

GT-8 – Informação e Tecnologia

ISSN 2177-3688

UMA PROPOSTA DE SISTEMA DE INFORMAÇÃO ABERTO E CONECTADO PARA ACERVOS DIGITAIS

AN OPEN AND CONNECTED INFORMATION SYSTEM FOR DIGITAL COLLECTION

Alison Carlos Filgueiras, Universidade Estadual de Goiás (UEG)

Feliz Alberto Ribeiro Gouveia, Universidade Fernando Pessoa (UFP/PT)

Paula Roberta Chagas, Universidade Estadual de Goiás (UEG)

Modalidade: Trabalho Completo

Resumo: A digitalização de objetos do patrimônio cultural é uma das formas de contribuir para a preservação patrimonial. A cada dia surgem instituições de memória com acervos digitalizados e esses espaços precisam de sistemas de informação para armazenar, organizar e realizar a recuperação dos objetos digitais sendo que, para tal, lançam mão de iniciativas que se esbarram em problemas relacionados à colaboração e ao reaproveitamento de informações de seus acervos digitais diversificados e multiformes. O objetivo deste artigo é apresentar um modelo e protótipo de sistema de informação semântico, aberto e conectado para facilitar diálogo tecnológico entre repositórios de forma semântica. O trabalho é resultado da investigação de doutoramento em Ciências da Informação na Universidade Fernando Pessoa (Porto, Portugal) que teve como foco o patrimônio cultural do Estado de Goiás e a metodologia utilizada para conduzi-la foi uma pesquisa-ação realizada em espaços de memória do estado. O processo metodológico culminou com a construção de um modelo de sistema de informação para que os vários espaços de memória possam colaborar uns com os outros por meio do intercâmbio de informações abertas sob várias perspectivas, em especial, a da *Linked Open Data* (LOD). Como resultado apresentamos um modelo ontológico apoiado em quatro dimensões (Pessoa-Fenômeno-Tempo-Espaço) e composto por três micro-ontologias de domínio (definição de acervo, definição de objetos e de dimensões) e um protótipo com um motor de interoperabilidade semântica, cujo algoritmo realiza ligações semânticas através de consultas SPARQL.

Palavras-chave: sistemas de informação do patrimônio cultural; interoperabilidade semântica; representação da informação; recuperação da informação.

Abstract: Digitization of cultural heritage objects is one of two ways to contribute to heritage preservation. Memory institutions with digitized collections emerge every day and these spaces require information systems to store, organize and create recovery mechanisms for digital objects and, for this, launch more initiatives that address problems related to collaboration and reuse of information from its diversified and multiform digital heritage. The objective of this article is to present a model of a semantic, open and connected information system for digital cultural heritage assets to facilitate technological dialogue between repositories in a semantic way. The work is a result of a PhD research in Information Sciences at the Fernando Pessoa University (Porto, Portugal) whose focus is the cultural heritage of the State of Goiás and the methodology used to conduct a research-action carried out in spaces of memory of state. The methodological process culminated in the construction of an information system model so that the various spaces of memory could

collaborate with each other through the exchange of open information on various perspectives, especially Linked Open Data (LOD). As a result, we present an ontological model supported by four dimensions (People-Phenomenon-Time-Space) and composed of three domain micro-ontologies (definition of collection, definition of objects and dimensions) and a prototype with a semantic interoperability engine, whose algorithm performs semantic bindings through SPARQL queries.

Keywords: cultural heritage information systems; semantic interoperability; information representation; information retrieval.

1 INTRODUÇÃO

A preservação do Patrimônio Cultural é um esforço contemporâneo da humanidade que envolve aspectos interdisciplinares políticos e científicos. Esse patrimônio, segundo a UNESCO (2020), compõe-se dos bens materiais e imateriais, naturais e culturais, móveis e imóveis herdados do passado e transmitidos às gerações futuras em virtude do seu valor insubstituível. Instituições como galerias de arte, bibliotecas, arquivos, museus, centros e outros espaços de memória e patrimônio (GLAMs), possuem importante papel quer seja na salvaguarda quer seja na difusão de conhecimento histórico ou contemporâneo em forma de cultura viva. A digitalização de objetos do patrimônio cultural é uma das maneiras de contribuir para a preservação patrimonial que, devido ao fenômeno da explosão digital, tem sido crescentemente incentivada e praticada em instituições de memória que utilizam Sistemas de Informação em uma miscelânea de áreas, comunidades e objetos que apresentam um amplo conjunto de desafios e oportunidades, sobretudo pelas capacidades da ciência aberta e dos dados abertos conectados.

Um desses desafios está relacionado à colaboração e ao reaproveitamento de informações de seus acervos digitais diversificados e multiformes. Um objeto digital de patrimônio como uma expressão artística, um quadro ou uma música popular, pode estar associado a eventos históricos, pessoas e lugares, que podem ser vistos como artefatos potencialmente relacionados em algum repositório digital na Web, em bancos de dados ou sistemas específicos, e quase sempre, sem o conhecimento dos curadores que ficam alheios quanto à sua existência. Esta é uma variação do clássico problema da interoperabilidade semântica (CANDELA *et al.*, 2019; MARTINS; CARMO; GERMANI, 2018). Como permitir que esses artefatos sejam conectados de forma múltipla e viva entre os diversos repositórios é um desafio que se estende desde a compreensão mais básica dos conceitos, como as definições de domínio, até as mais avançadas estratégias de integração e compartilhamento de conteúdo, como os Dados Abertos Conectados (DOERR, 2009; OLDMAN *et al.*, 2014).

No Estado de Goiás há diversas instituições de memória, dentre as quais, aquelas sob gestão e custódia da Secretaria de Estado da Cultura do Estado de Goiás (SECULT-GO). Esses espaços compostos de arquivos, bibliotecas e museus estão presentes em várias cidades do estado e vivenciam cotidianamente desafios tecnológicos nas mais diferentes áreas. Tais desafios passam por questões fundamentais, como a gestão de coleções, aquisições e inventários, e passam sobretudo, por questões relacionadas à interoperabilidade entre os acervos, que é o contexto de interesse do presente trabalho.

O objetivo deste artigo é apresentar um modelo e protótipo de sistema de informação para acervos digitais de patrimônio cultural com foco nos espaços de memória da SECULT-GO com vistas a facilitar o diálogo tecnológico entre os acervos destes espaços de forma semântica, aberta e conectada. O trabalho é resultado da investigação de doutoramento em Ciências da Informação na Universidade Fernando Pessoa (Porto, Portugal).

Além dessa seção de introdutória, este trabalho apresenta a seguinte composição. Na seção 2 é detalhada a metodologia utilizada. Na seção 3 é apresentada a fundamentação teórica e trabalhos relacionados, seguida da seção 4 que apresenta como resultados o modelo desenvolvido e realiza uma breve discussão acerca de alguns elementos do protótipo. A seção 5, por fim, apresenta as considerações finais.

2 METODOLOGIA

Esta pesquisa de abordagem qualitativa, foi conduzida em forma de pesquisa-ação e realizada nos espaços de memória da SECULT-GO, especificamente em 8 espaços da Superintendência do Patrimônio cujos museus, arquivos e bibliotecas estão geograficamente espalhados pelo estado. Foram realizadas entrevistas com curadores e análises documentais e tecnológicas nas instituições, além de uma revisão bibliográfica com foco na Ciência da Informação, nas tecnologias da Web Semântica e em modelos de referência como CIDOC-CRM e *Europeana Data Model* (EDM). O processo metodológico culminou com a construção de um modelo simplificado de sistema de informação para que os vários espaços de memória possam colaborar uns com os outros por meio do intercâmbio de informações abertas sob várias perspectivas, em especial, a da *Linked Open Data* (LOD).

A pesquisa-ação é uma modalidade de investigação-ação fortemente desenvolvida nas ciências sociais na busca de resultados concretos (THIOLENT, 2011). Já a Engenharia de

Software se apresenta sob forma de métodos técnicos e científicos voltados para extração, compreensão e construção de respostas (soluções) através de requisitos de sistemas, que também visam produzir resultados concretos (PRESSMAN, 2011).

As etapas do desenvolvimento metodológico foram as seguintes: 1) diagnóstico inicial com pesquisa nos documentos públicos e sites institucionais; 2) inquérito e entrevistas com curadores e colaboradores dos espaços de memória; 3) levantamento das tecnologias envolvidas nos acervos e sua organização; 4) engenharia de requisitos; 5) construção de modelos e teste protótipo.

3 FUNDAMENTAÇÃO TEÓRICA E TRABALHOS RELACIONADOS

Os Sistemas de Informação do Patrimônio Cultural geralmente atuam sobre domínio, guarda, processos de aquisição, interpretação, pesquisa e exposição de objetos digitais, bem como todas as atividades de gestão de acervos (DOERR, 2009; LIU; LIN, 2021). Segundo Doerr (2009), as ações destes sistemas podem ser classificadas de acordo com 4 funções gerais básicas: gestão de acervos, conservação, pesquisa e apresentação.

Belhi *et al.*, (2018) afirmam que adidas a essas funções estão as técnicas de digitalização de patrimônio, o enriquecimento de dados, o gerenciamento de conteúdo e a gestão da preservação (BELHI *et al.*, 2018). Isso porque estes sistemas precisam permitir a manipulação de informações com uma terminologia muito rica e em diferentes idiomas, bem como apresentar os conteúdos para usuários com uma ampla diversidade de interesses. Terminologias estas que tendem a atender especificidades de acordo com assuntos e apresentam, para além de questões técnicas, desafios intelectuais na aproximação e harmonização de conceitos que requerem estratégias semânticas (BOEUF *et al.*, 2018).

Os centros de memória, como espaço de formação e divulgação do conhecimento, proporcionam recolha, preservação e transmissão do patrimônio cultural com a missão principal de tornar informações acessíveis em diferentes contextos (CAMARGO; GOULART, 2015). Apesar da colaboração mútua possibilitar mais informações ricas e contextualizadas, estamos diante de um cenário que apresenta diversos problemas para sua implementação prática. A interoperabilidade semântica no contexto de patrimônio cultural se apresenta em um cenário que exige múltiplas conexões e a necessidade de interoperação entre repositórios digitais de tecnologias variadas que atendem a linguagens diversificadas (CANDELA *et al.*, 2019; MARTINS; CARMO; GERMANI, 2018). Este problema está

epistemologicamente ancorado no campo de estudos da representação da informação e recuperação da informação (ARAÚJO, 2014) e tem suas raízes nos primórdios da Web (LIMA *et al.*, 2016).

Na literatura, o uso de ontologias tem sido a estratégia mais utilizada para enfrentar o problema da interoperabilidade semântica (MARCONDES, 2016). As ontologias apresentam vantagens em relação a outros tipos de esquemas devido à sua capacidade de conceituação de domínios e seu poder de representar explicitamente a semântica dos dados (CALVANESE *et al.*, 2015; NIANG *et al.*, 2017).

Em linhas gerais, as ontologias computacionais são um meio de modelar formalmente a estrutura de sistemas em termos de entidades e relações relevantes que emergem de sua observação e que são úteis para um determinado universo do discurso (GUARINO; OBERLE; STAAB, 2009). A Web Semântica é uma extensão da Web composta de linguagens e padrões que permitem que ontologias sejam utilizadas para os mais diversos fins. Algumas dessas tecnologias são o *Resource Description Framework* (RDF), a *Ontology Web Language* (OWL) e a SPARQL, linguagem que permite manipular e recuperar grafos RDF. Berners-Lee, ao propor essa extensão tecnológica no início dos anos 2000, defendeu a necessidade expor dados na Web de uma forma que as máquinas possam compreender. Ao se realizar ligações entre esses dados utilizando-se as tecnologias da Web Semântica cria-se uma rede de dados colaborativa para máquinas e pessoas (BERNERS-LEE, 2006).

As três abordagens mais comuns para integração baseada em ontologias são a construção de ontologia única, a construção de múltiplas ontologias e a abordagem híbrida. Na primeira, as fontes de dados estão relacionadas a uma ontologia global e, portanto, tem mesma visão de domínio (ABU MUSA *et al.*, 2020). A abordagem de múltiplas ontologias requer que cada fonte de dados seja descrita com sua própria ontologia de forma local (BANNOUR *et al.*, 2018). Na abordagem híbrida cria-se um vocabulário comum de alto nível e cada ontologia local é mapeada para esta ontologia (DRAKOPOULOS *et al.*, 2018; FREIRE *et al.*, 2018; NIANG *et al.*, 2017). Esta última abordagem parece ser mais vantajosa, porém, custosa de se implementar, uma vez que requer a construção de um vocabulário global compartilhado além de ontologias locais a ele associadas.

Iniciativas mundialmente conhecidas como Europeana, DPLA, DigitalNZ, Trove e Biblioteca Digital da Índia, coletam grande quantidade de acervos distintos e fornecem acesso a bens culturais públicos digitalizados da Europa, Estados Unidos da América, Nova

Zelândia, Austrália e Índia, respectivamente (FREIRE *et al.*, 2018). Plataformas próprias e Sistemas de repositório institucionais são utilizados em larga escala nessas e em outras instituições. *Omeka-S*, *E-Print* e *DSpace* são exemplos dessas ferramentas.

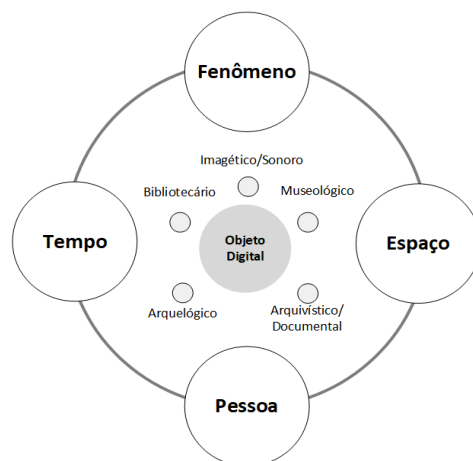
Em Martins, Carmo e Germani (2018) podemos encontrar uma ferramenta brasileira importante para que instituições de memória possam organizar, armazenar e permitir que seus acervos sejam expostos na Internet. O Tainacan é um Sistema de Informação do Patrimônio Cultural que visa ofertar uma tecnologia mais acessível e de fácil adoção social para a construção de repositórios brasileiros na área da cultura, fornecido como *plugin* do *Wordpress*, este último um Sistema Gerenciador de Conteúdo (CMS) popular fortemente utilizado na Web e tem em seu favor a fácil instalação e manutenção, bem como a comunidade de usuários (MARTINS; CARMO; GERMANI, 2018).

Os modelos CIDOC-CRM (DOERR *et al.*, 2020) e EDM (EUROPEANA, 2017) são exemplos popularizados de modelos de referência utilizados que buscam uma cobertura completa para objetos do patrimônio cultural em suas diversas disciplinas. O primeiro e mais conhecido, proposto e mantido pelo *International Council of Museums* (ICOM). O segundo, modelo utilizado na *Europeana*. Na mesma direção, a recente ontologia *Record in Context* (RiC) do *International Council of Archives* (MIRANDA, 2018) propõe um padrão conceitual para arquivos. Estas iniciativas contribuem para a disseminação de modelos e funcionam como arcabouço conceitual para construção e evolução de sistemas.

4 MODELO PROPOSTO

A partir da interação com os curadores da SECULT-GO e com base uma arquitetura genérica de funcionamento de um GLAM, chegou-se a uma proposta, conforme Figura 2, com 5 tipos de artefatos: arquivístico/documental, imagético/sonoro, arqueológico, museológico e bibliotecário, que atendem os diversos eixos (nichos) dos espaços de memória e ainda ampliam as possibilidades de classificação dos artefatos em cada um deles.

Figura 1 – Modelo em 4 dimensões



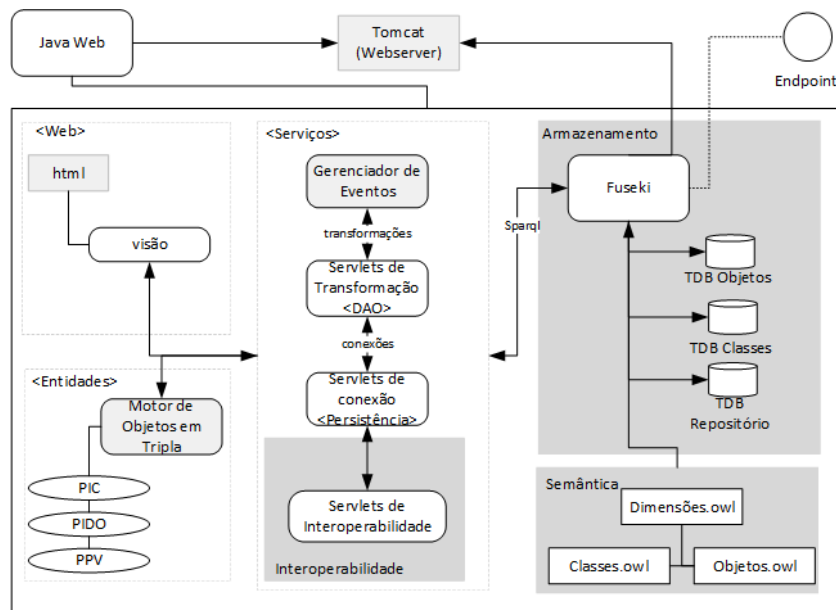
Fonte: elaborada pelos autores.

Desta forma, independentemente das especificidades dos espaços, as coleções podem dialogar dentro de uma perspectiva em que um objeto é associado a um ou mais eixos informacionais de um típico GLAM. Os eixos informacionais estão intrinsecamente associados a quatro dimensões conceituais: Pessoa, Fenômeno, Tempo e Espaço.

A construção do ambiente para o protótipo utilizou o *Apache Jena* como framework para modelagem de triplas RDF e a aplicação *Fuseki* como sistema gerenciador de banco de dados triplos (TDB) e *endpoint* para atender a chamadas SPARQL. Como servidor Web utilizou-se o *Apache Tomcat*.

O protótipo se apresenta em uma arquitetura composta de cinco módulos, sendo dois externos e três internos. Internamente, o Módulo Web (páginas web), o Módulo de Entidades, que funciona como um motor de manipulação de objetos em triplas RDF, o Módulo de Serviços contendo três submódulos: um Gerenciador de Eventos, os Serviços de Transformação (DAO) e Persistência, com um Motor de Objetos em Tripla (MOT), e ainda um Motor de Interoperabilidade (MI) com os Serviços de Interoperabilidade que acionam os algoritmos de busca nos repositórios. Os dois módulos externos dizem respeito ao armazenamento TDB e as ontologias da camada semântica que são mantidas pelo *Apache Jena Fuseki*. A Figura 2 apresenta uma visão arquitetural do protótipo.

Figura 2 – Visão arquitetural do protótipo



Fonte: elaborada pelos autores

Para construir as requisições e respostas *http* e transitar entre os módulos e camadas do protótipo, bem como para responder e realizar consultas externas, foram utilizados *Servlets Java*. *Servlets* são classes Java que estendem a classe *HttpServlet* e atendem aos verbos *http* (*GET*, *POST*, *PUT*, *PATCH*, *DELETE*). No Módulo de Serviços são utilizadas ainda classes especiais gerenciáveis para realizar a inversão de dependências das classes do modelo e integrações com as páginas da camada de páginas Web.

O funcionamento geral dessa arquitetura, de forma simplificada, porém alinhada ao que o modelo proposto preconiza, prevê que as Páginas Web sejam responsáveis pela interação com os usuários humanos que, por meio dos eventos disponíveis nos controles das páginas HTML, disparam ações que são transformadas em requisições enviadas para o Gerenciador de Eventos. Esse, por sua vez, aciona o Motor de Objetos em Triplas (MOT) que realiza a transformação de objetos para triplas RDF e vice-versa. Os objetos transformados são enviados aos *servlets* de transformação (DAO) que acionam os objetos de acesso ao banco de dados (TDB) após realizar as transformações necessárias para a linguagem SPARQL. Isso para que os grafos sejam pesquisados ou persistidos no TDB *Fuseki*, que mantém o serviço de Armazenamento (*RDF Store*) dos bancos de dados em tripla e das ontologias da Camada Semântica. No *Fuseki* há um *endpoint* disponibilizado para consultas externas em SPARQL.

4.1 Semântica e Interoperabilidade

A camada semântica da proposta é composta de três micro-ontologias integradas. A primeira para classificação do acervo, a segunda para definição dos objetos e a terceira para o dimensionamento. O Quadro 1 apresenta as propriedades da primeira micro-ontologia denominada *Classes.owl*.

Quadro 1 - Propriedades para classificação de acervo

Propriedade	Descrição	IRI
<i>owl:class</i>	Essa marcação indica que se trata de uma classe OWL.	https://www.w3.org/2002/07/owl#Class
<i>rdfs:label</i>	Onde se define um nome para a classe que está sendo definida.	https://www.w3.org/TR/rdf-schema/#ch_label
<i>rdfs:subClassOf</i>	Através dessa propriedade é possível criar uma relação de especialização ou generalização entre as classes.	https://www.w3.org/TR/rdf-schema/#ch_subclassof
<i>rdfs:comment</i>	Essa propriedade é usada para detalhar comentários, características gerais e outras informações sobre a classe que está sendo criada.	https://www.w3.org/TR/rdf-schema/#ch_comment

Fonte: elaborado pelos autores

A primeira propriedade, *owl:class*, é um marcador para informar que se trata de uma ontologia OWL. A segunda, *rdfs:label*, serve para definir um rótulo para cada elemento da classificação do acervo, como uma coleção ou um tipo. A terceira, *rdfs:subClassOf*, é um marcador para hierarquizar as divisões do acervo. Uma descrição textual simplificada de cada um destes elementos é realizada através da propriedade *rdfs:comment*. A segunda micro-ontologia, denominada *Objetos.owl*, é utilizada para definição dos objetos do acervo e suas propriedades podem ser vistas no Quadro 2.

Quadro 2 - Propriedades para definição dos objetos digitais

Propriedade	Descrição	IRI
<i>:id</i>	Identificador do objeto que é gerado automaticamente pelo protótipo criando um identificador único para o recurso.	http://200.137.241.247:8080/fuseki/objetos#id
<i>rdf:type</i>	Propriedade para definir que o objeto é de uma classe específica. Denota a relação “é um”, ou também o “tipo”	https://www.w3.org/TR/rdf-schema/#ch_type
<i>dc:title</i>	Conforme o vocabulário do padrão Dublin Core (DCMI), esta propriedade descreve o título (ou nome) de cada objeto.	http://purl.org/dc/elements/1.1/title
<i>dc:subject</i>	Esta propriedade, também do padrão Dublin Core (DCMI), apresenta o tópico ou assunto, sendo uma boa prática a utilização de URI, mas para todo o caso, texto livre quando não for possível.	http://purl.org/dc/elements/1.1/subject
<i>dc:description</i>	Essa propriedade do padrão Dublin Core (DCMI) permite uma descrição mais detalhada do objeto digital em questão, podendo envolver, mas não se	http://purl.org/dc/elements/1.1/description

	limita a, uma tabela de conteúdos, resumo, um texto livre uma representação gráfica	
dim:tipoFisico	Essa propriedade é definida na ontologia de definição das dimensões para classificar os objetos quanto ao tipo de mídia que relaciona-se especificamente ao GLAM.	http://200.137.241.247:8080/fuseki/dimensoes#tipoFisico
dim:tipoDimensional	Para o caso de objetos dimensionais, essa propriedade permite que estes sejam classificados em uma das quatro dimensões (tempo, espaços, fenômeno e pessoa)	http://200.137.241.247:8080/fuseki/dimensoes#tipoDimensional

Fonte: elaborado pelos autores

Uma regra básica para a criação de um novo objeto é que todos os dados da definição inicial digital estejam completos. As quatro propriedades básicas (mínimas) a serem informadas são *rdf:type*, *dc:title*, *dc:subject* e *dc:description*. A propriedade adicional *:id* é relacionada à referência inequívoca indica que seja criado automaticamente um id, que é o identificar único do recurso que o objeto representa. Há ainda duas propriedades importantes. A propriedade *dim:TipoFisico* é utilizada para tipificação do objeto quanto à mídia ou artefato a que este objeto representa dentro de uma visão de objetos de um GLAM. E por fim, para o caso de objetos dimensionais, há uma propriedade definida na ontologia específica (dimensionamento) que determina para qual dimensão um possível objeto dimensional se direciona.

A terceira micro-ontologia (*Dimensões.owl*) é responsável pela conexão entre as duas ontologias iniciais (definições de classes do acervo e de objetos do acervo). Esta fornece um conjunto de propriedades para permitir que os objetos digitais do acervo e os objetos dimensionais possam se interconectar. Deve permitir ainda que seja possível aplicar outras regras sobre os objetos digitais e sua tipificação quanto às características físicas ou dimensionais. Para isso, apresenta também um conjunto de propriedades que de fato trarão sentido às relações entre os objetos, conforme Quadro 3.

Quadro 3 - Propriedades para dimensionamento dos objetos digitais

Propriedade	Descrição	IRI
:relacionaPessoa	Permite associar qualquer coisa ao um Objeto Dimensional Pessoa.	http://200.137.241.247:8080/fuseki/dimensoes#relacionaPessoa
:relacionaFenômeno	Utilizada para explicitar uma relação de alguma coisa a um Objeto Dimensional Fenômeno.	http://200.137.241.247:8080/fuseki/dimensoes#relacionaFenômeno
:relacionaEspaço	Propriedade que associa uma coisa a um espaço.	http://200.137.241.247:8080/fuseki/dimensoes#relacionaEspaço
:relacionaTempo	Para que se possa associar uma coisa a um Objeto Dimensional Tempo.	http://200.137.241.247:8080/fuseki/dimensoes#relacionaTempo

:tipoFísico	Propriedade para classificar um Objeto Digital em um tipo físico, que envolve os artefatos físicos dos acervos	http://200.137.241.247:8080/fuseki/dimensoes#tipoFísico
:tipoDimensional	Propriedade para classificar um Objeto Digital em um tipo dimensional, que não envolve, necessariamente, os artefatos físicos dos acervos	http://200.137.241.247:8080/fuseki/dimensoes#tipoFísico
dim:tipoDimensional	Para o caso de objetos dimensionais, essa propriedade permite que estes sejam classificados em uma das quatro dimensões (Tempo, Espaços, Fenômeno e Pessoa)	http://200.137.241.247:8080/fuseki/dimensoes#tipoDimensional

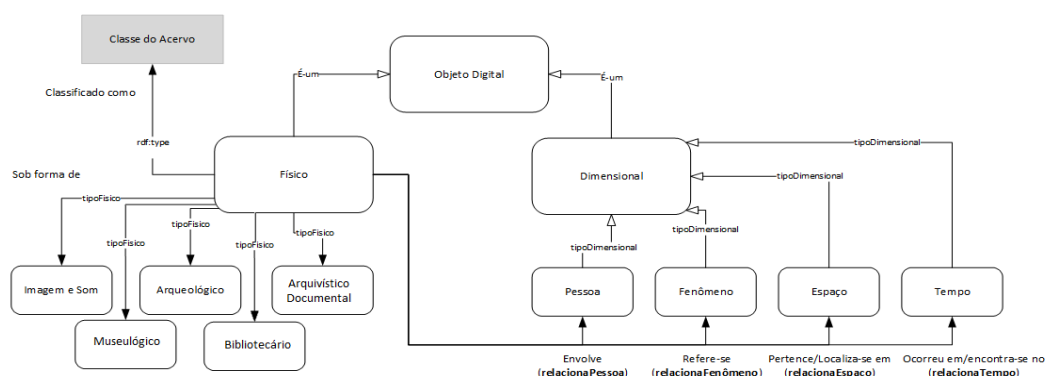
Fonte: elaborado pelos autores

Um objeto digital pode ser encontrado sob várias formas, referir-se a fenômenos e pode envolver pessoas em um determinado tempo, de acordo com as quatro dimensões apresentadas. Essa estrutura fornece classes e propriedades para que os objetos possam ser encontrados em duas formas principais – Objetos Físicos e Objetos Dimensionais – para um domínio nas quatro dimensões e nos cinco eixos fundamentais, respeitando a classificação dos acervos que são determinadas e mantidas por cada Curador de espaço.

Os objetos dimensionais dizem respeito às quatro dimensões e não são, necessariamente, ligados a um artefato físico, por exemplo, a descrição de uma pessoa ou de um espaço. Já os objetos físicos são os objetos digitais materializados presentes nos espaços, como uma fotografia, uma carta ou uma vestimenta. A Figura 3 apresenta uma visão ampliada da classificação dos objetos, que são divididos em Objetos Físicos e Objetos Dimensionais.

A interoperabilidade entre os objetos dos acervos é constituída através da associação entre os elementos presentes nas ontologias e que são persistidos nos bancos de dados em tripla (TTB). Um algoritmo de interoperabilidade é utilizado para captar relações possíveis em repositórios presentes na Web. Esses repositórios são adicionados pelo Curador do espaço em uma lista de repositórios amigos.

Figura 3 – Visão dos objetos digitais no acervo



Fonte: elaborada pelos autores

Esta interoperação ocorre entre os repositórios através da ligação de propriedades adicionadas pelos respectivos curadores, uma vez que os repositórios oferecem uma interface pública via SPARQL e/ou chamadas diretas por meio de *middlewares* de serviços Web. A camada de interoperabilidade é composta por dois elementos básicos. Um motor de interoperabilidade e um *end-point* SPARQL. O primeiro é responsável pelo intercâmbio entre os repositórios digitais e o segundo permite que requisições internas e externas possam ser respondidas por meio de chamadas SPARQL.

Quadro 4 - Algoritmo de interoperabilidade

1. Leia (Objeto_1)
2. Seja Dimensão={Pessoa,Fenômeno,Espaço,tempo}
3. Seja Relação={relacionaPessoa, relacionaFenômeno, relacionaEspaço, relacionaTempo}
4. Carregue a lista de *repositórios amigos*
5. Para cada repositório faça:
 - 5.1 Busque os Objetos cujo título, assunto ou descrição contêm as palavra-chave do Objeto_1.
 - 5.2 Busque os Objetos ?x onde exista <Objeto_1> <Relação> ?x.
6. Devolva todos os grafos ordenados por Dimensão e Repositório.
7. Apresente os dados em triplas.

Fonte: elaborado pelos autores

O motor de interoperabilidade é composto por um algoritmo principal (conforme Quadro 4) que realiza consultas SPARQL quando *eventos* são acionados em cada repositório, e algoritmo auxiliares que são acionados pelo algoritmo principal ou chamadas específicas. Por exemplo, se um Curador digital insere um novo objeto ou uma nova propriedade em um objeto já existente, o motor realiza uma busca nos *repositórios amigos* para encontrar informações nas quatro dimensões de interoperabilidade definidas no modelo.

4.2 Exemplos de interação do protótipo

O protótipo do sistema Guará encontra-se disponível no endereço <http://200.137.241.247:8080/guara>. Na Figura 4 é apresentado um exemplo de edição de classes. No primeiro quadro um acervo de arquivo, no segundo a estrutura de um museu.

Figura 4 – Página do protótipo com exemplo de edição de classes de acervos

Classe: (owl:Class)	Rótulo: (rdfs:label)	Comentário/Descrição: (rdfs:comment)	Filha de: (rdfs:subClassOf)	Ação
http://200.137.241.247:8080/fuseki/dicoseclass#Acervo	Acervo	Classe que agrupa todos os registros		Editar Excluir
http://200.137.241.247:8080/fuseki/dicoseclass#Batismo	Batismo	Classe que agrupa todos os registros referentes a batismo...	http://200.137.241.247:8080/fuseki/dicoseclass#Registro	Editar Excluir
http://200.137.241.247:8080/fuseki/dicoseclass#Registro	Registro	Classes dos registros	http://200.137.241.247:8080/fuseki/dicoseclass#Acervo	Editar Excluir
http://200.137.241.247:8080/fuseki/dicoseclass#Casamento	Casamento	Essa é a classe para agrupar os registros de casamento.	http://200.137.241.247:8080/fuseki/dicoseclass#Registro	Editar Excluir
http://200.137.241.247:8080/fuseki/dicoseclass#Obito	Obito	Classe para os registros relacionados às declarações de óbito.	http://200.137.241.247:8080/fuseki/dicoseclass#Registro	Editar Excluir

Classe: (owl:Class)	Rótulo: (rdfs:label)	Comentário/Descrição: (rdfs:comment)	Filha de: (rdfs:subClassOf)	Ação
http://200.137.241.247:8080/fuseki/mpclass#Acervo	Acervo	Classe que agrupa todos os registros		Editar Excluir
http://200.137.241.247:8080/fuseki/mpclass#Documento	Documentos	Classe para agrupar os documentos do acervo do Arquivo.	http://200.137.241.247:8080/fuseki/mpclass#Acervo	Editar Excluir
http://200.137.241.247:8080/fuseki/mpclass#DocumentoOficial	Documentos Oficiais	Coleção para todos os documentos oficiais	http://200.137.241.247:8080/fuseki/mpclass#Documento	Editar Excluir
http://200.137.241.247:8080/fuseki/mpclass#Manuscrito	Manuscritos	Coleção para todos os documentos manuscritos do arquivo	http://200.137.241.247:8080/fuseki/mpclass#Documento	Editar Excluir
http://200.137.241.247:8080/fuseki/mpclass#Cartas	Cartas	Coleção para as cartas que se referem aos documentos manuscritos.	http://200.137.241.247:8080/fuseki/mpclass#Documento	Editar Excluir

Fonte: elaborada pelos autores.

Na Figura 5 encontra-se a visualização de uma página do protótipo onde os curadores podem realizar inserção, modificação e exclusão de objetos dimensionais. Como é possível visualizar na imagem, cada objeto no protótipo recebe um identificador único. No caso, o objeto adicionado foi um objeto dimensional referente à Pessoa Pedro Ludovico Teixeira.

Figura 5 – Exemplo de manipulação de objetos dimensionais

Criar Novo Objeto Dimensional

Tipo: Pessoa

id: a6966469-1260-415d-a4c2-1452d899

Título: (dc:title) Pedro Ludovico Teixeira

Descrição: (Description) Pedro Ludovico Teixeira. Interventor e Governador de Goiás após a revolução de 1930. Idealizador e fundador da nova capital do estado, Goiânia, em 1933. "

Gravar Voltar

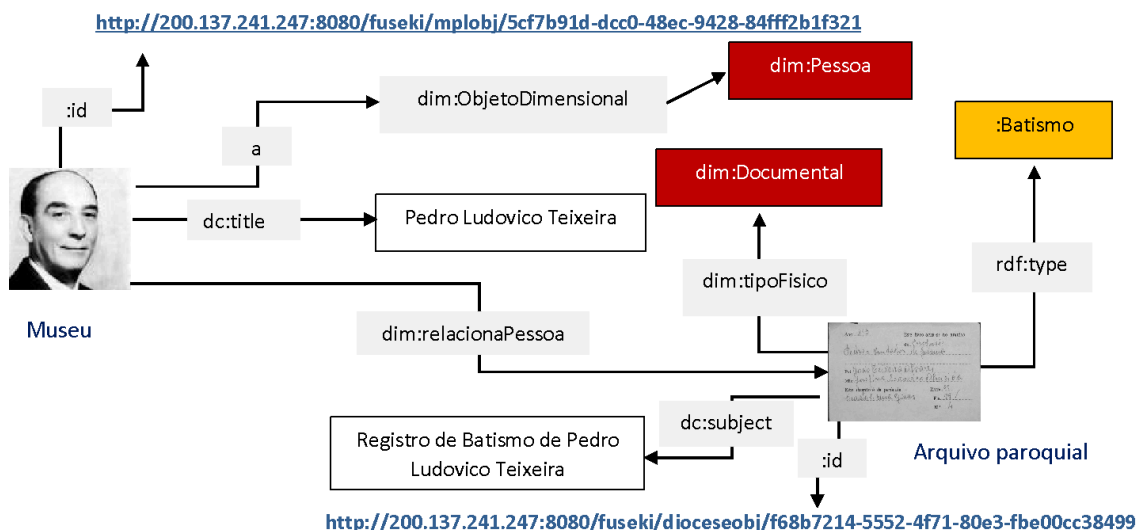
```
1 [...]
2 PREFIX obj: <http://200.137.241.247:8080/fuseki/objetos#>
3 PREFIX cls: <http://200.137.241.247:8080/fuseki/classes#>
4 PREFIX dim: <http://200.137.241.247:8080/fuseki/dimensoes#>
5 PREFIX owl: <http://www.w3.org/2002/07/owl#>
6 INSERT DATA{<http://200.137.241.247:8080/fuseki/mpobj/5cf7b91d-
7 dcc0-48ec-9428-84fff2b1f321> dc:title "Pedro Ludovico Teixeira";
8 a dim:ObjetoDimensional; dim:tipoDimensional dim:Pessoa;dc:subject
9 "Pedro Ludovico Teixeira";
10 dc:description "Pedro Ludovico Teixeira. Interventor e Governador de
11 Goiás após a revolução de 1930. Idealizador e fundador da nova capital do
12 estado, Goiânia, em 1933. "
13 }
```

Fonte: elaborada pelos autores

A Figura 6 representa, simbolicamente, o processo de interoperação entre espaços de memória. Em dois repositórios diferentes existem registros referentes a Pedro Ludovico Teixeira. No repositório do Museu encontra-se uma definição dimensional da Pessoa. No

repositório do Arquivo Paroquial encontra-se um objeto físico que é um registro de batismo de Pedro Ludovico. A chamada SPARQL apresentada na Figura 5 permite gerar o grafo abaixo.

Figura 6 – Exemplo de Grafo de Interoperabilidade



Fonte: elaborada pelos autores.

5 CONSIDERAÇÕES FINAIS

O principal objetivo deste trabalho foi apresentar um sistema de informação semântico, aberto e conectado para acervos digitais abordando o problema de integração e interoperação de repositórios de instituições de memória. Assumimos que os sistemas de informação na sociedade em rede precisam estar conectados à rede maior, a Internet. Também assumimos que a colaboração é essencial para que os sistemas possam crescer e fornecer, de forma útil, informações mais ricas contidas em outros repositórios. Por tal, entendemos que deve haver interoperabilidade semântica e, para que ela seja plena, que as informações patrimoniais sejam abertas. Certamente há um grande conjunto de desafios que precisam ser transformados em processos. Os espaços de memória têm características próprias, uma questão fundamental é como apoiar os vários modelos de organização da informação diante da diversidade de dados de memória combinados com as facetas exigidas pelas funções e as expectativas de uso.

Com o protótipo e modelo buscamos mostrar que é possível fornecer uma variedade de serviços e conectá-los a outros repositórios através do uso de ontologias e uma API SPARQL que permite que consultas remotas sejam executadas. Como pôde ser visto nos exemplos de interação, o protótipo do sistema Guará permite a integração múltipla de repositórios e as três micro-ontologias apresentadas são capazes de garantir que, de forma

muito simples, cada espaço de memória tenha sua própria organização e possa colaborar com outras coleções na Web, agregando assim mais valor aos seus objetos digitais. Como trabalho futuro, esperamos melhorar o algoritmo de interoperabilidade e treiná-lo para coletar automaticamente informações em repositórios de interesse. Há ainda a expectativa de integração e harmonização (mapeamento) ontologias e taxonomias como as preconizadas nos modelos de referência EDM, CIDOC-CRM e RiC, bem como a geração automática de entradas para estes modelos.

REFERÊNCIAS

- ABU MUSA, T. H. *et al.*, **OntoM: An Ontological Approach for Automatic Classification**. 2020 IEEE International Conference on Informatics, IoT, and Enabling Technologies. **Anais...**2020.
- ARAÚJO, C. A. Á. Fundamentos Da Ciência Da Informação: Correntes Teóricas e o Conceito de Informação. **Perspectivas em Gestão & Conhecimento**, v. 4, p. 57–79, 2014.
- BANNOUR, I. *et al.*, **CRMCR - a CIDOC-CRM extension for supporting semantic interoperability in the conservation and restoration domain**. 2018 3rd Digital Heritage International Congress (DigitalHERITAGE)
- BELHI, A. *et al.*, **Digitization and Preservation of Cultural Heritage Products**. 2018.
- BERNERS-LEE, T. **Linked Data**. [s.l: s.n.]. Disponível em: <<https://www.w3.org/DesignIssues/LinkedData.html>>.
- CALVANESE, D. *et al.*, **A ‘historical case’ of Ontology-Based Data Access**. 2015 Digital Heritage. **Anais...**2015.
- CAMARGO, A. M.; GOULART, S. **Centros de Memória: Uma proposta de definição**. 1a. ed. São Paulo: Sesc-SP, 2015.
- CANDELA, G. *et al.*, A linked open data framework to enhance the discoverability and impact of culture heritage. **Journal of Information Science**, v. 45, n. 6, p. 756–766, 2019.
- CARVALHO JUNIOR, J. M. C.; MARTINS, D. L.; GERMANI, L. B. GLAM e Instituições de Memória em Rede: uma ‘Infosfera’ de Floridi? **pragMatizes**, 2019.
- DOERR, M. Ontologies for Cultural Heritage. Em: STAAF, S.; STUDER, R. (Eds.). **Handbook on Ontologies**. Second Edi ed. New York, NY, USA: Springer, 2009. p. 463–486.
- DOERR, M. *et al.*, **Definition of the CIDOC Conceptual Reference ModelDefinition of the CIDOC Conceptual Reference Model Produced**. [s.l: s.n.].

DRAKOPOULOS, G. *et al.*, **Towards a framework for tensor ontologies over Neo4j: Representations and operations**. 2017 8th International Conference on Information, Intelligence, Systems and Applications, IISA 2017. **Anais...**2018.

EUROPEANA. **Definition of the Europeana Data Model**. Disponível em: <<https://pro.europeana.eu>>. Acesso em: 23 jul. 2018.

FREIRE, N. *et al.*, Aggregation of cultural heritage datasets through the Web of Data. **Procedia Computer Science**, v. 137, p. 120–126, 2018.

GUARINO, N.; OBERLE, D.; STAAB, S. What Is an Ontology? Em: STAAB, S.; STUDER, R. (Eds.). **International Handbooks on Information Systems**. 2. ed. [s.l: s.n.]. p. 1–17.

LE BOEUF, Patrick *et al.* (Ed.). **Definition of the CIDOC Conceptual Reference Model**: version 6.2.3: International Council of Museums (ICOM); International Committee for Documentation(CIDOC), 2018. Disponível em: <http://www.cidoccrm.org/Version/version-6.2.3-0>. Acesso em: 23 jul. 2018.

LIU, Y.; LIN, H. W. Construction of interpretation and presentation system of cultural heritage site: An analysis of the old city, Zuoying. **Heritage**, v. 4, n. 1, p. 316–332, 2021.

MARCONDES, C. H. Interoperabilidade entre acervos digitais de arquivos, bibliotecas e museus: Potencialidades das tecnologias de dados abertos. **Perspectivas em Ciência da Informação**, v. 21, n. 2, p. 61–83, 2016.

MARTINS, D. L.; CARMO, D. DO; GERMANI, L. B. **MUSEU DO ÍNDIO : ESTUDO DE CASO DO PROCESSO DE MIGRAÇÃO E ABERTURA DOS DADOS LIGADOS SEMÂNTICOS DO ACERVO MUSEOLÓGICO COM O SOFTWARE LIVRE TAINACAN**. XiX Enancib. **Anais...**2018.

MIRANDA, J. M. DE. Records in Contexts (RiC): Análise da sua aplicação em arquivos, à luz das tecnologias Linked Open Data (LOD). p. 1–26, 2018.

NIANG, C. *et al.*, Supporting semantic interoperability in conservation-restoration domain: The PARCOURS project. **Journal on Computing and Cultural Heritage**, v. 10, n. 3, 2017.

OLDMAN, D. *et al.*, Realizing lessons of the last 20 years: A manifesto for data provisioning and aggregation services for the digital **D-Lib Magazine**, v. 20, n. 7–8, 2014.

PRESSMAN, R. S. **Engenharia de Software: Uma Abordagem Profissional**. 7. ed. São Paulo: McGraw Hill Brasil, 2011.

SILVA, M.; MARTINS, D. L.; SIQUEIRA, J. Web semântica em repositórios: ontologia para representação de bibliotecas digitais. **Ciência da Informação em Revista**, v. 6, n. 1, p. 99–113, 2019.

THIOLENT, M. **METODOLOGIA DA PESQUISA-AÇÃO**. 18. ed. São Paulo: Cortez, 2011.