

Renovação teórica e inovação tecnológica em descrição de arquivos

O modelo conceitual Records in Contexts (RiC)

Theoretical renewal and technological innovation in archival description: Records in Contexts conceptual model / Renovación teórica e innovación tecnológica en la descripción de archivos: el modelo conceptual Records in Contexts (RiC)

Thiago Almeida Rodrigues Borges
Mestre em Ciência da Informação pela Universidade de Brasília (UnB). Arquivista da Câmara dos Deputados, Brasil.
tarborges@hotmail.com

Cynthia Roncaglio
Doutora em Meio Ambiente e Desenvolvimento pela Universidade Federal do Paraná (UFPR), com pós-doutorado na Facultad de Ciencias de la Documentación da Universidad Complutense de Madrid (UCM), Espanha. Professora do Curso de Arquivologia e do Programa de Pós-Graduação em Ciência da Informação da Universidade de Brasília (UnB), Brasil.
roncaglio@unb.br

RESUMO

Este artigo explora as inovações propostas pelo modelo conceitual Records in Contexts (RiC) na teoria e prática da descrição arquivística, incluindo a crítica à hierarquização na descrição, a adoção da descrição multidimensional e a expansão do princípio da proveniência. Além disso, examina a integração dos arquivos à web semântica e aos dados abertos conectados, evolução da web de documentos para a web de dados.

Palavras-chaves: Records in Contexts (RiC); descrição arquivística; web semântica; arquivos.

ABSTRACT

This article explores the innovations proposed by the Records in Contexts (RiC) conceptual model in the theory and practice of archival description, including criticism of hierarchical description, the adoption of multidimensional description, and the expansion of the principle of provenance. Additionally, it examines the integration of archives with the semantic web and linked open data, marking the evolution from the web of documents to the web of data.

Keywords: Records in Contexts (RiC); archival description; semantic web; archives.

RESUMEN

Este artículo explora las innovaciones propuestas por el modelo conceptual Records in Contexts (RiC) en la teoría y práctica de la descripción archivística, incluyendo la crítica a la jerarquización en la descripción, la adopción de la descripción multidimensional y la expansión del principio de procedencia. Además, examina la integración de los archivos con la web semántica y los datos abiertos conectados, marcando la evolución de la web de documentos a la web de datos.

Palabras clave: Records in Contexts (RiC); descripción archivística; web semántica; archivos.

Introdução

O avanço das tecnologias de informação e comunicação (TICs) tem transformado profundamente diversas atividades humanas, desde a indústria e o comércio até a pesquisa acadêmica. A ciência da informação, a arquivologia, a biblioteconomia e a museologia não ficam de fora dessa mudança, uma vez que todas têm a informação como eixo comum de estudo. O cenário atual mostra que a transformação digital da documentação e as inovações tecnológicas têm levado a revisões das práticas e teorias voltadas à gestão do patrimônio documental e cultural da humanidade.

No que diz respeito à arquivologia e aos arquivos, as funções fundamentais, como a descrição, estão adaptando-se ao mundo digital, possibilitando tanto a construção de novas perspectivas quanto de formas criativas e inovadoras de acesso aos documentos. Tal cenário requer um reexame dos conceitos e das práticas arquivísticas tradicionais. Com a expansão da internet e a evolução para a web semântica – que possibilita a integração de dados e sua interpretação por máquinas –, instituições responsáveis pelo patrimônio documental têm buscado modernizar seus sistemas e disponibilizar acervos em plataformas on-line. Marcondes (2016, p. 78) destaca a potencialidade das tecnologias semânticas e dos dados abertos na publicação de acervos na web.

Em escala global, nota-se que bibliotecas e museus se adaptaram à web semântica antes dos arquivos, com destaque para os modelos conceituais Cidoc CRM¹ dos museus e FRBR² das bibliotecas. Entretanto, os arquivos estão expandindo sua presença on-line, melhorando os serviços, especialmente na busca e recuperação de documentos.

O Conselho Internacional de Arquivos (ICA), através do Grupo de Especialistas em Descrição Arquivística (Egad), tem se dedicado à investigação da relação entre os arquivos e a web semântica. Isso culminou no desenvolvimento do modelo conceitual de dados destinados à descrição arquivística Records in Contexts (RiC-CM). O modelo tem como objetivo a integração das normas internacionais de descrição vigentes, bem como a consideração das práticas adotadas em áreas afins do patrimônio, tais como: bibliotecas, museus e centros de documentação. O foco principal deste artigo consiste na análise detalhada do RiC, com o propósito de identificar as inovações teóricas, práticas e tecnológicas que o modelo propõe.

1 Cidoc Conceptual Reference Model, do Comitê Internacional para a Documentação do Conselho Internacional de Museus (Icom).

2 Functional Requirements for Bibliographic Records, da Federação Internacional de Associações e Instituições de Bibliotecas.

O artigo consiste em identificar as principais modificações propostas pelo RiC no contexto da teoria arquivística, com ênfase na função de descrição, além de abordar as inovações tecnológicas preconizadas pelo modelo. A motivação para esta pesquisa se baseia na escassez de referências bibliográficas disponíveis no Brasil sobre o tema dos arquivos e da web semântica, bem como sobre o modelo conceitual RiC, conforme apurado durante a investigação.

Metodologia

No desenvolvimento da presente pesquisa adotou-se como metodologia a revisão bibliográfica – caracterizada pela análise crítica e reflexiva das publicações técnicas e científicas – e o próprio modelo conceitual Records in Contexts e materiais afins (os quais foram obtidos por meio do site do Egad). O *corpus* documental selecionado é considerado relevante para atingir os objetivos propostos neste estudo. Optou-se por uma metodologia de cunho exploratório, visando à sistematização e à interpretação aprofundada tanto do conteúdo inerente ao modelo RiC quanto dos documentos a ele relacionados, além de segmentos pertinentes de sua fundamentação teórica.

Na fase preliminar da pesquisa, definiu-se o escopo do material a ser analisado. Depois da publicação de versões preliminares em 2016, 2019 e 2021, a primeira versão oficial do modelo foi lançada em 2023. Dessa forma, para este artigo, foram utilizados o Modelo Conceitual Records in Contexts (RiC-CM *consultation draft v0.2*), publicado em julho de 2021, a versão oficial publicada em 2023 e o *Records in Contexts: Introduction to Archival Description (RiC-IAD consultation draft v.02)*, publicado em dezembro de 2021, juntamente a textos correlatos, disponíveis na página do grupo de especialistas Egad, relacionados à introdução ao modelo RiC e aos relatórios de progresso na revisão e harmonização das normas de descrição em arquivos.

Na fase de pré-análise, foi realizada a leitura das fontes selecionadas, a fim de se obter uma visão geral do conteúdo. Na etapa seguinte – que envolve a exploração do material – procedeu-se à leitura e à identificação das principais questões teóricas propostas pelo modelo em termos de inovação, além da identificação das tecnologias com as quais o modelo opera.

Por fim, na etapa de tratamento dos resultados, foram criados dois eixos de análise: o primeiro eixo está relacionado à teoria arquivística e o segundo aborda as questões tecnológicas. Para uma compreensão mais aprofundada dos conceitos propostos pelo modelo RiC, buscou-se também um referencial teórico complementar à análise dos conceitos.

Revisão teórica e prática de descrição arquivística

A análise crítica da teoria e da prática em descrição arquivística, particularmente a reflexão sobre a efetividade do uso das normas internacionais de descrição estabelecidas pelo Conselho Internacional de Arquivos e o questionamento da visão tradicional do princípio da proveniência, destaca-se como o pilar fundamental de renovação teórica proposto pelo modelo conceitual RiC.

Em relação às normas internacionais de descrição em arquivos, ao se elaborar um breve histórico sobre essa função arquivística, constata-se que uma das principais referências para as normas do ICA tem origem na comunidade arquivística do Canadá. Cook (2017) explora a experiência canadense na padronização da descrição, elucidando que, anteriormente à publicação da norma *Rules for Archival Description* (RAD) em 1990, os arquivistas engajados neste processo dedicaram-se à reflexão acerca das variadas concepções de fundo, oriundas tanto da tradição francesa quanto da anglo-saxônica. Cook (2017) esclarece que, dado o propósito do comitê de estabelecer um padrão de descrição, era imperativo que conceitos basilares e práticas associadas — como a definição de fundo — fossem compreendidos de forma precisa e aceitos, consensualmente, pela comunidade arquivística, pelo menos no âmbito canadense.

Com forte influência da RAD canadense, inicia-se nos arquivos o processo de separação dos elementos de descrição, como fundos/documentos das autoridades/produtores, lógica absorvida pelo ICA no início dos anos de 1990. Conforme Gueguen et al. (2013), as recomendações sobre as vantagens da separação foram introduzidas por membros do grupo de arquivistas logo no início dos esforços de formulação de uma norma geral internacional de descrição arquivística, visando especificamente a criação de “pontos de acesso” relacionados às autoridades vinculadas aos documentos. A princípio, a concepção de separação das descrições das autoridades dos demais elementos estava fortemente motivada também pela adoção de sistemas computacionais da década de 1980. Influenciado por tais perspectivas e almejando procedimentos mais econômicos e flexíveis para a descrição arquivística, bem como novas maneiras de explorá-la em sistemas de acesso, o ICA adotou a segmentação dos componentes.

Entre 1994 e 2008, o ICA disponibilizou quatro normas internacionais para o uso da comunidade arquivística, a saber: Isad(G), Norma Geral Internacional de Descrição Arquivística; Isaar(CPF), Norma Internacional de Registo de Autoridade Arquivística para Entidades Coletivas, Pessoas e Famílias; ISDF, Norma Internacional para Descrição de Funções; e Isdiah, Norma Internacional para Descrição de Instituições com Acervos Arquivísticos.

Quadro 1 – Cronologia do desenvolvimento das normas internacionais de descrição arquivística

Norma	Edição	Data de elaboração	Data de publicação
Declaração de princípios		(1988) 1989-1992	1992
Isad	1 ^a	1990-1993	1994
Isaar	1 ^a	1993-1995	1996
Isad	2 ^a	1996-2000	1999
Isaar	2 ^a	2000-2004	2004
ISDF	1 ^a	2005-2007	2007
Isdiah	1 ^a	2005-2008	2008

Fonte: Gueguen et al. (2013).

Após as fases de elaboração e aplicação das normas em numerosas instituições arquivísticas ao redor do mundo, iniciou-se, a partir de 2009, a fase sobre a aplicabilidade dos padrões internacionais de descrição. Há concordância entre os especialistas do ICA quanto à predominância do uso da Isad-G em relação às demais normas. Tal fato corrobora a tradição na descrição de fundos e documentos em arquivos. Em consequência, as normas Isaar(CPF) (autoridades/produtores), ISDF (funções/atividades) e Isdiah (entidade custodiadora), que têm seus elementos derivados da Isad(G) e que deveriam ser exploradas autonomamente em sistemas específicos de forma interconectada, com vistas a fomentar novas compreensões e usos dos documentos de arquivo, não atingiram tal propósito.

Um dos principais fatores que explicam a não assimilação e uso das normas internacionais pelas instituições arquivísticas, segundo os profissionais da Egad (ICA, 2021b), é a carência de manuais práticos com diretrizes claras sobre como as normas poderiam ser conjuntamente empregadas. A esse aspecto, adicionam-se as limitações funcionais dos sistemas e suas ferramentas, fazendo com que os componentes do sistema de descrição seguissem direções distintas.

A centralidade da Isad(G) é um dos aspectos mais criticados pelo grupo Egad no processo revisional da descrição arquivística. Entende-se que a priorização da descrição de fundos e documentos – em detrimento de outras entidades que poderiam ser descritas – seja prejudicial no complexo contexto atual do universo digital em que os documentos são produzidos, usados e preservados. As pessoas, instituições, funções, atividades e entidades custodiadoras relacionadas aos arquivos, apesar de possuírem normas próprias de descrição – Isaar(CPF), ISDF e Isdiah –, não tiveram suas descrições elaboradas na maioria das instituições de arquivo, tornando-se,

quando muito, apenas pontos de acesso para recuperar fundos e documentos do acervo. Somam-se ainda às entidades locais, assuntos e datas, que também são usados apenas como pontos de acesso, não possuindo normas específicas, nem metadados próprios, e sua descrição não é tradicionalmente feita pelos arquivistas. Conforme o ICA, a intenção é que no modelo conceitual RiC os arquivistas descrevam também esses elementos, considerados importantes para a compreensão da ampla rede de entidades em que os arquivos se relacionam:

Em vez de continuar tentando revisar a Isad(G) e suas normas complementares, uma a uma, para abordar essa questão, a ICA instruiu o Egad a desenvolver um esquema abrangente para a descrição arquivística baseado em um modelo conceitual que reconheça as realidades relacionais e contextuais dos arquivos e que considere a entidade “documento” como sendo apenas uma de uma série de entidades e relações que devem ser descritas para que os sistemas de controle arquivístico reflitam as realidades das diversas formas como os arquivos são feitos, mantidos e utilizados. (ICA, 2021b, p. 9, tradução nossa)

Nas justificativas da adoção para o modelo conceitual RiC, constatam-se inúmeras críticas adicionais às normas de descrição. Devido ao fato das normas do ICA terem tido forte influência em termos práticos e tecnológicos da época em que foram produzidas (os anos 1990), uma das críticas dá-se ao fato do registro de entrada da descrição elaborada por arquivistas ser idêntico, na maioria dos casos, à versão de saída da descrição disponibilizada aos usuários, mesmo que este padrão tenha sido rejeitado explicitamente, como consta na introdução da Isad(G): “A norma não define formatos de saída ou modos nos quais esses elementos são apresentados, por exemplo, em inventários, catálogos, listas etc.” (ICA, 2004, p. 11). Como resultado, caracterizada por sua natureza linear, a recuperação das descrições por usuários, em inúmeros sistemas de busca em arquivos criados com referência nas normas, se assemelha às buscas efetuadas nos tradicionais instrumentos de pesquisa impressos, lidos página a página, com alguns recursos adicionais, como índices, remissivas ou pontos de acesso.

Desse modo, tal como sugere o ICA (2021b), a relação entrada-saída de dados é uma das principais inovações do modelo conceitual RiC: trata-se de um modelo referencial apenas para a entrada de dados nos sistemas de registro – os dados separados e interligados em ampla rede de relações contextuais – com alta aderência às tecnologias atuais. Por outro lado, quanto à saída dos dados, sua apresentação e seu manejo pelos usuários, a proposta do RiC é que não haja formato definido, sendo mais aberta e com possibilidades infinitas de uso conforme os vieses de interesse dos pesquisadores.

Acrescentam-se questões relativas à noção das unidades de descrição, já que os campos eram os mesmos independentemente do nível a ser descrito: fundo, grupo, série, dossiê ou item. Não havia individualização dos metadados a serem descritos, supondo-se certa equivalência irreal entre fundos e itens, por exemplo. No modelo RiC, apesar de vários atributos compartilhados, há metadados individualizados por entidades, já que “tratá-los cada um como algo distinto, produzidos em momentos distintos e com propósitos diferentes, possibilita uma descrição inequívoca e clara” (ICA, 2021b, p. 10, tradução nossa).

Um dos aspectos mais inovadores do modelo conceitual RiC é a descrição dos arquivos para além da lógica multinível, em direção à descrição multidimensional. A descrição multinível é consagrada na área de arquivos e teve sua divulgação ampliada pela Isad(G) e, no Brasil, pela Norma Brasileira de Descrição Arquivística (Nobrade), elaborada pelo Conselho Nacional de Arquivos (Conarq) com base na primeira. O conceito de descrição multinível consiste em descrever o fundo como um todo, posteriormente, cada parte, consideradas as divisões hierárquicas existentes nos quadros de arranjo da documentação. Conforme consta na Isad(G), para a efetivação da descrição multinível existem regras básicas a serem cumpridas pelos arquivistas: a descrição do geral para o particular consiste em “representar o contexto e a estrutura hierárquica do fundo e suas partes componentes”; a informação relevante para o nível de descrição se baseia em “representar com rigor o contexto e o conteúdo da unidade de descrição”; a relação entre as descrições “torna explícita a posição da unidade da descrição na hierarquia”; e, por fim, a não repetição busca “evitar a redundância de informação em descrições hierarquicamente relacionadas” (ICA, 2004, p. 17).

Segundo o ICA (2021b), a descrição multidimensional proposta no modelo conceitual RiC aproxima-se mais de uma representação de dados dos arquivos em “redes” ou “grafos”, conforme se observa na Figura 1, não importando a hierarquia da descrição multinível, possibilitando uma descrição “mais viável para documentos com proveniência complexa e diversificada dentro de um fundo”.

Na abordagem de descrição multidimensional, os documentos e conjunto(s) de documentos, suas inter-relações, suas relações com pessoas, grupos e atividades, bem como as relações entre estes elementos, são representados como uma rede na qual um fundo específico pode ser identificado. O contexto imediato de cada fundo é estabelecido, embora suas fronteiras sejam permeáveis, pois ele existe dentro de uma rede de registros inter-relacionados e de conjunto(s) de documentos, indivíduos, grupos e atividades. (ICA, 2021b, p. 11, tradução nossa)

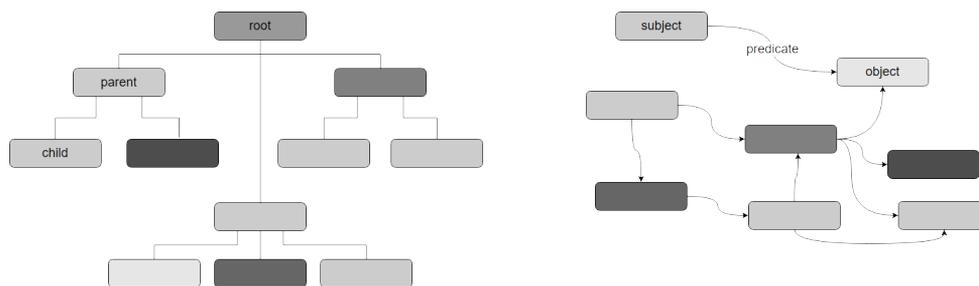


Figura 1 – Representação comparativa de dados em estrutura hierárquica, como XML (à esquerda) e estrutura em triplas, como grafos (à direita). Fonte: ICA (2021b)

Todas as críticas postuladas às normas de descrição do ICA e a seus conceitos relacionados somam-se às discussões teóricas mais amplas da teoria arquivística. Dessa forma, ao elaborar o modelo RiC, o ICA não só refletiu sobre os ganhos, as perdas e o aperfeiçoamento das normas de descrição, mas considerou as discussões teóricas sobre o princípio da proveniência, que vêm ocorrendo há décadas na comunidade arquivística, das quais retomaremos pontos centrais para melhor compreensão das propostas sugeridas no modelo conceitual RiC.

Em sua obra, Cook (2017) discorre sobre múltiplos aspectos inerentes ao conceito de fundo, dos quais destaca a natureza intrínseca do fundo e como tal noção, desde sua origem, visa garantir o caráter de prova dos arquivos, configurando-se ainda como o princípio que distingue os arquivos de bibliotecas, de museus e de centros de documentação. O autor também ressalta a origem francesa do princípio de respeito aos fundos, suas dimensões internas e externas, sua correlação com o princípio do respeito à ordem original e com a proveniência germânica, além da sua assimilação no cenário anglo-saxão por meio da noção de *records group*, uma temática de notável controvérsia, sobre a qual ele defende o abandono do referido conceito.

Das inúmeras reflexões teóricas salientadas por Cook, ganham destaque as perspectivas custodial e pós-custodial do conceito de fundo, bem como diferentes visões sobre a descrição dos arquivos. Por um lado, pratica-se na comunidade arquivística a descrição com base no entendimento do fundo como uma categoria conceitual, delineada nas classes mais gerais do quadro de arranjo, com ênfase nos produtores e seus respectivos contextos jurídico-administrativos. Em contrapartida, há a prática de descrição com destaque para os níveis mais concretos, delineados nos níveis de séries, processos/dossiês e documentos individuais, onde são localizados os documentos em sua essência (Cook, 2017, p. 61). Essa dicotomia é corroborada pelos especialistas do ICA:

O modelo descritivo dominante no período era o de usar uma única descrição para descrever um fundo arquivístico, começando com a descrição mais geral e, em crescente especificidade, as partes do fundo, as partes das partes, e assim por diante. Todos os atributos dos documentos em um fundo, e o contexto da produção e uso dos documentos, eram combinados em uma descrição, com os diferentes elementos descritivos entrelaçados para formar uma descrição compreensiva e completa. A descrição cobria, assim, os documentos como tais, bem como aquilo que com o tempo viemos a entender como o contexto do documento: o produtor, funções e atividades, e entidade custodiadora. (Gueguen et al., 2013, p. 103)

Conforme Cook (2017), a concepção clássica do campo, que interpreta o fundo como um conjunto físico de documentos oriundos de um único produtor, é categorizada como abordagem “custodial” ou “curatorial”.

Na perspectiva pós-custodial, a compreensão de fundo assume uma conotação mais abstrata, visto que reconhece os documentos contemporâneos como caracterizados por um processo de produção e acumulação coletiva, envolvendo múltiplos produtores. Tal dinâmica torna inviável a associação exclusiva entre uma série documental e um único produtor, como frequentemente observado em instituições do século XIX. Diversos fatores justificam a natureza da documentação originada por múltiplos produtores, tais como a dinâmica governamental atual, a intrincada burocracia, a presença de estruturas produtivas concomitantes, atuações de grupos de trabalho e envolvimento em equipes de projetos, entre outros aspectos (Cook, 2017, p. 46). A essas considerações adicionam-se as implicações das inovações tecnológicas emergentes nas últimas décadas do século XX, com a produção de documentos em sistemas eletrônicos ou digitais. As séries, os processos, os dossiês e os itens documentais muitas vezes possuem múltiplos produtores, resultando em documentos com “multiproveniência”. Cook argumenta a necessidade de uma libertação da perspectiva custodial no âmbito arquivístico (p. 13).

Outro pesquisador, Peter Scott, um dos mais proeminentes teóricos a abordar a natureza de multiprodução das séries documentais, fundamentado em sua vasta experiência no âmbito australiano, emerge como um crítico contundente da perspectiva custodial do fundo. Ele postula a série documental como o núcleo da descrição arquivística (Cook, 2017, p. 23). Nessa ótica, Scott sustenta que uma série pode ser gerada por diversos produtores, seja de forma concomitante, seja sucessiva.

Para exemplificar os conceitos no âmbito dos arquivos legislativos, utilizamos a tramitação ordinária de um projeto de lei que circula por diversas comissões em uma das casas legislativas (Câmara dos Deputados ou Senado Federal)

e, depois, é encaminhado à casa revisora para a análise de outras comissões. A depender da interpretação dos arquivistas, aplica-se a visão maximalista à noção de fundo – os produtores da série Projetos de Lei seriam a Câmara dos Deputados e o Senado Federal, portanto, as comissões seriam apenas autoras. Na visão minimalista, todas as comissões por onde o projeto de lei tramitou podem ser reconhecidas como suas entidades produtoras. Nesse caso, a série Projetos de Lei estaria descrita e associada a todos os fundos dos respectivos produtores (inclusive as comissões), sendo representada em seus quadros de arranjo e descrita nos instrumentos de pesquisa pertinentes desses fundos. Sob tal perspectiva, teóricos como Peter Scott argumentam que procedimentos de arranjo e descrição devem ser mais maleáveis do que aqueles preconizados pela abordagem custodial dos arquivos. Nessa visão, uma série não deve ser rigidamente associada a um produtor, limitada a única estrutura de arranjo e documentada exclusivamente em um instrumento de pesquisa, como um inventário de um fundo específico, como explicado por Cook (2017):

À medida que empresas e órgãos públicos nessas circunstâncias adotam novos modelos de informação baseados em planejamentos de dados corporativos e administração de recursos de dados, a ideia de um documento pertencente fisicamente a um lugar ou até mesmo a um sistema está desmoronando frente a novos paradigmas conceituais, em que a “produção” é um processo fluido de manuseio de informação de várias fontes em uma enorme quantidade de formas, ao invés de algo que conduz a um produto estático, fixo e físico. Para os arquivistas, isso sinaliza que a era custodial está dando passagem à era pós-custodial, quando a curadoria de objetos físicos definirá muito menos a profissão que uma compreensão das inter-relações conceituais entre estruturas produtoras, suas funções vitais e os documentos então resultantes. (Cook, 2017, p. 47)

No contexto da prática arquivística pós-custodial, Peter Scott evidencia que um dos principais desafios enfrentados pelos arquivistas reside em “determinar a qual produtor ou fundo uma dada série pertence” (Cook, 2017, p. 33). Como solução, sugere a implementação de remissivas nos instrumentos de pesquisa e a elaboração de diversas listagens específicas que estabeleçam referências cruzadas entre os produtores e as séries documentais inter-relacionadas, a exemplo das listas de controle de autoridades. De acordo com Cook, a abordagem proposta por Peter Scott, que contempla a concepção de “múltiplas entradas e a distinção entre as descrições dos contextos administrativos e dos documentos físicos, constitui-se em uma inovação verdadeiramente revolucionária” (p. 54).

Nas discussões teóricas na comunidade arquivística, durante a elaboração da primeira versão de consulta do modelo conceitual RiC, entre 2012 e 2016, é importante notar que os especialistas do ICA não apenas fizeram diversas críticas em relação às normas de descrição, mas também questionaram abordagens conceituais mais abrangentes com impacto direto na descrição. Isso implica em questionamentos aos entendimentos mais tradicionais relacionados aos princípios da proveniência e do respeito à ordem original.

Entre as críticas lançadas pelo Conselho Internacional de Arquivos, incluem-se aquelas dirigidas à visão clássica e restrita do princípio da proveniência,³ sendo considerada prejudicial à descrição dos arquivos, na perspectiva do Egad. Na concepção do modelo conceitual RiC, a abordagem tradicional tem como objetivo preponderante registrar e preservar a relação dos documentos com o produtor/acumulador, negligenciando outras pessoas, instituições e contextos relacionados aos registros, em seus momentos de produção, acumulação e uso. Dessa forma, o modelo RiC promove a visão de um princípio da proveniência “ampliado”, direcionando a descrição para outros contextos desaplicados nos quais os arquivos se relacionam e que vão além da mera descrição da proveniência ou da acumulação. O modelo RiC ambiciona estabelecer as relações destes documentos na intrincada camada de contextos, interações e interdependências existentes entre eles, de forma a refletir mais eficientemente a complexidade social e dos registros documentais (ICA, 2021b). Justifica-se que

todas essas observações levam à conclusão de que a origem dos documentos é muito mais complexa do que o entendimento de fundo documental há muito estabelecido. Os documentos em um fundo não existem apenas em relação a ele, nem o fundo documental existe isoladamente. Tanto o fundo quanto os documentos nele contidos existem em camadas de contextos interconectados do passado, do presente e do futuro. (ICA, 2021a, p. 4)

O segundo aspecto relacionado às novas propostas teóricas e práticas de descrição relaciona-se com o princípio do respeito à ordem original,⁴ que, em termos gerais, visa à manutenção da ordem original para fins de preservação do inter-relacionamento existente entre os documentos dentro do fundo

3 O princípio da proveniência, também chamado de princípio do respeito aos fundos, segundo o Dicionário Brasileiro de Terminologia Arquivística, é um “princípio básico da arquivologia segundo o qual o arquivo produzido por uma entidade coletiva, pessoa ou família não deve ser misturado aos de outras entidades produtoras”.

4 Princípio do respeito à ordem original: princípio segundo o qual o arquivo deveria conservar o arranjo dado pela entidade coletiva, pessoa ou família que o produziu.

documental. Segundo o ICA (2021b), esse princípio sempre foi controverso na comunidade arquivística, tanto no entendimento de original quanto de ordem. Em relação à ordem original, inúmeras questões são apresentadas, como a verificação de qual seria a ordem original definitiva em um contexto em que os documentos são acumulados dentro dos processos dinâmicos, fluidos e mutáveis, ocasionando, na maioria das vezes, reordenações do conjunto documental nas fases do ciclo documental. Ademais, à qual ordem o princípio se refere: a ordem intelectual entre os documentos ou a disposição física original? Somam-se ainda os muitos casos de documentos que entram nos arquivos sem uma ordem tangível – nem física, nem intelectual –, ocasião em que arquivistas reconstroem ou estabelecem uma ordenação.

Observa-se que, na perspectiva do ICA, a concepção ampliada do princípio da proveniência não descarta, em momento algum, a conexão primordial dos documentos com seus respectivos fundos de origem. Tal vínculo representa uma garantia fundamental para a preservação do contexto original de produção e acumulação dos documentos, sendo, ademais, um princípio identitário e essencial dos arquivos.

No entanto, adota-se uma visão mais inclusiva, na qual se incorporam outros contextos que não são tradicionalmente registrados nos sistemas de registro e descrição em arquivos. Além do contexto de origem, considera-se o contexto de uso dos documentos, por exemplo, as interações de pesquisadores nos arquivos, que muitas vezes são registradas em fichas de pesquisa ou em bancos de dados específicos. Estes bancos contêm uma variedade de dados relacionados às pesquisas, incluindo o nome do pesquisador, a instituição, o tema da pesquisa e os documentos selecionados, entre outros. É importante notar que, depois da conclusão da pesquisa, quando os documentos selecionados são utilizados em teses, dissertações, artigos científicos, reportagens, processos judiciais e outros produtos similares, geralmente não existe a tradição de registro unificado no sistema de descrição do acervo. Vale a pena mencionar o elemento “nota de publicação” que desempenhava essa função na Isad(G). No entanto, deve-se considerar que os sistemas de informações em arquivos que contêm os metadados sobre documentos e arquivos muitas vezes estão isolados dentro da instituição arquivística, seguindo a lógica de “silos de dados”.

O mesmo pode ser aplicado aos dados relacionados às ações de preservação do acervo. Em alguns casos, mesmo quando documentos analógicos passam por intervenções nos laboratórios de restauração, higienização, desinfestação ou recondicionamento, os dados gerados por esses procedimentos, embora mantenham referências de controle sobre os documentos, são mantidos em bases de dados e sistemas específicos, também seguindo a lógica de “silos de dados” sobre os arquivos.

Por fim, o viés crítico é reforçado por representantes do ICA, que salientam que a descrição arquivística é frequentemente percebida como autocontida e desvinculada do contexto externo dos arquivos (Pitti, 2016). Diante desse cenário, o modelo conceitual RiC sugere uma transição da descrição hierarquizada multinível atual, nos moldes das normas internacionais, para um modelo conceitual de dados, que se baseia em descrição multidimensional, considerada mais versátil e expressiva.

Desse modo, o modelo conceitual RiC sugere uma nova compreensão de conceitos arquivísticos, como o princípio da proveniência (ampliado) e a integração das quatro normas vigentes, que não ocorreu na prática por meio dos sistemas: um objetivo teórico já antevisto na teoria arquivística anterior à concepção das normas internacionais, quando a teoria em questão era revisitada à luz das emergentes tecnologias e da consolidada perspectiva pós-custodial dos arquivos. Conforme exemplifica Cook,

usando automação para combinar ou relacioná-las, qualquer combinação de produtores e séries é possível: múltiplos produtores a uma única série, múltiplas séries de um único produtor, múltiplos produtores com séries múltiplas, mas relacionadas, séries com séries, e produtores com produtores, seja para cima ou para baixo na tradicional cadeia mono-hierárquica vertical, ou para os lados, nas novas estruturas organizacionais horizontais. (Cook, 2017, p. 55)

Inovação tecnológica no modelo conceitual RiC: arquivos e web semântica

O modelo conceitual para descrição de arquivos RiC apresenta renovações teóricas e práticas na arquivologia, bem como inovações possibilitadas pelas novas tecnologias de informação e comunicação (TICs), mais precisamente as tecnologias semânticas. Dessa forma, ao analisarmos o modelo RiC é necessário entendermos que há por parte do Conselho Internacional de Arquivos a adesão a uma tecnologia específica: a web semântica.

O conceito foi introduzido por Tim Berners-Lee, James Hendler e Ora Lassila em 2001: “A web semântica não é uma web separada, mas uma extensão da atual, na qual a informação é dada com um significado bem definido, permitindo melhor que computadores e pessoas trabalhem em cooperação” (Berners-Lee et al., 2001, p. 28-37). Dessa forma, a web semântica visa tornar os dados na web mais compreensíveis e utilizáveis por humanos e máquinas.

Web semântica (Decker et al., 2000; Berners-Lee et al., 1999) é o nome genérico deste projeto, capitaneado pelo W3C, que pretende embutir inteligência e contexto nos códigos XML utilizados para confecção de páginas web, de modo a melhorar a forma com que programas podem interagir com estas páginas e também possibilitar um uso mais intuitivo por parte dos usuários. (Souza et al., 2004, p. 133)

Para uma compreensão mais abrangente da web semântica, é essencial entender as transformações pelas quais a internet passou. A world wide web, mais conhecida como web, foi fundada entre 1989 e 1991, período em que Tim Berners-Lee desenvolveu e aprimorou conceitos anteriores dos cientistas Bush e Ted Nelson, relacionados à ideia de hipertexto (Pickler, 2007).

No entendimento de Sharma (2022), na primeira fase da world wide web, conhecida como web 1.0 ou web de documentos, o conteúdo era predominantemente estático e compartilhado em páginas pessoais, com atualizações realizadas unicamente pelos proprietários dos sites. A estrutura se baseava em um computador principal conectado à rede, utilizado como servidor de arquivos, e as páginas eram construídas usando linguagens como Server Side Includes (SSI) e Common Gateway Interface (CGI).

A web 2.0 ou web social simboliza a transformação da internet em uma plataforma participativa, destacando-se pela geração de conteúdo pelos usuários e pela ênfase na interação e colaboração. Durante esse período, houve uma evolução na forma como a web era percebida e utilizada, com destaque para a personalização avançada através de APIs e a diversidade de ferramentas on-line, como podcasts e redes sociais.

A web 3.0, também conhecida como web de dados, representa a terceira geração. Segundo Sharma (2022), sua distinção reside na aplicação da inteligência artificial para organizar e recuperar dados, transcendendo a simples busca por páginas. Essa evolução abrange a extração de dados da web para formar bancos de dados, favorecendo o avanço das tecnologias de *back-end* (infraestrutura), que durante anos priorizaram inovações do *front-end* (interface). Uma característica notável da web 3.0 é a natureza aberta e compartilhada dos dados, permitindo que um único site ou conjunto de dados seja interpretado de diversas maneiras. Heath e Bizer (2011) afirmam: “A web de dados que está se formando através da publicação de *linked data* na web está possibilitando uma mudança de uma web de documentos para uma web de dados”.⁵

⁵ “The Web of Data that is taking shape via the publication of Linked Data on the Web is enabling a shift from a Web of Documents to a Web of Data”, Heath e Bizer (2011).

Um recurso central da web 3.0 é a web semântica, que busca organizar a informação do mundo de maneira mais estruturada, considerando a perspectiva das máquinas, e não apenas a humana. Utiliza-se, para isso, a Web Ontology Language (OWL), uma linguagem de ontologia específica, aplicada a domínios de conhecimento particulares, possibilitando que as máquinas compreendam e gerem informações de forma autônoma. Diferentemente das versões anteriores, a web semântica prioriza a melhoria da semântica das tecnologias de criação, compartilhamento e conexão de dados, realizando buscas e análises que consideram o significado real das palavras, em vez de se limitar a palavras-chave ou números (Sharma, 2022).

Portanto, torna-se necessário que os arquivos e os arquivistas estejam cientes de que existe um amplo ecossistema na web semântica no qual o modelo conceitual RiC e sua respectiva ontologia RiC-O se inserem. Esse modelo conceitual representa uma ferramenta central usada no processo de modelagem e enriquecimento semântico dos dados de arquivos, conforme esclarecido por Nazário e Dias.

O enriquecimento semântico pode ser compreendido como um processo de atribuição de maior significado aos dados e metadados, através do uso da semântica atribuída por vocabulários preexistentes, sinônimos e informações de proveniência. Esse processo tem por objetivo facilitar a compreensão, a integração e o processamento de dados por pessoas e máquinas. (Nazário; Dias apud Lira, 2021, p. 3)

De acordo com a visão de Santarém Segundo e Coneglian (2016), a web semântica tem como conceito central “construir uma rede de informações onde os nós estejam semanticamente ligados, formando um grande grafo global, com informações advindas de [...] fontes diferentes ao redor do planeta” (Santarém Segundo; Coneglian, 2016, p. 220). Assim, a web semântica é um conjunto de conceitos, modelos, metodologias e tecnologias em transformação, enriquecido por estudos de caso aplicados ao longo dos anos. Ela é particularizada pela existência do nível semântico, pela automação por máquinas e por mecanismos como serendipidade e motores de inferência.

Inicialmente, é essencial que se tenha um conjunto de dados disponibilizado conforme algum padrão de interoperabilidade, como a linguagem Extensible Markup Language (XML) ou o Resource Description Framework (RDF). Em seguida, destaca-se a importância da elaboração de modelos conceituais, ontologias e vocabulários, como o modelo RiC, no qual desempenha um papel crucial na organização e representação dos conceitos de domínios

específicos de conhecimento. Estes estabelecem axiomas que delineiam a lógica das inferências no processo de semantização. Após a construção dessas ontologias, torna-se necessário formalizá-las em uma linguagem computacional, sendo a Web Ontology Language (OWL) uma das opções mais contemporâneas e eficazes para essa finalidade (Santarém Segundo; Coneglian, 2016).

O modelo conceitual RiC é considerado de alto nível, sendo possível a elaboração de modelos conceituais mais específicos para segmentos de arquivos, como arquivos legislativos, arquivos judiciários, arquivos universitários etc. Ao mesmo tempo, o modelo RiC já apresenta sua especificação em linguagem OWL, o Records in Contexts-Ontology (RiC-O). Um dos recursos mais centrais da web semântica é a criação de Uniform Resource Identifiers (URIs) persistentes para os recursos na web, “usados para identificar qualquer coisa que tenha identidade. Não apenas documentos na web, mas também pessoas e objetos físicos” (Berners Lee, 2006).

Para consultar o conjunto de dados já semantizados pelo RiC, é possível utilizar a tecnologia específica de busca e recuperação da web semântica, o Protocol and RDF Query Language (SPARQL). Esses dados podem ser representados visualmente como grafos RDF. Essas ações, quando aplicadas aos dados, alinham-se à proposta geral do *linked data* ou do *linked open data*, conforme elucidado por Santarém Segundo e Coneglian (2016):

Essa característica torna os dados não apenas acessíveis e processáveis por máquinas, mas passíveis de processos de organização que podem facilitar a geração de novos dados, apresentação de resultados, relação com outros grupos de dados, aumento do conhecimento para tomadas de decisão, novos modelos de dados gerados a partir do relacionamento e cruzamento de dados de várias esferas governamentais, além da geração de novos modelos mentais de apresentação da informação de forma a facilitar o acesso dos dados pela sociedade civil. (Santarém Segundo; Coneglian, 2016, p. 223)

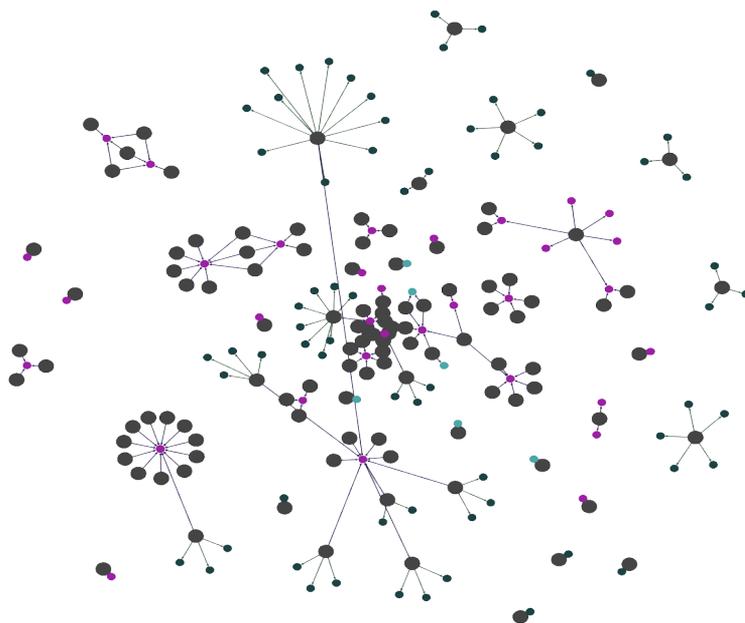


Figura 2 – Exemplo de grafo de dados do Arquivo Nacional da França com fundos, documentos e autoridades relacionadas. Fonte: Projeto PIAAF/Arquivo Nacional da França

Conforme proposto por Tim Berners-Lee, existe uma escala de cinco estrelas que ilustra o grau de disponibilidade de um dado de acordo com os padrões do *linked open data* (LOD). Essa escala (apresentada no Quadro 2) avalia diversos aspectos, incluindo a presença de formatos com licenças abertas, o nível de estruturação dos dados, a legibilidade dos dados por máquinas, a publicação de dados na web em formatos não proprietários e a utilização do formato RDF. Além disso, verifica-se que os metadados em RDF contêm referências de URIs (Uniform Resource Identifier) para outras bases de dados relacionadas.

Quadro 2 – 5 estrelas de dados abertos

Estrelas	Descrição
1	Dados publicados na web em qualquer formato sob uma licença aberta
2	Dados publicados de forma estruturada e em um formato que seja legível por máquinas
3	Dados publicados de forma estruturada na web e utilizando um formato não proprietário
4	Dados publicados na web em formato RDF
5	O metadado no formato RDF indica URIs de outras bases de dados relacionadas

Fonte: Paula; Deus; Barbosa (2020).

Ao analisar as tecnologias predominantes nas últimas décadas, os especialistas do Egad/ICA destacam que, a partir dos anos 1980, houve um deslocamento em direção à adoção de dois padrões essenciais: o Structured Query Language (SQL) e o Standard Generalized Markup Language (SGML), que mais tarde deu origem ao Extensible Markup Language (XML). Conforme apontam Gueguen et al. (2013), existe uma distinção crucial entre esses padrões. As tecnologias de marcação, exemplificadas pelo SGML/XML, são categorizadas como centradas em documentos. Elas foram concebidas primordialmente para documentos textuais tradicionais, tais como ensaios, artigos e livros. De maneira geral, esses documentos apresentam características comuns, como a presença de subpartes em número variável, a importância da sequência ou ordenação, uma estrutura semirregular ou hierárquica, uma combinação diversificada e marcada de tipos de dados e uma multiplicidade de inter-relações entre elas.

As tecnologias voltadas para bases de dados, a exemplo do SQL, são frequentemente descritas como orientadas a dados. Distinguindo-se por sua segmentação em campos interconectados de maneira consistente, elas proporcionam aos usuários do sistema ampla versatilidade de configurações, adaptando-se aos variados usos e abordagens. Em geral, essas tecnologias apresentam campos em quantidade regular, cuja sequência não é de suma importância. Cada campo é preenchido com dados específicos, campos inter-relacionados seguem uma hierarquia predefinida e simples, e o conteúdo obedece a critérios rígidos quanto a sua forma e estruturação. Conforme elucidam os especialistas do Egad:

As duas tecnologias, base de dados e de marcação, usaram diferentes abordagens para representar os dados. Cada uma se aplica melhor em usos específicos. Os usuários tiveram que analisar os seus dados, determinar suas características, ao mesmo tempo que tiveram que considerar como eles queriam usar os dados para então decidir qual das duas tecnologias adotar. (Gueguen et al., 2013, p. 571, tradução nossa)⁶

Ao abordar a implementação das duas tecnologias em arquivos, especialmente no que tange à descrição ao nível do fundo, os autores destacam que não se observou uma adesão integral e exclusiva a uma determinada tecnologia. Conforme aponta Gueguen et al. (2013), não prevaleceu uma tecnologia em detrimento da

⁶ “The two technologies, database and markup, take different approaches to representation of data. Each is best suited to particular uses. Users thus have needed to analyze their data, to ascertain their features, while also bearing in mind how they want to use the data in order to decide which of the two technologies to use” (Gueguen et al., 2013, p. 571).

outra. Ao invés disso, o que se evidenciou foi uma integração, uma convergência entre elas. As tecnologias baseadas em banco de dados, como o SQL, foram predominantemente adotadas em sistemas voltados para o registro e a atualização de descrições. Por outro lado, as tecnologias ancoradas em linguagens de marcação, como o XML, mostraram-se mais relevantes para a comunicação e transmissão de dados, seja entre máquinas ou entre máquinas e usuários finais.

O modelo RiC introduz uma inovação ao propor um terceiro padrão para a representação de informações e dados em sistemas, baseado nas tecnologias de grafos. Este padrão é representado pelo citado Resource Description Framework (RDF), desenvolvido pelo World Wide Web Consortium (W3C). O RDF estabelece a representação de dados em triplas, seguindo a estrutura lógica de sujeito-predicado-objeto. Em contraste com padrões anteriores, como o XML, que utilizava grafos hierárquicos ou estruturas em árvore, o RDF emprega grafos em configurações de teias. Esse formato proporciona abordagens mais expressivas e inovadoras para representar a realidade, percebendo-a como uma extensa e dinâmica rede de indivíduos e objetos interligados, contextualizados temporal e espacialmente (Gueguen et al., 2013, p. 572, tradução nossa). Assim, o RDF fortalece as tecnologias semânticas, aspirando à construção de uma rede coesa de informações interconectadas, articuladas de maneira lógica e com clareza semântica.

Como observado anteriormente, a web semântica, em sua essência, contempla procedimentos e padrões que visam transformar a web em uma ampla rede semântica de declarações interligadas sobre o mundo e suas intrincadas relações. Conforme apontado por Gueguen et al. (2013), há uma crescente tendência em diversos domínios da vida humana de adotar a perspectiva de dados abertos e interconectados. Essa metodologia permite que máquinas explorem a rede, propiciando a emergência de novos conhecimentos e padrões, nos quais a informação coletiva transcende a mera agregação de seus componentes individuais. Vale destacar que um pilar fundamental da web semântica é a disponibilização gratuita de dados na internet, promovendo sua reutilização em sistemas distintos daquele de sua origem. Como citado previamente, tal movimento é reconhecido como *linked open data* (LOD).

No âmbito das tecnologias semânticas, destaca-se a teoria geral sobre modelos conceituais adotada pelo RiC, e outro modelo conhecido como modelo entidade-relacionamento (ER ou MER). Oriundo da engenharia de software, esse modelo é amplamente reconhecido na teoria e prática de bancos de dados. Sua concepção remonta ao artigo seminal de Peter Chen, publicado em 1976, no qual são introduzidos conceitos fundamentais, como entidades, atributos e relacionamentos.

Com o surgimento de base de dados de objetos relacionais, a abordagem ER foi adaptada para refletir os métodos de representação de dados orientados a objetos. Com o surgimento de tecnologias semânticas baseadas em grafos, mais métodos de modelagem e ferramentas surgiram para dar suporte a eles. O W3C, baseado no RDF, desenvolveu o Web Ontology Language (OWL) como um método de formalização de “representação de conhecimento” para ontologias de domínios específicos. (Gueguen et al., 2013, p. 574, tradução nossa)⁷

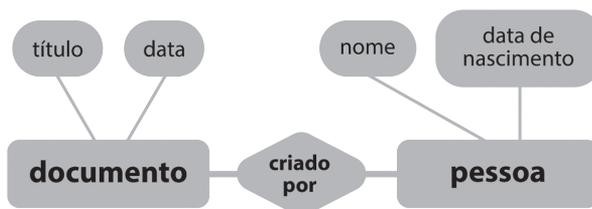


Figura 3 – Modelo entidade-relacionamento (ER) aplicado aos arquivos no modelo RiC. Fonte: Adaptada de ICA (2021b, tradução nossa)

De acordo com Pitti (2016), para efetivar a representação do conhecimento em domínios específicos, como arquivos, bibliotecas e museus, é imprescindível a elaboração de ontologias. Contudo, para que essas ontologias atendam aos propósitos da semantização e permitam o processamento computacional, é fundamental que estejam em uma linguagem formal. Nesse contexto, o W3C, embasado no padrão RDF, introduziu a Web Ontology Language (OWL) para formalizar tais ontologias e avaliar a lógica e conformidade dos modelos de dados. O desafio reside na transição de tecnologias tradicionais, como XML ou bancos de dados relacionais SQL, para bases de dados não relacionais (NoSQL) e tecnologias em grafos, alinhadas aos padrões de metadados para web, como o RDF.

Em síntese, diante de tanta inovação, recomenda-se que arquivistas incorporem tecnologias semânticas e *linked open data* (LOD) aos dados descritivos dos acervos arquivísticos. Apesar da complexidade, essas tecnologias, já adotadas por bibliotecas, museus, setores comerciais, industriais e científicos, apresentam vantagens significativas, como expressividade e flexibilidade na manipulação de dados (Pitti, 2016).

⁷ “With the emergence of object relational databases, the ER approach was adapted to reflect object-oriented data representation methods. With the emergence of graph-based semantic technologies, more modeling methods and tools have emerged to support them. The W3C, based on RDF, developed the Web Ontology Language (OWL) as a method for formal knowledge representation of domain-specific ontologies” (Gueguen et al., 2013, p. 574).

O modelo conceitual Records in Contexts

Após uma década de dedicação dos especialistas do Egad, o ICA publicou três versões de consulta do modelo conceitual Records in Contexts entre 2016 e 2021, buscando apreciação da comunidade arquivística. Em novembro de 2023, a versão oficial desse modelo foi divulgada. Como abordado neste artigo, o referido modelo inova em diversos aspectos. Ele incorpora discussões teóricas e práticas recentes, considerando conceitos como a expansão do princípio da proveniência, bem como críticas às normas de descrição estabelecidas pelo ICA e a necessidade de atualização tecnológica com ênfase em tecnologias semânticas. O modelo ressalta que seu público-alvo é consideravelmente mais abrangente do que o das normas de descrição estabelecidas pelo ICA. Ele engloba não somente os arquivistas responsáveis por arquivos permanentes, mas também aqueles dedicados à gestão documental, considerando as fases corrente e intermediária. Essa inclusão visa alinhar as práticas de descrição desde a origem, dada a complexidade inerente ao ambiente arquivístico digital. Adicionalmente, abarca profissionais de áreas correlatas ao patrimônio documental, como biblioteconomia e museologia, visando projetos de integração de dados. Considera também os desenvolvedores de sistemas, tendo em vista a intrincada natureza do modelo RiC e os desafios de sua implementação. Por fim, contempla os pesquisadores dos arquivos, permitindo a utilização e reutilização dos dados sob variadas perspectivas acadêmicas (ICA, 2021b).

Como mencionado, o modelo conceitual RiC é fundamentado no modelo entidade-relacionamento (ER) e na abordagem da descrição multidimensional, englobando a definição de entidades, atributos e relacionamentos. As entidades representam os elementos centrais de interesse para profissionais de arquivologia envolvidos na gestão de documentos e nos arquivos permanentes, uma vez que “fornecem o contexto intelectual que atende às necessidades de gerenciamento físico, preservação, localização, utilização e compreensão dos documentos ao longo de sua trajetória histórica” (ICA, 2021b, p. 17, tradução nossa). No total, foram estabelecidas 19 entidades, as quais estão organizadas em quatro níveis hierárquicos, conforme apresentado no Quadro 3.

Quadro 3 – Entidades do modelo RiC em quatro níveis de hierarquia

1º nível	2º nível	3º nível	4º nível	
RiC-E01 coisa ou entidade	RiC-E02 fonte documental	RiC-E03 conjunto documental		
		RiC-E04 documento		
		RiC-E05 parte de documento		
	RiC-E06 registro documental (manifestação documental ou inscrição)			
	RiC-E07 agente	RiC-E08 pessoa		
		RiC-E09 grupo		RiC-E10 família
				RiC-E11 entidade coletiva
		RiC-E12 cargo		
		RiC-E13 mecanismo		
	RiC-E14 evento	RiC-E15 atividade		
RiC-E16 regra	RiC-E17 atribuição			
RiC-E18 data				
RiC-E22 localidade				

Fonte: ICA (2023, tradução nossa).

As entidades foram especificadas em 19 quadros contendo ID, nome, definição, nota de escopo, exemplos e comentários, quando necessário, conforme exemplo no Quadro 4 para entidade “documento”.

Quadro 4 – Exemplo de quadro explicativo de entidade no modelo RiC

ID	RiC-E04
Nome	Documento
Definição	Informação registrada por pelo menos um método em um suporte durável e recuperável, por um agente, no curso de sua vida ou atividades
Nota de escopo	Documento é um tipo de fonte documental (RiC-E02). Um documento pode ser, ou tem que ter, ao menos um registro. Um documento pode ter mais de um registro. Um registro originado de outro registro de documento pode ser considerado um registro do mesmo documento ou um registro de novo documento, a depender do contexto. Um documento pode servir para várias finalidades, apesar de serem sempre documentos ou provas de uma atividade.
Exemplos	Ato de nomeação de John Bambridge, xerife de Leicestershire, com o 3º Grande Selo do Charles I anexado Esboço do mapa da Península do Qatar Mensagem de e-mail relativa ao acordo de participação no Congresso do ICA em Seul (2016), contendo dois anexos e um sinal digital Registro de entrada do trabalhador alemão Jacob Schwarz na Hospedaria de Imigrantes da Ilha das Flores, vindo da Antuérpia no navio Graf Bismark e de sua saída para Porto Alegre
Comentários	-

Fonte: ICA (2023, tradução nossa).

Cada entidade apresenta características distintas, denominadas atributos, que desempenham um papel essencial na definição da identidade da própria entidade. Conforme indicado nas diretrizes do ICA (2021b), o processo de descrição das entidades envolve a ênfase nas suas características intrínsecas. No entanto, ao contrário das normas tradicionais do ICA, que possuíam conjuntos uniformes de elementos de descrição independentemente do tipo de entidade a ser descrita – seja ela um fundo, grupo, série, dossiê, item (conforme a Isad-G), uma entidade coletiva, pessoa ou família (de acordo com a Isaar-CPF), ou ainda uma função, atividade ou tarefa (segundo a ISDF) –, no modelo RiC, os atributos para cada entidade são predefinidos e individualizados. Embora alguns atributos sejam compartilhados por várias entidades, também existem aqueles exclusivos de entidades específicas. O ICA (2021b) reconhece que a distinção entre o que deve ser considerado uma entidade e o que deve ser classificado como um atributo não é uma tarefa trivial. No entanto, o Egad optou por designar como entidades apenas aqueles fenômenos que são fundamentais no contexto da gestão de documentos e arquivos permanentes.

Além disso, os atributos no modelo RiC são acompanhados de um esquema de valores, o qual não é prescritivo nem proibitivo, mas indicativo. Em outras palavras, cada entidade está vinculada a um atributo por meio de uma afirmação, como exemplificado a seguir: o documento (entidade) “A” possui o idioma (atributo) “B”, em que “B” representa o valor do idioma. Os esquemas de valores dos atributos no modelo conceitual RiC podem ser categorizados em quatro tipos distintos: texto livre, texto baseado em modelo, valor controlado e valor baseado em regra (ICA, 2021b).

O Egad identificou um total de 41 atributos no modelo RiC, que abrangem uma variedade de características, incluindo tipo de atividade, nota de autenticidade, dimensões, tipo de suporte, classificação, condições de acesso, condições de uso, formato, tipo de entidade coletiva, entre outras. Esses atributos estão associados às 19 entidades definidas no modelo. Para facilitar a compreensão desses atributos, o Egad desenvolveu um quadro padronizado para cada um deles (Quadro 5), contendo informações essenciais, como número de identificação, nome, definição, entidades às quais se aplicam, especificações, possibilidade de extensões, repetibilidade, escopo e exemplos ilustrativos.

Quadro 5 – Exemplo de quadro explicativo do atributo classificação

ID	RiC-A07
Nome	Classificação
Definição	Um termo, número ou sequência alfanumérica geralmente proveniente de um vocabulário ou esquema de classificação externo, que qualifica o registro ou parte do registro
Domínio de entidade	Conjunto de documentos; documento; parte do documento
Especificações	
Extensibilidade	Pode ser estendido com qualquer número de atributos específicos
Repetibilidade	Repetível
Esquema de valor	Valor controlado; texto baseado em modelo
Escopo	Pode ser utilizado na descrição de um conjunto de registros quando o valor do atributo é compartilhado por alguns ou todos os membros do conjunto de registros. O atributo não deve ser confundido com o atributo identificador, embora, em alguns casos, as informações possam ser as mesmas
Exemplos	BUD-01-F002 [exemplo do número de classificação de um plano de classificação]; gestão de recursos humanos; registro de estudantes; assuntos financeiros; itens digitalizados

Fonte: ICA (2021b, tradução nossa).

Após fornecer as definições dos 41 atributos por meio de quadros explicativos, o modelo RiC disponibiliza a lista das entidades acompanhadas de seus atributos correspondentes (Quadro 6). A coloração atribuída aos atributos indica aqueles que são compartilhados com entidades superiores, em conformidade com a estrutura hierárquica das 22 entidades, enquanto aqueles sem codificação de cores são atributos exclusivos da entidade em questão.

Quadro 6 – Exemplo de quadro de atributos da entidade documento

RiC-E04	Documento	RiC-E04	Documento
ID do atributo	Nome do atributo	ID do atributo	Nome do atributo
RiC-A16	Nota descritiva	RiC-A08	Condições de acesso
RiC-A22	Identificador	RiC-A09	Condições de uso
RiC-A28	Nome	RiC-A10	Gênero documental
RiC-A21	Histórico	RiC-A17	Espécie documental
RiC-A35	Dimensão	RiC-A24	Integridade
RiC-A38	Âmbito e conteúdo	RiC-A25	Idioma
RiC-A40	Estrutura	RiC-A26	Status legal
RiC-A03	Nota de autenticidade	RiC-A39	Status (tradição documental)
RiC-A07	Classificação		

Fonte: ICA (2021b, tradução nossa).

Por último, o modelo conceitual RiC apresenta o conjunto de relacionamentos, que tem como objetivo conectar as entidades. Esses relacionamentos desempenham um papel central no modelo RiC, pois descrevem todos os contextos nos quais os documentos são produzidos, mantidos e utilizados, o que é uma das principais propostas do modelo. Além disso, eles são considerados “mais simples e flexíveis”, pois utilizam o modelo de grafos e o Resource Description Framework (RDF), padrões voltados para a web semântica.

A versão do modelo RiC publicada em novembro de 2023 inclui um total de 78 relacionamentos. Eles podem ser agrupados em 13 categorias, e, de acordo com as orientações do ICA (2021b), compreender o tipo de relacionamento pode ajudar a esclarecer seu papel no sistema de descrição. As categorias incluem os tipos de relacionamentos possíveis: todo-parte, sequencial, por assunto, entre uma fonte documental e seu registro, proveniência, entre tipos de registro, de gestão, de agente para agente, de eventos, de regras, de datas e espaciais. Dessa forma, eles podem variar de mais genéricos a mais específicos.

Os relacionamentos seguem uma hierarquia que vai do geral ao específico. O relacionamento mais amplo, ou mais geral, é “está relacionado com”, que pode conectar qualquer entidade RiC-CM e não faz declaração específica sobre como ou porque essas entidades estão conectadas. Conforme você desce na hierarquia, cada relacionamento se torna mais específico, herdando as propriedades dos relacionamentos imediatamente acima dela na hierarquia. (ICA, 2021b, p. 73, tradução nossa)

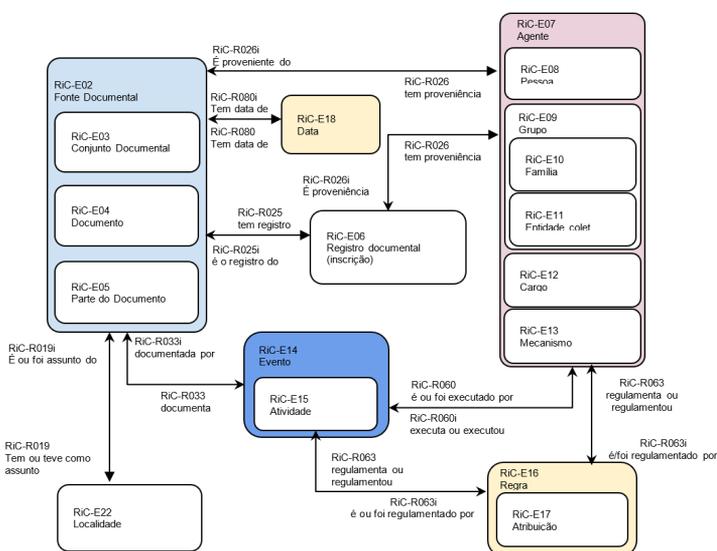


Figura 4 – Relacionamentos entre entidades do modelo RiC. Fonte: ICA (2023, tradução nossa)

Conclusão

O modelo conceitual RiC introduz uma série de mudanças significativas na teoria e prática arquivísticas. O princípio da proveniência, que é a pedra angular dos arquivos, passa a ser compreendido de maneira mais abrangente, reconhecendo os contextos complexos nos quais os documentos de arquivo são produzidos, mantidos e utilizados. Isso implica não apenas a intenção de preservar os diversos contextos, mas também de modificar a forma como entendemos o acesso aos arquivos. O Conselho Internacional de Arquivos defende ainda que a entrada de dados, que tradicionalmente era hierárquica, deve agora ser multidimensional e incluir novos elementos descritivos, como *normas*, *locais* e *datas*. Além disso, a disponibilização e o compartilhamento dos dados arquivísticos com os usuários devem ser feitos de forma aberta e em conformidade com os padrões potencializadores de seu uso e reúso, como os da web semântica e tecnologias afins.

Uma das inovações mais marcantes é a integração de bases de dados dos acervos, seja dentro da mesma instituição arquivística ou entre instituições que compartilham entidades e mantêm relacionamentos. Isso promove a disseminação da ideia de integração dos dados culturais ou do patrimônio cultural e documental compartilhados por arquivos, bibliotecas e museus (LAMs).

Através do modelo entidade-relacionamento (ER) e da ontologia propostos, denominados RiC-CM e RiC-O, os arquivistas irão aprimorar a estruturação dos dados referente aos arquivos e documentos. Com isso, alcançam-se diversas possibilidades, tais como: interoperabilidade entre sistemas; inferência e raciocínio automático pelas máquinas; enriquecimento semântico e integração de dados; buscas refinadas e reutilização dos dados pelos usuários.

Apesar das inovações tecnológicas, o modelo RiC é uma reafirmação da visão do RiC sobre a descrição arquivística, que foi consolidada em 1989 e continua relevante até hoje. A visão se baseia na Declaração de Princípios de Descrição Arquivística, que busca uma descrição arquivística consistente, relevante e explícita, facilitando o acesso e a troca de informações sobre os acervos arquivísticos e buscando integrar descrições de diferentes instituições em sistemas de informação unificados (Gueguen et al., 2013, p. 567, tradução nossa).

Segundo Chenard (2020), a web semântica e o modelo RiC seriam muito mais que uma mudança de formato, propondo uma evolução mais profunda na forma de descrever nossos fundos documentais: do instrumento de pesquisa a ser lido (inventários) para o instrumento de pesquisa a ser consultado (amplos bancos de dados conectados); das simples descrições à estruturação

de dados; dos modelos de normas (Isad(G), Isaar, ISDF, Isdiah) e dos instrumentos de pesquisa aos modelos atuais de dados de diversas origens conectados; da possibilidade limitada de pesquisa para uma infinidade de possibilidades de detalhamentos. E o autor completa:

Atomizar ainda um pouco mais a informação em múltiplos campos, segundo um vocabulário controlado, indexar em grande escala, registrar informações sistematicamente em certos campos até o momento um pouco esquecidos. Em síntese, preparar a mudança para o RiC e para a web semântica é continuar realizando o que fazemos desde sempre: aperfeiçoar sem parar a qualidade de nossas descrições. (Chenard, 2020, tradução nossa)

De acordo com Cook, é crucial estabelecer um sistema “dinâmico” em vez de um sistema “estático” na descrição arquivística. Isso implica substituir os inventários físicos por sistemas de gerenciamento de bancos de dados relacionais (ou, atualmente, também não relacionais), com o objetivo de preservar o princípio da proveniência e evitar distorções. O autor destaca a importância de os arquivistas manterem a atenção no “mapeamento das relações” entre os segmentos dos fundos para elucidar o conjunto, complementados pela visão inovadora de registros dos demais contextos. Cook enfatiza que “tais conexões são cruciais para desvendar toda a potencialidade contextual da proveniência” (Cook, 2017, p. 55; 59).

Não se pode afirmar com certeza que a relação entre arquivos e a web semântica será completamente consolidada. Contudo, é evidente que os modelos conceituais e as tecnologias semânticas se configuram como recursos potenciais e aliados tecnológicos essenciais para os arquivistas em sua missão fundamental de preservar e proporcionar um acesso amplo aos documentos de arquivo. Isso não se limita apenas à preservação do contexto de produção dos documentos, mas se estende à preservação de diversos outros contextos que podem estar intrinsecamente relacionados a esses documentos, contribuindo para sua compreensão e utilização futura.

Referências

- BERNERS-LEE, T. Linked data. Design issues. World Wide Web Consortium (W3C). jul. 2006. Disponível em: <https://www.w3.org/DesignIssues/LinkedData.html>. Acesso em: 3 dez. 2023.
- BERNERS-LEE, T.; HENDLER, J.; LASSILA, O. The semantic web. *Scientific American*, v. 284, n. 5, p. 28-37, may 2001. Disponível em: https://www-sop.inria.fr/acacia/cours/essi2006/Scientific%20American_%20Feature%20Article_%20The%20Semantic%20Web_%20May%202001.pdf. Acesso em: 3 dez. 2021.
- CHENARD, G. Des points du vue multiples: RiC et moi. *Archivistes! Paris*, n. 136, janv./mars 2021, p. 27.
- CONARQ. Conselho Nacional de Arquivos. ISDF: norma internacional para descrição de funções. Brasília: Conarq, 2007. Disponível em: <https://www.gov.br/conarq/pt-br/centrais-de-conteudo/publicacoes/ISDF.pdf>. Acesso em: 10 out. 2023.
- CONARQ. Conselho Nacional de Arquivos. Norma brasileira de descrição arquivística (Nobrade). Brasília: Conarq, 2006. Disponível em: <https://www.gov.br/conarq/pt-br/centrais-de-conteudo/publicacoes/nobrade.pdf>. Acesso em: 10 out. 2023.
- CONARQ. Conselho Nacional de Arquivos. *Isaar*(CPF): norma internacional sobre os registros de autoridade arquivística para entidades coletivas, pessoas e famílias. 2. ed. Brasília: Conarq, 2003. Disponível em: https://www.gov.br/conarq/pt-br/centrais-de-conteudo/publicacoes/isaar_cpf.pdf. Acesso em: 10 out. 2023.
- COOK, T. O conceito de fundo arquivístico: teoria, descrição e proveniência na era pós-custodial. Rio de Janeiro: Arquivo Nacional, 2017.
- GUEGUEN, G.; FONSECA, V. M. M.; PITTI, D. V.; GRIMOÛARD, C. S. Para um modelo conceitual internacional de descrição arquivística. *Acervo*, Rio de Janeiro, v. 26, n. 2, p. 100-116, 2013a. Disponível em: <https://revista.arquivonacional.gov.br/index.php/revistaacervo/article/view/517>. Acesso em: 3 dez. 2023.
- GUEGUEN, G.; FONSECA, V. M. M.; PITTI, D. V.; GRIMOÛARD, C. S. Toward an international conceptual model for archival description: a preliminary report from the International Council on Archives' Experts Group on Archival Description. *The American Archivist*, [s. l.], v. 76, n. 2, p. 566-583, fall/winter, 2013b. Disponível em: https://www.ica.org/sites/default/files/EGAD_English.pdf. Acesso em: 3 dez. 2023.
- HEATH, T., BIZER, C. *Linked data: evolving the web into a global data space. Synthesis lectures on the semantic web: theory and technology*. Kentfield: Morgan & Claypool, 2011. Disponível em: <https://www.iro.umontreal.ca/~nie/IFT6255/Books/LinkedData.pdf>. Acesso em: 3 dez. 2023.
- ICA. International Council on Archives. Experts Group on Archival Description. *Records in Contexts Conceptual Model 1.0*. 2023. Disponível em: <https://www.ica.org/en/records-in-contexts-conceptual-model>. Acesso em: 3 dez. 2023.
- ICA. International Council on Archives. Experts Group on Archival Description. *Introduction to Archival Description (RiC-IAD Consultation Draft v0.2)*. 2021a. Disponível em: https://www.ica.org/sites/default/files/ric-cm-02_july2021_0.pdf. Acesso em: 3 dez. 2023.
- ICA. International Council on Archives. Experts Group on Archival Description. *Records in Contexts Conceptual Model 0.2 (RiC-CM Consultation Draft v0.2)*. 2021b. Disponível em: https://www.ica.org/sites/default/files/ric-cm-02_july2021_0.pdf. Acesso em: 25 abr. 2022.
- ICA. International Council on Archives. *Records in Contexts: a conceptual model for archival description. Consultation draft v. 1*, Sept. 2016. Disponível em: <https://www.ica.org/sites/default/files/RiC-CM-0.1.pdf>. Acesso em: 3 dez. 2023.
- ICA. International Council on Archives. Committee on best practices and standards. *Progress report for revising and harmonizing ICA descriptive standards*. 2012. Disponível em: <https://www.ica.org/en/cbps-progress-report-revising-and-harmonizing-ica-descriptive-standards>. Acesso em: 30 out. 2022.
- ICA. International Council on Archives. *Isad(G): norma geral internacional de descrição arquivística*. 2. ed. Rio de Janeiro: Arquivo Nacional, 2004.
- MARCONDES, C. H. Interoperabilidade entre acervos digitais de arquivos, bibliotecas e museus: potencialidades das tecnologias de dados abertos interligados. *Perspectivas em Ciência da Informação*, Belo Horizonte, v. 21, n. 2, p. 61-83, jun. 2016. Disponível em: http://www.scielo.br/scielo.php?script=sci_arttext&pi

d=S1413-99362016000200061&lng=pt&nrm=iso.

Acesso em: 3 dez. 2023.

NAZÁRIO, C. E.; DIAS, C. C. Avaliação do Europeana Data Model para o enriquecimento semântico de objetos. *Múltiplos Olhares em Ciência da Informação*, Belo Horizonte, n. especial, 2021. Disponível em: <https://periodicos.ufmg.br/index.php/moci/article/view/37189>. Acesso em: 3 dez. 2023.

PICKLER, M. E. V. Web semântica: ontologias como ferramentas de representação do conhecimento. *Perspectivas em Ciência da Informação*, Belo Horizonte, v.12, n.1, abr.2007. Disponível em: http://www.scielo.br/scielo.php?script=sci_arttext&pid=S1413-99362007000100006&lng=en&nrm=iso. Acesso em: 10 maio 2022.

SANTAREM SEGUNDO, J. E.; CONEGLIAN, C. S. Web semântica e ontologias: um estudo sobre construção de axiomas e uso de inferências. *Informação & Informação*, Londrina, v. 21, p. 217-244, 2016. Disponível em: <http://www.uel.br/revistas/uel/index.php/informacao/article/view/26417>. Acesso em: 3 dez. 2023.

SHARMA, M. Difference between Web 1.0, Web 2.0, and Web 3.0. *Geeks for Geeks*, 27 jan. 2022. Disponível em: <https://www.geeksforgeeks.org/web-1-0-web-2-0-and-web-3-0-with-their-difference/>. Acesso em: 3 dez. 2023.

SOUZA, R. R.; ALVARENGA, L. A web semântica e suas contribuições para a ciência da informação. *Ciência da Informação*, Brasília, v. 33, n. 1, p. 132-141, jan./abr. 2004. Disponível em: <https://www.scielo.br/j/ci/a/sp3XpmZhXw384H5Fw9H89Y-L/?format=pdf>. Acesso em: 18 out. 2023.

Recebido em 30/10/2023

Aprovado em 31/1/2024