Museus está disponível para acesso no endereço eletrônico <http://integracaoibram.tainacan.org/>. A interface com o usuário é denominada "Acervo em Rede" e, por isso, tanto o site quanto as capturas de tela apresentadas neste artigo contêm este nome.

⁶⁰ <http://200.137.217.144:6601/goto/613663b72fc306a908e86a6233e5219c>. Acesso em 11 jan. 2023

Figura 42. Portal Brasileira Museus (Acervo em Rede) - página principal

ACERVO EM REDE INÍCIO SOBRE MUSEUS CURADORIAS INVENTÁRIO CONTATO

ACERVO EM REDE

Agregador do patrimônio histórico cultural brasileiro

CURADORIAS SELECIONADAS

Interiores Religiosidade Trabalho Cozinha Vestime

NAVEGUE PELO INVENTÁRIO

17.317 itens digitalizados de 20 museus brasileiros

Acessar Inventário

POR MATERIAL E TÉCNICA

Madeira Ouro Bronze Argila

MOBILIÁRIO & INTERIORES

Ver curadoria

Porcelana chinesa Séc. XVII
Porcelana japonesa Séc. XIX
Porcelana chinesa Séc. XVII
Porcelana japonesa Séc. XIX
Porcelana chinesa Séc. XVII
Porcelana japonesa Séc. XIX

MUSEUS AGREGADOS

Museu da Inconfidência Guro Preto - MG
Museu Villa Lobos Rio de Janeiro - RJ
Museu Casa de Hera Vassouras - RJ
Museu de Arqueologia de Itaipu Niterói - RJ
Museu Casa de Benjamin Constant Rio de Janeiro - RJ
Museu do Diamante Diamantina - MG
Museu Histórico Nacional Rio de Janeiro - RJ
Museu Casa Histórica de Alcântara Alcântara - MA

Fonte: SIQUEIRA; MARTINS; LEMOS (2022, p. 9)

A tecnologia Tainacan oferece diferentes tipos de visualizações, que por si só entregam uma lista mais organizada dos objetos de acordo com as necessidades do usuário, que pode escolher quais metadados são apresentados na lista e pelos quais a lista deve ser ordenada (ordem crescente ou decrescente) e forma de visualização (tabelas, listas).

Para o subsistema de busca e recuperação da informação, o Tainacan oferece diferentes estratégias para a localização do item, sendo uma delas denominada facetar, isto é, pontos de acesso (IFLA, 2016) pré-concebidos pelo catalogador ou responsável pelo processo de indexação dos documentos pertencentes ao acervo. Na interface (

Figura 42), filtros podem ser conjugados com as facetar como meio de refinar o processo de busca e recuperação da informação a partir da necessidade do usuário (LANCASTER, 2004).

No projeto foram definidas as facetar de nome dos museus, Unidade Federativa (UF) do museu, classificação e estado de conservação, porém, qualquer metadado, controlado ou não, pode ser transformado em uma facetar. A Figura 43 ilustra parte dos resultados da busca utilizando as facetar com o nome do museu e a classificação do item. Nesse exemplo, agrupamos dois metadados, mas podem ser agrupados quantos forem necessários para refinar a busca.

Figura 43. Resultado da busca por itens com a Classificação “05 interiores”, nos museus Regional do Caeté, Arte Religiosa e Tradicional e das Missões



Fonte: SIQUEIRA; MARTINS; LEMOS (2022, p. 10)

No caso do Brasileira Museums, a busca por facetas é ainda mais eficiente quando se utiliza campos preenchidos com vocabulários controlados, pois resultam em índices consistentes e normalizados. Dos 15 elementos descritivos do INBCM, alguns podem ser considerados pequenos vocabulários, como por exemplo, Estado de conservação (Ótimo, Bom, Regular e Ruim) ou ainda os que utilizam de fato taxonomias, como é o caso do campo Classificação, preenchido com valores de dados oriundos do Thesaurus para Acervos Museológicos (FERREZ; BIANCHINI, 1987). O Tainacan permite a modelagem de qualquer metadado em vocabulário controlado do tipo taxonomia. O objetivo é facilitar o agrupamento de itens normalizados pelos termos utilizados no preenchimento dos metadados.

A partir das taxonomias, o Portal oferece a página de inventário de todos os itens por categoria, sendo possível visualizar os itens agrupados por museu, classificação, estado de conservação e UF. Dessa forma, o campo Classificação é um importante metadado para criação de curadorias. Por exemplo, recuperar as pinturas de todos os museus federais digitais. A partir da Classificação “02 artes visuais/cinematográfica > 02.6 pinturas” todas as pinturas são reunidas, conforme Figura 44, além de apresentar quais museus as possuem.

Figura 44. Itens museais referenciados com a classificação ‘pinturas’



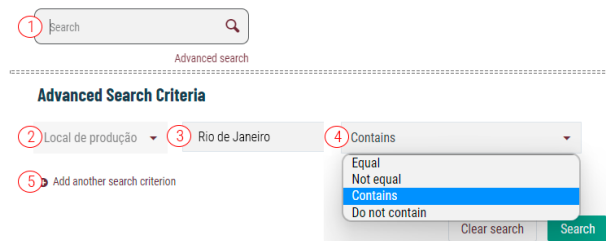
Fonte: SIQUEIRA; MARTINS; LEMOS (2022, p. 11)

Curadorias mais complexas também podem ser feitas a partir da seleção manual de diferentes combinações de pesquisa como, por exemplo, a curadoria

Vestimentas, que reúne as classificações Objetos Pessoais > Peça de Indumentária e Trabalho > Equipamento de Fiação/Tecelagem. A página Curadoria apresenta algumas coleções.

Para buscas, o Tainacan oferece uma busca geral, que pesquisa por palavra-chave em todos os metadados do acervo e uma busca avançada. A Figura 45 apresenta as opções.

Figura 45. Busca simples e avançada



Fonte: SIQUEIRA; MARTINS; LEMOS (2022, p. 12)

Na Figura 45, 1, é possível realizar busca por palavras-chave em todos os campos dos itens. Em 2, na busca avançada, o usuário pode escolher o metadado no qual deseja pesquisar; na sequência, em 3, qual texto deseja procurar e, em 4, qual operador lógico deseja usar, se igual, diferente, contém ou não contém. Caso necessário, em 5, o usuário pode criar critérios de busca e realizar combinações complexas entre essas possibilidades.

O sistema também constrói URLs expressivas que incorporam os critérios de busca, permitindo que os usuários salvem o link de suas buscas, compartilhem por mensagens em e-mail ou mídias sociais, facilitando a difusão e extroversão dos acervos. A título de exemplo, apresenta-se no Quadro 36, a URL da busca realizada pelo termo "cadeiras", com filtro específico para mostrar apenas itens do "Museu Casa da Princesa".

Quadro 36. URL de um critério combinado de busca

--

https://integracaoibram.tainacan.org/acervos?view_mode=masonry&perpage=12&paged=1&order=ASC&orderby=date&fetch_only=thumbnail%2Ccreation_date%2Ctitle%2Cdescription&fetch_only_meta=&taxquery%5B0%5D%5Btaxonomy%5D=tn_c_tax_34&taxquery%5B0%5D%5Bterms%5D%5B0%5D=85&taxquery%5B0%5D%5Bcompare%5D=IN&search=cadeira

Fonte: SIQUEIRA; MARTINS; LEMOS (2022, p. 12)

Com o desenvolvimento do protótipo, foi possível comprovar que o Elastic Stack junto ao Tainacan alcança excelentes resultados, pois permite consultas extremamente rápidas e escalabilidade horizontal, além de ser capaz de lidar com petabytes de dados. O modelo provou ser de média complexidade e customização, sendo uma possibilidade viável para agregação e para análise de dados.

Por fim, a agregação de dados, embora utilize o repositório Tainacan, pode ser implantando com quaisquer repositórios que disponibilizem seus dados em formato aberto.

5 DISCUSSÕES

Os objetos que compõem o patrimônio cultural das nações vêm gradualmente sendo digitalizados e disponibilizados na Web, alterando a forma como interagimos, visto terem sido quebradas as barreiras do tempo-espaço, de horários de visita, de possíveis pagamento de entradas etc. Além disso, os acervos mantêm-se com um lugar de entretenimento, mas ocupam um novo espaço, o de permitir a busca e recuperação da informação, tanto no que diz respeito ao patrimônio local quanto ao global.

Tanto dinamismo trouxe consigo a necessidade de otimizar a localização de objetos digitais por meio de ferramentas para a busca e a recuperação de informações eficientes. Dessa forma, uma nova realidade surgiu, a de que os museus digitais poderiam cumprir o papel de provedores de dados para serviços especializados na agregação de acervos que permitiriam uma interface única para a busca e recuperação da informação.

Este modelo de serviço é uma realidade em países como Europa e Estados Unidos da América, com comprovadas qualidades e benefícios, porém, no Brasil, estamos caminhando a passos curtos na mesma direção. Dessa forma, este estudo se propôs a contribuir com a área de estudo por meio da construção de um serviço de busca e recuperação da informação, averiguando a hipótese de que a criação de interfaces únicas de acesso à informação melhoraria substancialmente a busca e a recuperação de objetos digitais museais para diferentes categorias de usuários.

O serviço Brasileira Museus é fruto de diversas linhas de pesquisa, dada sua complexidade e interdisciplinaridade e, por isso, embora tenha sido desenvolvido considerando a realidade dos museus do Ibram, tem aplicabilidade em outras realidades.

Entende-se que uma das possibilidades de construção da pesquisa científica envolvendo o tema "tecnologia" no campo da Ciência da Informação se traduz em pesquisa aplicada que se apropria ou desenvolve tecnologias de forma a desenhar processos de circulação da informação de maneira a se adaptarem aos contextos informacionais locais. Por contextos locais deve-se entender não apenas os padrões de metadados, as regras de catalogação e as linguagens documentárias utilizadas ou não por uma comunidade, mas também o histórico sociotécnico da documentação dos

objetos informacionais em questão, os atores envolvidos e os sistemas de informação utilizados.

O serviço intitulado *Brasiliana Museus* é composto por dois produtos, o Portal *Brasiliana Museus* e o Painel de Visualização Analítica que se servem da agregação automatizada de metadados museais, disponíveis em formato aberto nos repositórios digitais dos acervos do Ibram.

Para construção do protótipo e verificação da veracidade da hipótese foram traçados seis objetivos específicos. O primeiro deles, identificar e caracterizar as etapas utilizadas por serviços de agregação de objetos digitais culturais disponíveis, foi alcançado por meio do uso de diferentes técnicas de pesquisa, que resultaram em 23 projetos agregadores, dos quais foram extraídas as principais etapas para agregação e as tecnologias utilizadas, além de um contexto breve da realidade na qual estão inseridos, no que diz respeito a investimentos, política, dentre outros.

Dos 23, dois tiveram sua documentação estudada a fundo, a DPLA e Europeia, projetos robustos que possuem vasta documentação disponível, para compreensão do processo de agregação por completo e na prática, visando conhecer a rotina necessária à manutenção do processo de agregação.

A segunda etapa contemplou dois objetivos específicos: definir quais são as etapas e as tecnologias necessárias ao desenvolvimento do agregador brasileiro, além de determinar quais seriam as condições necessárias e suficientes para agregação.

A escolha das tecnologias priorizou a busca por ferramentas utilizadas por outros agregadores, porém constatou-se que elas não atendiam a demanda brasileira e, na sequência, buscou-se por ferramentas livres, gratuitas e atuais, optando-se pelo uso da pilha *Elastic Stack*, que possui um conjunto de características e funcionalidades adequadas à necessidade do projeto.

Com a compreensão da ferramenta, definiram-se as etapas mínimas necessárias para a agregação, de forma que o serviço requeresse o mínimo possível de intervenções na rotina de trabalho dos museus.

Foram definidas seis etapas: i) mapeamento dos metadados para o INBCM; ii) coleta automatizada dos metadados; iii) agregação e transformação destes dados; iv) envio e armazenamento dos metadados no Tainacan, para criação do Portal

Brasiliiana Museus; v) envio e armazenamento dos metadados para o Elasticsearch e; vi) criação do painel analítico.

Dessa maneira, a única função atribuída aos museus, inclusive, essencial ao processo de agregação, é o mapeamento de seus dados para o formato INBCM. Porém, para tal, foi necessário incorporar esta funcionalidade ao software Tainacan, demandando da equipe de desenvolvimento a criação do plugin *INCBM mapper*, que adicionou uma interface simplificada de mapeamento. Destaca-se que o fato de o Ibram ter determinado o uso desses elementos descritivos pelos museus, com certeza, foi um facilitador.

O quarto objetivo visava a criação do workflow completo, com as etapas e tecnologias necessárias à manutenção do serviço Brasiliiana Museus. Nessa etapa é relevante destacar que a decisão de receber os dados como eles estão, sem interferir ou modificar os processos instaurados nos museus de origem, impactou diretamente no desenho do workflow. Como pode ser observado nos workflows apresentados no estudo, é uma preocupação latente e extremamente importante com a qualidade dos dados, pois dados de origem de qualidade são revertidos em dados agregados de qualidade e, por consequência, em melhores resultados na busca e recuperação da informação, o que é essencial para um bom serviço.

No entanto, o esforço necessário à melhoria dos dados nos museus de origem requer mais do que a vontade dos pesquisadores ou até mesmo da validação da eficácia da proposta. Ela depende de mudanças nos modelos de governança e em incentivo político e econômico, fatores que estão aquém do escopo da proposta. Dessa forma, embora cientes da importância de dados de qualidade para um serviço de agregação eficiente, este ponto não pode ser considerado.

Por outro lado, acredita-se que a visualização dos dados agregados, tanto pelo Ibram quanto por cada museu individualmente, mas em comparação aos demais, abre inúmeras possibilidades de análise de dados, permitindo uma melhor gestão da documentação museológica; a demonstração de inconsistências sintáticas ou semânticas; a localização de padrões e conexões de interesse, além de indicadores que possam garantir melhor planejamento estratégico, produtividade e agilidade no acesso aos dados e; o enriquecimento de fontes de dados de pesquisa, gerando efeitos nas dimensões administrativa, cultural e educacional.

Assim, de forma otimista, acredita-se que os resultados deste trabalho possam incentivar a busca por uma maior qualidade dos dados, de forma que futuramente se reverta em incentivo político e econômico necessários à criação e execução de diretrizes para qualidade de dados, assim como ocorre com renomados agregadores de dados.

Os últimos dois objetivos tratam da implementação do serviço, ou seja, da validação da proposta por meio do desenvolvimento dos protótipos, que são dois: o Painel de Visualização Analítico e o Portal Brasileira Museus de busca e recuperação, que se beneficiam, cada um a seu modo, dos dados agregados, que são atualmente 17.315 itens de 20 museus federais.

A efetivação do protótipo teve início na validação do processo de coleta, tratamento e armazenamento dos dados agregados para validação da tecnologia, afinal, era preciso confirmar se o Elastic Stack realmente atenderia às demandas. Nesse processo, a grata surpresa foi a Kibana, que oferecia uma possibilidade além da agregação, permitindo a criação de painéis interativos, apresentando os dados dos museus de forma sistematizada e interativa. Além disso, o uso do Kibana possibilitou a confirmação da relativa simplicidade da ferramenta que, com interface intuitiva, possibilita para profissionais de qualquer área desenvolver seus próprios painéis obtendo o máximo de benefício dos dados.

No Painel as inconsistências nos dados ficam nítidas e, em contrapartida, também ficam nítidos os benefícios dos dados de qualidade, pois permitem que gráficos ou qualquer outro tipo de visualização retornem informações confiáveis e consistentes, além de permitir a busca muito mais eficiente, reafirmando a discussão sobre a importância de diretrizes para qualidade de dados.

O Kibana permite a criação de quantos painéis forem necessários, dessa maneira, além do Painel principal criou-se, como parte do serviço, o painel que armazena os dados que apresentaram qualquer tipo de erro na submissão para o Tainacan, descomplicando o processo de identificação e correção de erros. Além do mais, caso cada museu quisesse ter sua própria visualização ou o Ibram novos painéis, seria possível e fácil de ser realizado.

O Elasticsearch e o Kibana apresentam ferramentas importantes para a busca e recuperação dos dados e, por si só, trariam inúmeros benefícios aos museus e ao Ibram. No entanto, o último objetivo específico - implementar o protótipo do Portal

Brasiliiana Museus - visa a uma interface única de busca e recuperação para todos os usuários interessados, muito inspirados por projetos de sucesso como a Europeana e a DPLA.

Assim, buscou-se identificar uma nova ferramenta para criação do Portal Web. Embora outras ferramentas tenham sido cogitadas, entendeu-se que o Tainacan poderia realizar esse papel, visto oferecer ferramentas de busca e recuperação, ter o WordPress como base, permitindo a estruturação de páginas Web e comporia uma arquitetura coesa, com o mínimo de softwares, facilitando o processo. Também tornaria mais fácil encontrar profissionais capacitados, sempre que necessário, pois os próprios museus e o Ibram estariam habituados a ela. O Tainacan e o Elastic Stack se mostraram de média complexidade, tornando-se uma opção viável de aprendizado e customização.

6 PREMIAÇÕES E PUBLICAÇÕES

As etapas do presente estudo foram publicadas em diferentes periódicos e anais e estão listadas a seguir. Além destas, são apresentadas as premiações.

6.1 Premiações

- ENANCIB 2021 - Melhor trabalho completo do GT 8 – Informação e Tecnologia: Painel de visualização analítica dos acervos digitais integrados do Instituto Brasileiro de Museus: o uso das tecnologias Tainacan e Elastic Stack.
- ENANCIB 2021 - Melhor trabalho completo do GT 9 – Museu, Patrimônio e Informação: Requisitos de Qualidade para dados de agregação em Museus: O Caso do Ibram.

6.2 Publicações

SIQUEIRA, Joyce; CARMO, Danielle do; MARTINS, Dalton Lopes; LEMOS, Daniela Lucas da Silva; MEDEIROS, Vinicius Nunes; OLIVEIRA, Luís Felipe Rosa de. Elements for Constructing a Data Quality Policy to Aggregate Digital Cultural Collections: Cases of the Digital Public Library of America and Europeana Foundation. In: Bisset Álvarez, E. (eds) Data and Information in Online Environments. DIONE 2021. Lecture Notes of the Institute for Computer Sciences, Social Informatics and

Telecommunications Engineering, vol 378. Springer, Cham.
https://doi.org/10.1007/978-3-030-77417-2_8

SIQUEIRA, Joyce; MARTINS, Dalton Lopes. Análise e mapeamento da organização técnica dos acervos digitais das instituições federais vinculadas ao Ministério da Cultura: fontes de informação para o desenvolvimento das Humanidades Digitais. 2021a, In: PIMENTA, Ricardo M.; ALVES, Daniel. Humanidades digitais e o mundo lusófono. Local: FGV, 2021. p. 283-309. Disponível em: <https://editora.fgv.br/produto/humanidades-digitais-e-o-mundo-lusofono-3625>. Acesso em 10 jan. 2023

SIQUEIRA, Joyce; MARTINS, Dalton Lopes. LEMOS, Daniela Lucas da Silva. Brasiliana Museus: serviço de busca e recuperação da informação agregada dos acervos digitais do Instituto Brasileiro de Museus. In: XXII Encontro Nacional de Pesquisa e Pós-graduação em Ciência da Informação, 2022, Porto Alegre. XXII Enancib, 2022. Disponível em: <https://enancib.ancib.org/index.php/enancib/xxiienancib/paper/view/712>. Acesso em 13 jan. 2023

SIQUEIRA, Joyce; MARTINS, Dalton Lopes. Painel de visualização analítica dos acervos digitais integrados do instituto brasileiro de museus: o uso das tecnologias Tainacan e Elastic Stack. In: XXI Encontro Nacional de Pesquisa e Pós-graduação em Ciência da Informação, 2021b, Rio de Janeiro. XXI Enancib, 2021. Disponível em: <https://enancib.ancib.org/index.php/enancib/xxienancib/paper/view/95>. Acesso em 12 jan. 2023

SIQUEIRA, Joyce; MARTINS, Dalton Lopes. Recuperação de informação: descoberta e análise de workflows para agregação de dados do patrimônio cultural. Ciência da Informação, [S. l.], v. 49, n. 3, 2020. Disponível em: <https://revista.ibict.br/ciinf/article/view/5399>. Acesso em: 12 jan. 2023.

SIQUEIRA, Joyce; MARTINS, Dalton Lopes. Serviços de descoberta em escala na web voltados a repositórios digitais utilizados por GLAM's: Uma revisão sistemática de literatura, 2019a. In Encontro nacional de pesquisa em ciência da informação. Programa de Pós-graduação em Ciência da Informação da UFSC. Disponível em: <https://conferencias.ufsc.br/index.php/enancib/2019/paper/view/588>. Acesso em 10 jan. 2023

SIQUEIRA, Joyce; MARTINS, Dalton Lopes. Workflow de agregação de dados: Processos para criação de uma interface de busca integrada do patrimônio cultural, 2019b. Workshop de informação, dados e tecnologia (3: 2019: Brasília). Workshop de informação, dados e tecnologia (WIDAT 2019): anais do evento/Dalton Lopes Martins ... [et al.], organizadores. Brasília: Universidade de Brasília, Faculdade de Ciência da Informação, 2019. ISBN: 978-65-86503-01-2. Disponível em: <http://widat2019.fci.unb.br/index.php/anais-widat-2019>. Acesso em 10 jan. 2023

SIQUEIRA, Joyce; MARTINS, Dalton Lopes. Workflow models for aggregating cultural heritage data on the web: A systematic literature review, 2021c. Journal of the Association for Information Science and Technology, 1–21, 2021. DOI: <https://doi.org/10.1002/asi.24498>

SIQUEIRA, Joyce; MARTINS, Dalton Lopes; MEDEIROS, Vinícius Nunes. Brasileira Museus: teste funcional do agregador de dados museais do Instituto Brasileiro de Museus. Workshop de informação, dados e tecnologia (2022: Espírito Santo). Workshop de informação, dados e tecnologia (WIDAT 2022): anais do evento/Daniela Lucas da Silva Lemos ... [et al.], organizadores. Vitória: Universidade Federal do Espírito Santo. Faculdade de Ciência da Informação, 2022a. Disponível em: anais ainda não estão disponíveis. Acesso em 10 jan. 2023

SIQUEIRA, Joyce; MARTINS, Dalton Lopes; MEDEIROS, Vinícius Nunes. Metadata mapping in Tainacan: new functionality for digital museums linked to the Brazilian Institute of Museums. En T.M.R. Dias (Ed.), Informação, Dados e Tecnologia. Advanced Notes in Information Science, volume 2 (pp. 182-191), 2022b. Tallinn, Estonia: ColNes Publishing. DOI: <https://doi.org/10.47909/anis.978-9916-9760-3-6.92>

SIQUEIRA, Joyce; MARTINS, Dalton Lopes; MEDEIROS, Vinícius Nunes. Workflow with Emphasis in Technologies for the Construction of an Aggregator and a Dashboard of Brazilian Museums Digital Collections. In: Pinto, A.L., Arencibia-Jorge, R. (eds) Data and Information in Online Environments. DIONE 2022. 2022c. Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering, vol 452. Springer, Cham. https://doi.org/10.1007/978-3-031-22324-2_8

MARTINS, Dalton Lopes; LEMOS, Daniela Lucas da Silva; CARMO, Danielle do; SIQUEIRA, Joyce; OLIVEIRA, Luís Felipe Rosa de. Requisitos de Qualidade para dados de agregação em Museus: O Caso do Ibram. In: XXI Encontro Nacional de Pesquisa e Pós-graduação em Ciência da Informação, 2021, Rio de Janeiro. XXI Enancib, 2021. Disponível em: <https://enancib.ancib.org/index.php/enancib/xxienancib/paper/view/575>. Acesso em 10 jan. 2023

7. CONSIDERAÇÕES FINAIS

A agregação de metadados culturais permite a criação de interfaces únicas de busca e recuperação da informação, melhorando substancialmente a interação dos usuários com os objetos digitais museais, tanto os especialistas na área, até quaisquer usuários leigos interessados em conhecer a cultura brasileira por meio do seu patrimônio cultural.

Dessa forma, a pesquisa elaborou um modelo conceitual para um novo serviço de busca e recuperação da informação, a partir da agregação automatizada de metadados, voltado a realidade do Ibram. O modelo composto por dois diferentes serviços, o Portal Brasileira Museus e o Painel de Visualização Analítica, foi prototipado, desenvolvido e avaliado para fins de prova de conceito.

O desenvolvimento validou a viabilidade da proposta tecnológica e suas ferramentas. Sua primeira versão foi avaliada a partir de uma técnica de Engenharia de Software, intitulado teste funcional, visando a identificar se ele realiza corretamente as principais funções às quais se propõe. O teste resultou em algumas melhorias, porém, a versão final ainda apresenta uma importante limitação: a exclusão dos dados não é feita corretamente e requer maiores pesquisas para ser implementado.

As diferentes visualizações dos metadados agregados representam um importante avanço pois permite aos gestores maior compreensão dos seus acervos. Se apenas forem identificadas e corrigidas as inconsistências sintáticas ou semânticas, já haverá uma evolução importante na catalogação dos acervos. Quando se vislumbra a possibilidade localizar padrões e conexões de interesse abre espaçosas portas para estudos na área da organização da informação.

Quanto mais os metadados forem analisados, mais ficará evidente a importância de se buscar a mais alta qualidade dos dados nas bases de origem. Assim, essa pesquisa destaca esse tema como sendo merecedor de um estudo amplo e complementar, visando a investigar se, e como as práticas de catalogação influenciam nos resultados do agregador. E, se influenciam, quais alterações são necessárias à sua melhoria.

Também pode-se observar os benefícios sociais da solução. Facilitar acesso ao patrimônio cultural brasileiro, valorar esse acervo, utilizá-lo em sala de aula, por exemplo, em diferentes representações e contextos. Usufruir do Portal Brasileira Museus, na perspectiva de um espaço aberto, acessível e com foco na comunicação.

Por fim, o serviço atualmente agrega 17.315 objetos culturais de 20 museus e se torna uma porta de entrada sistematizada e agregada para visitaç o pelos usu rios, sejam eles professores, estudantes, pesquisadores ou o p blico leigo em geral, da cultura musealizada brasileira podendo funcionar como ferramenta de apoio a diversas iniciativas futuras de educa o, media o e diferentes formas de literacia informacional. Entende-se que o projeto cumpre um importante papel no contexto da pol tica p blica cultural no pa s e uma contribui o do campo da Ci ncia da Informa o para a  rea.

Referências

ABADAL, Ernest; CODINA Luis. Bases de datos documentales: características, funciones y métodos, 2005. Madrid: Síntesis. ISBN: 9788497562638. Disponível em: <http://hdl.handle.net/2445/141978>. Acesso em 10 jan. 2023.

ARTINI, Michele; BARDI, Alessia; BIAGINI, Federico; DEBOLE, Franca; LA BRUZZO, Sandro; MANGHI, Paolo; MIKULICIC, Marko; SAVINO, Pasquale; ZOPPI, Franco. Data interoperability and curation: The European film gateway experience. In M. Agosti, F. Esposito, S. Ferilli, e N. Ferro (Eds.), Digital libraries and archives, 2013. IRCDL 2012. Communications in computer and information science (Vol. 354, p. 33–44). Springer. DOI: https://doi.org/10.1007/978-3-642-35834-0_6

AVAZPOUR, Iman; GRUNDY, John; ZHU, Liming. Engineering complex data integration, harmonization, and visualization systems. Journal of Industrial Information Integration, 2019, volume 16, ISSN 2452-414X. DOI: <https://doi.org/10.1016/j.jii.2019.08.001>

BAEZA-YATES, Ricardo; RIBEIRO-NETO, Berthier. 2011. Modern information retrieval: The concepts and technology behind search. Addison Wesley.

BAEZA-YATES, Ricardo; RIBEIRO-NETO, Berthier. Recuperação de Informação: Conceitos e Tecnologia das Máquinas de Busca. Bookman: Grupo A, 2013. E-book. ISBN 9788582600498.

Bardin, L.: Análise de Conteúdo. 3ª Reimpressão da 1, vol. 70. Edições, São Paulo (2016)

BARBOSA, Alice Príncipe. Classificações facetadas. Ciência da Informação, Rio de Janeiro, v. 1, n. 2, p. 73-81. 1972.

BARDI, Alessia; MANGHI, Paolo; ZOPPI, Franco. Aggregative data infrastructures for the cultural heritage. In J. M. Doderó, M. Palomo-Duarte, e P. Karampiperis (Eds.), Metadata and semantics research, 2012. MTSR 2012. Communications in computer and information science (Vol. 343, p. 239–251). Springer. DOI: https://doi.org/10.1007/978-3-642-35233-1_24

BARDI, Alessia; MANGHI, Paolo; ZOPPI, Franco. Coping with interoperability and sustainability in cultural heritage aggregative data infrastructures, 2014. International Journal of Metadata, Semantics Ontologies, 9(2), 138–154. DOI: <https://doi.org/10.1504/IJMSO.2014.060341>

BELLINI, Pierfrancesco; BRUNO, Ivan; NESI, Paolo; PAOLUCCI, Michela. IPR centered institutional service and tools for content and metadata management, 2015 International of Journal of Software Engineering and Knowledge Engineering, 25(8), 1237–1270. DOI: <https://doi.org/10.1142/S0218194015500242>

BORKO, Harold. Information science: What is it?. Amer. Doc., 19: 3-5, 1968. DOI: <https://doi.org/10.1002/asi.5090190103>

BRASIL. Instituto Brasileiro de Museus, Resolução Normativa IBRAM n. 6, de 31 de agosto de 2021. Diário Oficial. Disponível em: <https://www.in.gov.br/web/dou/-/resolucao-normativa-ibram-n-6-de-31-de-agosto-de-2021-342359740>. Acesso em 10 jan. 2023.

BRASILIANA FOTOGRAFICA. Objetivos. 2023a. Disponível em: https://brasilianafotografica.bn.gov.br/?page_id=96. Acesso em 10 jan. 2023

BRASILIANA FOTOGRAFICA. Quem somos. 2023b. Disponível em: https://brasilianafotografica.bn.gov.br/?page_id=119. Acesso em 10 jan. 2023

BRASILIANA ICONOGRÁFICA. Sobre o Projeto. 2017. Disponível em: <https://www.brasilianaiconografica.art.br/sobre-o-projeto>. Acesso em 10 jan. 2023

BRIGHAM, Tara J.; FARRELL, Ann M.; TRZASKO, Leah C. Osterhaus; ATWOOD, Carol Ann; WENTZ, Mark W.; ARP, Kelly A. Web-scale discovery service: Is it right for your library? 2016, Mayo Clinic libraries experience. Journal of Hospital Librarianship, 16(1), 25– 39. DOI: <https://doi.org/10.1080/15323269.2016.1118280>

BUCKLAND, M. K. Information as thing, 1991. Journal of the American Society for Information Science, [s.n.], v. 42, n. 5, p. 351-360

CHAN, Lois Mai; ZENG, Marcia Lei. Metadata Interoperability and Standardization – A Study of Methodology Part I. Achieving Interoperability at the Schema Level, 2006. D-Lib Magazine, Volume 12, Number 6. ISSN 1082-9873. DOI: <https://doi.org/10.1045/june2006-chan>

CHARLES, Valentine; ISAAC, Antoine; HILL, Timothy (ed.): Discovery - User scenarios and their metadata requirements v3, 2015. Europeana Foundation. Disponível em: https://pro.europeana.eu/files/Europeana_Professional/EuropeanaTech/EuropeanaTech_WG/DataQualityCommittee/DQC_DiscoveryUserScenarios_v3.pdf. Acesso em 10 jan. 2023

CLAYPHAN, Robina; CHARLES, Valentine; ISAAC, Antoine. Europeana Data Model – Mapping Guidelines v2.3. Europeana Think Culture, 2016. Disponível em: https://pro.europeana.eu/files/Europeana_Professional/Share_your_data/Technical_requirements/EDM_Documentation/EDM_Mapping_Guidelines_v2.3_112016.pdf. Acesso em 10 jan. 2023

CUNHA, Bruna Maria Campos da; ALVES, Rachel Cristina Vesu. O Protocolo OAI-PMH e os Provedores de Dados e Provedores de Serviços, 2019. VIII SECIN - Seminário de Ciência da Informação. Paraná - PR. Disponível em: <http://www.uel.br/eventos/cinf/index.php/secin2019/secin2019/paper/viewFile/579/405>. Acesso em 10 jan. 2023

DE SOUZA, Érika Madeira; MACHADO, Elenora Nobre. Rede Web de Museus: Relato de experiência na gestão e acesso aos acervos culturais do Estado do Rio de Janeiro. PragMATIZES-Revista Latino-Americana de Estudos em Cultura, p. 74-90, 2019. Disponível em: <https://periodicos.uff.br/pragmatizes/article/view/27513>. Acesso em 10 jan. 2023

DIGITAL NEW ZEALAND. Our History. 2023. Disponível em: <https://digitalnz.org/about/our-history>. Acesso em 10 jan. 2023

DIGITAL NEW ZEALAND. This is Digital New Zealand. 20 dez. 2018. YouTube. Disponível em: <https://www.youtube.com/watch?v=UWbIDwsaA4o>. Acesso em 10 jan. 2023

DPLA. About us. 2023a. Disponível em: <https://dp.la/about>. Acesso em 10 jan. 2023

DPLA. Becoming a Service Hub. 2023b. Disponível em: <https://pro.dp.la/prospective-hubs/becoming-a-service-hub>. Acesso em 10 jan. 2023

DPLA. Collection Development Guidelines. 2023c. Disponível em: <https://pro.dp.la/hubs/collection-development-guidelines>. Acesso em 10 jan. 2023

DPLA. Documentation and Tools. 2023d. Disponível em: <https://pro.dp.la/hubs/documentation>. Acesso em 10 jan. 2023

DPLA. DPLA Data Exchange Agreement. 2017a. Disponível em: <https://digitalpubliclibraryofamerica.atlassian.net/wiki/download/attachments/85406202/DataExchangeAgreement2017March.docx?version=1&modificationDate=1494427927703&cacheVersion=1&api=v2>. Acesso em 10 jan. 2023

DPLA. DPLA Partners Crosswalk, MAP v4.0. 2023e. Disponível em: <https://docs.google.com/spreadsheets/d/1BzZvDOf4fgas3TD21xF40lu2pk2XW0k2pTGJKIt6438/edit#gid=1453046017>. Acesso em 10 jan. 202

DPLA. DPLA Standardized Rights Statements Implementation Guidelines. 2017b. Disponível em: <https://docs.google.com/document/d/1alnokOllsgf-B4iMTXU33qYN5B2iA3s91KgWoh7DZ7Q/edit?usp=sharing>. Acesso em 10 jan. 2023

DPLA. Hub Re-ingest Schedule. 2022. Disponível em: <https://digitalpubliclibraryofamerica.atlassian.net/wiki/spaces/CT/pages/84969744/Hu+b+Re-ingest+Schedule>. Acesso em 10 jan. 2023

DPLA. Our Hubs, 2023f. Disponível em: <https://pro.dp.la/hubs/our-hubs>. Acesso em 10 jan. 2023

DPLA. Portal Digital Public Library of America. 2023g. Disponível em: <https://dp.la/>. Acesso em 10 jan. 2023

DSPACE. About DSpace. 2023. Disponível em: <https://duraspace.org/dspace/about/>. Acesso em 11 jan. 2023

ELASTIC. Elastic Stack. 2023a. Disponível em: <https://www.elastic.co/elastic-stack/>. Acesso em 13 jan. 2023

ELASTIC. Filebeat overview | Filebeat Reference [8.6] | Elastic. 2023b. Disponível em: <https://www.elastic.co/guide/en/beats/filebeat/current/filebeat-overview.html>. Acesso em 13 jan. 2023

ELASTIC. Logstash reference [8.6] | Filter plugins. 2023c. Disponível em: <https://www.elastic.co/guide/en/logstash/current/filter-plugins.html>. Acesso em 13 jan. 2023

ELASTIC. Logstash reference [8.6] | Filter plugins. Prune filter plugin. 2023d. Disponível em: <https://www.elastic.co/guide/en/logstash/current/plugins-filters-prune.html>. Acesso em 13 jan. 2023

ELASTIC. Logstash reference [8.6] | Getting started with Logstash. Stashing your first event. 2023e. Disponível em: <https://www.elastic.co/guide/en/logstash/current/first-event.html>. Acesso em 13 jan. 2023

ELASTIC. Logstash reference 8.6] | Logstash introduction. 2023f. Disponível em: <https://www.elastic.co/guide/en/logstash/current/introduction.html>. Acesso em 13 jan. 2023

ELASTIC. Logstash reference 8.6] | Filter plugins. Fingerprint filter plugin. 2023g. Disponível em: <https://www.elastic.co/guide/en/logstash/current/plugins-filters-fingerprint.html>, last accessed on 07/09/2021.

ELASTIC. Elastic Stack. O que é o Elasticsearch? 2023h. Disponível em: <https://www.elastic.co/pt/what-is/elasticsearch>. Acesso em 13 jan. 2023

ELASTIC. Logstash Reference 5.5 | Working with plugins. 2023i. Disponível em: <https://www.elastic.co/guide/en/logstash/5.5/working-with-plugins.html>. Acesso em 13 jan. 2023

EUROPEANA. About us. 2023a. Disponível em: <https://www.europeana.eu/en/about-us>. Acesso em 11 jan. 2023

EUROPEANA. Europeana Collections. 2023b. <https://www.europeana.eu/>. Acesso em 11 jan. 2023

EUROPEANA FOUNDATION. Europeana Data Exchange Agreement. 2023. Disponível em: https://pro.europeana.eu/files/Europeana_Professional/Publications/Europeana%20Data%20Exchange%20Agreement.pdf. Acesso em 11 jan. 2023

EUROPEANA FOUNDATION. The more you give, the more you get. Europeana Publishing Framework 2.0. 2019. Disponível em: https://pro.europeana.eu/files/Europeana_Professional/Publications/Publishing_Framework/Europeana%20Publishing%20Framework%20V2.0%20English.pdf. Acesso em 11 jan. 2023

EUROPEANA PRO. Europeana Aggregators. 2023a. Disponível em: <https://pro.europeana.eu/page/aggregators>. Acesso em 11 jan. 2023

EUROPEANA PRO. Europeana Foundation. About us, 2023b. <https://pro.europeana.eu/about-us/foundation#europeana-foundation>. Acesso em 11 jan. 2023

EUROPEANA PRO. Rights statements, 2023c. Disponível em: <https://rightsstatements.org/page/1.0/>. Acesso em 11 jan. 2023

FERREIRA, Nuno Miguel Antunes. Visualização de informação e visualização analítica: mapa de visualização gráfica da informação agregada do país, um sistema de apoio à decisão. 2012. Tese de Doutorado. Instituto Universitário de Lisboa. Escolas de Tecnologias e Arquitetura. Departamento de Ciências e Tecnologias da Informação. Portugal. Disponível em: <https://repositorio.iscte-iul.pt/handle/10071/5906>. Acesso em 11 jan. 2023

FERREZ, Helena Dodd. Tesouro de objetos do patrimônio cultural nos museus brasileiros. Rio de Janeiro: Fazer Arte. Gerência de Museus da Secretaria Municipal de Cultura, 2016. Disponível em: <https://www.tesauromuseus.com.br/download/tesouro.pdf>. Acesso em 10 jan. 2023

FERREZ, Helena Dodd; BIANCHINI, Maria Helena S. Thesaurus para acervos museológicos. Ministério da Cultura, Secretaria do Patrimônio Histórico e Artístico Nacional, Fundação Nacional Pró-Memória, Coordenadoria Geral de Acervos Museológicos, 1987.

FINK, Eleanor E. Overview and Recommendations for Good Practices. American Art Collaborative. Linked Open Data Initiative, 2018. Disponível em: <https://s3.amazonaws.com/assets.saam.media/files/documents/2020-07/OverviewandRecommendationsAccessible.pdf>. Acesso em 11 jan. 2023

FREIRE, Nuno. The Data Aggregation Lab Software: Experimentation for Linked Data Aggregation in Cultural Heritage. In E. Garoufallou, F. Fallucchi, e E. William de Luca (Eds.), 2019. Metadata and semantic research. MTSR 2019. Communications in computer and information science (Vol. 1057, pp. 419– 424). Springer. DOI: https://doi.org/10.1007/978-3-030-36599-8_38

FREIRE, Nuno; MANGUINHAS, Hugo; ISAAC, Antoine; ROBSON, Glen; HOWARD, John Brooks. Web Technologies: A Survey of Their Applicability to Metadata Aggregation in Cultural Heritage. Expanding Perspectives on Open Science: Communities, Cultures and Diversity in Concepts and Practices, 2017, p. 235 - 244. DOI: <https://doi.org/10.3233/978-1-61499-769-6-235>

FROSINI, Lucas; BARDI, Alessia; MANGHI, Paolo; PAGANO, Pasquale. An Aggregation Framework for Digital Humanities Infrastructures: The PARTHENOS Experience, 2018. Scientific Research and Information Technology, v. 8, n. 1, 11 Jul. 2018. DOI: <https://doi.org/10.2423/i22394303v8n1p33>

GÄNGLER, Thomas. Semantic Federation of Musical and Music-Related Information for Establishing a Personal Music Knowledge Base. 2011. Tese (Doutorado em Ciência da Computação) - Department of Computer Science, Institute for Systems Architecture, Chair of Computer Networks, Technische Universität Dresden, Dresden – Alemanha, 2011. Disponível em: <https://nbn-resolving.org/urn:nbn:de:bsz:14-qucosa-72434>. Acesso em 11 jan. 2023

GAVRILIS, Dimitris; DALLAS, Costis; ANGELIS, Stravos. A curation-oriented thematic aggregator. 2013. In T. Aalberg, C. Papatheodorou, M. Dobрева, G. Tsakonas, e C. J. Farrugia (Eds.), Research and advanced technology for digital libraries. TPDL 2013. Lecture notes in computer science (Vol. 8092, pp. 132– 137). Springer. DOI: https://doi.org/10.1007/978-3-642-40501-3_13

GILLILAND, A. J. Setting the Stage. In: BACA, M. (Ed.). Introduction to metadata. 3. ed. Los Angeles: Getty Research Institute, 2016.

GITHUB. Europeana - Metis Framework. 2023. Disponível em: <https://github.com/europeana/metis-framework>. Acesso em 11 jan. 2023

GLUSHKO, Robert J. The discipline of organizing. BULLETIN of the American society for information science and technology, v. 40, n. 1, p. 21-27, 2013.

GOV.BR. Ministério do Turismo. Acervo em Rede e Projeto Tainacan. 2021d. Disponível em: <https://www.gov.br/museus/pt-br/aceso-a-informacao/acoes-e-programas/acervo-em-rede-e-projeto-tainacan>. Acesso em 12 jan. 2023.

GOV.BR. Ministério do Turismo. Museu Casa da Hera. 2023. Disponível em: <https://www.gov.br/museus/pt-br/museus-ibram/museu-casa-da-hera-ibram>. Acesso em 09 jan. 2023

GOV.BR. Ministério do Turismo. Museu Casa da Princesa. 2022. Disponível em: <https://www.gov.br/museus/pt-br/museus-ibram/museu-casa-da-princesa-ibram>. Acesso em 09 jan. 2023

GOV.BR. Ministério do Turismo. Museu Casa de Benjamin Constant. 2023a. Disponível em: <https://www.gov.br/museus/pt-br/museus-ibram/museu-casa-de-benjamin-constant-ibram>. Acesso em 09 jan. 2023

GOV.BR. Ministério do Turismo. Museu Casa Histórica de Alcântara. 2023b. Disponível em: <https://www.gov.br/museus/pt-br/museus-ibram/museu-casa-historica-de-alcantara-ibram>. Acesso em 09 jan. 2023

GOV.BR. Ministério do Turismo. Museu da Abolição. 2021a. Disponível em: <https://www.gov.br/museus/pt-br/museus-ibram/museu-da-abolicao-ibram>. Acesso em 09 jan. 2023

GOV.BR. Ministério do Turismo. Museu da Inconfidência. 2023c. Disponível em: <https://www.gov.br/museus/pt-br/museus-ibram/museu-da-inconfidencia>. Acesso em 09 jan. 2023

GOV.BR. Ministério do Turismo. Museu das Bandeiras. 2022a. Disponível em: <https://www.gov.br/museus/pt-br/museus-ibram/museu-das-bandeiras-ibram>. Acesso em 09 jan. 2023

GOV.BR. Ministério do Turismo. Museu das Missões. 2023d. Disponível em: <https://www.gov.br/museus/pt-br/museus-ibram/museu-das-missoes-ibram>. Acesso em 09 jan. 2023

GOV.BR. Ministério do Turismo. Museu de Arqueologia de Itaipu. 2023e. Disponível em: <https://www.gov.br/museus/pt-br/museus-ibram/museu-de-arqueologia-socioambiental-de-itaipu>. Acesso em 09 jan. 2023

GOV.BR. Ministério do Turismo. Museu de Arte Religiosa e Tradicional. 2023f. Disponível em: <https://www.gov.br/museus/pt-br/museus-ibram/museu-de-arte-religiosa-e-tradicional>. Acesso em 09 jan. 2023

GOV.BR. Ministério do Turismo. Museu de Arte Sacra da Boa Morte. 2022b. Disponível em: <https://www.gov.br/museus/pt-br/museus-ibram/museu-de-arte-sacra-da-boa-morte-ibram>. Acesso em 09 jan. 2023

GOV.BR. Ministério do Turismo. Museu do Diamante. 2021b. Disponível em: <https://www.gov.br/museus/pt-br/museus-ibram/museu-do-diamante-ibram>. Acesso em 09 jan. 2023"

GOV.BR. Ministério do Turismo. Museu do Ouro. 2022c. Disponível em: <https://www.gov.br/museus/pt-br/museus-ibram/museu-do-ouro>. Acesso em 09 jan. 2023

GOV.BR. Ministério do Turismo. Museu Histórico Nacional. 2023g. Disponível em: <https://www.gov.br/museus/pt-br/museus-ibram/museu-historico-nacional>. Acesso em 09 jan. 2023

GOV.BR. Ministério do Turismo. Museu Regional Casa dos Ottoni. 2023h. Disponível em: <https://www.gov.br/museus/pt-br/museus-ibram/museu-regional-casa-dos-ottoni>. Acesso em 09 jan. 2023

GOV.BR. Ministério do Turismo. Museu Regional de Caeté. 2022d. Disponível em: <https://www.gov.br/museus/pt-br/museus-ibram/museu-regional-de-caete>. Acesso em 09 jan. 2023

GOV.BR. Ministério do Turismo. Museu Regional de São João del-Rei. 2023i. Disponível em: <https://www.gov.br/museus/pt-br/museus-ibram/museu-regional-de-sao-joao-del-rei>. Acesso em 09 jan. 2023

GOV.BR. Ministério do Turismo. Museu Victor Meirelles. 2022e. Disponível em: <https://www.gov.br/museus/pt-br/museus-ibram/museu-victor-meirelles-ibram>. Acesso em 09 jan. 2023

GOV.BR. Ministério do Turismo. Museu Villa-Lobos. 2023j. Disponível em: <https://www.gov.br/museus/pt-br/museus-ibram/museu-villa-lobos>. Acesso em 09 jan. 2023

GOV.BR. Ministério do Turismo. Museus Ibram. 2023k. Disponível em: <https://www.gov.br/museus/pt-br/museus-ibram>. Acesso em 11 jan. 2023

GOV.BR. Ministério do Turismo. Sobre o órgão. 2023l. Disponível em: <https://www.gov.br/museus/pt-br/aceso-a-informacao/institucional/sobre-o-orgao>. Acesso em 11 jan. 2023

GOVERNO DO RIO DE JANEIRO. Secretaria de Estado da Cultura. Museus. Instituições que fazem parte da Rede Web de Museus. 2023a. Disponível em: <http://www.museusdoestado.rj.gov.br/rede-web-de-museus/instituicoes-que-fazem-parte-da-rede-web-de-museus/>. Acesso em 11 jan. 2023

GOVERNO DO RIO DE JANEIRO. Secretaria de Estado da Cultura. Museus. Rede Web de Museus. 2023b. Disponível em: <http://www.museusdoestado.rj.gov.br/rede-web-de-museus/>. Acesso em 11 jan. 2023

GUEGUEN, Gretchen et al. Metadata Application Profile, version 5.0. DPLA - Digital Public Library of America, 2017. Disponível em: https://drive.google.com/file/d/1fJEWWhnYy5Ch7_ef_-V48-FAViA72OieG/view. Acesso em 11 jan. 2023

HARPRING, Patricia. Introduction to controlled vocabularies: terminology for art, architecture, and other cultural works. Getty Publications, 2013.

HJORLAND, Birger. Semantics and knowledge organization. Annual Rev. Inf. Sci. Technol., v. 41, n. 1, p. 367-405, 2007.

IBICT. Sobre os Repositórios Digitais. 2012. Disponível em: <http://sitehistorico.ibict.br/informacao-para-ciencia-tecnologia-e-inovacao%20/repositorios-digitais>. Acesso em 11 jan. 2023

IBRAM (Instituto Brasileiro de Museus). Acervos digitais nos museus: manual para realização de projetos. Instituto Brasileiro de Museus; Universidade Federal de Goiás - Brasília, DF: Ibram, 2020. Disponível em: <https://www.museus.gov.br/wp-content/uploads/2021/05/Acervos-Digitais-nos-Museus.pdf>. Acesso em 10 jan. 2023

INNOCENTI, Perla; VULLO, Giuseppina; ROSS, Seamus. Towards a Digital Library Policy and Quality Interoperability Framework: The DL.org Project. 2010. New Rev. Inf. Netw. 15, 1 (May 2010), 29–53. DOI: <https://doi.org/10.1080/13614571003751071>

ISOTANI, Seiji; BITTENCOURT, Ig Ibert. Dados abertos conectados: em busca da web do conhecimento. Novatec Editora, 2015.

KIRSTEN, Toralf; KIEL, Alexander; RÜHLE, Mathias; WAGNER, Jonas. Metadata Management for Data Integration in Medical Sciences. Datenbanksysteme für Business, Technologies, and Web (BTW 2017), 2017. Disponível em: <https://dl.gi.de/handle/20.500.12116/627>. Acesso em 11 jan. 2023

KITCHENHAM, Barbara. Procedures for performing systematic reviews (Technical Report TR/SE-0401), 2004 Department of Computer Science, Keele University. Disponível em: <https://www.inf.ufsc.br/~aldo.vw/kitchenham.pdf>. Acesso em 11 jan. 2023

KITCHENHAM, Barbara; CHARTERS, Stuart. Guidelines for performing systematic literature reviews in software engineering (Technical Report EBSE 2007-001), 2007. Department of Computer Science, Keele University. Disponível em: <https://userpages.uni-koblenz.de/~laemmel/esecourse/slides/slr.pdf>. Acesso em 11 jan. 2023

KOLLIA, Ilianna; TZOUVARAS, Vassilis; DROSOPOULOS, Nasos. STAMOU, Giorgos. A systemic approach for effective semantic access to cultural content. Semantic Web, v. 3, n. 1, p. 65–83, 2012. DOI: <https://doi.org/10.3233/SW-2012-0051>

KUHLTHAU, Carol Collier. Inside the search process: Information seeking from the user's perspective. J. Am. Soc. Inf. Sci., 42: 361-371, 1991. DOI: [https://doi.org/10.1002/\(SICI\)1097-4571\(199106\)42:5<361::AID-ASL6>3.0.CO;2-#](https://doi.org/10.1002/(SICI)1097-4571(199106)42:5<361::AID-ASL6>3.0.CO;2-#)

LANCASTER, Frederick Wilfrid. Indexação e resumos: teoria e prática. Brasília: Briquet de Lemos, 2004.

LANCASTER, Frederick Wilfrid. Vocabulary control for information retrieval. 2ª ed. Virgínia: IRP, 1986.

LAPA, Rodrigo Amaral. Museu, Arte e Tecnologia: As transformações dos museus contemporâneos influenciadas pelas TIC's. Dissertação de Mestrado - Programa de Pós-Graduação em Arquitetura e Urbanismo e Área de Concentração em Teoria e História) - Instituto de Arquitetura e Urbanismo de São Carlos da Universidade de São Paulo, 2011. Disponível em: https://www.teses.usp.br/teses/disponiveis/18/18142/tde-01062012-141408/publico/rodrigo_lapa.pdf. Acesso em 11 jan. 2023

LARIVIÈRE, Vincent; SUGIMOTO, Cassidy R. Mesurer la Science, 2018, pp. 1– 176. Montréal, Canadá: Les Presses de l'université de. Disponível em: <https://www.pum.umontreal.ca/catalogue/mesurer-la-science/fichiers/1e1f0ce3-63b3-4a72-a29d-c9d84efafd65/9782760639522.pdf>. Acesso em 11 jan. 2023

LE MOS, Daniela Lucas da; SOUZA, Renato Rocha. Knowledge organization systems for the representation of multimedia resources on the web: A comparative analysis. Ko Knowledge Organization, v. 47, n. 4, p. 300-319, 2020. DOI: <https://doi.org/10.5771/0943-7444-2020-4-300>

MAKRIS, Konstantinos; SKEVAKIS, Giannis; KALOKYRI, Varvara; ARAPI, Polyxeni; CHRISTODOULAKIS, Stavros; STOITSIS, John; MANOLIS, Nikos; ROJAS, Sarah Leon. Federating natural history museums in natural Europe. 2013. In E. Garoufallou e J. Greenberg (Eds.), *Metadata and semantics research. MTSR 2013. Communications in computer and information science* (Vol. 390, pp. 361– 372). Springer. DOI: https://doi.org/10.1007/978-3-319-03437-9_35

MANGHI, Paolo; ARTINI, Michele; ATZORI, Claudio; BARDI, Alessia; MANNOCCI, Andrea; LA BRUZZO, Sandro; CANDELA, Leonardo; CASTELLI, Donatella; PAGANO, Pasquale. The D-NET software toolkit: A framework for the realization, maintenance, and operation of aggregative infrastructures. 2014. *Program*, v. 48, n. 4, p. 322–354, 27 ago. 2014. DOI: <https://doi.org/10.1108/PROG-08-2013-0045>

MANNOCCI, Andrea; CASAROSA, Vittore; MANGHI, Paolo; ZOPPI, Franco. (2014). The Europeana network of ancient Greek and Latin epigraphy data infrastructure. 2014. In S. Closs, R. Studer, E. Garoufallou, e M. A. Sicilia (Eds.), *Metadata and semantics research. MTSR 2014. Communications in computer and information science* (Vol. 478, pp. 286– 300). Springer. DOI: https://doi.org/10.1007/978-3-319-13674-5_27

MARCONDES, Carlos Henrique. Interoperabilidade entre acervos digitais de arquivos, bibliotecas e museus: potencialidades das tecnologias de dados abertos interligados. *Perspectivas em Ciência da Informação*, [S.l.], v. 21, n. 2, p. 61-83, jun. 2016. ISSN 19815344. Disponível em: <https://www.scielo.br/j/pci/a/8svGtzqw5HZCrfrPjBRypsb/?lang=pt&format=html>. Acesso em 11 jan. 2023

MARCONDES, Carlos Henrique; BETTENCOURT, Angela Maria Monteiro. Elementos para uma política brasileira de acesso integrado, utilização e preservação de acervos digitais em memória e cultura. 2019. *PragMATIZES - Revista Latino-Americana de Estudos em Cultura*, n. 16, p. 44-61, 3 jun. 2019. DOI: <https://doi.org/10.22409/pragmatizes.v0i16.27518>

MARTÍN-MARTÍN, Alberto; ORDUNA-MALEA, Enrique; THELWALL, Mike; DELGADO LÓPEZ-CÓZAR, Emílio. Google Scholar, Web of Science, and Scopus: A systematic comparison of citations in 252 subject categories. 2018, *Journal of Informetrics*, 12(4), 1160– 1177. DOI: <https://doi.org/10.1016/j.joi.2018.09.002>

MARTINS, Dalton Lopes; LEMOS, Daniela Lucas da Silva, OLIVEIRA, Luís Felipe Rosa de; SIQUEIRA, Joyce, CARMO, Danielle do, & MEDEIROS, Vinícius Nunes, 2022. Information organization and representation in digital cultural heritage in Brazil: Systematic mapping of information infrastructure in digital collections for data science applications. *Journal of the Association for Information Science and Technology*. DOI: <https://doi.org/10.1002/asi.24650>

MARTINS, Dalton Lopes; LEMOS, Daniela Lucas da Silva; CARMO, Danielle do; SIQUEIRA, Joyce; OLIVEIRA, Luís Felipe Rosa de. Requisitos de Qualidade para dados de agregação em Museus: O Caso do Ibram. In: XXI Encontro Nacional de Pesquisa e Pós-graduação em Ciência da Informação, 2021, Rio de Janeiro. XXI Enancib, 2021. Disponível em: <https://enancib.ancib.org/index.php/enancib/xxienancib/paper/view/575>. Acesso em 10 jan. 2023

MARTINS, Dalton Lopes; MARTINS, Luciana Conrado. Desafios e Aprendizados na Implantação do Tainacan nos Museus do Instituto Brasileiro de Museus. Revista Eletrônica Ventilando Acervos - Florianópolis: MVM, 2021 – ISSN 2318-6062. Disponível em: <https://ventilandoacervos.museus.gov.br/wp-content/uploads/2021/08/A6-Luciana-e-Dalton.pdf>. Acesso em 11 jan. 2023

MARTINS, Dalton Lopes; SILVA, Marcel Ferrante; CARMO, Danielle. Acervos em rede: perspectivas para as instituições culturais em tempos de cultura digital. Em Questão, v. 24, n. 1, p. 194-216, 2018. DOI: <https://doi.org/10.19132/1808-5245241.194-216>

MARTINS, Luciana Conrado; MARTINS, Dalton Lopes. Experimentações sociotécnicas para organização e difusão de coleções digitais universitárias: o caso do projeto Tainacan. Revista CPC, [S. l.], v. 15, n. 30esp, p. 34-61, 2020. DOI: <https://doi.org/10.11606/issn.1980-4466.v15i30espp34-61>

MARTINS, Vinicius Pontes. BNDigital–10 anos: retrospectiva e perspectivas para os próximos 10 anos. Seminário Serviços de Informação em Museus, p. 159-169, 2018. Disponível em: <http://biblioteca.pinacoteca.org.br:9090/bases/biblioteca/322707.pdf>. Acesso em 10 jan. 2023

MATIENZO, Mark A.; RUDERSDORF, Amy. The Digital Public Library of America ingestion ecosystem: Lessons learned after one year of large-scale collaborative metadata aggregation. 2014, In Proceedings of the 2014 international conference on Dublin Core and Metadata Applications (DCMI'14). Dublin Core metadata initiative (pp. 12– 23). Springer. DOI: <https://dl.acm.org/doi/10.5555/2771234.2771236>

MÉXICO. SECRETARÍA DE CULTURA. Mexicana Repositorio del Patrimonio Cultural de México. Ciudad de México: Secretaría de Cultura, 2018. Disponível em: <https://mexicana.cultura.gob.mx/work/models/repositorio/Resource/126/2/images/documentacion.pdf>. Acesso em 11 jan. 2023

MILES, Alistair; ZHAO, Jun; KLYNE, Graham, WHITE-COOPER, Helen; SHOTTON, David. OpenFlyData: an exemplar data web integrating gene expression data on the fruit fly *Drosophila melanogaster*. Journal of biomedical informatics, 43(5), 752–761, 2010. DOI: <https://doi.org/10.1016/j.jbi.2010.04.004>

MINISTÉRIO DO TURISMO. Plano de Integridade do Instituto Brasileiro de Museus, 2019. Disponível em: https://www.gov.br/museus/pt-br/aceso-a-informacao/institucional/plano_de_integridade_ibram_v_17-12-19.pdf. Acesso em 11 jan. 2023

MUCHACHO, Rute. O Museu Virtual: as novas tecnologias e a reinvenção do espaço museológico, 2005. Biblioteca online de ciências da comunicação. Disponível em: <http://bocc.ufp.pt/pag/muchacho-rute-museu-virtual-novas-tecnologias-reinvencao-espaco-museologico.pdf>. Acesso em 11 jan. 2023

NATIONAL LIBRARY OF AUSTRALIA. Trove Help Center. Trove System Architecture Diagram, 2010. Disponível em: www.nla.gov.au/trove/marketing/Trove%20architecture%20diagram.pdf. Acesso em 14 set. 2019.

NAVARRETE, Trilce. Europeana as online culturais information service: study report. [S. l.]: Europeana, 2016. Disponível em:

https://pro.europeana.eu/files/Europeana_Professional/Publications/europeana-benchmark-report-sep-2016.pdf. Acesso em 11 jan. 2023

NISO (National Information Standards Organization). Understanding Metadata: What is Metadata, and What is it For?: A Primer. Bethesda, MD: NISO Press. 2004. Disponível em: <http://www.niso.org/standards/resources/UnderstandingMetadata.pdf>. Acesso em 11 jan. 2023

OFOEDA, Joshua; BOATENG, Richard; EFFAH, John. Application Programming Interface (API) Research: A Review of the Past to Inform the Future, 2019. International Journal of Enterprise Information Systems, 15(3), 76–95. DOI: <https://doi.org/10.4018/ijeis.2019070105>

OOMEN, Johan; TZOUVARAS, Vassilis. Publishing Europe's television heritage on the web: The EUscreen project. 2012, In C. Grana & R. Cucchiara (Eds.), Multimedia for cultural heritage. MM4CH 2011. Communications in computer and information science (Vol. 247, pp. 136– 142). Springer. DOI: https://doi.org/10.1007/978-3-642-27978-2_12

OOMEN, Johan; TZOUVARAS, Vassilis; Hyypää, Kati. Linking and visualizing television heritage: The EUscreen virtual exhibitions and the linked open data pilot. 2013, In Proceedings of the 22nd international conference on world wide web (WWW'13 companion) (pp. 481–484). Association for Computing Machinery. DOI: <https://doi.org/10.1145/2487788.2487974>

PAPATHEODOROU, Christos; DALLAS, Costis; ERTMANN-Christiansen, Christian; FERNIE, Kate; GAVRILIS, Dimitris; MASCI, Maria Emilia; CONSTANTOPOULOS, Panos; ANGELIS, Stravos. A new architecture and approach to asset representation for Europeana aggregation: The CARARE way, 2011. In E. Garcia-Barriocanal, Z. Cebeci, M. C. Okur, e A. Öztürk (Eds.), Metadata and semantic research. MTSR 2011. Communications in computer and information science (Vol. 240, pp. 412– 423). Springer. DOI: https://doi.org/10.1007/978-3-642-24731-6_41

PAVÃO, Caterina Marta Groposo; CAREGNATO, Sônia Elisa. Serviços de descoberta em rede: a experiência do modelo Google para os usuários de bibliotecas universitárias. Em Questão, v. 21, n. 3, p. 130-149, 2015. Disponível em: <https://www.seer.ufrgs.br/EmQuestao/article/view/58410>. Acesso em 11 jan. 2023

PIPERIDIS, Stelios; PAPAGEORGIOU, Harris; SPURK, Christian; REHM, Georg; CHOUKRI, Khalid; HAMON, Oliver; CALZOLARI, Nicoletta; GRATTA, Riccardo Del; MAGNINI, Bernardo; GIRARDI, Christian. Meta-share: One year after, 2014. In N. C. Chair, K. Choukri, T. Declerck, H. Loftsson, B. Maegaard, J. Mariani, A. Moreno, J. Odijk, e S. Piperidis (Eds.), Proceedings of the ninth international conference on language resources and evaluation (LREC'14) (pp. 1532– 1538). European Language Resources Association (ELRA). Disponível em: http://www.lrec-conf.org/proceedings/lrec2014/pdf/786_Paper.pdf. Acesso em 11 jan. 2023

PLANO B. Brasileira Iconográfica - Portfólio. 2023a. Disponível em: <https://plano-b.com.br/portfolio/1238/brasiliانا-iconografica>. Acesso em 11 jan. 2023

PLANO B. Shiro 3.0 - Plataforma para Gestão Colaborativa de Acervos Digitais. 2023b. Disponível em: <https://plano-b.com.br/portfolio/1238/brasiliانا-iconografica>. Acesso em 11 jan. 2023

HISTÓRIA CIÊNCIAS SAÚDE MANGUINHOS. BN e IMS lançam portal Brasileira Fotográfica, 2015. Disponível em: <https://www.revistahcsm.coc.fiocruz.br/bn-e-ims-lancam-portal-brasiliana-fotografica/>. Acesso em 11 jan. 2023

RONZINO, Paola; FELICETTI, Achille; GIORGIO, Sara Di. Integrating archaeological datasets: The Ariadne Portal, 2016. In A. S. Luciana, F. Bordoni, e F. Mele (Eds.), International workshop on artificial intelligence for cultural heritage co-located with the 15th international conference of the Italian association for artificial intelligence (AI*IA 2016) (pp. 42–49). CEUR-WS. Disponível em: <http://ceur-ws.org/Vol-1772/paper6.pdf>. Acesso em 11 jan. 2023

SANTAREM SEGUNDO, José Eduardo; SILVA, Marcel Ferrante; MARTINS, Dalton Lopes. Revisitando a interoperabilidade no contexto dos acervos digitais, 2019. Informação & Sociedade: Estudos, [S. l.], v. 29, n. 2. Disponível em: <https://periodicos.ufpb.br/ojs2/index.php/ies/article/view/38107>. Acesso em 11 jan. 2023

SARACEVIC, Tefko. A natureza interdisciplinar da ciência da informação. Ciência da Informação, [S. l.], v. 24, n. 1, 1995. Disponível em: <https://www.brapci.inf.br/index.php/res/v/22344>. Acesso em 11 jan. 2023

SARACEVIC, Tefko. Information Science. Journal of the American Society for Information Science. 50(12): 1051–1063, 1999. Disponível em: https://ils.unc.edu/courses/2013_spring/inls285_001/materials/Saracevic.T.1999.Information_science.JASIS.50.12.1051-1063.pdf. Acesso em 10 jan. 2023

SARIKAYA, Alper; CORRELL, Michael; BARTRAM, Lyn; TORY, Melanie e FISHER, Danyel. What Do We Talk About When We Talk About Dashboards? IEEE Transactions on Visualization and Computer Graphics, v. 25, n. 1, p. 682-692, 2019. DOI: <https://doi.org/10.1109/TVCG.2018.2864903>

SCHOLZ, Henning. EPF - Publishing guidelines, 2022. Europeana Knowledge Base. Disponível em: <https://europeana.atlassian.net/wiki/spaces/EF/pages/2059763713/EPF+-+Publishing+guidelines>. Acesso em 10 jan. 2023

SCHOLZ, Henning. Europeana Publishing Guide v1.8. Europeana Foundation. 2019. https://pro.europeana.eu/files/Europeana_Professional/Publications/Europeana%20Publishing%20Guide%20v1.8.pdf. Acesso em 10 jan. 2023

SIQUEIRA, Joyce; CARMO, Danielle do; MARTINS, Dalton Lopes; LEMOS, Daniela Lucas da Silva; MEDEIROS, Vinicius Nunes; OLIVEIRA, Luís Felipe Rosa de. Elements for Constructing a Data Quality Policy to Aggregate Digital Cultural Collections: Cases of the Digital Public Library of America and Europeana Foundation. In: Bisset Álvarez, E. (eds) Data and Information in Online Environments. DIONE 2021. Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering, vol 378. Springer, Cham. https://doi.org/10.1007/978-3-030-77417-2_8

SIQUEIRA, Joyce; MARTINS, Dalton Lopes. Análise e mapeamento da organização técnica dos acervos digitais das instituições federais vinculadas ao Ministério da Cultura: fontes de informação para o desenvolvimento das Humanidades Digitais. 2021a, In: PIMENTA, Ricardo M.; ALVES, Daniel. Humanidades digitais e o mundo lusófono. Local: FGV, 2021. p. 283-309. Disponível em:

<https://editora.fgv.br/produto/humanidades-digitais-e-o-mundo-lusofono-3625>.

Acesso em 10 jan. 2023

SIQUEIRA, Joyce; MARTINS, Dalton Lopes. LEMOS, Daniela Lucas da Silva. *Brasiliانا Museums: serviço de busca e recuperação da informação agregada dos acervos digitais do Instituto Brasileiro de Museus*. In: XXII Encontro Nacional de Pesquisa e Pós-graduação em Ciência da Informação, 2022, Porto Alegre. XXII Enancib, 2022. Disponível em: <https://enancib.ancib.org/index.php/enancib/xxiienancib/paper/view/712>. Acesso em 13 jan. 2023

SIQUEIRA, Joyce; MARTINS, Dalton Lopes. *Painel de visualização analítica dos acervos digitais integrados do instituto brasileiro de museus: o uso das tecnologias Tainacan e Elastic Stack*. In: XXI Encontro Nacional de Pesquisa e Pós-graduação em Ciência da Informação, 2021b, Rio de Janeiro. XXI Enancib, 2021. Disponível em: <https://enancib.ancib.org/index.php/enancib/xxienancib/paper/view/95>. Acesso em 12 jan. 2023

SIQUEIRA, Joyce; MARTINS, Dalton Lopes. *Recuperação de informação: descoberta e análise de workflows para agregação de dados do patrimônio cultural*. *Ciência da Informação*, [S. l.], v. 49, n. 3, 2020. Disponível em: <https://revista.ibict.br/ciinf/article/view/5399>. Acesso em: 12 jan. 2023.

SIQUEIRA, Joyce; MARTINS, Dalton Lopes. *Serviços de descoberta em escala na web voltados a repositórios digitais utilizados por GLAM's: Uma revisão sistemática de literatura*, 2019a. In *Encontro nacional de pesquisa em ciência da informação. Programa de Pós-graduação em Ciência da Informação da UFSC*. Disponível em: <https://conferencias.ufsc.br/index.php/enancib/2019/paper/view/588>. Acesso em 10 jan. 2023

SIQUEIRA, Joyce; MARTINS, Dalton Lopes. *Workflow de agregação de dados: Processos para criação de uma interface de busca integrada do patrimônio cultural*, 2019b. *Workshop de informação, dados e tecnologia (3: 2019: Brasília)*. *Workshop de informação, dados e tecnologia (WIDAT 2019): anais do evento/Dalton Lopes Martins ... [et al.]*, organizadores. Brasília: Universidade de Brasília, Faculdade de Ciência da Informação, 2019. ISBN: 978-65-86503-01-2. Disponível em: <http://widat2019.fci.unb.br/index.php/anais-widat-2019>. Acesso em 10 jan. 2023

SIQUEIRA, Joyce; MARTINS, Dalton Lopes. *Workflow models for aggregating cultural heritage data on the web: A systematic literature review*, 2021c. *Journal of the Association for Information Science and Technology*, 1–21, 2021. DOI: <https://doi.org/10.1002/asi.24498>

SIQUEIRA, Joyce; MARTINS, Dalton Lopes; MEDEIROS, Vinícius Nunes. *Brasiliانا Museums: teste funcional do agregador de dados museais do Instituto Brasileiro de Museus*. *Workshop de informação, dados e tecnologia (2022: Espírito Santo)*. *Workshop de informação, dados e tecnologia (WIDAT 2022): anais do evento/Daniela Lucas da Silva Lemos ... [et al.]*, organizadores. Vitória: Universidade Federal do Espírito Santo. Faculdade de Ciência da Informação, 2022a. Disponível em: anais ainda não estão disponíveis. Acesso em 10 jan. 2023

SIQUEIRA, Joyce; MARTINS, Dalton Lopes; MEDEIROS, Vinícius Nunes. *Metadata mapping in Tainacan: new functionality for digital museums linked to the Brazilian*

Institute of Museums. En T.M.R. Dias (Ed.), Informação, Dados e Tecnologia. Advanced Notes in Information Science, volume 2 (pp. 182-191), 2022b. Tallinn, Estonia: ColNes Publishing. DOI: <https://doi.org/10.47909/anis.978-9916-9760-3-6.92>

SIQUEIRA, Joyce; MARTINS, Dalton Lopes; MEDEIROS, Vinícius Nunes. Workflow with Emphasis in Technologies for the Construction of an Aggregator and a Dashboard of Brazilian Museums Digital Collections. In: Pinto, A.L., Arencibia-Jorge, R. (eds) Data and Information in Online Environments. DIONE 2022. 2022c. Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering, vol 452. Springer, Cham. https://doi.org/10.1007/978-3-031-22324-2_8

SOUTHWICK, Silvia B. A Guide for Transforming Digital Collections Metadata into Linked Data Using Open-Source Technologies, 2015, Journal of Library Metadata, v. 15, n. 1, p. 1–35, 2. DOI: <https://doi.org/10.1080/19386389.2015.1007009>

STATHOPOULOU, Ioanna-Ourania; STATHOPOULOS, Panagiotis; GEORGIADIS, Haris; HOUSSOS, Nikos; BANOS, Vangelis; SACHINI, Evi. An open cultural digital content infrastructure, 2014, pp. 285– 288, IEEE/ACM Joint Conference on Digital Libraries. DOI: <https://doi.org/10.1109/JCDL.2014.6970181>

SUPPLEJACK. Architecture. Documentation (Version 0.1), 2023. Disponível em: <http://digitalnz.github.io/supplejack/architecture.html>. Acesso em 10 jan. 2023

SVENONIUS, Elaine. The intellectual foundation of information organization. MIT press, 2000.

TAINACAN. Casos de uso do Tainacan, 2023. Disponível em: <https://tainacan.org/casos-de-uso-do-tainacan/>. Acesso em 10 jan. 2023

TAINACAN. Conceitos Gerais. 2022. Disponível em: <https://tainacan.github.io/tainacan-wiki/#/pt-br/general-concepts>. Acesso em 09 jan. 2023.

TAINACAN. Indexando Conteúdos em PDF. 2022a. Disponível em: <https://tainacan.github.io/tainacan-wiki/#/pt-br/indexar-pdf>. Acesso em 09 jan. 2023.

TAINACAN.ORG. Documentação da API. 2023. Disponível em: <https://tainacan.org/api-docs/>. Acesso em 09 jan. 2023

TAINACAN.ORG. Início. 2023. Disponível em: <https://tainacan.org/>. Acesso em 09 jan. 2023

TAINACAN.ORG. Plugin: PDF Exposer. 2022. Disponível em: <https://tainacan.github.io/tainacan-wiki/#/plugin-pdf-exposer?id=introduction>. Acesso em 11 jan. 2023

TROVE. About. 2023. Disponível em: <https://trove.nla.gov.au/about>. Acesso em 10 jan. 2023

WORDPRESS. Conheça o Wordpress. 2023. Disponível em: <https://br.wordpress.org/>. Acesso em 10 jan. 2023

YIN, Robert K. Estudo de Caso: Planejamento e métodos. Editora: Porto Alegre: Bookman, 2015. 5. ed. 290 p. ISBN: 9788582602317.

ZIÉBELIN, Danielle; HOBUS, Kim; GENOUD, Philippe; BOUVERET, Sylvain. Heterogeneous Data Integration Using Web of Data Technologies, 2017. In: Brosset D., Claramunt C., Li X., Wang T. (eds) Web and Wireless Geographical Information Systems. W2GIS 2017. Lecture Notes in Computer Science, v. 10181. Springer. DOI: https://doi.org/10.1007/978-3-319-55998-8_3