

XXIV ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO – XXIV ENANCIB

ISSN 2177-3688

GT-8 – Informação e Tecnologia

ARQUIVOS QUE SE COMPLETAM: UMA PROPOSTA DE INTEGRAÇÃO DE ACERVOS DIGITAIS DO PATRIMÔNIO CULTURAL NO ESTADO DE GOIÁS

COMPLEMENTARY ARCHIVES: A PROPOSAL FOR THE INTEGRATION OF DIGITAL COLLECTIONS OF CULTURAL HERITAGE IN THE STATE OF GOIÁS

Alison Carlos Filgueiras – Universidade Estadual de Goiás (UEG)

Feliz Ribeiro Gouveia – Universidade Fernando Pessoa (UFP/PT)

Juliana Vasconcelos Braga – Universidade Estadual de Goiás (UEG)

Modalidade: Trabalho Completo

Resumo: Arquivos históricos e centros de documentação, assim como outros espaços de memória, são elementos fundamentais para história social do conhecimento. Os processos de digitalização dos acervos diversificados em tais espaços geram desafios e oportunidades no que diz respeito à curadoria, estruturação, recuperação e uso da informação, e em especial, complementariedades de fatos, fenômenos, personalidades e outros agentes cujas informações estão presentes em suas coleções e catálogos. O objetivo deste trabalho é propor um Sistema de Informação (S.I.), denominado Guará, que visa facilitar a gestão e o uso de acervos digitais de memória e patrimônio em Goiás. O sistema busca promover a integração semântica de acervos diversificados, ampliando as possibilidades de organização, exposição e intercâmbio de informações entre diferentes espaços de memória. O referido sistema tem sido desenvolvido de forma colaborativa para que variados espaços de memória, institucionais ou particulares, possam colaborar uns com os outros realizando intercâmbio de informações abertas sob várias perspectivas, contribuindo assim para uma rede de repositórios. Como resultado apresentamos um modelo ontológico e exemplos de descrição que fazem uso das quatro dimensões (Pessoa-Evento-Tempo-Lugar) estabelecidas no que chamamos de micro-ontologias de domínio (definição de acervo, definição de objetos e de dimensões) de forma que o protótipo permite ligações semânticas através de consultas SPARQL por meio de APIs e consultas federadas.

Palavras-chave: Sistemas de informação do Patrimônio Cultural; interoperabilidade semântica; representação da informação; recuperação da informação.

Abstract: Historical archives and documentation centers, as well as other memory institutions, are crucial for the social history of knowledge. The digitization of diverse collections in these spaces presents challenges and opportunities for curation, structuring, retrieval, and use of information, particularly in terms of complementing facts, events, personalities, and other agents represented within their collections and catalogs. The objective of this paper is to propose an Information System (IS), named Guará, aimed at facilitating the management and utilization of digital memory and heritage

collections in Goiás. The system is designed to promote semantic integration of diverse archives, enhancing possibilities for organization, exhibition, and information exchange among different memory spaces. Developed collaboratively, the system enables various memory institutions, both institutional and private, to engage in the exchange of open information from multiple perspectives, thus fostering a network of interconnected repositories. As a result, we present an ontological model and examples of descriptions using the four fundamental dimensions (Person, Event, Time, Place) established in domain micro-ontologies, which allow for semantic linking through SPARQL queries via APIs and federated queries

Keywords: Cultural Heritage Information Systems; semantic interoperability; information representation; information retrieval.

1 INTRODUÇÃO

Os centros de memória são importantes fontes de informação do patrimônio cultural. No estado de Goiás esses espaços, independente da institucionalidade, estão espalhados por várias cidades. As diferentes nomenclaturas e realidades encontram significado comum na preservação na memória e na necessidade de garantir acesso às pessoas a seus acervos de forma facilitada. Grande parte desses espaços possui a custódia ou salvaguarda de coleções em forma de patrimônio arquivístico-documental, museal, fotográfico, entre outros. Apesar de muitos destes espaços – institucionais ou não – experimentarem uma realidade comumente difícil em relação aos recursos financeiros e tecnológicos, a digitalização do patrimônio é inevitável e urgentemente necessária.

Necessária por que, para além de auxiliar na preservação – uma vez que objetos materiais deixam de ser manipulados, por exemplo – amplia as possibilidades de organização, exposição e uso. Inevitável porque, em uma era digital onde surgem e, consequentemente, exigem-se cada vez mais das chamadas tecnologias disruptivas, o foco se desloca da custódia física para a gestão eficiente da informação digital, do acesso centralizado para o acesso aberto. As possibilidades de construção e reconstrução do conhecimento, aumentadas por esse novo foco, abrem espaço, por um lado, para melhoria do uso localizado, e por outro, para a integração de acervos diversificados, naquilo que aqui são chamados de espaços de memória que se completam. Nesse contexto, os sistemas de informação possuem especial valor no sentido de facilitar a organização, o armazenamento, a exposição e possibilidades de uso e intercâmbio de informação entre diferentes espaços e iniciativas. Sem nenhuma pretensão de reduzir a problemática do patrimônio cultural a questões tecnológicas, muito menos com intenção de substituir práticas arquivísticas, museais, dentre outras, este trabalho

discute uma proposta de Sistema de Informação (S.I) que amplia as possibilidades de gestão e uso de acervos digitais de memória e patrimônio.

Além dessa seção introdutória apresenta a seguinte composição. Na seção 2 é detalhada a metodologia utilizada. Na seção 3 é apresentada a fundamentação teórica e os trabalhos relacionados, seguida da seção 4 e 5, são apresentados e discutidos resultados a partir do modelo desenvolvido. A seção 6, por fim, apresenta as considerações finais.

2 METODOLOGIA

A pesquisa qualitativa foi conduzida em várias etapas sequenciais para o desenvolvimento do sistema de informação Guará. Inicialmente, foi realizada uma análise documental de fontes públicas dos espaços de memória em Goiás, com o objetivo de compreender as práticas correntes e os desafios enfrentados por esses espaços.

Na fase de seleção de tecnologias, adotou-se como referência o estado da arte e as práticas de portais líderes como a Europeia, com foco em interoperabilidade e acessibilidade de dados. Uma revisão bibliográfica específica sobre linguagens e ferramentas de Web Semântica foi conduzida para fundamentar as escolhas tecnológicas.

A engenharia de ontologias seguiu a Metodologia 101, uma abordagem estruturada e rigorosa para a criação de ontologias. A análise de requisitos foi embasada em entrevistas com curadores e colaboradores dos espaços de memória, fornecendo um entendimento das necessidades dos usuários.

Na etapa de modelagem conceitual, os requisitos coletados foram traduzidos em um modelo que representa as entidades e relações do domínio de interesse. A prototipação foi realizada com uma abordagem ágil, permitindo iterações rápidas e ajustes com base em *feedback*.

Por fim, estudos de sistemas semelhantes foram conduzidos para identificar lições aprendidas e melhores práticas aplicáveis ao desenvolvimento do Guará. Essa abordagem iterativa assegurou a relevância e utilidade do sistema proposto para os espaços de memória em Goiás.

3 CURADORIA DIGITAL E SISTEMAS DE INFORMAÇÃO DO PATRIMÔNIO CULTURAL

A curadoria digital, a gestão, a preservação do patrimônio cultural, os espaços de memória e os sistemas de informação são tópicos basilares e com interrelações disciplinares no que se discute nesse trabalho. O âmago dos sistemas de informação está nos serviços que podem oferecer para uma área especializada, nesse caso, uma área inter e transdisciplinar que envolve a ciência da informação, da computação, do patrimônio, humanidades digitais, dentre outras. No domínio do patrimônio cultural são fundamentais para a gestão eficiente de dados complexos e interligados, facilitando a preservação e o acesso ao conhecimento cultural (Doerr, 2009). A implementação de padrões de metadados e ontologias computacionais, como o CIDOC-CRM, permite a criação de um modelo de dados coerente que pode representar as nuances dos objetos culturais e seus contextos históricos e sociais. Esses padrões são essenciais para garantir a interoperabilidade entre diferentes sistemas de informação, o que é vital para a colaboração entre instituições culturais e para o enriquecimento de dados através da integração de múltiplos acervos.

Abordamos a curadoria digital como um campo em desenvolvimento dentro da Ciência da Informação, situado na intersecção entre gestão da informação, gestão do conhecimento, comunicação e tecnologia (Abby Clobridge, 2013). A estreita relação entre a curadoria digital e a gestão da informação se confirma em Nonato e Aganette (2022), ao postular que a gestão da informação está relacionada à gestão do ciclo de vida da informação. Uma abordagem de ciclo de vida para o gerenciamento de materiais digitais permite a visualização dos processos, atividades e relacionamentos necessários para uma curadoria bem-sucedida e preservação de longo prazo (Higgins, 2018).

A curadoria digital, como um campo de pesquisa e prática, visa contemplar as fases do ciclo de vida da informação e compreende tarefas como seleção, preservação, manutenção, coleta, arquivamento e reavaliação de ativos digitais (Fermann; Lima; Chardulo, 2020). Adicionar valor aos conjuntos de dados e objetos digitais, através da inserção de metadados ou anotações adicionais, para que então estes possam ser reutilizados, é uma característica central, visto que envolve uma ampla variedade de partes interessadas transpondo fronteiras disciplinares: organizações de patrimônio cultural como bibliotecas, arquivos e museus, agências de financiamento, órgãos governamentais, data centers, repositórios institucionais, e organizações da sociedade civil.

Especificamente os arquivos, enquanto estruturas sociais, ensejam grande responsabilidade e devem seguir preceitos de confiabilidade, autenticidade e integridade (Furtado; Almeida; Santos, 2022). A ausência de qualquer um desses preceitos pode resultar em grande prejuízo coletivo quer seja pela construção de narrativas com intenções ou leituras equivocadas, ou por aspectos mais profundos como o não reconhecimento de direitos a populações excluídas. Para Gak *et al.* (2023) “o arquivo deve ser pensado não apenas pelo seu conteúdo, mas também pelo seu contexto, sua função e atuação na comunidade em que está inserido” (Gak *et al.*, 2021, p. 5). Esse mesmo cuidado da arquivologia deve ser empregado nas demais instituições de memória, entretanto, sem perder de vista a característica central destacada por Harvey (2016), que enseja uma descrição rica por metadados, facilidades de acesso e interoperabilidade.

É nesse espírito de fornecer aos curadores ferramentas e capacidades tecnológicas de gestão, e ao mesmo tempo, de permitir que as informações presentes nesses espaços sejam efetivamente acessadas, utilizadas e reutilizadas pela comunidade, que sistemas, cada um à sua maneira, contribuem para o patrimônio cultural. Os S.I. do Patrimônio Cultural têm interesse em gerir o domínio, a guarda, a inserção de objetos digitalizados ou nato-digitais, a interpretação, a pesquisa e exposição, bem como todas as atividades de gestão de acervos (Doerr, 2009; Liu; Lin, 2021). Suas ações podem ser classificadas de acordo com 4 funções gerais básicas: gestão de acervos, conservação, pesquisa e apresentação. As técnicas de digitalização do patrimônio, segundo Belhi *et al.*, (2018), estão relacionadas à gestão dos acervos e a curadoria digital, da mesma maneira que o enriquecimento de dados, o gerenciamento de conteúdo e a gestão da preservação (Belhi *et al.*, 2018).

O tratamento e o enriquecimento de dados podem ser abordados sobe vários aspectos, um dos quais, com a aplicação de técnicas de descrição de artefatos patrimoniais que utilizam taxonomias e outras formas de classificação, com utilização de vocabulários e padrões de metadados, normalmente controlados ou fornecidos pelas comunidades especializadas. A ligação (isto é, reutilização) com artefatos presentes em acervos externos não é muito encontrada e nem de fácil implementação (*Linked Open Data*). Este último sugere que informações presentes em um acervo são complementadas por informações presentes em outros, criando uma rede colaborativa que amplia o enriquecimento de dados. Nessa direção, os padrões CIDOC/CRM e EDM, relacionados a museus, *Records in Context/CM*,

relacionado a arquivos, com suporte de padrões de metadados utilizados, ampliam as oportunidades nos acervos digitais.

A riqueza dos padrões descritivos, como os arquivísticos e museológicos, possibilita aos curadores um manuseio de informações igualmente ricas e em diferentes idiomas, sugerindo, de forma adicional uma expectativa maior de ganhos informacionais através de ligação multiforme entre acervos. Expectativa esta que apresenta desafios técnicos e intelectuais na aproximação e harmonização de conceitos (Boeuf *et al.*, 2018). Há estratégias variadas para realizar tal feito. A utilização das tecnologias da Web Semântica e da iniciativa dos dados abertos conectados é uma das abordagens que têm sido utilizadas no domínio do patrimônio cultural e que tem no epicentro de seu arcabouço tecnológico as ontologias computacionais.

Ontologias são estruturas computacionais dotadas de mecanismos para modelar formalmente um domínio de conhecimento 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).

Como suporte às ontologias, há outras linguagens e ferramentas no que se conhece como Web Semântica, uma extensão da Web composta de linguagens e padrões que permitem que tais ontologias sejam utilizadas para os mais diversos fins. O *Resource Description Framework* (RDF), a *Ontology Web Language* (OWL) e a SPARQL, linguagem que permite manipular e recuperar grafos RDF são algumas das partes importantes desse arcabouço.

Em Filgueiras *et al.* (2023) é possível visualizar que as três abordagens mais comuns para integração baseada em ontologias são, em primeiro lugar, a construção de ontologia única, seguida da construção de múltiplas ontologias e da abordagem híbrida. Na primeira, as fontes de dados estão relacionadas a uma ontologia global que une a mesma visão de domínio (Abu Musa *et al.*, 2020; Stanković *et al.*, 2024). A abordagem de múltiplas ontologias requer que cada fonte de dados seja descrita localmente com sua própria ontologia (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). Embora a abordagem híbrida possa oferecer vantagens significativas, como a interoperabilidade entre ontologias locais e um vocabulário global compartilhado, sua

implementação pode ser custosa devido à necessidade de construir e manter tanto o vocabulário global quanto as ontologias locais associadas a este.

No domínio do patrimônio cultural há, da mesma maneira, um número expressivo de iniciativas que fornecem modelos e ontologias, para este trabalho são destacados os três já mencionados: CIDOC-CRM, no domínio dos museus, EDM, com uma visão mais aplicada aos museus e outros espaços europeus e o *Record in Context* (RiC/DM) no campo dos arquivos.

O modelo CIDOC-CRM pode ser visualizado, de uma maneira geral, através das funções suportadas pelos seus relacionamento que são: i) identificação de itens do mundo real através de nomes do mundo real; classificação de itens do mundo real; decomposição de artefatos materiais e imateriais, entidades temporais, grupo de pessoas, lugares e tempos; participação de itens persistentes em unidades temporais; localização de entidades temporais no espaço-tempo e coisas materiais no espaço; influência de objetos em atividades e produtos e vice-versa; e referências de objetos de informação a qualquer item do mundo real (Boeuf *et al.*, 2018).

A construção dos termos do CIDOC-CRM segue uma terminologia relativamente conhecida para quem transita nas áreas das ciências da informação e computação, em especial quem conhece o modelo RDF e outros padrões de metadados.

As principais classes do CIDOC-CRM incluem a *E1 CRM Entity*, que é a entidade mais genérica no CIDOC-CRM. Tem-se ainda *E2 Temporal Entity* que representa qualquer entidade com uma extensão temporal, *E3 Condition State* que descreve o estado de algo durante um período de tempo, *E4 Period* que representa um intervalo de tempo, *E5 Event* representa ocorrências nas entidades físicas ou conceituais, *E7 Activity* que representa atividades humanas intencionais, *E18 Physical Thing* que representa qualquer objeto físico, *E19 Physical Object* refere-se a qualquer objeto físico fixo e *E21 Person* refere-se a seres humanos. Há ainda *E24 Physical Human-Made Thing* que inclui objetos físicos feitos por humanos, *E28 Conceptual Object* que se refere a objetos imateriais como textos e conceitos, *E31 Document* que inclui registros documentais, *E52 Time-Span* um intervalo de tempo e a *E53 Place* que se refere a um local definido.

O modelo EDM (Europeana Data Model), que utiliza em grande parte elementos do CIDOC-CRM, define um conjunto de classes e propriedades para descrever diferentes aspectos do patrimônio cultural. As principais classes incluem *E2:Agent*: Uma pessoa ou organização

que desempenha um papel na criação, posse ou uso de um objeto patrimonial; *E2:Place*: Um lugar físico ou virtual associado a um objeto patrimonial; *E2:Event*: Um evento que está relacionado a um objeto patrimonial, como sua criação, exibição ou destruição; *E2:Concept*: Um conceito ou ideia que está relacionado a um objeto patrimonial.; *E2:Manifestation*: Uma representação física ou digital de um objeto patrimonial, como um livro, uma pintura ou uma gravação; *E2:PhysicalThing*: Um objeto físico que é parte do patrimônio cultural, como uma obra de arte, um artefato ou um monumento.

A classe *NonInformation Resource* se refere a todos os recursos que não são de natureza da informação como pessoas, lugares e objetos tangíveis. Todos esses recursos estão associados a eventos (Event), agentes (Agent), lugares (Places), intervalo de tempo (Time Span) que se referem a conceitos, definidos pela classe externa *skos:Concept*. Estes conceitos são utilizados para contextualização dos recursos a partir de ideias ou significados específicos de acordo a descrição de sua estrutura conceitual dentro um sistema de organização do conhecimento.

Já o modelo *Record In Context*, (RiC/DM), o mais recente dos três, segundo seu manual de referência (ICA/EGAD, 2023), define um conjunto de entidades principais que representam os conceitos fundamentais do mundo arquivístico em torno de Entidades Centrais, Entidades de Suporte e Entidades Conceituais. Segundo o manual, esse conjunto de entidades e propriedades “permite uma descrição mais completa, embora sempre incompleta, dos contextos em que os documentos surgem e existem, a fim de permitir múltiplas perspectivas e múltiplas vias de acesso” (ICA/EGAD, 2023, p. 7).

As entidades se relacionam em uma hierarquia conceitual e extensível que tem como raiz entidade *Thing* que representa todas as coisas (*RiC-E01 Thing*). Na hierarquia estão as coisas nas quais curadores e arquivistas se concentram para cumprir suas responsabilidades de preservar e fornecer acesso aos registros. A entidade *RiC-E02 Record Resource* corresponde ao conceito de Recurso e engloba três tipos distintos de registros arquivísticos, em um terceiro nível hierárquico, quais sejam: *RiC-E01 Record*, uma representação documentada de um evento ou atividade, *RiC-E03 Record Set*: um grupo de registros reunidos por um agente com base em critérios específicos, como temática, função ou período de tempo e *RiC-E04 Record Part*: um componente de um registro, como uma página de um documento, um quadro de

uma pintura ou um trecho de um vídeo. *RiC-E06 Instantiation* representa a informação inscrita por um agente *RiC-E07 Agent*, que desempenha um papel crucial na geração e modelagem da paisagem documental representa as entidades responsáveis os associadas a um registros, e pode ser um indivíduo, organização ou sistema.

A classe *RiC-E08 Person* e *RiC-E09 Group* representam pessoas e têm *RiC-E10 Family*, *RiC-E11 Corporate Body*, *RiC-E12 Position* e *RiC-E13 Mechanism* em hierarquia inferior. *RiC-E14 Event* representa os acontecimentos. A entidade *RiC-E15 Activity* é uma especialização de evento em um nível hierárquico imediatamente menor. As outras entidades importantes são *RiC-E16 Rule* representa as leis, manuais, etc e tem como entidade de nível inferior *RiC-E17 Mandate*, que confere a responsabilidade ou autoridade a um agente em uma atividade. *RiC-E18 Date* representa as datas ou eventos cronológicos (uma data específica, um século, um período) e *RiC-E22 Place*, que representa os lugares geográficos. Uma jurisdição, um rio, uma estrutura, uma cidade etc.

4 MODELO CONCEITUAL PARA CENTROS DE MEMÓRIA

Em edição anterior do Enancib apresentamos um modelo base que parte de uma arquitetura genérica de funcionamento de um centro de memória conforme Figura 1. Esse modelo prevê cinco 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. Na nova versão do modelo ontológico, alguns elementos forma modificados em termos de vocabulário. Primeiro optamos por adotar o termo “Lugar” ao invés de espaço, e o segundo, o termo “Evento” ao invés de fenômeno. Essa mudança sutil tem a intenção de evitar algumas confusões com a semântica das palavras envolvidas e estar mais orientada ao que se utiliza com vocabulários na Web.

Figura 1 – Modelo em 4 dimensões



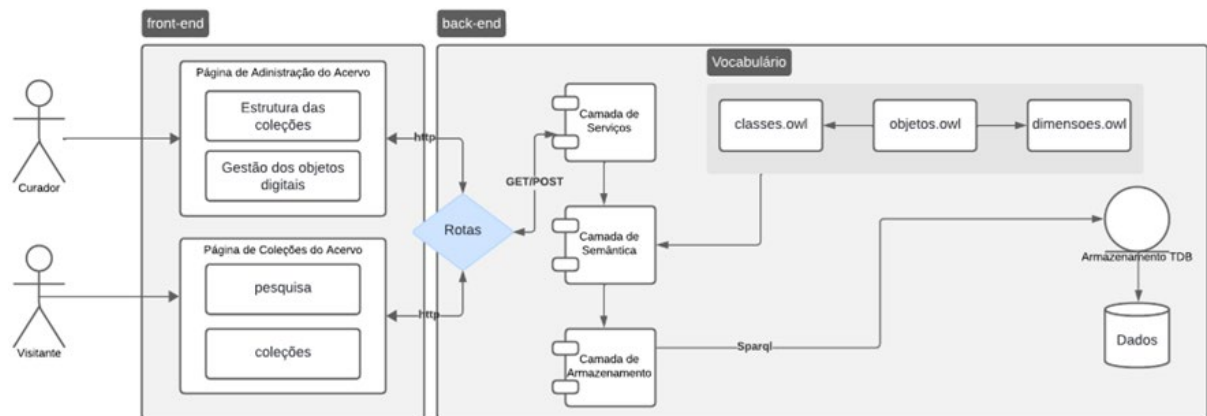
Fonte: *adaptado de Filgueiras et al. (2023).*

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 espaço de memória, isto é, pode se tratar de um artefato que é ao mesmo tempo patrimônio arquivístico documental, museal e imagético-sonoro. Os eixos informacionais estão intrinsecamente associados a quatro dimensões conceituais: Pessoa, Evento (Fenômeno), Tempo e Lugar (Espaço), e nesse caso, um mesmo objeto dimensional não pode assumir duas ou mais dimensões. Isto quer dizer, que uma Pessoa não pode ser um Evento, ou que um Tempo não pode ser um Lugar.

O protótipo que sustenta o modelo, evoluiu em vários aspectos, mas mantém uma arquitetura modular dividida, inicialmente, em *front-end* e *back-end*, sendo que o cerne da organização, armazenamento e tratamentos semânticos ocorre nessa última, enquanto a interação com os usuários é tratada na primeira. A camada servidor (*back-end*) fornece os serviços que são consumidos pela camada externa (isto é, cliente), que podem ser páginas da Web ou aplicativos específicos que acionam tais serviços. Em nível de funcionalidade há 3 camadas internas: camada de serviços, camada de armazenamento e camada semântica. A camada de serviços possui as funcionalidades que são acionados nas rotas correspondentes. Nessa camada, utilizando das ontologias presentes na camada semântica (Ontologia de definição de classes, de objeto e de dimensões), são processadas as requisições e construídas as respostas aos serviços com base nos bancos de dados em tripla (TDB) presentes no RDF

Store da camada de armazenamento. A camada de armazenamento realiza as inserções e alterações de grafo com base em chamadas SPARQL. A Figura 1 ilustra a arquitetura genérica do protótipo Guará.

Figura 2 – Arquitetura geral do protótipo Guará



Fonte: elaborada pelos Autores

O trânsito de informações é realizado por requisições e respostas *http*. Os serviços são acionados por *API Restful* com os métodos Http (*GET, POST, PUT, PATCH, DELETE*). A abordagem utilizada previu criar serviços específicos relacionados às classes e objetos

Essa arquitetura, apesar de simples, atende ao modelo proposto, permitindo que os curadores acessem os repositórios de cada espaço e realizem as atividades de gestão dos acervos. As APIs, bem como os demais artefatos do *back-end* foram construídos em Python, como parte da evolução do protótipo anterior desenvolvida em Java com *Apache Jena* (Filgueiras; Gouveia; Chagas, 2023). Para interação com o usuário, os artefatos em *Javascript* foram desenvolvidos utilizando a tecnologia *Vue* com o framework *Quasar*, que amplia as possibilidades de uso por meio de aplicativos móveis e com interações mais adaptadas à navegação Web moderna.

4.1 Abordagem em micro-ontologias e armazenamento RDF

O protótipo não utiliza instalação de bancos de dados relacionais. A camada semântica da proposta é composta de três micro-ontologias principais integradas. A primeira para classificação do acervo, a segunda para definição dos objetos e a terceira para o dimensionamento. Todo o armazenamento das micro-ontologias e, conseqüentemente dos

acervos dos espaços, se dá por meio de TDB/RDF. A opção por micro-ontologias separadas se dá pelo princípio da separação de responsabilidades, emprestado da engenharia de software.

Entendemos que a abordagem em micro-ontologias permite que os esforços sejam concentrados em áreas específicas do domínio de interesse, o que traz facilidades para o desenvolvimento, a compreensão e a manutenção das ontologias. À medida que os modelos amadurecem através das experiências absorvidas, em se ganham novos tipos de objetos digitais, as ontologias podem se estender gradualmente, adicionando novas classes, propriedades e relações conforme necessário.

A primeira ontologia, denominada *classes.owl*, possui três propriedades principais: o rótulo, que é determinado por *rdfs:label*, serve para definir um nome para cada elemento da classificação do acervo, como uma coleção ou um tipo. Cada classe recebe um identificador que será utilizado como URI (por exemplo, #CartasPessoais), porém o nome, que é definido no rótulo, pode sugerir que o nome seja “Cartas Pessoais”. A segunda propriedade, também definida no espaço de nomes do RDFS (*rdfs:comment*) é utilizada para acrescentar uma descrição mais elaborada, porém não muito extensa da classe, como por exemplo: “Representa o conjunto de cartas pessoais enviadas e recebidas por um determinado agente”. A terceira propriedade, *rdfs:subClassOf*, é um importante marcador para hierarquizar as divisões do acervo e amplamente utilizada nos modelos RDF/RDF-S. A segunda micro-ontologia, denominada *Objetos.owl*, é utilizada para definição dos objetos do acervo.

A ideia é que para objetos digitais sejam adicionados e mantidos, é necessário que seja descrito através das quatro propriedades básicas (mínimas) a serem informadas são *rdf:type*, *dc:title*, *dc:subject* e *dc:description* com pelo menos uma instância. A propriedade adicional *:id* é relacionada à referência criada automaticamente (URI) do objeto representa, e a propriedade *schema:contentUrl* permite que sejam adicionados objetos de mídia em qualquer objeto físico ou dimensional. Outras propriedades para descrição dos objetos digitais estão presentes nessa ontologia. Duas outras são centrais na classificação e semântica dos objetos.

A propriedade *dim:ObjetoFisico* é utilizada para tipificação do objeto quanto à mídia ou artefato a que este objeto representa dentro de uma visão de objetos de um acervo. Para o caso de objetos dimensionais, *dim:ObjetoDimensional* é uma propriedade específica para esse grupo de objetos, e para quais das quatro dimensões os objetos se direcionam (relacionam). No sentido de classificação, as dimensões são únicas para cada objeto, enquanto ao tipo físico, são múltiplas.

A terceira micro-ontologia (*Dimensoes.owl*), agrega à definição de objetos e de classes dos acervos, 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 que seja possível aplicar outras regras sobre os objetos digitais e sua tipificação quanto às características físicas ou dimensionais. Suas principais propriedades são *:relacionaPessoa* que permite associar qualquer coisa a um Objeto Dimensional Pessoa, *:relacionaEvento*, que explicita uma relação de algo a um Objeto Dimensional Evento, *:relacionaLugar*, propriedade que associa uma coisa a um espaço ou lugar, *:relacionaTempo*, para que se possa associar uma coisa a um Objeto Dimensional Tempo. E as propriedades *:tipoFisico* para classificar um Objeto Digital que envolve os artefatos físicos dos acervos e *:tipoDimensional* para objetos dimensionais, que permite que estes sejam classificados em uma das quatro dimensões mencionadas.

Nessa abordagem, o objeto digital pode assumir uma forma diversa. Fatos, eventos, pessoas, fotografias, um determinado tempo ou um lugar podem estar associados em vários contextos, de acordo com as quatro dimensões apresentadas e o tipos materiais encontrados. Essa estrutura fornece classes e propriedades para que os objetos possam ser encontrados em nas 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. O quadro abaixo apresenta um comparativo simples com a representação aproximada das principais entidades do modelo proposto em relação aos modelos apresentados na seção anterior.

Quadro 1 – Comparativo de representação dos principais elementos

Elemento		EDM	CIDOC-CRM	RiC-CM
Pessoa	:Pessoa	edm:Agent	crm:E39 Person	ric:E14 Agent
Evento	:Evento	edm:ProvidedCHO	crm:E13 Event	ric:E14 Event
Lugar	:Lugar	edm:Place	crm:E53 Place	ric:E53 Place
Tempo	:Tempo	edm:TimeSpan	crm:E52 Time-Span	ric:E52 Time-Span
Objeto	:objetoDigital	edm:ProvidedCHO	crm:E70 Document	ric:E13 Record

Fonte: elaborado pelos autores.

Com a utilização do modelo ontológico, a colaboração entre acervos distintos é possível através de relações entre os elementos presentes nas ontologias que são persistidos nos bancos de dados em tripla (TTB). Isto é, cada espaço, com base nas definições gerais, tem suas ontologias armazenadas no repositório. 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. Uma lista de repositórios amigos e, bem como o controle de acesso e os dados do espaço de memória também estão presentes no TDB.

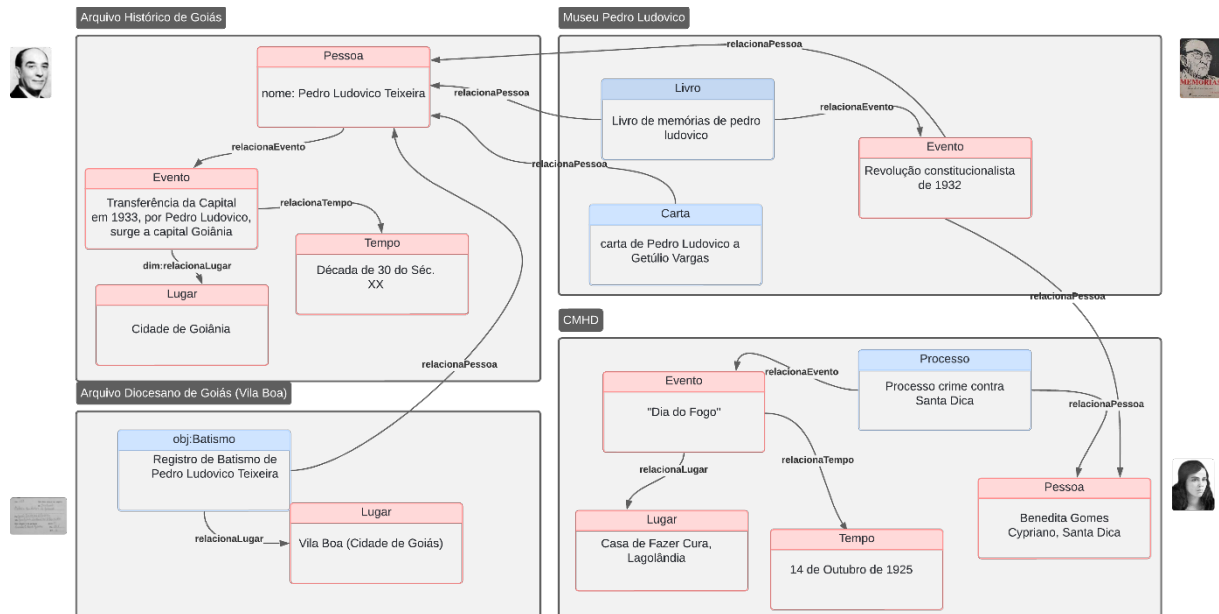
O processo de colaboração é disparado entre os repositórios no momento em são adicionados ou pesquisados itens dos acervos. Um algoritmo simples de interoperabilidade realiza consultas SPARQL quando *eventos* são acionados em cada repositório. Por exemplo, se um Curador digital insere um novo objeto ou uma nova propriedade em um objeto já existente, o algoritmo realiza uma busca nos *repositórios amigos* para encontrar informações e sugere essas novas possibilidades, as quais o curador pode ou não adicionar ao objeto em destaque.

5 EXPERIÊNCIA COM OS ARQUIVOS HISTÓRICOS E CENTROS DE DOCUMENTAÇÃO

Nossa experiência foi realizada com artefatos de acervo de quatro espaços de memória 1) Arquivo Histórico de Goiás 2) Museu Pedro Ludovico Teixeira; 3) Arquivo Diocesano da Cidade Goiás e 4) Centros de Documentação da UEG de Morrinhos (CMHD). Figura 3 representa, simbolicamente, o processo de interoperação entre espaços de memória através de ilustração do grafo que engloba esses elementos.

O objeto dimensional Pedro Ludovico Teixeira, do tipo *dim:Pessoa* está presente no Arquivo Histórico de Goiás e tem informações complementares em outros dois espaços. No Museu Pedro Ludovico há dois objetos físicos: Livro de Memórias de Pedro Ludovico e uma Carta enviada por Pedro Ludovico a Getúlio Vargas. Já Arquivo Diocesano de Goiás há um objeto físico referente a seu Registro de Batismo na Cidade de Goiás (então Vila Boa).

Figura 3 – Exemplo de Grafo de Interoperabilidade



Fonte: elaborada pelos autores.

No exemplo há ainda objetos dimensionais *dim:Evento* (o evento histórico referente à transferência da capital Vila Boa para Goiânia), pertencente ao primeiro espaço, e o evento histórico denominado “Dia do Fogo”, referente ao dia do cerco policial à Benedita Gomes Cypriano, a Santa Dica, cujo processo crime de outubro de 1925, em forma de objeto físico, bem como outras informações encontram-se no CDHM. Outro evento histórico é simbolizado referente à Revolução Constitucionalista de 1932 e envolve as duas personagens, Pedro Ludovico Teixeira, e o objeto dimensional *dim:Pessoa* referente à Santa Dica.

6 CONSIDERAÇÕES FINAIS

Este trabalho buscou apresentar os resultados, ainda parciais, do projeto de pesquisa “Arquivos que conectam”, em andamento na Universidade Estadual de Goiás (PPG Estudos Culturais, Memória e Patrimônio), especificamente avanços em torno do protótipo Guará, que possibilita a colaboração tecnológica e informacional entre espaços de memória e, à sua maneira, contribui para a preservação, acesso e reuso do patrimônio cultural, isto é, ações de curadoria.

Entendemos que a digitalização do patrimônio cultural é urgente e necessária, e discutimos, na medida do possível, questões tecnológicas relacionadas à urgência e necessidade em um recorte específico. Com o protótipo buscamos reafirmar que é possível

fornece uma ampla variedade de serviços e conectá-los por meio das tecnologias e linguagens da Web contemporânea, como as APIs *Restful*, e outras não tão exploradas como as ontologias, grafos e armazenamento RDF. Nossa abordagem com micro-ontologias, sustentamos, sugere algumas vantagens, como agilidade, pois permite adaptação a necessidades emergentes do domínio de forma fácil e incremental, flexibilidade, já que permite evoluir ontologias total ou parcialmente com base em novos conhecimentos e requisitos mantendo a capacidade de resposta às mudanças e, sobretudo, a gestão simplificada, já que permite iniciar do micro pro macro, e assim compreender o escopo de cada parte do domínio, o que facilita a manutenção e a evolução ao longo do tempo.

REFERÊNCIAS

ABBY CLOBRIDGE. The digital curation toolkit: strategies for adding value to word-related social systems. *In*: [S. l.]: O'Reilly, 2013.

ABU MUSA, T. H. *et al.* OntoM: an ontological approach for automatic classification. *In*: IEEE International Conference on Informatics, IoT, and Enabling Technologies, Doha, Catar. **2020 IEEE [...]**. [S. l.]: IEEE, 2020. p. 329–334.

BANNOUR, I. *et al.* CRMCR - a CIDOC-CRM extension for supporting semantic interoperability in the conservation and restoration domain. *In*: Digital Heritage International Congress, 3.; International Conference on Virtual Systems & Multimedia, São Francisco, CA, USA, 2018. **2018 3rd digital [...]**. [S. l.]: IEEE, 2018. p. 1–8.

BELHI, Abdelhak *et al.* Digitization and preservation of cultural heritage products. *In*: JOSE RIOS (org.). **Product lifecycle management and the industry of the future**. [S. l.]: Springer International Publishing, 2018. p. 241–253.

BOEUF, Patrick Le *et al.* **Definition of the CIDOC conceptual reference model**. [S. l.: s. n.], 2018.

DOERR, Martin. Ontologies for cultural heritage. *In*: STAAF, Steffen; STUDER, Rudi (org.). **Handbook on ontologies**. 2nd. ed. New York, NY, USA: Springer, 2009. p. 463–486.

DRAKOPOULOS, G. *et al.* Towards a framework for tensor ontologies over Neo4j: Representations and operations. *In*: INTERNATIONAL CONFERENCE ON INFORMATION, INTELLIGENCE, SYSTEMS & APPLICATIONS, 8., 2017, Larnaca, Chipre. **2017 8th international [...]**. [S. l.]: IEEE, 2018. p. 1–6.

FERMANN, Angélica Cintra; LIMA, Assahi Pereira; CHARDULO, Suzana Mayumi Iha. Gestão da Informação: desafios e competências para preservação digital. **Prisma.com**, Porto, n. 44, p. 123–144, 2020.

FILGUEIRAS, Alison Carlos; GOUVEIA, Feliz Ribeiro; CHAGAS, Paula Roberta. Uma proposta de sistema de informação aberto e conectado para acervos digitais. *In: Encontro Nacional de Pós-Graduação em Ciência da Informação*, 23., 2023, São Cristóvão, SE. **Anais [...]**. São Cristóvão: Enancib, 2023.

FREIRE, Nuno *et al.* Aggregation of cultural heritage datasets through the Web of Data. **Procedia Computer Science**, [s. l.], v. 137, p. 120–126, 2018. Trabalho apresentado na 14th International Conference on Semantic Systems, entre 10 e 13 set. 2018, Viena, Áustria. Disponível em: <https://doi.org/10.1016/j.procs.2018.09.012>.

FURTADO, Renata Lira; SANTOS, Maria de Nazaré Coelho dos; SANTOS, Felipe Cesár Almeida dos. Precisamos falar sobre os fenômenos informacionais contemporâneos no contexto arquivístico: um mapeamento da produção bibliográfica sobre pós-verdade, desinformação e fake news. **Inf. Pauta**, Fortaleza, v. 7, p. 1–28, 2022.

GAK, Igor *et al.* Arquivos e decolonialidade: breves considerações acerca de uma abordagem necessária em pesquisa e extensão. **Raízes e rumos**, Rio de Janeiro, v. 9, n. 2, p. 104–113 2021.

GUARINO, Nicola; OBERLE, Daniel; STAAB, Steffen. What is an Ontology?. *In: STAAB, Steffen; STUDER, Rudi (org.). International handbooks on information systems*. 2. ed. [S. l.: s. n.], 2009. p. 1–17.

HIGGINS, Sarah. Digital curation: the development of a discipline within information science. **Journal of Documentation**, [s. l.], v. 74, n. 6, p. 1318–1338, 2018.

ICA/EGAD. **Records in contexts conceptual model**. [S. l.: s. n.], 2023. Disponível em: <https://github.com/ICA-EGAD/RiC-CM/issues>.

LIU, Ying; LIN, Hui Wen. Construction of interpretation and presentation system of cultural heritage site: an analysis of the old city, Zuoying. **Heritage**, [s. l.], v. 4, n. 1, p. 316–332, 2021.

NIANG, Cheikh *et al.* Ontology-based data integration system for conservation-restoration data (OBDIS-CR). **Proceedings of the 20th International Database Engineering & Applications Symposium**. New York, NY, USA: Association for Computing Machinery, 2016. p. 218–223. Realizado em Montreal, Canadá. Disponível em: <https://doi.org/10.1145/2938503.2938545>.

STANKOVIĆ, Ranka *et al.* **Towards semantic interoperability**: parallel corpora as linked data incorporating named entity linking. [S. l.: s. n.], 2024. Disponível em: <https://www.w3.org/TR/>.