

XX ENANCIB

21 a 25 Outubro/2019 – Florianópolis

A Ciência da Informação e a era da Ciência de Dados

ISSN 2177-3688

GT-8 – Informação e Tecnologia

FUNDAÇÃO NACIONAL DE ARTES: ESTUDO DE CASO DA MIGRAÇÃO E PUBLICAÇÃO DOS DADOS DO ACERVO DIGITAL COM O SOFTWARE LIVRE TAINACAN

NATIONAL ARTS FOUNDATION: CASE STUDY ON DIGITAL COLLECTION DATA MIGRATION AND PUBLICATION WITH TAINACAN FREE SOFTWARE

Luis Felipe Rosa de Oliveira - UnB

Dalton Lopes Martins - UnB

Modalidade: Trabalho Completo

Resumo: Este artigo expõe um estudo de caso de migração e reformulação de acervo do projeto Funarte Digital, realizado a partir de uma parceria entre o Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT) e a Fundação Nacional de Artes (Funarte). O estudo parte de preceitos técnicos aplicados de tratamento e organização de dados, como análise descritiva e pré-processamento de dados, aliados a um aporte metodológico que é constituído das etapas de análise, coleta, tratamento, migração, validação e formação, e difusão e colaboração em rede. O projeto utilizou como software de disponibilização de publicação do acervo a plataforma livre Tainacan, e como resultados, são caracterizadas as ações tomadas em cada uma das etapas metodológicas. Como conclusão, foi expressa a importância da aplicação das técnicas de tratamento e organização de dados, que possibilitaram reestruturar o acervo centralizando-o, ao identificar que inconstâncias na estrutura da instituição foram refletidas no modo como o acervo foi estruturado, dividido em nichos formando barreiras ao seu acesso e uso. Além disso, a reestruturação do acervo abre caminho para que etapas de modelagem conceitual e enriquecimento semântico dos dados sejam aproveitados efetivamente pela instituição.

Palavras-Chave: Migração de Acervo Digital; Análise Descritiva de Dados; Pré-Processamento de Dados; Funarte; Tainacan

Abstract: This paper presents a case study of migration and reformulation of the Funarte Digital project collection, conducted through a partnership between the Brazilian Institute of Information on Science and Technology (IBICT) and the National Arts Foundation (Funarte). The study departs from applied technical precepts of data processing and organization, such as descriptive analysis and data preprocessing, combined with a methodological approach that consists of the stages of analysis, collection, treatment, migration, validation and formation, and dissemination and dissemination and network collaboration. The project used as software for publication of the collection the free platform Tainacan, and as a result, the actions taken in each of the methodological steps are characterized. In conclusion, it was expressed the importance of applying the data processing and organization techniques, which continued the restructuring of the collection by centralizing it, by identifying that inconsistencies in the structure of the institution were reflected in the way the collection was structured, divided into niches forming barriers to access and use. In addition, the restructuring of the collection opens the way for conceptual modeling and semantic data enrichment steps to be effectively used by the institution.

Keywords: Digital Collection Migration; Descriptive Data Analysis; Data Preprocessing; Funarte; Tainacan.

1 INTRODUÇÃO

Historicamente o domínio sobre os acervos culturais e de memória serviram de condição para reforçar o poder do estado, e com as possibilidades tecnológicas atuais a sociedade tem a chance de reverter essa situação. (MARTINS; CARVALHO, 2017). Não somente exigindo a publicação dos acervos digitais on-line, mas sim uma real condição de acesso e usabilidade desses acervos de forma livre e estruturada, sem controle de empresas privadas, ou limitações de formatos proprietários. Assim é inicialmente contextualizado o estudo de caso aqui proposto, em um anseio científico por uma pesquisa aplicada que não almeja contribuir somente com mais um potencial técnico-metodológico, mas sim que tenha o intuito de colaborar com outras iniciativas que buscam fornecer condições equivalentes de acesso real à informação.

A partir dessas premissas é proposto o estudo de caso do projeto de organização e migração do acervo da Fundação Nacional de Artes (Funarte) realizado em parceria com o Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT) denominado Funarte Digital. Sob perspectivas de tratamento e organização de dados, busca-se descrever quais os aspectos do método de migração utilizado no projeto como estratégia de centralização do acervo digital de relevância da Funarte, objetivando contribuir com experiências práticas de aplicações de migração em acervos digitais da cultura brasileira.

Esse estudo de caso se justificava no esforço de responder a preocupação sobre facilitar o acesso estruturado de acervos digitais de instituições culturais no Brasil. Dessa forma, documentar cientificamente as etapas da estratégia de migração do acervo digital da Funarte pode contribuir para a comunidade inerente como uma produção aplicada no contexto cultural.

O projeto Funarte Digital nasce da ideia da Instituição de reformular seu site, tornando-o um ambiente mais inclusivo, participativo e com melhores estratégias para acesso dinâmico aos relevantes conteúdos de seu acervo sobre diversas linguagens artísticas. A proposta desse projeto, e o aspecto em que esse estudo é focado, não é apenas a resposta ao desejo de construção de um novo portal de informações, mas a possibilidade de diagnosticar, organizar e reformular os acervos da Funarte na nova visão de um repositório digital integrado, constituindo uma fonte de informação organizada para desenvolvimento de serviços e produtos de informação para um órgão que se lança nos desafios da cultura digital do século XXI.

Dessa forma, são utilizadas aplicações de tratamento e organização de dados como a análise descritiva e o pré-processamento de dados, técnicas que auxiliaram na compreensão e estruturação das informações da instituição a partir de suas bases de dados, e permitiram a reestruturação do ambiente informacional.

Considera-se ainda que esse ambiente em sua totalidade permitirá a integração, o acesso e o compartilhamento de conhecimentos, sendo o seu principal objetivo o de promover o amplo acesso ao acervo do Centro de Documentação da Funarte (CEDOC), além daqueles que já se encontram em domínio público ou que tenham a sua divulgação devidamente autorizada.

A plataforma utilizada no processo de migração do acervo digital da Funarte foi o Tainacan, uma plataforma livre de publicação de acervos digitais, desenvolvido a partir da necessidade de uma “solução tecnológica livre de fácil utilização e capaz de desmistificar o exercício da interoperabilidade entre modelos de dados dos diferentes domínios de acervos culturais ” (MARTINS; CARVALHO JÚNIOR; GERMANI, 2019, p.60). O Tainacan responde à problemáticas como o armazenamento e publicação de conteúdos em múltiplos formatos como vídeos, áudio, textos e imagens em um só repositório, o que dá condições de integração

entre os objetos através de seus metadados, tendo como otimização do acesso uma de suas contribuições.

O método utilizado para a migração do acervo utilizando o Tainacan é baseado no modelo aplicado no Museu do Índio e proposto por Martins, Carmo e Germani (2018), que consiste em um conjunto de sete etapas, desde a análise contextual da instituição e de seu acervo, até sua publicação e colaboração em rede.

Desse modo, o estudo se estrutura nos próximos tópicos a partir da apresentação do contexto institucional das duas organizações parceiras no projeto, bem como aspectos do tratamento e organização de dados, além de um diagnóstico da situação inicial do acervo digital da Funarte, seguido dos aspectos da metodologia utilizada como referência, e finda na descrição dos resultados das etapas de migração realizadas no projeto.

2 CONTEXTO INSTITUCIONAL

A Funarte,

“Criada em 1975, é o órgão do Governo Federal brasileiro cuja missão é promover e incentivar a produção, a prática, o desenvolvimento e a difusão das artes no país. É responsável pelas políticas públicas federais de estímulo à atividade produtiva artística brasileiras; e atua para que a população possa cada vez mais usufruir das artes. [...] vinculada ao Ministério da Cidadania, alcança as áreas de circo, dança e teatro; de música, de concerto, popular e de bandas; e de artes visuais; e também a preservação da memória das artes e a pesquisa na esfera artística. É a única instituição no Estado brasileiro com as atribuições e especialidades necessárias para tratar desses campos de atividade” (FUNARTE, 2019).

Já o IBICT tem por missão promover a competência, o desenvolvimento de recursos, a infraestrutura de informação e a transferência de tecnologias de informação por meio dos fundamentos da Ciência da Informação. Destaca-se que a transferência de tecnologias de informação é uma das ações que consolidam o IBICT como referência na área no Brasil e no exterior, sendo o Instituto demandado por diferentes instituições para a realização de pesquisas com foco na prospecção, avaliação, customização e capacitação nas diversas soluções tecnológicas voltadas para as mais diferentes áreas do conhecimento.

É no escopo do projeto Funarte Digital que as duas instituições – IBICT e Funarte – congregam com o intuito de realização de pesquisas para a geração de conhecimentos e sua aplicação para a disponibilização aos usuários de um sistema de informação. Este ambiente propõe o compartilhamento do acervo da instituição compreendido por suas diversas

manifestações artísticas: música, dança, teatro, artes visuais, circo e literatura. O conteúdo a ser organizado, armazenado e disseminado abertamente contribuirá para a pesquisa de professores, artistas, estudantes de artes, pesquisadores e todo cidadão interessado na memória cultural brasileira.

3 TRATAMENTO E ORGANIZAÇÃO DO ACERVO

O escopo que delimita o foco do então estudo consiste na utilização de um conjunto de técnicas e métodos de tratamento de dados que possibilitaram a centralização do acervo e a facilitação de seu acesso. Para isso, as principais técnicas utilizadas foram a análise descritiva e o pré-processamento dos dados.

Para buscar esclarecimento sobre como a instituição lida com o armazenamento de suas informações ao estruturar suas bases de dados, a técnica de análise descritiva de dados (ADD) se mostra uma aplicação útil. A ADD “é utilizada para descrever, simplificar ou sumarizar as principais características de uma base de dados” (DE CASTRO; FERRARI, 2017, p.60), o que revela tanto o dimensionamento do ambiente informacional, quanto provê indicadores de síntese que auxiliam na tomada de decisões sobre o escopo informacional.

O processo de aplicação da ADD pode ser compreendido em três passos: organizar os dados utilizando distribuições de frequência; apresentar os dados usando técnicas de visualização; calcular medidas de tendência central, variação e associação (DE CASTRO; FERRARI, 2017, p.62). Esse processo dá condições para os gestores da instituição compreenderem inicialmente as características estruturais de suas bases de dados, além de situar o campo de ação em que as técnicas de organização e pré-processamento de dados serão aplicadas.

Outra técnica que também é inerente ao tratamento do banco de dados é o pré-processamento de dados. Essa técnica lida com possíveis problemas na qualidade dos dados, como a incompletude, que consiste em ausência de valores, atributos ou objetos; também o ruído, como dados com valores que apresentam muita variação; e as inconsistências, que ocorrem quando um mesmo dado aparece em diferentes locais configurando a violação de domínio e a discrepância de valores (DE CASTRO; FERRARI, 2017, p.26-27).

Aplicar o pré-processamento de dados é importante para evitar interpretações equivocadas em resultados analíticos, ou na reformulação dos fluxos de informação. Na

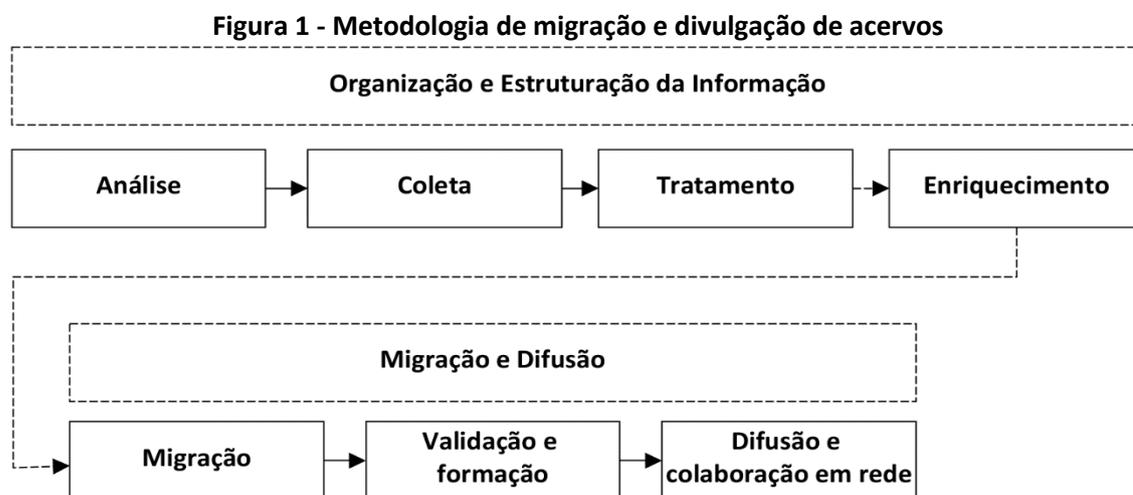
medida que existem problemas na qualidade dos dados armazenados, a possibilidade de se apropriar do conhecimento que esses dados podem oferecer fica comprometida, resultando em decisões pouco efetivas e limitação de acesso à informação.

Ambas técnicas apresentadas, tanto de análise descritiva quanto o pré-processamento de dados compõe um conjunto de outras técnicas propostas pela temática da ciência de dados, que objetiva fornecer técnicas e ferramentas com o foco na preocupação e apropriação do contexto informacional constituído sob dados de determinado ambiente. Neste estudo são descritas essas duas técnicas especificamente pelas limitações do desafio técnico de abertura e integração dos acervos em um único repositório, de modo que a instituição não apresentou necessidade do uso de técnicas mais avançadas para alcançar seus principais objetivos com o projeto.

De modo que, utilizando desse escopo de tratamento e organização de dados, foi possível promover a centralização do acervo, e a facilitação de seu acesso. A ADD permitiu diagnosticar a situação do contexto de armazenamento de informações da instituição, e o pré-processamento possibilitou o tratamento destes dados armazenados permitindo obtê-los de diferentes nichos e integrá-los em um só ponto interligado de acesso.

4 METODOLOGIA

A metodologia do processo de migração e divulgação do acervo da Funarte compartilha dos mesmos princípios da metodologia proposta por Martins, Carmo e Germani (2018), adaptada a este caso e reproduzida no diagrama na figura 2.



Fonte: Adaptado de Martins, Carmo e Germani (2018, p. 5144 e 5145)

O atual projeto com a Funarte como já mencionado, foi dividido em duas etapas, a primeira em que parte dos itens do acervo digitalizado foram migrados para o Tainacan, e a segunda onde mais itens serão selecionados para a migração e uma nova forma de apresentação do acervo será proposta, bem como procedimentos semânticos serão aplicados aos dados, nos moldes do que foi executado com o Museu do Índio (MARTINS; CARMO; GERMANI, 2018). Dessa forma, o caso atual percorreu todos os passos metodológicos apresentados com exceção do enriquecimento, que constitui a etapa de aplicação semântica sobre os dados e será executada na segunda etapa do projeto no primeiro semestre de 2020.

Quanto as características técnicas de aplicação da metodologia no caso do acervo da Funarte, na etapa de **análise** foram realizadas reuniões com os gestores das áreas informacionais em julho de 2018. Na primeira visita foram apresentados os acervos físicos, os materiais digitais, e coletados os catálogos e estruturas de arquivos que pudessem indicar o dimensionamento do acervo, informação até então superficial e imprecisa entre os gestores. Para isso, os documentos textuais, PDFs, pastas, imagens e tabelas foram transferidas para um HD externo para posterior análise, já para obter os quantitativos dos registros armazenados no *Sophia Web* (sistema proprietário de gestão de materiais bibliográficos e até então utilizado para catalogação de acervo arquivístico e documental), foi coletado o banco de dados ainda na instituição, no formato proprietário *SQL Server*, e convertido para o formato aberto *MySQL* utilizando a ferramenta de conversão do *MySQL Workbench*.

Os dados obtidos foram apropriados utilizando a análise descritiva de dados e novamente apresentados aos gestores, onde foi definido qual o recorte prioritário de itens do acervo a serem migrados na primeira etapa do projeto.

A **coleta** ocorreu em dois contextos diferentes, já tendo definido o recorte do acervo para migração, que foi composto pelos itens disponíveis no site e por parte dos objetos digitalizados do CEDOC. No caso do site obtivemos acesso de administrador do *Wordpress*, tecnologia utilizada para o site da Funarte, e acesso remoto ao servidor onde os dados estavam armazenados, e foram utilizados scripts na linguagem PHP¹ para recuperar os itens dispersos nos diversos pontos do site, e organizá-los no formato de coleções em um só ponto

¹ Scripts de migração em PHP disponibilizados em: https://github.com/medialab-ufg/wp-funarte/tree/master/migration_scripts

de acesso pelo Tainacan. Vale dizer que o site era utilizado para publicação de vários objetos digitais de diferentes coleções da Funarte, mas não ainda de forma sistematizada e integrada.

O segundo contexto da coleta foi aplicado nos itens digitalizados no *Sophia Web* selecionados pelo CEDOC (Centro de Documentação), em que foi fornecida uma tabela com informações de identificação dos itens a serem migrados, um HD externo com as imagens digitalizadas de todos os itens, e a informação de que os metadados dos itens estavam disponíveis online pelo Sophia Web.

O primeiro passo foi recuperar os dados pelo banco de dados coletado na etapa de diagnóstico, porém como as informações de identificação repassadas não correspondiam com os dados que foram obtidos do banco de dados, foi necessário desenvolver um script na linguagem *Python*² para a raspagem de dados das páginas dos itens a partir do identificador no *Sophia Web* disponível na tabela repassada, perfazendo assim um processo semiautomático, em que a partir dos identificadores eram acessadas as páginas individuais com metadados de cada item, e pelo script esses metadados foram coletados e armazenados em planilha para posterior tratamento e relacionamento com as imagens.

A etapa de **tratamento** foi executada utilizando aplicações da técnica de processamento de dados, e também ocorreu em duas frentes, uma com foco no conteúdo disponível no site da instituição e outra com foco no acervo CEDOC disponível on-line pelo SophiaWeb.

Nos conteúdos coletados do site, como a quantidade de metadados para as coleções de itens era pequena, foi possível fazer as alterações direto pelo script de migração, centrando os esforços na divisão das informações em metadados, se atentando aos problemas de inconsistência e domínio dos dados. Por exemplo no título dos programas de rádio do “Estúdio F”, que continha o nome do artista em foco no programa, assim foi criado um metadado artista com o nome proveniente do título, que constituiu uma taxonomia de agentes, que conecta essa coleção às demais pelo nome do agente.

Já nos dados coletados da raspagem das páginas do Sophia Web, foi preciso tratar primeiro a planilha obtida do script, organizando os metadados em colunas, já que houve ruído pela inexistência de uma padronização de preenchimento dos metadados, além da

² Script de raspagem de dados do Sophia Web disponível em: https://github.com/medialab-ufg/wp-funarte/blob/master/migration_scripts/sophiaweb_scrape.py

incompletude identificada por alguns itens apresentaram informações a mais que outros no resultado final. Posteriormente a partir da classificação dos arquivos disponível na planilha repassada, as imagens para cada item foram recuperadas do HD externo obtido, e ainda como o nome das coleções dos itens obtidos continham nome dos artistas envolvidos, também foram separados em outros metadados de referência à taxonomia de agentes.

A **migração**, acompanhando a estrutura das etapas de coleta e tratamento ocorreu em dois momentos, o primeiro em que ao reformular o site, ao mesmo tempo foram aplicados os mesmos scripts utilizados na coleta para a migração do conteúdo do site antigo, para não perder nenhuma atualização e nem necessitar de uma pausa na produção de conteúdo no processo, dessa forma cada conjunto de conteúdo disperso no site foi migrado para o Tainacan e disponibilizado em um único ponto de acesso conectados pelos metadados em comum. Já os itens repassados pelo CEDOC foram migrados utilizando a ferramenta de importação do Tainacan, primeiramente a coleção foi criada, depois seus metadados, assim a migração ocorreu com o mapeamento das colunas da tabela de dados no formato CSV (valores separados por vírgula) coletada e tratada para o Tainacan via ferramenta de importação, as imagens como representavam um conteúdo muito grande em termos de armazenamento (115 Gb), foram alocadas no servidor da Funarte e tiveram seu caminho no servidor mencionadas na tabela de importação, assim também foram importadas como anexos dos itens.

O processo de **validação e formação** ocorreu a partir da disponibilidade das coleções no site após a migração, elas foram apresentadas aos gestores e técnicos que confirmaram a presença de todos os itens, bem como a adequação dos metadados. Os técnicos foram acompanhados via lista de e-mail nas primeiras semanas posterior a migração, e foram auxiliados a inserir conteúdo e classificá-lo de acordo com os metadados existentes na estrutura da plataforma Tainacan.

O processo de **difusão e colaboração em rede** ocorreu em conjunto com os técnicos e gestores do acervo, que acompanham como o acervo é utilizado e requisitado, tanto internamente quanto por usuários externos.

5 RESULTADOS

O processo de migração foi desenvolvido entre Julho de 2018 e Julho de 2019, e é importante situar que o projeto não trata de todo o acervo disponível na Funarte como será apresentado a seguir na etapa de análise. Uma das especificações do projeto foi a redução do escopo na escolha de partes do acervo mais relevantes com base na experiência dos gestores da instituição. Dessa forma o acervo migrado constitui de coleções de objetos presentes no site da Funarte e de parte do acervo do CEDOC.

A etapa de **análise**, descrita na metodologia, utilizou da ADD e serviu de diagnóstico do acervo junto aos gestores, o que permitiu identificar as relações de estruturação do acervo com a forma como a cultura organizacional é desenvolvida na instituição, além de definir quais as porções prioritárias do acervo seriam migradas para o Tainacan.

A primeira tarefa dessa etapa foi buscar por um inventário do acervo da instituição. Como não houve documentação que explicitasse o dimensionamento do acervo completo, os gestores de conteúdo compartilharam diversos documentos digitais em texto e planilhas de cada área com catálogos preenchidos a mão pelos técnicos de curadoria. A partir organização dos dados utilizando distribuição de frequência desses documentos foi possível traçar a dimensão aproximada do acervo da Funarte.

A matriz representada na tabela 1 já indica o esforço de apresentar os dados visualmente com medidas de síntese. A matriz é composta pela quantidade de itens por tipo de material com dados obtidos a partir dos documentos descritivos do acervo de acordo com as áreas informacionais da Funarte. São ao todo aproximadamente 379.547 itens, em que foram identificados 13 tipos de materiais, sendo os registros obtidos a partir de metadados (36,57%), fotos de artistas e espetáculos (25,89%), e documentos como roteiros e programações (entre outros não especificados nos catálogos) (23,65%), os três tipos de material com maior proporção do acervo, somando 86,11% do total. As áreas de maior representação do acervo são o arquivo privado, o audiovisual, e a biblioteca, representando 97,16% da quantidade total de itens, já o site, forma de acesso digital da instituição contém 2,83% do acervo.

A forma de disponibilização e estruturação destes itens é dispersa, são diferentes tecnologias utilizadas para a curadoria do acervo, mesma situação identificada no caso do Museu do Índio (MARTINS; CARMO; GERMANI, 2018), o que gera problema na recuperação

de informações pela fragmentação do acervo em pequenos nichos de conteúdo com diferentes formas de organização dos dados.

Tabela 1 - Matriz da quantidade de itens identificados no acervo Funarte

Tipo de Material	Arquivo Privado	Audiovisual	Biblioteca	Site	TOTAL
Registros	-	-	134.600	4.205	36,57%
Fotos	-	98.060	191	-	25,89%
Documentos	84.684	-	220	4.861	23,65%
Folhas	28.228	-	-	-	7,44%
Cartazes	-	14.369	-	-	3,79%
Referências	-	4.165	-	-	1,10%
Pastas	3.467	-	-	-	0,91%
Partituras	-	-	-	1.200	0,32%
Mídias	-	810	-	-	0,21%
Programas	-	-	-	343	0,09%
Arquivos Musicais	-	-	-	70	0,02%
Projetos	-	-	-	52	0,01%
Livros	-	-	-	22	0,01%
TOTAL	30,66%	30,93%	35,57%	2,83%	379.547

Fonte: Dados da pesquisa (2018)

Como apresenta a Tabela 2, os itens do arquivo privado estão armazenados em estrutura de arquivos locais e são compostos de imagens e catálogos, disponíveis somente para acesso dos técnicos curadores do acervo, já o áudio visual é composto em sua maioria de materiais físicos, com representação digital somente do quantitativo de itens a partir de catálogos preenchidos manualmente em documentos textuais, os itens de biblioteca, são compostos além dos metadados de livros, de dossiês com conjuntos de fotos unidas no formato PDF (sem acesso ao formato individual de imagem) com seus respectivos metadados, disponíveis pelo software privado Sophia Web com acesso pelo site, já o site contém áudios, metadados e imagens armazenados no próprio sistema de gerenciamento de conteúdo (CMS) *Wordpress*, que possui como banco de dados o *MySQL*, em diferentes pontos de acesso.

Tabela 2 - Tecnologias e tipos de acesso para o acervo Funarte

Área	Tecnologia	Tipo de Acesso
Arquivo Privado	Estrutura de Arquivos	Interno
Audiovisual	Físico/Documento de Texto	Interno
Biblioteca	Sophia Web/SQL Server	Público/On-Line
Site	<i>Wordpress/MySQL</i>	Público/On-Line

Fonte: Dados da Pesquisa (2018)

Além dessas questões estruturais, foi identificado a partir de reuniões, que as principais formas de acesso ao conteúdo do acervo são on-line via site, e via contato pessoal ou telefone com os técnicos da Funarte. Essas práticas limitam as possibilidades de explorar e recuperar as informações dos usuários, que como denotado pelos gestores dos acervos, são compostos de uma parcela de pesquisadores, os quais têm um potencial comprometido de recuperação e utilização desses dados pelo fracionamento do acervo e a dificuldade de acesso aos itens.

A situação dos materiais disponíveis pelo site por exemplo, apresentava oito pontos de acesso diferentes, o “Acervo Sérgio Britto” com metadados e documentos relacionados ao artista em uma outra instância do Sophia Web, o “Brasil Memória das Artes” com áudios, imagens, vídeos e textos referentes ao conjuntos de artistas selecionados, o “Catálogo CEDOC” com o material da biblioteca de responsabilidade do centro de documentação disponível por outra instância do Sophia Web, as “Sinopses” com material descritivo sobre as produções bibliográficas da Funarte, as “Edições On-line” com os livros em formato aberto disponíveis em PDF para *download*, os “Vídeos” com material audiovisual de entrevistas e artísticos produzidos ou relacionados à Funarte, o “Música nas Américas”, com uma seleção de áudios com entrevistas e músicas temáticas, e o “Estúdio F”, com programas de rádio produzidos pela própria instituição.

Figura 2 – Exemplos de pontos de acesso ao acervo no site da Funarte



Fonte: funarte.gov.br (2018)

Essa dispersão das condições de acesso (Figura 1), além de estabelecer uma barreira para o potencial de utilização das informações do acervo, exprime fragmentação na forma de perceber o acervo dentro da instituição.

Essas características das formas como as informações são tratadas dentro da instituição e que constituem o acervo, revelam que este projeto além de o organizar e divulgá-lo, também tem como intuito auxiliar na gestão da informação institucional, guiando à unicidade no tratamento da informação, independentemente de sua área de gestão, o que permite otimizar o processo de busca e uso tanto do usuário interno quanto externo.

É a partir desse contexto informacional e institucional que foi construído um primeiro cenário do dimensionamento do acervo da instituição, que posteriormente foi apresentado em reunião aos gestores, que até então desconheciam tal aproximação, permitindo a deliberação de quais as porções mais determinantes do acervo seriam migradas para o Tainacan.

O que resulta então na escolha do conjunto do material difuso no site, constituído de sete dos oito conjuntos de itens apresentados no dimensionamento do acervo, além de 1.407 objetos provenientes do acervo CEDOC, disponíveis até então no software Sophia Web com acesso restrito às imagens (somente por PDF), e excluindo neste momento o “Acervo Sérgio Britto”. Consistindo por fim em 2.107 objetos escolhidos para a migração prioritária (Tabela 3).

Tabela 3 - Coleções selecionadas para migração

Tipo de Material	Nº de Itens	Proporção
Coleções CEDOC	1.407	65,23%
Estúdio F	392	18,17%
Brasil Memória das Artes	172	7,97%
Vídeos	79	3,66%
Sinopses	54	2,50%
Música nas Américas	29	1,34%
Edições On-line	24	1,11%
TOTAL	2.157	-

Fonte: Dados da Pesquisa

Quanto à **coleta**, ocorreu em duas etapas, no caso do site de forma automatizada através de scripts em PHP que permitiram a recuperação de todos os itens de seis dos sete conjuntos migrados, já que os itens do conjunto “Brasil Memória das Artes” apresentaram somente os áudios disponíveis, e foram coletados manualmente e posteriormente migrados via importador por tabela de metadados no Tainacan. No caso do acervo CEDOC, o processo

de coleta foi semiautomático, com raspagem de metadados a partir das páginas do Sophia Web dos itens selecionados, juntamente com o relacionamento das imagens repassadas pelos técnicos da Funarte e resultou na formação das “Coleções CEDOC” no Tainacan.

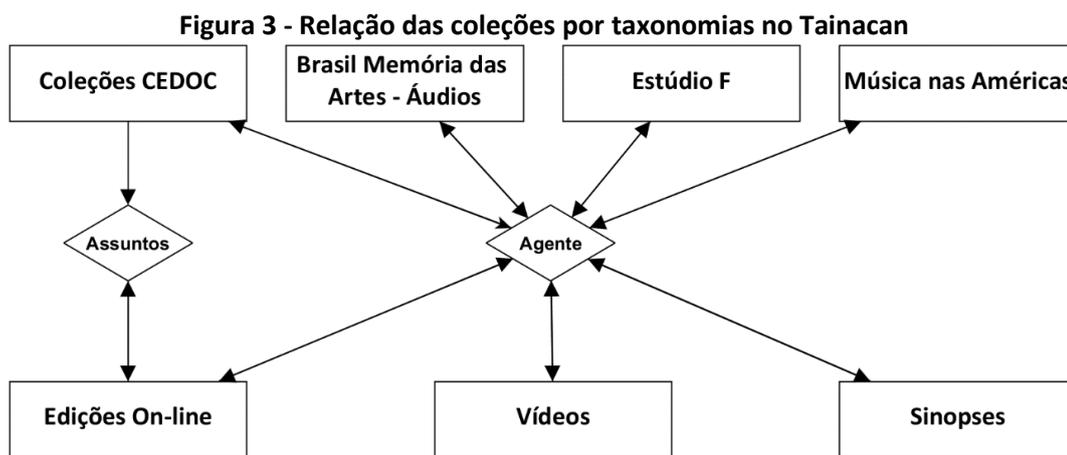
A etapa de **tratamento** no contexto dos itens provenientes do site, foi marcada pela aplicação do pré-processamento de dados diretamente no script de coleta, cujo resultado foi principalmente a divisão de metadados a partir das informações obtidas dos itens.

Utilizando métodos como *explode*, *trim*, e combinação de *strings*, informações foram extraídas dos textos de descrição ou título dos itens, como o nome dos artistas relacionados nos itens, retirados do título no caso do “Estúdio F” e dos áudios do “Brasil Memória das Artes”, da descrição no caso dos “Vídeos”, e do metadado de autoria no caso de “Sinopses” e “Edições On-line”. Essa abordagem de transformação permitiu a criação de uma taxonomia de agentes no Tainacan, que permite conectar todos os itens pelo nome do agente relacionado, possibilitando assim a recuperação de itens por agente, condição que não existia anteriormente.

Ainda no caso das coleções de “Sinopses” e “Edições On-line” todos os metadados bibliográficos estavam dispostos no texto da descrição do livro, e todos foram divididos em metadados individuais para otimizar a recuperação e filtragem específica de informações no acervo.

No contexto das coleções CEDOC, o tratamento ocorreu no nível de organização dos metadados após recuperação do portal Sophia Web, além dos títulos das coleções também serem aproveitados para configurar o metadado agentes.

Esses tratamentos iniciais nas bases de dados obtidas do acervo parcial migrado resultaram em um conjunto de coleções que, além de se encontrarem no mesmo ponto de acesso no site, permitem a busca livre textual, e a navegação por filtros, e ao selecionar as taxonomias em comum das coleções é possível fazer a listagem de itens pertencentes a uma mesma categoria, como assuntos ou agentes por exemplo, no contexto atual das coleções disponíveis, a taxonomia agente é compartilhada por todas as coleções, e a taxonomia assuntos só tem metadados em comum nas “Coleções CEDOC” e nas “Edições On-line” (Figura 3), ponto a ser melhorado na modelagem conceitual do acervo.



Fonte: Dados da Pesquisa (2019)

A etapa de **migração** que ocorreu principalmente através de scripts em PHP no caso dos itens do site, e através do importador de tabela no Tainacan quanto às “Coleções CEDOC”, e resultou em um processo adaptado ao contexto da Funarte, a possibilidade de reformulação do site e acesso ao sistema de gerenciamento nos direcionou para a opção mais estratégica de coleta, tratamento e migração automática dos itens do site. Já no caso dos itens do acervo CEDOC, que estavam fragmentados em locais diferentes, no HD de imagens, nos metadados do Sophia Web, e na tabela de informações de identificação repassada, todo o processo de coleta, tratamento e migração foi dividido se assemelhando a um processo comum de migração de acervo para o Tainacan. Esse tipo de abordagem aponta o quão flexível essa metodologia e o processo de migração precisa ser para atender a realidade organizacional e situacional da transformação dos dados.

A etapa de **validação e formação** ocorreu com as coleções já migradas para o repositório do Tainacan, e resultou na avaliação dos metadados, disposição das informações e documentos anexados dos itens do acervo, em que os técnicos mencionam alterações que devem ser realizadas, e confirmam a qualidade e representatividade precisa dos dados dispostos na plataforma com suas informações do acervo.

Além disso, uma lista de e-mails é compartilhada entre a equipe de migração e implantação do Tainacan e os técnicos curadores do acervo, para resolução de eventuais problemas, sanar dúvidas e orientar na manutenção dos itens na nova plataforma. No caso específico desse acervo parcial migrado, o processo de divisão das informações dos itens precisou ser separado em metadados, e alguns deles que não existiam antes precisam ser preenchidos pela equipe, como a cultura anterior não exigia essas técnicas, é algo que está

sendo orientado e acompanhado de perto na formação dos profissionais que cuidam das coleções através de reuniões e orientações por e-mail.

Essa etapa garante que a barreira de interconexão dos dados que existia anteriormente não persista, de modo que agora metadados que garantem a interconexão mínima precisam ser preenchidos, como as taxonomias de agente e assunto.

Por último, a etapa de **difusão e colaboração em rede**, que no caso da Funarte ainda não foi liberada a permissão para interação direta com o acervo por meio de comentários, as coleções e as funcionalidades do Tainacan ainda estão sendo assimiladas pelos técnicos, e como resultado principal entrega uma interface de acesso unificado, em que o usuário não precisa navegar ostensivamente pelo site em busca dos acervos disponíveis, atualmente, através do item “Acervos” no menu principal se tem acesso direto ao conjunto de coleções disponíveis do acervo (Figura 4).

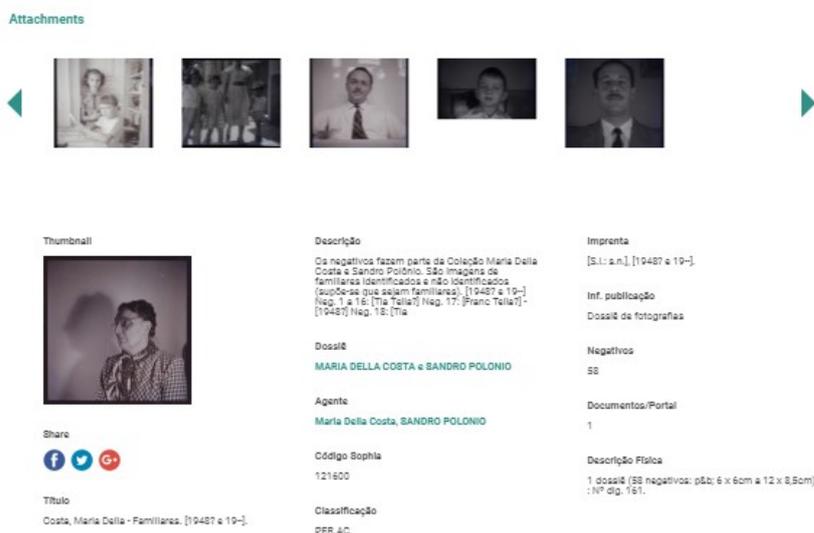


Fonte: <http://www.funarte.gov.br/colecoes/> (2019)

Outro ponto a ser destacado no contexto da difusão é ao acesso aos itens completos, como no caso do acervo CEDOC, em que só se tinha acesso externo aos metadados pelo Sophia Web, e internamente as imagens anexas aos itens foram reunidas em documentos no formato PDF para atender as estruturas do software, limitando seu uso. No caso atual, as imagens estão anexadas aos itens e podem ser acessadas separadamente (Figura 5) e perfazem atualmente a instalação do Tainacan com maior conjunto de imagens, são 14.666

imagens ao todo em itens com mais de 500 anexos, todas essas imagens estão disponíveis no formato JPG e em alta qualidade de resolução (300dpi).

Figura 5 - Exemplo da interface de acesso ao item das Coleções CEDOC



Fonte: <http://www.funarte.gov.br/colecao/maria-della-costa-sandro-polonio/#/> (2019)

A difusão e organização do acervo parcial finda os resultados esperados para essa primeira parte do projeto, permitindo que a equipe da Funarte assimile a transição de plataformas e estrutura informacional de forma gradual, e que os processos de organização e transformações informacionais posteriores sejam executados em conjunto, assegurando que as práticas de curadoria anteriormente utilizadas sejam lapidadas em busca da conexão dos acervos e das áreas informacionais da instituição.

6 CONSIDERAÇÕES FINAIS

Uma das reflexões que resultam do estudo é que a forma como a organização se comporta mediante ao tratamento de seus fluxos informacionais influencia diretamente em como ela estrutura seu acervo. No caso da Funarte, que passou por reformulações institucionais durante períodos de diferentes perspectivas políticas (FUNARTE, 2019), isso fica mais evidente ao analisar a situação do acervo, em que chega a se perder sua dimensão pela dificuldade de proceder com a curadoria do material.

Dessa forma o uso das técnicas de análise descritiva na etapa de análise se mostra válido, pois auxiliou no esclarecimento de como o acervo institucional estava estruturado

quanto suas bases de dados. Foi de fundamental importância a habilidade técnica de alcançar uma aproximação do dimensionamento do acervo, informação que não existia na instituição.

O dimensionamento do acervo não foi o único desafio no caso da Funarte, foi somente o primeiro deles. A tecnologia de disponibilização utilizada para cada tipo de material do acervo também é um grande ponto de reflexão para a instituição em direção à publicação de seus itens. O último software utilizado pela Funarte para divulgação dos itens do CEDOC foi o Sophia Web, que tem grandes contribuições para materiais bibliográficos, mas no armazenamento da parte cultural digitalizada do acervo peca em permitir o acesso mais otimizado aos itens como imagens, áudios e vídeos. Dessa forma os problemas de escolha tecnológica também são comuns nesse tipo de instituição, como também aponta (MARTINS; CARMO; GERMANI, 2018, p. 5142).

Além disso deve-se notar a importância da técnica de pré-processamento de dados aplicada na etapa de tratamento. As transformações de dados, explicitadas pela divisão de informações em metadados, foram estrategicamente desenvolvidas para fomentar a integração entre os acervos através da garantia de padrão entre eles, de forma que é possível navegar entre as coleções filtrando pela taxonomia de agente por exemplo, função que aprimora o acesso e uso à informação, e inexistente na estrutura anterior.

Manter essa organização da base de dados também é um desafio, e demandará o desenvolvimento de políticas informacionais, bem como conscientização institucional do acervo.

Já na perspectiva da apropriação institucional da nova estrutura informacional, foi observado que a migração do acervo digitalizado relevante foi bem recebida pela gestão da Funarte, que juntamente com a reformulação do site proporcionou um modo mais efetivo de busca e uso da informação disponível no portal online. Atualmente a equipe de desenvolvimento e migração segue acompanhando a adaptação a nova estrutura, e a segunda parte do projeto está em processo de início, que versa sobre a construção de uma modelo conceitual para o acervo digitalizado, e aplicações semânticas de enriquecimento dos dados.

Além das contribuições atuais de reestruturação informacional que a migração do acervo entregou, esse processo contribuiu para proceder de forma gradual, respeitando a adaptação dos curadores ao novo suporte do acervo. De modo que se abre caminho para a modelagem conceitual e o enriquecimento semântico dos dados, que objetiva promover a

interoperabilidade do acervo Funarte com outros acervos digitais, o que não envolve somente alinhar padrões de metadados, mas também uma proposta complexa de promover o enriquecimento e integração em rede das informações (MARCONDES, 2016, p.68-71).

REFERÊNCIAS

FERRARI, D. G.; DE CASTRO, L. N. **Introdução a mineração de dados**. São Paulo: Saraiva, 2017.

FUNARTE. **Institucional**. Disponível em: <<http://www.funarte.gov.br/a-funarte/>>. Acesso em: 07 jul. 2019.

MARCONDES, C. H. 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, 2016.

MARTINS, D. L.; CARMO, D.; GERMANI, L. Museu do índio: estudo de caso do processo de migração e abertura dos dados ligados semânticos do acervo museológico com o software livre Tainacan. In: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO. 19., 2018. **Anais do ENANCIB**, Londrina, 2018.

MARTINS, D. L.; CARVALHO, J. M. C. Memória como prática na cultura digital. In: **CETIC.BR. TIC CULTURA Pesquisa Sobre o Uso das Tecnologias de Informação e Comunicação nos Equipamentos Culturais Brasileiros**. Comitê Gestor da Internet no Brasil, São Paulo, 2017.

MARTINS, D. L.; CARVALHO, J. M. C.; GERMANI, L. Projeto Tainacan: Experimentos, Aprendizados e Descobertas da Cultura Digital no Universo dos Acervos das Instituições Memoriais. In: **TIC CULTURA Pesquisa Sobre o Uso das Tecnologias de Informação e Comunicação nos Equipamentos Culturais Brasileiros — 2018**. Comitê Gestor da Internet no Brasil, São Paulo, 2019.