

Plano de Trabalho SIMPLIFICADO IBICT – FUNDEP

Projeto de Pesquisa: Acervo digital semântico da Funarte: aplicação de novos modelos de organização e interconexão da informação dos acervos digitalizados do CEDOC

1) APRESENTAÇÃO

Este Projeto de Pesquisa é uma iniciativa conjunta do Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT) e da Fundação Nacional de Artes (Funarte).

Desde a sua criação, em 1954, à época denominado como Instituto Brasileiro de Bibliografia e Documentação (IBBD), o Instituto Brasileiro de Informação em Ciência e Tecnologia (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 nesse contexto 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 continuidade da disponibilização aos usuários de um sistema de informação no escopo do chamado Projeto Funarte Digital. Este ambiente propôs o compartilhamento do acervo da instituição compreendido por suas diversas manifestações artísticas: música, dança, teatro, artes visuais, circo e literatura. No âmbito da primeira parceria estabelecida entre a Funarte e o IBICT, foi possível reformular o site da instituição e iniciar a disponibilização de acervos digitais, bem como mapear e identificar possibilidades de integração e tratamento.

No presente projeto, visando ampliar as ações direcionadas aos acervos já digitalizados do CEDOC da Funarte, tem-se por objetivo desenvolver um modelo conceitual de interoperabilidade semântica das coleções e acervos já constituídos, bem como ampliar o reuso do ambiente da Funarte Digital para difusão e gestão das coleções.

Assim sendo, este Projeto buscará a continuidade e ampliação do desenvolvimento de pesquisa para a organização da informação aos novos modelos de reuso de conteúdos digitais a serem aplicados na memória cultural brasileira reunida nos acervos da Fundação Nacional de Artes.

2) JUSTIFICATIVA

O patrimônio cultural digitalizado tem se tornado um ativo de grande importância para diversas frentes do conhecimento, pesquisa e desenvolvimento da sociedade. Por um lado, a formação de grandes bases de dados de objetos culturais tem criado a possibilidade do desenvolvimento de novos métodos de pesquisa para a compreensão da cultura e suas dinâmicas sociais, métodos estes hoje agrupados e disseminados sob o nome de Humanidades Digitais e Ciências Sociais Computacionais. Novos indicadores, métricas, técnicas de análise estatística para a descoberta de padrões, análise de imagem, seja estática ou em movimento, e diversas aplicações se valendo de importantes inovações no âmbito da aprendizagem de máquina e da inteligência artificial se beneficiam da disponibilização de grandes bases de dados curadas e organizadas de forma sistemática da produção cultural. Por outro lado, vem se descobrindo a importância da digitalização da cultura na ativação de novos processos econômicos, fortalecendo o reuso de objetos digitalizados para o fomento de inovação na indústria criativa, no turismo e mesmo na educação.

Nessa faceta econômica, o trabalho de Tessler (2013) demonstra um modo de estudo do retorno financeiro ao investimento feito na British Library, chegando a demonstrar que para cada 1 libra investida na biblioteca a mesma retorna 4.9 libras para a sociedade britânica. O estudo também é pioneiro por avaliar pela primeira vez o retorno econômico dos serviços web prestados pela biblioteca, conseguindo demonstrar que a soma desses serviços contribui com 19.5 milhões de libras/ano para a economia. Mostrando diferentes formas de uso, seja para a educação, para a pesquisa, para o turismo, para o design, entre outros, a disponibilização do acervo da biblioteca em formato digital passa a ser reconhecido como um elemento estratégico de geração de valor não apenas cultural, mas também financeiro para uma sociedade da informação altamente centrada no desenvolvimento de serviços e produtos imateriais. Já o estudo de Poort et al. (2013) é focado essencialmente na avaliação do retorno econômico e do perfil do público de uso do portal Europeana, o grande agregador da União Européia de acervos digitais publicados por galerias, arquivos, bibliotecas e museus. O estudo demonstra que os usuários são agrupados em 5 grupos, sendo eles as instituições memoriais, que se valem dos serviços em escala de tecnologia da informação fornecidos pela iniciativa; o público geral interessado em arte, cultura e memória; turistas, que usam os sites, aplicativos e guias criados pela iniciativa para contribuir na organização de suas viagens; a indústria criativa, como editoras, jornalistas, designers e desenvolvedores de jogos, que usam os objetos digitais para inspirar novas criações e as instituições educacionais. O estudo conclui sua avaliação demonstrando que para um orçamento de 57.3 milhões de euros, a iniciativa dá um retorno base de 21.5 milhões de euro para a economia da União Europeia.

Torna-se evidente, a partir da análise apresentada, que grandes bases de dados de objetos digitalizados do patrimônio cultural de uma sociedade, disponibilizadas em rede e de forma agregada, podem trazer inúmeros ganhos e benefícios para essa sociedade, chegando mesmo a contribuir para a ativação de sua economia. No entanto, um dos elementos centrais dessa estratégia é a possibilidade de gerar uma infraestrutura informacional que permita a

interoperabilidade entre diversas instituições e seus sistemas de informação, garantindo que diferentes museus, arquivos, bibliotecas, cinematecas e centros culturais possam publicar seus acervos de forma descentralizada, serem coletados e agregados em grandes serviços que integrem essas bases de dados e forneçam um ponto de acesso unificado para facilitar a vida do usuário.

Há diversas iniciativas que têm realizado esse tipo de integração de bases de dados de objetos culturais digitalizados ao redor do mundo, conforme apresentado em Martins, Silva e Carmo (2018). Uma das principais, a Europeia¹, chega a disponibilizar na Internet mais de 58 milhões de objetos digitalizados de obras de arte, livros, filmes e áudios de toda a Europa, integrando mais de 3.500 instituições e gerando uma infraestrutura que permite a ampla interoperabilidade de diferentes tipos de sistemas de informação, envolvendo diferentes modelos conceituais, linguagens documentárias, regras de catalogação e padrões de metadados. A iniciativa se vale da criação de um modelo de dados chamado EDM (*Europeana Data Model*), baseado em diferentes padrões da web semântica e do movimento LOD (*Linked Open Data*), conforme Coneglian e Segundo (2017).

Uma dessas iniciativas, que apresentam grande destaque e importância, é a Europeia. A Europeia é um projeto de uma biblioteca virtual de toda Europa, que reúne e integra mais de 50 milhões de objetos culturais, de bibliotecas, de museus e de arquivos. O projeto da Europeia de LOD contém um ambiente de SPARQL Endpoint, que possibilita a recuperação dos dados estruturados através de uma linguagem de chamada de SPARQL Protocol and RDF Query Language (SPARQL). Vale ressaltar que, o desenvolvimento da Europeia está inserido no contexto das Humanidades Digitais (CONEGLIAN e SEGUNDO, 2017).

É preciso compreender o cenário dos acervos digitais das instituições memoriais brasileiras para que se possa avançar em experiências de integração de suas bases de dados, quando disponíveis, e gerar avanços tanto na pesquisa científica brasileira a respeito da cultura produzida no país, quanto oferecer serviços que apoiem e estimulem o reuso desses objetos em ampla escala. Segundo dados da pesquisa TIC Cultura 2016 (COMITÊ GESTOR DA INTERNET, 2016), em torno de 58% dos museus, 29% das bibliotecas e 74% dos arquivos brasileiros digitalizam parte de seu acervo. No entanto, apenas 37% dos museus, 18% das bibliotecas e 61% dos arquivos disponibilizam esse acervo para o público. Há, como se pode deduzir destes números, uma diferença significativa entre os acervos digitalizados e aqueles disponíveis para acesso ao público. Diversos são os fatores que levam a isso, mas em grande medida se pode destacar a dificuldade de armazenamento e fornecimento de acesso ao público desse conteúdo digital, bem como dos custos operacionais e técnicos, de difícil manutenção por parte das instituições, sobre como fornecer esses serviços, como garantir sua preservação, como configurar sistemas de informação, entre outros.

¹ <https://www.europeana.eu/portal/pt>

Mesmo quando os acervos digitais estão disponíveis em rede em iniciativas brasileiras, há diversos problemas técnicos e conceituais para que se possa agregar essa informação e gerar um ponto de acesso único ao patrimônio cultural brasileiro digitalizado.

Arquivos, bibliotecas e museus disponibilizam seus catálogos e/ou coleções digitais na Internet, tornando-os acessíveis a qualquer usuário, em qualquer lugar, a qualquer momento. Há, no entanto, grande dificuldade em tornar interoperáveis acervos hospedados em diferentes instituições, uma vez que esses acervos são dependentes de softwares específicos para publicar seus registros na Web. Os atuais softwares de gestão de catálogos - usados em arquivos, bibliotecas e museus -, também não permitem que sejam feitos links para um registro específico, excluindo seus registros de toda a economia da Web, com base na publicação de conteúdos e de links entre eles. (SANTOS NETO et al., 2013, p. 81)

Durante anos essas instituições têm desenvolvido semânticas, modelos, formatos e padrões próprios. Todo o potencial de integração e acesso proporcionado pela Web não pode ser aproveitado pelos sistemas convencionais que gerenciam esses catálogos, impedindo que tais acervos possam aproveitar suas sinergias e potenciais culturais, e que se beneficiem, e também que outras instituições possam fazer links para seus acervos. (MARCONDES, 2012, p. 181).

Logo, um dos importantes problemas de pesquisa ainda não resolvido em âmbito brasileiro é compreender sistematicamente como os repositórios digitais de cultura representam a informação descritiva de seus objetos e como essas diferentes formas de representação podem ser agregadas e mapeadas para um padrão em comum.

Sabe-se também que instituições culturais no país, como a Funarte, muitas delas possuem diferentes sistemas de informação (ferramenta), padrões de metadados, linguagens documentárias para indexação de objetos, regras de catalogação para descrição dos metadados e, eventualmente, modelos conceituais de como a informação deve ser representada. A partir desse contexto, evidencia-se o problema a ser trabalhado no presente projeto de pesquisa: **como agregar a informação e oferecer um ponto único de acesso para busca e recuperação da informação dos repositórios digitais de objetos culturais das instituições memoriais brasileiras?**

Visando minimizar essa problemática, este projeto tem como objetivo principal estudar as diferentes formas de interoperabilidade entre os repositórios digitais da Funarte e integrá-los em único ambiente denominado Funarte Digital.

Como resultado, tem-se como meta/ contribuir em duas frentes de ação. A primeira diz respeito a área de organização e representação da informação, a partir do estudo e sistematização da

forma como esses acervos são hoje representados e como podem ser mapeados para novos padrões semânticos visando sua integração e normalização para busca e recuperação. Nessa frente, espera-se estudar modelos conceituais que propõem o uso de ontologias como forma de conectar diferentes representações de acervos e gerar uma mesma rede semântica que garanta sua interoperabilidade mínima. Segundo Alexiev (2018), são referenciais para essa frente os modelos conceituais CIDOC-CRM², linked.art³, Schema.org⁴ e Wikidata⁵. Logo, o projeto tem por objetivo estudar esses modelos e avaliar qual melhor se adaptaria para facilitar a integração dos modelos de dados dos repositórios digitais a serem integrados em uma mesma rede semântica. Essa integração, uma vez consolidada, poderia se tornar uma referência para repositórios digitais de cultura no Brasil, facilitando com que novas instituições possam criar seus repositórios digitais assim como publicarem seus dados já catalogados, encontrando um caminho para se integrarem a essa mesma rede semântica de forma padronizada, aumentando seu valor de uso coletivamente por ampliação da escala de representação da rede.

A segunda contribuição esperada do projeto é um **modelo tecnológico** que permita a coleta, análise, curadoria, agregação, disponibilização e difusão para uso dos dados dos repositórios digitais estudados por usuários interessados em seus acervos. Entende-se esse resultado como uma importante contribuição para diversas áreas do conhecimento que tenham como foco os estudos culturais, bem como outras áreas que tenham interesse no reuso desses objetos digitais para inovação e geração de novos conhecimentos. Espera-se, com esse modelo tecnológico, ampliar o potencial de acesso à cultura brasileira digitalizada, ofertando um único ponto de acesso que agregue diferentes repositórios, categorias de acervo, tipos de mídias e temas culturais de amplo interesse.

3) OBJETO

Desenvolvimento de pesquisa para disponibilização, difusão e organização dos conteúdos já digitalizados do CEDOC da Funarte no ambiente do site institucional Funarte Digital.

4) OBJETIVO GERAL

Disponibilizar os metadados e os objetos completos do acervo tratado e já digitalizado do CEDOC da Funarte no ambiente Funarte Digital, ampliando o impacto social do seu alcance e beneficiando a área de artes no Brasil com informações de qualidade e curadas ao longo de décadas

² <http://www.cidoc-crm.org/>

³ <https://linked.art/>

⁴ <https://schema.org/>

⁵ <https://www.wikidata.org/>

5) OBJETIVOS ESPECÍFICOS

- Desenvolver modelo conceitual de interoperabilidade semântica da informação dos acervos;
- Tratar e normalizar os metadados dos acervos já digitalizados;
- Carregar e disponibilizar no ambiente da Funarte Digital os acervos tratados;
- Desenvolver estudos teóricos sobre os processos, modelos e tecnologias de informação aplicadas ao contexto do Projeto

6) METODOLOGIA

Para a coleta e análise de dados, a abordagem metodológica empregará diferentes técnicas de pesquisa qualitativa e quantitativa. O Projeto aplicará princípios de ciência de dados que visam a análise descritiva das estruturas de informação identificadas.

O desenvolvimento de tecnologia utilizará o modelo de trabalho online, por meio de plataforma de disponibilização e gestão de código GitHub. Os produtos desenvolvidos podem ser disponibilizados em licença livre em análise conjunta com a FUNARTE.

7) RESULTADOS ESPERADOS:

- Modelo conceitual de interoperabilidade semântica dos acervos da Funarte;
- Metadados dos acervos digitalizados tratados e normalizados;
- Carregamento e disponibilização dos acervos tratados no ambiente da Funarte Digital;
- Disseminação dos resultados por meio de apresentação das produções teóricas realizadas e demais atividades de pesquisa, como workshops, voltados para a área da Ciência da Informação, realizados durante o Projeto.

8) CRONOGRAMA DE EXECUÇÃO

RESULTADOS ENTREGUES	CRONOGRAMA (MESES)														
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Meta 1: Modelo conceitual de interoperabilidade semântica dos acervos da Funarte	X	X	X	X	X	X									
Meta 2: Metadados dos acervos digitalizados tratados e normalizados				X	X	X	X	X	X	X	X	X			
Meta 3: Carregamento e disponibilização dos acervos tratados no ambiente da Funarte Digital					X	X	X	X	X	X	X	X	X		
Meta 4: Disseminação dos resultados por meio de apresentação das produções teóricas realizadas e demais atividades de pesquisa, como workshops, voltados para a área da Ciência da Informação, realizados durante o Projeto											X	X	X	X	X

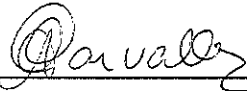
9) CRONOGRAMA DE ENTREGAS DOS RELATÓRIOS TÉCNICOS

RELATÓRIOS	ENTREGA PREVISTA
Meta 1 - Modelo conceitual	Dez/2019
Metas 2 e 3 - Acervos disponibilizados no Funarte Digital	Jul/2020
Meta 4 - Resultados técnicos científicos	Set/2020

10) ORÇAMENTO SIMPLIFICADO

Rubrica	Valor (R\$)
Bolsas	156.000,00
Diárias	12.573,00
Passagens	18.000,00
Custo Operacional	24.755,00
Mat. Consumo	1.000,00
STPJ	7.672,00
STPF	30.000,00
TOTAL	250.000,00

11) PLANO DE APLICAÇÃO DE RECURSOS – Previsão Orçamentária (memória de cálculo) (ANEXO)


Washington Luis R. de Carvalho Segundo
Coordenador do Projeto - IBICT

Ana Lúcia de Carvalho Gouveia
Substituta da Coordenação
de Atendimento à Comunidade
PO 28/16

De acordo,


Bianca Amaro de Melo

Bianca Amaro de Melo
Coordenadora-Geral de
Pesquisa e Manutenção de Produtos Consolidados

Coordenadora-Geral de Pesquisa e Manutenção de Produtos Consolidados – CGPC- IBICT

Brasília, 17 de julho de 2019.