

CLASSIFICAÇÃO ARQUIVÍSTICA MULTIFACETADA ORIENTADA À INTELIGÊNCIA ARTIFICIAL

JULIANNE TEIXEIRA E SILVA*
JOSEMAR HENRIQUE DE MELO**
ROSA ZULEIDE LIMA DE BRITO***

Resumo: O presente estudo, desenvolvido pelo Grupo de Pesquisa EADRD, propõe uma Classificação Arquivística Multifacetada (CAM) como alternativa ao modelo de classificação funcional. O objetivo geral da pesquisa é analisar a sistemática da CAM, aplicada às práticas de gestão e orientada à Inteligência Artificial (IA). A CAM integra elementos funcionais, orgânico-estruturais e temporais, considerando o contexto histórico, a missão institucional e a estrutura organizacional, buscando representar a organicidade dos documentos desde sua ontogênese. A metodologia da investigação é de natureza teórica e aplicada, com abordagem qualitativa. A CAM é fundamentada na Ontogênese Documental, na Teoria Geral dos Sistemas de Bertalanffy e no Pensamento Complexo de Edgar Morin. Adota a classificação facetada de Ranganathan, utilizando sete facetas inter-relacionadas, incluindo a Configuração Histórico Contextual do Todo e da Parte, e a Vinculação Institucional, além das mais conhecidas orgânicamente-estruturais e funcionais. Sua estrutura é baseada em dados, informações, metadados e paradados, utilizando o framework DAMA-DMBOK, visando garantir a confiabilidade e acurácia dos dados para o uso responsável e ético da IA. Testes preliminares, desenvolvidos em Python e Django, indicam resultados promissores.

Palavras-chave: Arquivística; Classificação; Documento digital; Inteligência Artificial.

Abstract: This study, developed by the EADRD Research Group, proposes a Multifaceted Archival Classification (CAM) as an alternative to the functional classification model. The general objective of the research is to analyze the CAM's systematic approach, applied to management practices and oriented towards Artificial Intelligence (AI). The CAM integrates functional, organic-structural, and temporal elements, considering the historical context, institutional mission, and organizational structure, aiming to represent the organic nature of documents from their ontogenesis. The methodology of the investigation is of a theoretical and applied nature, with a qualitative approach. The CAM is founded on Documental Ontogenesis, Bertalanffy's General Systems Theory, and Edgar Morin's Complex Thought. It adopts Ranganathan's faceted classification, utilizing seven interrelated facets, including the Historical Contextual Configuration of the Whole and the part and the Institutional Linkage, in addition to the more widely known organic-structural and functional ones. Its structure is based on data, information, metadata, and paradigm, using the DAMA-DMBOK framework, aiming to ensure data reliability and accuracy for the responsible and ethical use of AI. Preliminary tests, developed in Python and Django, indicate promising results.

Keywords: Archival Science; Classification; Digital document; Artificial Intelligence.

* Universidade Federal da Paraíba (UFPB). Email: julianne.teixeira@gmail.com. ORCID: <https://orcid.org/0000-0002-4959-0844>.

** Universidade Estadual da Paraíba (UFPB). Email: josemarhenrique@gmail.com. ORCID: <https://orcid.org/0000-0002-8586-518X>.

*** Universidade Federal da Paraíba (UFPB). Email: rosazuleide@hotmail.com. ORCID: <https://orcid.org/0009-0004-0280-1597>.

INTRODUÇÃO

A investigação aqui apresentada foi motivada a partir dos trabalhos e discussões do Grupo de Pesquisa e Estudos Arquivísticos em Documentos e Registros Digitais (EADRD), cujo uma das linhas de pesquisa tem se dedicado à classificação arquivística aplicada em sistemas de negócios e SIGAD. Por meio de revisão de literatura, identificou-se, na Arquivologia, um conjunto de trabalhos que vêm sendo desenvolvidos e debatidos a respeito das classificações funcionais. Desta forma, o trabalho ora em tela é fruto dos primeiros resultados dessas pesquisas realizadas ao longo de quatro anos do referido Grupo da Universidade Federal da Paraíba. Trata-se de proposta de classificação arquivística multifacetada voltada ao contexto de produção de documentos arquivísticos em ambientes digitais ou sistemas com a perspectiva de que seja facilitada por meio de modelos ou agentes de Inteligência Artificial (IA).

De acordo com Orr (2005), Bak (2012), Foscarini (2012), Duranti e Franks (eds., 2015), Packalen e Henttonen (2016), Mokhtar e Yusof (2017) e Sousa (2022), a classificação funcional tem se mostrado pouco eficiente frente à complexidade de representar o vínculo arquivístico dos documentos, principalmente em se tratando de sistemas informáticos produtores de documentos arquivísticos. Sob esse aspecto é comum às instituições elaborarem classificações arquivísticas funcionais por estas serem menos mutáveis que as orgânicas, contudo as mudanças, ao longo do tempo, irão exercer algum tipo de interferência até mesmo nas funções, o que afeta o contexto de produção dos documentos.

Além das questões relacionadas à funcionalidade e/ou organicidade no âmbito da gestão dos documentos arquivísticos digitais, outro questionamento preponderante surgiu nas discussões do grupo de pesquisa. Percebeu-se que as classificações arquivísticas apresentam um inquietante problema de anacronismo.

Deste modo, identificou-se que o fator «tempo» tem um papel crucial na elaboração das classificações arquivísticas. Desconsiderar «o tempo» do contexto da criação dos documentos nos leva a cometer anacronismos preocupantes ao organizá-los, principalmente quando estamos lidando com documentos de diversas épocas e em estruturas orgânicas distintas.

Os argumentos pontuados acima causam reflexos em várias direções, em que uma delas diz respeito à organicidade, dado que o uso da classificação funcional, usualmente, negligencia a estrutura organizacional. Com isso desconsideram-se também as transformações orgânicas que as instituições passam no decorrer do tempo. Tal fenômeno ocorre de maneira natural, uma vez que as instituições fazem parte do tecido social e, por mais que os gestores planejem, apliquem métodos e técnicas inovadoras na intenção de manterem suas corporações sob controle, é preciso considerar as coisas passadas, a impermanência do momento presente e as incertezas do devir, pois essas condições que afetam a conformação de suas estruturas orgânico-funcionais.

Neste sentido, começamos a perceber que a classificação funcional, tão propagada e defendida na literatura e normas, como uma forma de solucionar os problemas relacionados às mudanças organizacionais, tornou-se um impasse no afã de estabelecer cartesianamente a racionalização, a ordem e o controle da gestão dos documentos. Desse modo, geralmente, negligencia-se a estrutura orgânica por sua impermanência, desconsiderando a sua complexidade que aponta elementos necessários para entender a dinâmica das instituições.

O tempo, entendido aqui como uma construção social, pode ser explorado como outro ponto referencial na gestão de documentos arquivísticos, quando se quer identificar algo em um determinado momento cronológico. Vejamos o exemplo das imagens em movimento. Quando precisamos indexar ou encontrar uma cena num determinado vídeo, precisamos mapear todo o vídeo e registrar as marcações a partir do *timer* em que as «cenas indexáveis» ocorrem. Marcamos o tempo das cenas e suas respectivas descrições. De certa forma, essa analogia nos ajuda a compreender o «tempo» como um elemento marcador, referencial e indispensável ao estabelecimento da organicidade dos documentos, sendo necessária sua representação na classificação arquivística como meio imprescindível à organização e recuperação da informação.

Essas discussões nos levaram a um questionamento. Como considerar, numa classificação arquivística, os elementos prementes ao contexto de produção dos documentos nato digitais sendo eles: funcional, orgânico estrutural e temporal? Essa problemática tem nos orientado na elaboração da Classificação Arquivística Multifacetada (CAM).

A formulação dessa pergunta acionou pesquisadores que estavam dedicados à outra linha de pesquisa do Grupo EADRD sobre *Documentos nato digitais e as complexidades contextuais da gênese documental*. Este segundo grupo de pesquisadores estava dedicado às investigações sobre a gênese de documentos e os requisitos arquivísticos em sistemas de negócios. Esta pesquisa, em especial, propiciou o desenvolvimento da noção de *Ontogênese Documental*, a qual trouxe maior clareza e condições teórico-metodológicas para elucidar o problema de pesquisa, anteriormente explicado, a respeito da classificação arquivística.

A abordagem da Ontogênese Documental está sendo desenvolvida pelo grupo de pesquisa EADRD e é fundamentada na metodologia da identificação arquivística proposta por Bellotto (2014) e Rodrigues (2016) e nos estudos e conceitos do projeto InterPARES (Jansen e Duranti, 2013) e orientada à governança de dados. A referida abordagem defende que os documentos nato digitais, produzidos em sistemas de negócios, devem ser gerenciados antes mesmo de serem criados. Esta perspectiva está vinculada ao processo de preparação de uma infraestrutura de requisitos, metadados, dados, paradados e informações determinantes à elaboração de documentos arquivísticos melhor consubstanciados, com a intenção de propiciar diretrizes teóricas e

boas práticas quanto ao gerenciamento de documentos nativos digitais e digitalizados, no intuito de lhes conferir confiabilidade, integridade e presunção de autenticidade desde sua gênese. Isto suscitou-nos a possibilidade de também trabalhar, no âmbito da ontogênese documental, com o elemento «tempo» no contexto de criação e classificação dos documentos nato digitais.

A estrutura da CAM baseada em dados, metadados e paradados entre outras instâncias de informação que tratam do contexto de produção, em sua interação com diversos aspectos da Arquivologia, principalmente com as demandas pautadas pelas novas tecnologias digitais, nos levou ao seguinte questionamento para a construção deste texto: de que maneira a sistemática da classificação arquivística multifacetada pode corroborar o uso da Inteligência Artificial em práticas de gestão arquivística?

Desta forma, o objetivo geral de nossa investigação é: analisar a sistemática da classificação arquivística multifacetada, aplicada às práticas de gestão, orientada à Inteligência Artificial. Isto posto, nos aponta para os seguintes objetivos específicos: (a) Descrever os elementos que compõem a classificação multifacetada; (b) Identificar os pontos críticos da gestão de documentos digitais em que a CAM pode ser aplicada e (c) Identificar os meios para o uso da CAM por modelos de IA.

1. METODOLOGIA

Encontrar um percurso teórico-metodológico para pensar uma classificação arquivística que englobasse os elementos funcionais, orgânico-estruturais e temporais foi desafiador. Inicialmente pensou-se em trabalhar com a *Grounded Theory* dos sociólogos Barney G. Glaser e Anselm L. Strauss (1998), contudo, além da pesquisa observacional e teórica, havia a necessidade de testar a aplicabilidade da classificação às realidades arquivísticas, o que demandaria prototipagem e testes por meio do desenvolvimento de bancos de dados e sistemas informatizados.

No sentido de compreender o tamanho da pesquisa e da diversidade de caminhos a seguir, a Teoria Geral dos Sistemas de Bertalanffy (2013) pavimentou trilhas as quais possibilitaram enxergar o problema de pesquisa de modo mais abrangente e inter-relacionado. Segundo o autor,

É necessário estudar não somente partes e processos isoladamente, mas também resolver os decisivos problemas encontrados na organização e na ordem que os unifica, resultante da interação dinâmica das partes, tornando o comportamento das partes diferente quando estudado isoladamente e quando tratado no todo (Bertalanffy 2013, p. 55).

Ao aprofundarmo-nos na teoria dos sistemas chegamos à epistemologia do pensamento complexo de Edgar Morin e nesse pensador encontramos não apenas o

aporte epistemológico, mas um norte para o desenvolvimento teórico que a pesquisa demanda, bem como a concepção de método para traçarmos o desenvolvimento da pesquisa como um todo e cada uma de suas partes. Para Morin (2023, p. 335):

A palavra método deve ser concebida fielmente em seu sentido original, e não em seu sentido derivado, degradado, na ciência clássica; com efeito, na perspectiva clássica, o método não é mais do que um corpus de receitas, de aplicações quase mecânicas, que visa a excluir o sujeito de seu exercício. O método degrada-se em técnica porque a teoria se tornou um programa. Pelo contrário, na perspectiva complexa, a teoria é engrama, e o método, para ser estabelecido precisa de estratégia, iniciativa, invenção, arte. Estabelece-se uma relação recorrente entre método e teoria. [...] Assim, a teoria não é o fim do conhecimento, mas um meio-fim inscrito em permanente recorrência (Morin 2023, p. 335).

Todos os elementos acima descritos nos aproximam também da perspectiva sistêmica para os arquivos e a arquivística, que foi desenvolvida pelos pesquisadores portugueses Armando Malheiro, Fernanda Ribeiro, Júlio Ramos e Manuel Real em sua obra inaugural, intitulada *Arquivística: teoria e prática de uma ciência da informação* (1999). Nesta obra, apresentaram a evolução histórica dos arquivos e da Arquivologia e traçaram, como lastro teórico para a área, a Teoria Sistêmica, postulando os arquivos como sistemas semifechados, caracterizados pela natureza orgânica e funcional em que está vinculada também a memória. A perspectiva sistêmica permite não só estudar as partes que compõem o todo como também resolver os decisivos problemas encontrados na organização e na ordem que unifica cada um dos elementos que se encontram em uma interação dinâmica, tendo em vista que «A condição sistêmica dos arquivos pressupõe não apenas combinações intrínsecas, mas também uma articulação com outros sistemas dos campos político-econômico, jurídico-administrativo, cultural etc.» (Silva et al. 1999, p. 214).

Os autores portugueses criticaram o princípio de proveniência «o arquivo não é uma mera soma de “fundos”. Ao proporem a teoria sistêmica entendem o arquivo como [...] uma unidade integral e aberta ao(s) contexto(s) dinâmico(s) e histórico(s)» (Silva et al. 1999, p. 213). Todos esses elementos se coadunam e integram na proposta, tanto da ontogênese quanto da classificação multifacetada, proposta pelo grupo de pesquisa EADR.

Somando-se a isso, o presente trabalho trata de investigação de natureza teórica e aplicada, de abordagem qualitativa, tendo como base epistemológica o Paradigma Complexo de Edgar Morin, em consonância com os métodos da classificação facetada (Ranganathan 1967) e dos fundamentos arquivísticos da organicidade (vínculo arquivístico) e da gênese documental.

A gênese do documento de arquivo, de acordo com Bellotto (2014, p. 329), «implica na relação entre a natureza da sua finalidade e a fórmula diplomática capaz de conferir autenticidade a ele e a confiabilidade ao seu conteúdo». Nessa mesma perspectiva, o InterPARES Project 2 (2021) apresenta, em seu glossário, o termo «*archival bond*», com a noção similar ao princípio da organicidade, como conhecemos no Brasil, traduzido como «vínculo arquivístico», o qual refere-se a rede de relacionamentos que cada documento possui com os demais documentos pertencentes a uma mesma agregação (tradução nossa).

Segundo Bellotto (2014, p. 329), um dos pilares da arquivística está na indissolubilidade entre «a informação, o meio documental no qual ela está vinculada, o suporte, a proveniência e, sobretudo, o vínculo entre os documentos do mesmo contexto genético».

Neste percurso, faltava-nos ainda um aporte para o desenvolvimento da classificação, tanto no seu embasamento teórico quanto na elaboração do instrumento em si, com seu sentido verbal e notacional. Situados na epistemologia da complexidade, o construto teórico e prático para a classificação precisaria ser menos rígido, menos dicotômico. Nas buscas, por meio de revisão de literatura de escopo (*scoping review*), chegamos à classificação facetada do matemático indiano Ranganathan, que nos apresentou a possibilidade de desenvolver uma classificação flexível e adaptável, bastante alinhada com nosso problema de pesquisa.

Ao se deparar com a proposta da Classificação Arquivística Multifacetada (CAM), podem ocorrer associações imediatas com algumas correntes teóricas, normas, modelos e boas práticas arquivísticas, já existentes, inclusive como algumas dimensões e eixos do *Records Continuum*, recomendações do projeto InterPARES, conformidades das normas ISO para gestão de documentos, modelos de padrões de metadados, *Records in Contexts*, dentre outras vertentes arquivísticas. Todo este complexo de perspectivas arquivísticas denominado como «multiverso» por Gilliland, McKemmish e Lau (eds., 2017), está, de alguma forma, contemplado na classificação ora apresentada. Cabe ressaltar que o detalhamento teórico-metodológico sobre a CAM, sinteticamente aqui apresentada, está sendo registrado em livro, que em breve será publicado.

2. CLASSIFICAÇÃO FACETADA E CATEGORIZAÇÃO

A conformação do tipo de classificação mais utilizada no campo das ciências que lidam com a organização da informação e do conhecimento é a do tipo hierárquica, a qual tem suas bases na lógica aristotélica e reforçada pela proposta da árvore dicotômica de Porfírio, em que fica patente a estrutura subordinativa e dualista do processo classificatório no sentido de que, se um ente se encaixa numa classe, não pode pertencer à outra.

Este é um quesito de apego imediato ao pensamento arquivístico que opera sob a perspectiva do vínculo institucional e da unicidade documental. A divisão e hierarquização como forma de classificar têm sido a base para os sistemas de classificação arquivísticos, sejam eles funcionais e/ou orgânicos estruturais. Em que pese as classificações hierárquicas do tipo gênero/espécie (do geral para o específico) comportam em seu ordenamento uma composição de níveis e classes, também utilizados na função arquivística de descrição por meio de suas normas e instrumentos de pesquisas. A Classificação Arquivística Multifacetada apresenta, por sua complexidade, a necessidade de operar, concomitantemente, o mecanismo classificatório partitivo (Todo – Parte) e o hierárquico (Gênero – Espécie).

O matemático e bibliotecário indiano Shiyali Ramamrita Ranganathan (1892-1972) elaborou uma teoria da classificação que, dentre outras características, destaca-se pelo agrupamento de termos estruturados com base na análise de um tema, coordenando conceitos, para identificação de suas facetas (categorias) a partir dos aspectos nele contidos. Segundo Barbosa (1972), a grande contribuição de Ranganathan foi a ideia de dividir os temas por facetas (categorias), isto é, reunir grupos de classes por um mesmo princípio de divisão (critérios).

Para Ranganathan (1967), uma faceta é um aspecto ou característica de um objeto ou conceito que é distinto e separável de outros aspectos ou características. Piedade (1983, p. 25) nos esclarece sobre o significado de faceta e nos direciona ao processo de construção de um esquema facetado. Para a autora, «a palavra faceta é empregada indiretamente como sentido de característica, porque a base do agrupamento em classes resulta da divisão por determinada qualidade ou característica e, muitas vezes, o nome desta qualidade é usado para designar a faceta».

A classificação facetada criada por Ranganathan é composta por cinco categorias principais, conhecidas como PMEST – Personalidade, Matéria, Energia, Espaço e Tempo. Estas categorias principais referem-se às facetas mais amplas, que agrupam os objetos por suas características ou aspectos mais gerais, que, por sua vez, permitem subdividir, chegando ao nível mais específico possível, formando cadeias.

Diante do exposto, a investigação ora em curso busca apresentar um sistema de classificação arquivística com facetas inter-relacionadas que contemplam o contexto histórico; a missão institucional; a estrutura organizacional e funcional, representando a organicidade dos tipos documentais, o que proporciona uma gestão arquivística com base na ideia do vínculo arquivístico e o contexto de produção. De acordo com Sousa (2009, p. 95), delimitar o objeto a ser utilizado no processo classificatório coincide com o objeto da própria arquivística que, dependendo da perspectiva do classificador, pode ser o arquivo ou os documentos. Neste estudo, o objeto considerado é o tipo documental.

Após testes iniciados em duas instituições públicas brasileiras, a saber: Autarquia de Proteção e Defesa do Consumidor do Estado da Paraíba (PROCON-PB) e o Arquivo Público do Estado da Paraíba, cujos planos de classificação desenvolvidos foram baseados na estrutura multifacetada, proposta pelo grupo de pesquisa EADRD, foi possível verificar resultados parciais bastante substanciais, o que nos motivou a divulgar a referida proposta de classificação arquivística.

3. GOVERNANÇA ARQUIVÍSTICA DE DADOS E INTELIGÊNCIA ARTIFICIAL (IA)

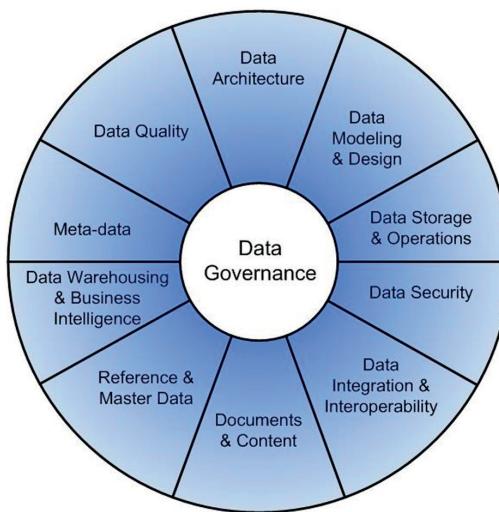
A formalização de uma classificação arquivística facetada exige a composição das facetas, por meio de dados estruturados, tendo como princípio a noção de ontogênese documental, notadamente os dados relacionados ao contexto de produção dos documentos arquivísticos.

Sob tal prospectiva, e com a preocupação da confiabilidade e qualidade dos dados, foi definido que haveria necessidade de se eleger um modelo (*framework*) para a estruturação dos dados. Nesse caso o modelo escolhido foi o DAMA-DMBOK (Dama International 2017), por ser de fácil acesso e por ser um modelo já reconhecido e bem difundido.

O DAMA-DMBOK é um guia com orientações sobre a governança e gestão de dados, um termo abrangente, que descreve os processos utilizados para planejar, especificar, habilitar, criar, adquirir, manter, usar, arquivar, recuperar, controlar e purgar (limpar/eliminar) dados, uma vez que os dados são considerados um ativo valioso para as instituições e devem ser gerenciados adequadamente.

O guia é organizado sob a orientação de 10 áreas do conhecimento que balizam o gerenciamento, a integração e a interoperabilidade de dados, inclusive uma dessas áreas é a gestão da documentação pautada pela norma ISO 15.489. As áreas são: Arquitetura de Dados, Modelagem e *Design* de Dados, Armazenamento e Operações de Dados, Segurança de Dados, Integração e Interoperabilidade de Dados, Documentos e Conteúdo, Dados de Referência e Mestre, *Data Warehousing* e *Business Intelligence*, Metadados e a Qualidade de Dados. Conforme representado na Figura 1. O *framework* detalha uma estrutura padronizada para a governança e gestão de dados, definindo áreas de conhecimento, processos, papéis e melhores práticas, as quais orientam as instituições a gerenciar seus dados como um ativo estratégico.

O que se pretende é desenvolver uma classificação arquivística, que não seja estritamente funcional, mas que atenda a perspectiva das diferentes esferas do contexto institucional, as quais precisam estar representadas numa classificação. Nesse sentido, é preciso considerar que se trata de um significativo volume de dados complexos sobre o contexto institucional, os quais precisam relacionar-se entre

**Fig. 1**

Roda das áreas do
DAMA-DMBOK

Fonte: Dama International 2017

si, representar fielmente o contexto de produção dos documentos arquivísticos e serem seguros e confiáveis.

O modelo DAMA-DMBOK não trata de modo específico dos paradados, contudo dedica-se com atenção redobrada à segurança e qualidade dos dados. Desse modo ao desenhar a Classificação Arquivística Multifacetada, embora os paradados não estejam explicitamente contemplados no *framework*, este recurso foi utilizado, pois naturalmente os paradados foram sendo necessários à medida que o *design* da CAM era modelado e prototipado.

A respeito dos paradados, Couper (2017) explica que se refere a dados auxiliares que descrevem um determinado processo. Este autor cunhou o termo em 1998 focado nos subprodutos automáticos de sistemas assistidos por computador. O propósito original de capturar esses dados era detectar e recuperar erros de sistema. Porém, rapidamente foi percebida sua utilidade para fornecer informações sobre o processo de coleta de dados em pesquisas científicas. Inicialmente, o enfoque estava na usabilidade dos instrumentos de pesquisa computadorizados, ajudando a identificar questões relacionadas em como as coletas de dados eram realizadas, documentando/registrando esse processo, no sentido de conferir evidência às pesquisas ou em outras demandas computacionais, distinguindo-os dos metadados, que descrevem os dados.

Corroborando com esse escopo, Feliciati e Duranti (2025) destacam a importância dos paradados sob a perspectiva da Inteligência Artificial. Segundo os autores, em ambientes arquivísticos a Inteligência Artificial deve ser utilizada de forma responsável. Destacam estes autores o papel essencial dos paradados na documentação e na garantia da responsabilidade e acurácia dos dados e informações.

Como fenômeno informacional, paradados diferem dos metadados tanto qualitativamente, pois visam demonstrar o contexto de produção e gestão em vez da identidade e integridade dos documentos, quanto quantitativamente, pois incluem informações mais extensas e menos detalhadas do que os metadados (Feliciati e Duranti 2025, p. 2, tradução livre).

A Inteligência Artificial (IA) apresenta inúmeras possibilidades de transformação digital no campo da arquivística e dos arquivos. Neste sentido, Dorneles (2024) apresenta revisão de literatura em bases bibliográficas internacionais e seus resultados apontam estudos e boas práticas utilizando-se a IA em acervos arquivísticos. Ao mesmo tempo o autor ressalta que, embora a IA apresenta potencialidades para soluções de problemas arquivísticos, como por exemplo o dilema das massas documentais acumuladas, as aplicações de IA devem ser estudadas pelos profissionais dedicados à Arquivística de forma meticulosa e ética para garantir conformidade e segurança no uso dessa ferramenta.

reafirma-se a necessidade de que, essas tecnologias sejam desenvolvidas e aperfeiçoadas de forma responsável e ética. Além de que, sejam estabelecidas redes de desenvolvimento científico para soluções de manutenção de registros arquivísticos auxiliados pelas tecnologias da IA que possam beneficiar o maior número de pessoas e instituições do mundo (Dorneles 2024, p. 102).

A representação do conhecimento e da informação é um dos pilares da Inteligência Artificial, pois permite que os sistemas de IA compreendam, processem e utilizem informações de maneira eficiente e eficaz. Davis, Shrobe e Szolovits (1993), assim como Norvig e Russell (2013), entendem que a representação do conhecimento é central porque agentes de IA necessitam que os objetos, o tempo e os fatos sobre o mundo estejam devidamente representados, por meio de sintaxes, fórmulas lógicas, matemáticas, dados, informações, notações, linguagens naturais e artificiais, categorias, ontologias, dentre outras formas de representação da informação e do conhecimento que possam ser estruturadas e adaptadas para aprendizagem de máquinas.

Essa representação de mundo passa pela gestão dos dados. Nesse panorama Feliciati e Duranti (2025) trazem uma distinção crucial, feita pelo projeto InterPARES Trust AI, que se dá entre a IA responsável (*accountable*) e a IA explicável (*explainable*). A IA explicável visa apenas tornar um instrumento comprehensível, ou seja, saber como um algoritmo ou ferramenta funciona, não atribui explicitamente responsabilidade. Em contrapartida, a IA responsável «deve documentar claramente por que, como, por quem e com que resultado um determinado sistema ou instrumento foi usado em um determinado contexto» (Feliciati e Duranti 2025, p. 3, tradução livre).

Sob esse aspecto e considerando a área da Arquivística, entende-se que operacionalizar dados relacionados à classificação, bem como de outras funções arquivísticas, deve ser realizado com ética e responsabilidade. Dados de qualidade e confiáveis, por meio dos paradados, são cruciais para que modelos de aprendizagem de máquinas e agentes de IA possam performar com eficiência.

Destarte, o modo como os dados da CAM são estruturados possibilitam maior confiabilidade e menos riscos de vieses no decurso do uso de algoritmos, agentes, *prompts* e modelos de IA. O grupo de pesquisa EADRD entende que há pertinência no uso da IA no contexto da ontogênese e da CAM a fim abrir novas perspectivas para a gestão contínua de documentos arquivísticos. Ante o exposto e considerando que a CAM é composta por dados de gestão, uma das possibilidades de lidar com esses dados é por meio de *dashboards* de dados do contexto de produção e gestão dos documentos. Dessa maneira, abre-se a interessante possibilidade para os arquivistas atuarem como gestores de dados, metadados, requisitos, informações descritivas e paradados de caráter arquivístico.

4. CLASSIFICAÇÃO ARQUIVÍSTICA MULTIFACETADA (CAM): RESULTADOS PRELIMINARES

Um dos objetivos da classificação arquivística multifacetada é garantir aos documentos a representação de sua organicidade desde sua gênese, ou melhor dizendo, desde sua ontogênese. Este modelo foi pensado para representar os documentos nato digitais, produzidos em contextos organizacionais. Desse modo, não foram pensados a partir de documentos, capturados, digitalizados, acervos legados, ou arquivos pessoais e de famílias. Contudo, testes iniciais já demonstram que a CAM também se adequa a estes e outros conjuntos documentais.

A CAM foi estruturada de forma facetada, conforme o pensamento de Ranganathan, a partir das seguintes facetas inter-relacionadas: o contexto histórico do todo e das partes; a missão/vínculo institucional; a estrutura organizacional e a configuração funcional. Esta ordem facetada possibilita representar a organicidade dos tipos documentais em sua integralidade.

O processo para elaboração de uma classificação arquivística multifacetada fundamenta-se na interdisciplinaridade da Arquivística com a Administração, História, Biblioteconomia, Computação, Filosofia e Linguística. Sua composição demanda várias etapas, as quais exigem diferentes técnicas e métodos para coleta de dados e informações específicas para cada faceta. Este arcabouço de informações categorizadas visa subsidiar diferentes estruturas em camadas, conforme observa-se na Figura 2. A partir do estabelecimento das relações entre os dados, tem-se a viabilidade da elaboração do plano de classificação, bem como da Tabela de Temporalidade, do controle de vocabulário, descrição, dentre outras possibilidades, não apenas de criação de

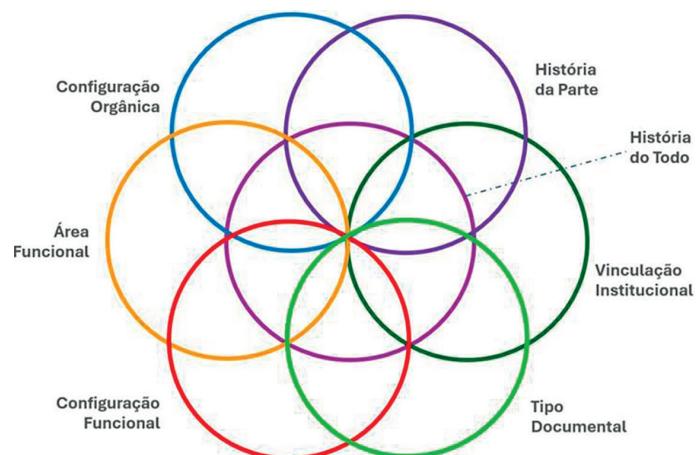


Fig. 2
Representação das camadas de estruturas de dados da CAM
Fonte: Elaborado pelos autores, 2025

instrumentos arquivísticos, mas, sobretudo, para gerenciar adequadamente os dados arquivísticos para eficiência administrativa e preservação digital, o que também viabiliza o uso e reuso dos dados e informações arquivísticas ao que cabe a governança de dados corporativos de modo geral.

A CAM é um modelo sistêmico relacional, construído por uma estrutura de dados, devidamente modelados e categorizados com significados implícitos, os quais caracterizam os contextos de criação e organicidade dos documentos arquivísticos. Os dados categorizados de cada faceta são recursos que podem e devem ser utilizados e reutilizados em todo e qualquer sistema informatizado que produzem documentos e registros nato digitais de caráter arquivístico, num âmbito institucional público ou privado.

Parte dos aportes metodológicos da CAM está baseado no delineamento proposto por Sousa (2009). Segundo o autor,

A partir da apropriação dos pressupostos da teoria da classificação, entendemos que os princípios de divisão do conjunto (arquivo) devem ser procurados na própria gênese documental, no processo de criação, nas relações estabelecidas e na reprodução das conexões originais (Sousa 2009, p. 114).

No que diz respeito ao escopo da CAM, o tipo documental é o objeto a ser classificado, podendo também ser considerada a série documental. Sousa (2009, p. 95) esclarece que para entender o objeto classificado é necessário compreender «sua amplitude, peculiaridade, funcionamento, extensão, pois acreditamos que o método de organização deve refletir com a maior exatidão possível a natureza do próprio objeto».

A classificação arquivística aqui proposta possui 7 (sete) facetas básicas relacionadas. Devido à sua complexidade esse instrumento deve ser operacionalizado por ferramenta computacional. Contudo para facilitar o entendimento é possível ser representada notacionalmente por algarismos arábicos, estruturados em casas decimais de 2 dígitos, separados por ponto (.), conectadas entre si pelo sinal de dois pontos (:). As facetas são denominadas da seguinte forma: Configuração Histórico-Contextual do Todo; Vinculação Institucional; Configuração Histórico-Contextual da Parte; Configuração Orgânico-Estrutural; Área Funcional; Configuração Funcional e Tipo Documental, conforme Figura 3.

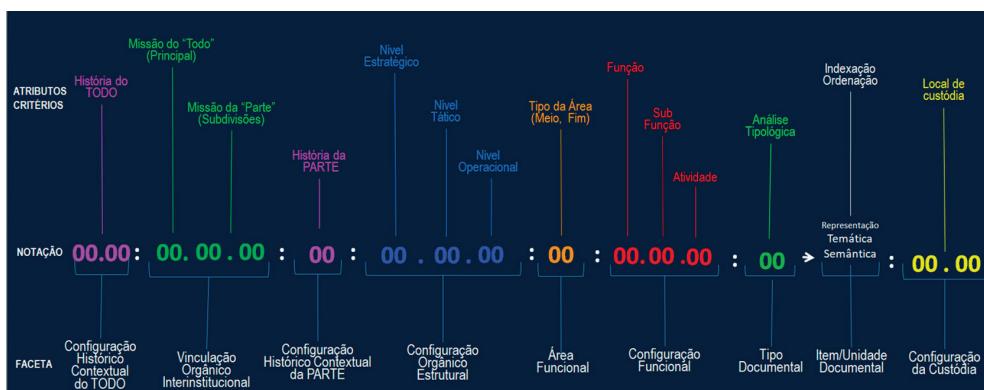


Fig. 3. Estrutura Notacional da Classificação Arquivística Multifacetada

Fonte: Elaborado pelos autores, 2025

A faceta orgânico-estrutural e as facetas funcionais são bem difundidas na Arquivologia. Um dos aspectos originais desta proposta está relacionado ao ingresso das facetas da Configuração Histórico-Contextual do Todo, da Vinculação Institucional e da Configuração Histórico-Contextual da(s) Parte(s). A necessidade desta inserção dá-se para uma maior compreensão do contexto de produção dos tipos documentais ao longo do tempo.

É interessante observar que, muitas vezes, trabalhos arquivísticos, realizados em âmbito institucional, principalmente a classificação arquivística, estão focados em resolver questões pontuais e/ou isoladas. Entretanto, quando observamos o conjunto dessas atividades, deparamo-nos com uma intrincada cadeia de eventos, interligados e sistêmicos, com variáveis mutuamente dependentes.

Normalmente a área não considera os acervos em sua complexidade, isto é, numa perspectiva de um «todo» sistêmico. Amiúde, detém-se apenas em casos arquivísticos particulares, muitas vezes desconexos dos seus vínculos institucionais originários, isto é, acervos documentais, geralmente, não são considerados como parte de um todo maior e complexo. Morin (2023) defende que é preciso compreender o princípio da

organização, entendido como algo que estabelece inter-relações sistêmicas, o qual pode combinar vários tipos de ligações: «liga os elementos entre eles, os elementos em uma totalidade, os elementos à totalidade, a totalidade aos elementos, ou seja, une entre si todas as ligações e constitui a ligação das ligações» (Morin 2013, p. 165). Nesse sentido, é fundamental compreender que a complexidade dos problemas só pode ser tratada a partir de uma perspectiva também complexa.

Assim, ao que compete a primeira faceta devemos observar o sentido macro das instituições produtoras de arquivos, ou seja, as relações de vinculação entre o todo e as partes, em outras palavras, devemos conhecer organismos produtores de documentos em sua amplitude. No processo de organização arquivístico não se deve deixar de entender a dinâmica de relações entre entidades, observando quais as partes que formam um todo. Como exemplo podemos destacar a estrutura administrativa de uma empresa privada com suas filiais ou de um estado ou município, formado por administração direta e indireta.

Definido o todo e as partes do processo classificatório, destaca-se um elemento que os perpassam, que os atravessam, já por nós descrito: o aspecto histórico-temporal, na medida em que os documentos não são apenas oriundos das funções e atividades de uma determinada instituição, eles são também produtos do momento histórico em que se encontram as entidades públicas ou privadas. Esta faceta orienta a produção documental para além dos aspectos já conhecidos: orgânicos e funcionais.

Representar o tempo no processo classificatório é instância necessária, como produtora de sentido para entendermos a complexa relação do documento com as unidades produtoras, tendo em vista que a estrutura temporal permite representar as coisas simultânea e/ou sucessivamente, pois «tudo que podemos conhecer está submetido às condições espaço-temporais» (Aranalde 2009, p. 96). A faceta temporal tem recebido atenção especial por parte dos pesquisadores do grupo EADRD, visto que é um tema pouco explorado pela área. Uma pesquisa de abordagem teórica a esse respeito está em andamento, em breve os dados serão publicados no repositório de dados abertos.

A fase dos testes de validação da CAM vem sendo realizada a partir do desenvolvimento de um sistema utilizando a linguagem Phyton e o *framework* Jango. O desenvolvimento em Python permite verificações no sentido de identificar problemas no código durante o processo de criação, dentre outras inconformidades de desenvolvimento e das relações entre as facetas. O que corrobora os ajustes necessários na estrutura conceitual, nos requisitos e metadados do projeto da CAM.

Nesse trabalho foram apresentados resultados parciais referentes à primeira etapa de uma pesquisa em andamento. Com ele, buscamos contribuir para o avanço das discussões e para a revisão necessária dos processos de classificação dos documentos arquivísticos produzidos em ambientes digitais, assim como daqueles presentes em acervos legados.

CONSIDERAÇÕES FINAIS

A investigação apresentada, fruto de quatro anos de pesquisa do Grupo EADRD da Universidade Federal da Paraíba, propõe a Classificação Arquivística Multifacetada (CAM) como uma solução inovadora para superar as ineficiências das classificações funcionais, que se mostram limitadas na representação do vínculo arquivístico de documentos digitais e enfrentam o problema do anacronismo ao negligenciar o fator «tempo» e as contínuas transformações organizacionais. O estudo apresenta a sistemática da CAM, aplicada às práticas de gestão e orientada à Inteligência Artificial (IA).

Destaca-se que a CAM se distingue de outros métodos de classificação arquivística por integrar elementos funcionais, orgânico-estruturais e temporais, considerando o contexto histórico, a missão institucional e a estrutura organizacional, buscando representar a organicidade dos documentos desde a fase da produção dos documentos. Fundamentada na noção de Ontogênese Documental, na Teoria Geral dos Sistemas e no Pensamento Complexo, a CAM adota os princípios da classificação facetada de Ranganathan (1967). A utilização de sete facetas inter-relacionadas, incluindo a Configuração Histórico-Contextual do Todo e da Parte e a Vinculação Institucional, além das conhecidas orgânico-estruturais e funcionais, permite uma representação complexa e dinâmica. A faceta temporal, em particular, é um aspecto original e crucial, pouco explorado na área.

A metodologia empregada é de natureza teórica e aplicada, com abordagem qualitativa, buscando não apenas a construção conceitual, mas também a validação da proposta por meio de desenvolvimento de sistema. A estrutura da CAM baseia-se na composição de dados, informações descritivas, metadados e paradados, utilizando o *framework* DAMA-DMBOK para a governança e gestão desses dados complexos. Além disso cabe a distinção entre IA explicável e IA responsável, o estudo enfatiza a necessidade de documentar «por que, como, por quem e com que resultado um determinado sistema ou instrumento foi usado em um determinado contexto» para garantir a responsabilidade na aplicação da IA. Nesse sentido, os dados estruturados são condição essencial para garantir a confiabilidade e a acurácia necessárias ao uso responsável e ético da Inteligência Artificial em práticas arquivísticas.

A proposta da CAM e sua integração com a IA posicionam os arquivistas como gestores de dados de caráter arquivístico. Os testes preliminares, desenvolvidos em Python e Django, já indicam resultados promissores, demonstrando a aplicabilidade e o potencial da CAM para lidar com volumes significativos de documentos digitais e legados. A Classificação Arquivística Multifacetada representa um olhar inovador no âmbito da Arquivística, ao propor um modelo sistêmico e relacional que comprehende os arquivos não como somas isoladas, mas como uma unidade integral e aberta a contextos dinâmicos e históricos. Ao considerar a complexidade dos acervos documentais em uma perspectiva de «todo» sistêmico e integrar múltiplas facetas de

classificação, o trabalho contribui para a revisão necessária dos processos de classificação e abre novas perspectivas para a gestão contínua e eficiente de documentos arquivísticos no ambiente digital.

REFERÊNCIAS

- ARANALDE, Maria Mercedes, 2009. Reflexões sobre os sistemas categoriais de Aristóteles, Kant e Ranganathan. *Ciência da Informação*. 38(1), 86-108.
- BAK, Greg, 2012. Continuous classification: capturing dynamic relationships among information resources. *Archival Science*. 12, 287-318.
- BARBOSA, Aurora Pimentel, 1972. *Classificações facetadas*. *Ciência da Informação*. 1(2), 73-81.
- BELLOTTTO, Heloísa Liberalli, 2014. *Arquivo: estudos e reflexões*. Belo Horizonte: Ed. UFMG.
- BELLOTTTO, Heloísa Liberalli, 2006. *Arquivos permanentes: tratamento documental*. 4.ª ed. Rio de Janeiro: Fundação Getúlio Vargas.
- BERTALANFFY, Ludwig von, 2013. *Teoria geral dos sistemas: fundamentos, desenvolvimento e aplicações*. Trad. Francisco M. GUIMARÃES. 7.ª ed. Petrópolis: Vozes.
- COUPER, MICK P, 2017. Birth and Diffusion of the Concept of Paradata. *Advances in Social Research*. (18), 14-26.
- COZMAN, Fabio Gagliardi, 2021. No canal da Inteligência Artificial – nova temporada de desgrenhados e empertigados. *Estudos Avançados*. 35, 7-20.
- DAMA INTERNATIONAL, 2017. *DAMA-DMBOK: data management body of knowledge*. 2.ª ed. Westfield: Technics Publications.
- DAVIS, Randall, Howard SHROBE, e Peter SZOLOVITS, 1993. What is a knowledge representation? *AI Magazine*. 14(1), 17-33.
- DORNELES, Sânderson Lopes, 2024. Aplicações de Inteligência Artificial na gestão documental: revisão sistemática na LISTA, Emerald e Scopus. Em: Elisa CERVEIRA, e Zeny DUARTE, org. *Integração e Convergência das Instituições de Informação e Cultura*. Porto: CITCEM, 85-104.
- DURANTI, Luciana, 2005. The long-term preservation of accurate and authentic digital data: the InterPARES Project. *Data Science Journal*. 4, 106-118.
- DURANTI, Luciana, e Patricia C. FRANKS, eds., 2015. *Encyclopedia of archival science*. Lanham: Rowman & Littlefield.
- FELICIATI, Pierluigi, e Luciana, DURANTI, 2025. The responsible use of Artificial Intelligence in archives through the use of paradata. *JLIS. it*. 16(2), 1-9.
- FOSCARINI, Fiorella, 2012. Understanding functions: an organizational culture perspective. *Records Management Journal*. 22(1), 20-36.
- GILLILAND, Anne J., Sue McKEMMISH, e Andrew J. LAU, eds., 2017. *Research in the archival multiverse*. Melbourne: Monash University Publishing.
- GLASER, Barney G., e Anselm L. STRAUSS, 1998. *Grounded theory: strategien qualitativer Forschung*. Berna: Huber.
- INTERPARES PROJECT 2, 2021. *Glossary* [Em linha] [consult. 2025-06-01]. Disponível em: http://www.InterPARES.org/ip2/ip2_terminology_db.cfm.
- JANSEN, Adam, e Luciana DURANTI, 2013. The InterPARES Trust Project – trust and digital records in an increasingly networked society. Em: Anne GILLILAND, et al. *The Future of Information Science: Information Governance*. Zagreb: Department of Information and Communication Sciences, Faculty of Humanities and Social Sciences, University of Zagreb, Croatia, pp. 63-68.
- LEWRICK, Michael, Patrick LINK, e Larry LEIFER, 2018. *The design thinking playbook: mindful digital transformation of teams, products, services, businesses and ecosystems*. New Jersey: John Wiley & Sons.

- MOKHTAR, Umi, e Zuraidah Mohd YUSOF, 2017. *Records classification: concepts, principles and methods, information, systems, context*. Kidlington: Elsevier.
- MORIN, Edgar, 2023. *Ciência com consciência*. 21.ª ed. Rio de Janeiro: Bertrand Brasil.
- MORIN, Edgar, 2013. *O método 1: a natureza da natureza*. 3.ª ed. Porto Alegre: Sulina.
- NORVIG, Peter, e Stuart RUSSELL, 2013. *Inteligência artificial*. 3.ª ed. Rio de Janeiro: Elsevier.
- ORR, Susan A., 2005. *Functions-based classification of records: is it functional?* Master of Science in Records Management, Northumbria University, School of Informatics.
- PACKALEN, Susanna, e Pekka HENTTONEN, 2016. Recordkeeping professionals' understanding of and justification for functional classification: Finnish public sector organizational context. *Archival Science*. **16**, 403-419.
- PIEDADE, Maria Antonietta Requião, 1983. *Introdução à teoria da classificação*. Rio de Janeiro: Interciência.
- RANGANATHAN, Shiyali Ramamrita, 1967. *Prolegomena to library classification*. Bombaim: Asia Publishing House.
- RODRIGUES, Ana Célia, 2016. Diplomática e arquivística: diálogos para a construção do método de identificação da tipologia documental. Em: *Anais do XVII Encontro Nacional de Pesquisa em Ciência da Informação, 20 a 25 de novembro de 2016, Salvador (BA)*. João Pessoa: UFPB, pp. 869-881.
- SILVA, Armando Malheiro da, et al., 1999. *Arquivística: teoria e prática de uma ciência da informação*. Porto: Edições Afrontamento.
- SOUSA, Renato Tarciso Barbosa de, 2022. A classificação funcional de documentos de arquivo é uma abstração intelectual ou um instrumento prático? *Acervo*. **35**(2), 1-21.
- SOUSA, Renato Tarciso Barbosa de, 2009. A classificação como função matricial do que-fazer arquivístico. Em: Vanderlei Batista SANTOS, Humberto Celeste INNARELLI, e Renato Tarciso Barbosa de SOUSA. *Arquivística: temas contemporâneos*. 3.ª ed. São Paulo: Senac, pp. 79-173.