

Linked Open Archives (LOA): Uma experiência de aplicação da ontologia Records in Contexts (RiC-O) e das tecnologias Linked Open Data (LOD) em acervos arquivísticos

Linked Open Archives (LOA): An application experience of the Records in Contexts Ontology (RiC-O) and Linked Open Data (LOD) Technologies in archival records

Jair Martins de Miranda

Universidade Federal do Estado do Rio de Janeiro (UNIRIO), Av. Pasteur, 458, Urca, Rio de Janeiro, RJ, jairmm@unirio.br

Resumo

O presente artigo analisa o novo padrão internacional de descrição arquivística Records in Contexts (RiC), proposto pelo Experts Group on Archival Description do *International Council on Archives* (Egad/ICA), por meio da sua aplicação experimental na série de documentos Escravidão, à luz das tecnologias Linked Open Data (LOD) e através da plataforma Linked Open Archives (LOA). Conclui que o RiC terá grande impacto nos princípios e práticas arquivísticas e no acesso aos arquivos.

Palavras-chave: Records in Contexts; Ontologia; Linked Open Data (LOD); Linked Open Archives (LOA); Escravidão.

Abstract

This article analyzes the new international standard for archival description Records in Contexts (RiC), proposed by the Experts Group on Archival Description of the *International Council on Archives* (Egad/ICA), through its experimental application in the Slavery document series and in the light of Linked Open Data technologies (LOD) and by the digital platform Linked Open Archives (LOA). It concludes that the RiC will have a major impact on archival principles and practices and access to archives.

Keywords: Records in Contexts; ontology; Linked Open Data (LOD); Linked Open Archives (LOA); Slavery.

1 INTRODUÇÃO

Ao apresentar, em 2016, uma consulta pública à comunidade arquivística internacional sobre o novo modelo conceitual de descrição Records in Contexts (RiC-CM), visando a criação da futura ontologia (RiC-O), o Experts Group on Archival Description (Egad) [1], vinculado ao International Council on Archives (ICA), lançou não só um grande desafio para o tratamento e acesso aos documentos de arquivo na era digital, mas também anunciou uma nova concepção do basilar princípio de proveniência, quando apresentou uma proposta de mudança significativa na forma como, desde 1994, com o ISAD(G), os padrões de descrição arquivística eram abordados (ICA, 2016).

Ao sugerir, com o novo padrão, uma linguagem documentária “semanticamente mais expressiva” (Marcondes, 2012) para os usuários de arquivos na *web* e mais acessível a outros contextos ainda inexplorados tecnologicamente nos acervos arquivísticos, incorporando e redefinindo os seus antigos padrões de descrição (ISAD, ISAAR, ISDF e ISDIAH), o documento do Egad (aberto à consulta pública em outubro de 2016), anunciou não só as mudanças nas abordagens teórica e prática, mas também apontou para significativas mudanças ao fundamentá-las nas tecnologias da *web* semântica [2] e ao flexibilizar a solidez dos princípios da proveniência, do respeito aos fundos e da ordem original, tão caros e basilares à ciência dos arquivos.

Diante de uma possível crise ontológica da arquivologia [3], com tanta inovação tecnológica e mudanças conceituais e estruturais tão significativas, duas questões se apresentaram de imediato, como 1 – Quais mudanças significativas no domínio dos arquivos ficaram refletidas no modelo conceitual RiC-CM-0.2, para além dos padrões anteriores? 2 – Como a ontologia Records in Contexts (RiC-O-v0.1) pode ampliar e representar os muitos contextos que se apresentam na realidade atual dos documentos de arquivos? Trata-se de questões complexas para a comunidade arquivística, por exigirem estudos interdisciplinares. Dessa forma, o presente artigo tem o objetivo de propor reflexões sobre o tema e apresentar elementos que auxiliem nas respostas a essas questões. Os resultados obtidos por meio de pesquisa bibliográfica, documental, experimental e entrevistas com membros do Egad, foram organizados e são apresentados nas seções seguintes.

2 O MULTINÍVEL E O MULTIDIMENSIONAL

Ao questionar os membros do Egad sobre os fundamentos teórico-metodológicos que influenciaram na concepção do modelo conceitual RiC-CM, é possível apreender em seus relatos, assim como observar na versão preliminar do modelo conceitual Records in Contexts, que para refletir sobre a nova realidade dos arquivos na era digital eles tiveram a influência de algumas correntes científicas, tanto no campo da arquivologia como no da ciência da computação, já iniciadas em diferentes países nos últimos anos [4]. Criado em 2012 com essa finalidade, o Egad realizou, até 2016, reuniões virtuais e presenciais e, nesse período, concebeu o novo padrão de descrição, a partir de um modelo conceitual, tendo como premissas a ampliação do conceito de contexto em arquivo e a inclusão da abordagem multidimensional.

No campo da arquivologia, a noção de múltiplos contextos em que os documentos são

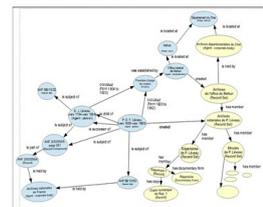
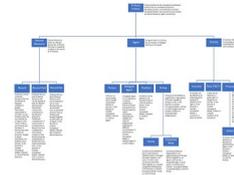
criados e acumulados é antiga e percebida na prática. No entanto, para os membros do Egad, com o advento dos documentos digitais e a rede mundial de computadores, a representação analógica e hierárquica, subjacente nos padrões de descrição antecedentes — ISAD(G), ISAAR(CPF), ISDF e ISDIAH —, já não atendia o estado da arte dos arquivos na era digital. Esse anacronismo ensejou uma crise nessa representação e a necessidade de uma releitura no basilar princípio da proveniência, assim como nos seus princípios correspondentes “respeito aos fundos” e “ordem original”, tão caros aos arquivistas e à arquivologia. Ao adotar a noção de contextos, e não somente contexto [5], e uma abordagem multidimensional (ou multicontextual) e não mais multinível para o novo padrão de descrição, os membros do Egad partiram da premissa de que é possível representar a nova realidade dos documentos de arquivo não somente descrevendo eles em unidades de arquivamento estanques, isoladas, focadas prioritariamente no produtor dos documentos e representadas analogicamente em pastas e estantes, como concebido originalmente pelo ISAD(G) e pelos padrões posteriores, mas nas relações entre essas unidades de arquivamento e seus agentes institucionais e sociais, como também entre as atividades que os criaram. Essa nova representação multidimensional, não mais hierárquica, mas apresentada como uma rede de relações em grafos, mudou substancialmente o enfoque multinível inicial usado na descrição dos documentos de arquivo e ficou evidente nas novas entidades e suas relações do modelo conceitual RiC-CM e, posteriormente, no RiC-O.

No campo da ciência da computação, a representação dessas premissas no modelo conceitual RiC-CM e na consequente ontologia RiC-O, feita de forma a permitir que os dados desse novo padrão estejam disponíveis em formato RDF (Resource Date Framework), passíveis de serem usados como dados abertos conectados, também demonstra claramente a opção dos membros do Egad por fundamentar o padrão nesse novo universo (ecossistema) onde a arquitetura da informação é representada como uma rede de grafos, diferentemente dos formatos dos padrões anteriores de descrição, nos quais as informações eram representadas na *web* de forma hierárquica, (Quadro 1) para a segunda geração da *web*, em grande parte, em formato XML. [6]

Quadro 1. Arquiteturas do multinível e do multidimensional

O Multinível e o Multidimensional

- A Abordagem Multinível da Descrição
 - Tradicional
 - Hierárquico
 - Centrada no Produtor
 - Com restritas formas de acesso à informação/documento
 - Tecnologias: Web 2.0, Banco de Dados Relacionais
- A Abordagem Multidimensional da Descrição
 - Contemporâneo
 - Rizomático, Facetada.
 - Centrada nas relações sociais e de colaboração
 - Com múltiplas formas de acesso à informação/documento
 - Tecnologias: Web 3.0, Ontologias, RDF, Linked Open Data (LOD)



Fonte: Elaborado pelo autor (2022).

Portanto, para além desses fundamentos teórico-metodológicos da arquivologia e da ciência da computação que influenciaram os membros do Egad na concepção dessa nova noção de “contextos” e da abordagem “multidimensional” no modelo conceitual RiC-CM e na ontologia RiC-O, observa-se que com esse novo padrão, um novo cenário se abre para os arquivos, onde a revisão e a flexibilização dos princípios arquivísticos e o uso das tecnologias da *web* semântica prometem um novo protagonismo dos arquivos e dos arquivistas na era digital.

3 O MODELO CONCEITUAL RECORDS IN CONTEXTS (RiC-CM)

O RiC-CM, em sua segunda versão 0.2, é resultado de mudanças significativas no domínio dos arquivos, sugeridas pelo Egad após consulta pública à comunidade internacional de arquivistas. É um modelo desenvolvido para também fundamentar a criação da ontologia RiC-O, com a proposta de substituir os padrões de descrição vigentes: ISAD(G), ISAAR(CPF), ISDF e ISDIAH.

Dado que os modelos conceituais são baseados em entidades, atributos e relações, o Egad atribuiu para a última versão do seu modelo (RiC-CM-0.2) uma hierarquia de 22 entidades e, a partir delas, 41 atributos e 78 relações, que resumidamente descrevemos a seguir.

3.1 Entidades no RiC-CM

As entidades são os principais objetos de interesse dos profissionais que gerenciam documentos, tanto no contexto de origem e uso (gestão de documentos) como aqueles retidos para preservação e acesso a longo prazo (gestão de arquivos). Da perspectiva dos arquivistas, as entidades identificadas são consideradas essenciais para fornecer o contexto intelectual que serve ao gerenciamento físico, preservação, busca, uso e compreensão dos documentos ao longo de sua história.

Quadro 2. Hierarquia das entidades do RiC

RiC Entities Hierarchy				
First Level	Second Level	Third Level	Fourth Level	
RiC-E01 Thing	RiC-E02 Record Resource	RiC-E03 Record Set		
		RiC-E04 Record		
		RiC-E05 Record Part		
	RiC-E06 Instantiation			
	RiC-E07 Agent	RiC-E08 Person		
		RiC-E09 Group		RiC-E10 Family
				RiC-E11 Corporate Body
		RiC-E12 Position		
		RiC-E13 Mechanism		
	RiC-E14 Event	RiC-E15 Activity		
	RiC-E16 Rule	RiC-E17 Mandate		
	RiC-E18 Date	RiC-E19 Single Date		
		RiC-E20 Date Range		
		RiC-E21 Date Set		
RiC-E22 Place				

Fonte: International Council on Archives (2019).

3.2 Os Atributos no RiC-CM

Atributos são as características das entidades. Os atributos de uma entidade, juntamente com as relações que ela mantém com outras entidades, constituem a sua identidade, ou seja: o que ela é. Descrever uma entidade envolve necessariamente observação e análise de evidências para identificar as suas características. Cada atributo é descrito com base em um modelo e, associado a outros, caracterizam as entidades, a exemplo dos atributos específicos para a entidade Record, apresentados no Quadro 3, a seguir.

3.3 As Relações no RiC-CM

Para compreender e descrever os documentos de arquivo é essencial documentar o contexto em que esses recursos de registro foram criados, acumulados e gerenciados ao longo do tempo e do espaço. O papel das relações no RiC-CM é descrever as conexões entre as entidades, uma vez que contribuem para conhecer o contexto da criação e manutenção de documentos e, como resultado, expressam características significativas da história e gestão dos documentos arquivísticos. As relações no RiC-CM estão focadas nas conexões e terão um grande impacto na descrição arquivística. Elas divergem conceitualmente dos padrões

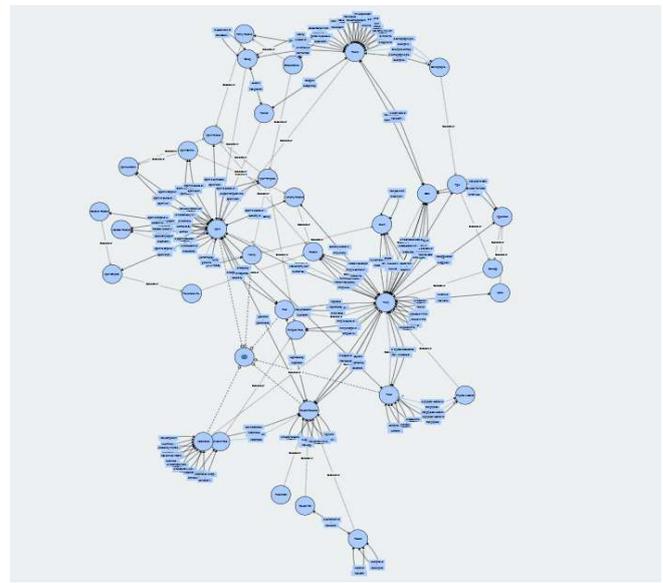
baseados em XML, como EAD e EAC, que dependem da hierarquia estrutural para definir conexões entre, por exemplo, registros para séries e fundos. Em vez disso, as relações RiC-CM são baseadas em um modelo de grafo, semelhante ao Resource Description Framework (RDF), que permite conexões mais simples e flexíveis, como ilustrado na Figura 1.

Quadro 3. Atributos específicos da entidade Record

RiC-E04	Record
Attribute ID	Attribute Name
RiC-A16	Descriptive Note
RiC-A22	Identifier
RiC-A28	Name
RiC-A21	History
RiC-A35	Record Resource Extent
RiC-A38	Scope and content
RiC-A40	Structure
RiC-A03	Authenticity Note
RiC-A07	Classification
RiC-A08	Conditions of Access
RiC-A09	Conditions of Use
RiC-A10	Content Type
RiC-A17	Documentary Form Type
RiC-A24	Integrity
RiC-A25	Language
RiC-A26	Legal Status
RiC-A39	State

Fonte: International Council in Archives (2019)

Figura 1. Grafo do modelo conceitual RiC criado pelo software WebVowl no editor de ontologias Protégé



Fonte: Elaborado pelo autor (2022).

4 ESTUDO DE APLICAÇÃO DO RiC NA SÉRIE ESCRAVIDÃO

O conjunto documental “série Escravidão” é constituído por documentos em papel, de várias espécies, tais como requerimentos, cartas, telegramas, ofícios etc., possuindo um carácter unitário, referindo-se a um só assunto: o trabalho compulsório urbano no Rio de Janeiro. Os documentos encontram-se encadernados sob a forma de códice, respeitando a origem do órgão acumulador, a Câmara Municipal... São registros referentes à abolição, africanos livres, capitães do mato, à licenças para arrematação, compra e venda de escravos, à leilões, escravos ladinos, matrículas de escravos, escravos de ganho, posturas municipais, tráfico de escravos, sisas, mercadores de escravos etc., acumulados pela administração pública da cidade do Rio de Janeiro, desde os tempos coloniais até a promulgação da Lei Áurea. (ARQUIVO GERAL DA CIDADE DO RIO DE JANEIRO, 2007, p. 4).

A escolha da série Escravidão para o presente estudo de aplicação do novo padrão de descrição Records in Contexts (RiC) deve-se à proposta de analisá-la sob a ótica das possíveis entidades, atributos e relações do RiC, que é baseada nos conteúdos dos elementos (ou campos) de descrição da Norma Brasileira de Descrição Arquivística (Nobrade), na forma como indica as suas regras e como a série foi originalmente descrita.

Essa análise foi dividida pelas oito áreas de descrição da Nobrade e busca fazer, de forma interpretativa, uma transcrição desses campos e seus conteúdos para as possíveis entidades, atributos e relações do novo padrão RiC. Trata-se, portanto, de uma projeção da forma como se prevê o exercício da descrição arquivística nos próximos anos, por entender que este será o grande desafio que o novo padrão RiC traz para os arquivos e arquivistas. Ou seja: a transposição de uma abordagem multinível para uma multidimensional de descrição.

Sendo assim, dado que o exercício da descrição, mesmo com normas, objetivos e regras, é um ato subjetivo de diálogo do arquivista-descritor com os documentos de arquivo para interpretar, inferir e conhecer os seus múltiplos contextos e melhor representá-los para os futuros usuários, apresentamos a seguir o nosso estudo de transposição da Nobrade para o novo padrão RiC [7], cientes de que essa será nos próximos anos uma tarefa de fundamental importância para a descrição dos documentos, que como sabemos será sempre incompleta.

4.1 Projeção dos campos da Nobrade para as possíveis entidades, atributos e relações do RiC

Os campos da área de identificação têm como foco descrever o próprio documento arquivístico e são considerados essenciais por reunir os cinco dos sete campos obrigatórios de todos os 28 da Nobrade. Vistos sob a ótica do novo padrão RiC, esses campos representariam principalmente os atributos e relacionamentos das entidades Record Resource e suas sub-entidades Record Set, Record e Record Part, além da entidade Instantiation. Ao analisarmos os conteúdos dos campos da série Escravidão do AGCRJ, projetamos uma transposição dos campos da Nobrade para as possíveis entidades, atributos e relacionamentos do RiC, considerando as seguintes especificidades de cada campo:

O conteúdo do campo 1.1 – código de referência, como é regra da Nobrade, foi formado pelos códigos do país, BR (Brasil), da cidade RJ (Rio de Janeiro), da entidade custodiadora AGCRJ (Arquivo Geral da Cidade do Rio de Janeiro) e dos códigos específicos das unidades de descrição CM (fundo Câmara Municipal do Rio de Janeiro) e ESC (série Escravidão). Portanto, como se observa, sob a ótica do RiC, além da correlação da série BR.RJ.AGCRJ.CM.ESC com a entidade Record Set e o atributo *identifier* (RiC-A22), é

possível projetar, apenas neste campo identificação, várias relações possíveis no RiC, como *has subject* (RiC-R019), *contains* (RiC-R007), *is jurisdiction of* (RiC-R076), *is holder of* (RiC-R039), *has constituinte* (RiC-R003) e *has provenance* (RiC-R026).

Os conteúdos dos campos 1.2 – título (Escravidão), 1.3 – data (s) (1794-1880) e 1.4 – nível de descrição (Série N3) podem ser entendidos, respectivamente, como os conteúdos dos atributos *name* (RiC-A28), *expressed date* (RiC-A19) e *record set type* (RiC-A36) do RiC e, em princípio, não sugerem relações nesse novo padrão de descrição.

O conteúdo do campo 1.5 – dimensão e suporte da série Escravidão apresenta os três gêneros documentais (textual, micrográfico e eletrônico/digital) do seu acervo e as três espécies (códices, microfilme 35 mm e DVD) e os respectivos quantitativos (101 códices, 30 microfilmes e 20 DVDs) a eles associados. Visto sob a ótica das entidades do RiC, esses gêneros remetem à entidade Instantiation (RiC-E06), a seus respectivos atributos *carrier type* (RiC-A05) e *carrier extent* (RiC-A04) e às possíveis relações *is instantiation associated with instantiation* (RiC-R034) e *has derived instantiation* (RiC-R014).

Previendo uma futura transição do exercício de descrição de um padrão para outro, apresentamos no Quadro 4, por associação, uma projeção desses campos da área identificação da série Escravidão para as possíveis entidades, atributos e relações do novo padrão RiC.

Quadro 4. Projeção de alguns campos da Nobrade para o RiC – Área de identificação

Projeção da série Escravidão (descrita pelos campos da Nobrade) para as possíveis entidades, atributos e relações do novo padrão RiC							
Área de identificação							
Campo	Conteúdo	Entidades no RiC	Atributos no RiC	Possíveis relações no RiC			Tipo de relação
				Sujeito	Propriedade	Objeto	
1.1 – Código de referência	BR.RJ.AGCRJ.CM.ESC	RiC-E03 Record Set	<i>identifier</i> (RiC-A22)	BR.RJ.AGCRJ.CM.ESC	<i>has subject</i> (RiC-R019)	Escravidão	Subject relations
	BR = Brasil	RiC-E22 Place (BR)		BR	<i>contains</i> (RiC-R007)	RJ	Spatial relations
	RJ = Rio de Janeiro	RiC-E22 Place (RJ)		RJ	<i>is jurisdiction of</i> (RiC-R076)	AGCRJ	Spatial relations
	AGCRJ = Arquivo Geral da Cidade do Rio de Janeiro	RiC-E07 Agent (AGCRJ)		AGCRJ	<i>is holder of</i> (RiC-R039)	CM	Management relations
	CM = Fundo Câmara Municipal do Rio de Janeiro	RiC-02 Record Resource		Fundo CM	<i>has constituinte</i> (RiC-R003)	Série ESC	Record Resource relations
	ESC = Série Escravidão	RiC-E03 Record Set		Série ESC	<i>has provenance</i>	Fundo CM	Provenance relations

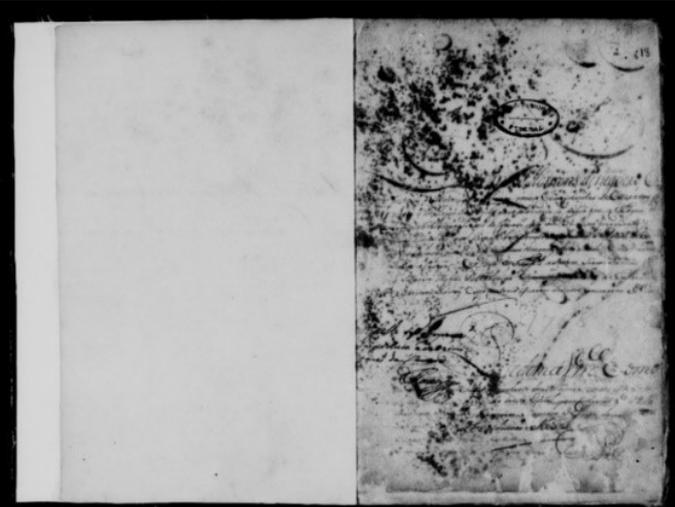
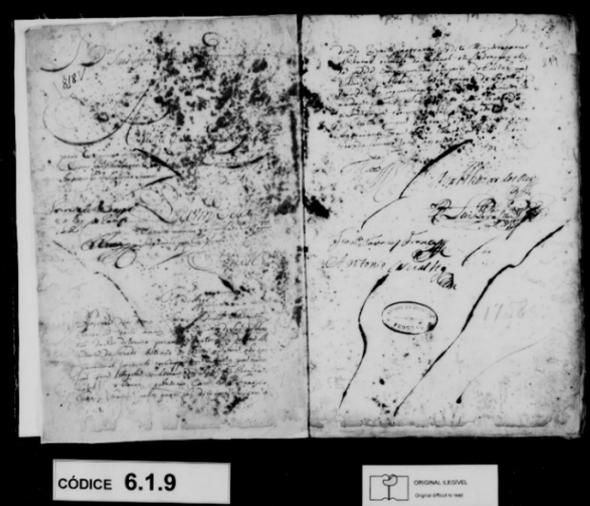
					(RiC-R026)		
1.2 – Título	Escravidão	RiC-E03 Record Set	<i>name</i> (RiC-A28)				
1.3 – Data (s)	1794-1888	RiC-E20 Date Range	<i>expressed date</i> (RiC-A19)				
1.4 – Nível de descrição	Série (N3)	RiC-E03 Record Set	<i>record set type</i> (RiC-A36)				
1.5 – Dimensão e suporte							
1.5.1 – Gênero	textual	RiC-E06 Instantiation		Gênero textual	<i>is instantiation associated with instantiation</i> (RiC-R034)	Gênero micrográfico	Instantiation to Instantiation relations
Espécie	códices		<i>carrier type</i> (RiC-A05)				
Quantificação/Mensuração	101 códices/dossiês ou 4,5 metros lineares		<i>carrier extent</i> (RiC-A04)				

Fonte: Elaborado pelo Autor (2022).

4.2 Análise complementar de um manuscrito da série Escravidão

A análise do manuscrito BR.RJ.AGCRJ.CM.ESC.6.1.9.M.1 [8] (integrante do código BR.RJ.AGCRJ.CM.ESC.6.1.9) e da série Escravidão BR.RJ.AGCRJ.CM.ESC é incluída aqui de forma complementar. Serve como exemplo para ilustrar apenas um dos muitos manuscritos dos 101 códices da série. Serve também para projetar as possíveis entidades, atributos e relações do RiC, por meio de alguns campos considerados obrigatórios e principais da Nobrade, para esta análise, conforme o Quadro 5.

Quadro 5. Formulário de transcrição paleográfica do manuscrito BR.RJ.AGCRJ.CM.ESC.6.1.9.M1

Formulário de transcrição paleográfica Manuscrito BR.RJ.AGCRJ.CM.ESC.6.1.9.M1 [9] Origem: série Escravidão do Arquivo Geral da Cidade do Rio de Janeiro (AGCRJ) Petição ao juiz de paz pelos homens de negócio da cidade do Rio de Janeiro contra a transferência do local de comércio dos escravos novos		
Imagens e transcrições		
Transcrição folha 818r  <p>CÓDICE 6.1.9</p> <p>ORIGINAL LEGÍVEL</p>		Folha 818r Dizem os Homens de negócio, e [ilegível] de Navios, [ilegível] e mais carregantes de escravos que [ilegível] do Reino de Angolla para esta cidade que a sua noticia chegou a cujo Nobilíssimo Senado da Câmara de que [ilegível] para presidente hum edital pelo qual lhe parece cumprir [ilegível] [linha parcialmente ilegível] trazem Lado a Carioca. Nós os suplentes assistem ser [ilegível] que se fazem dignas de atenção e [ilegível] ao mesmo edital cuja matéria querem deduzir por mesmo [rubrica] [ilegível] pelle se [ilegível] peça da perdura e determinação da Câmara A ver Pedem a Vossa Mercê como [ilegível] presidente deste officimo Senado [ilegível] dar vida de dito edital [ilegível] os de por irem [ilegível] e efeitto do mês <p style="text-align: right;">athea a direção do Mês</p>
Imagem folhas 818v e 819r		
Transcrição folhas 818v e 819r  <p>CÓDICE 6.1.9</p> <p>ORIGINAL LEGÍVEL</p>	Folha 818v Senhor Doutor Juiz de fora, Presidente do Senado da Câmara A vista que os [ilegível] pedem [ilegível] para em porem de fez [ilegível] e como vossa mercê lho denegar [ilegível]. Tomão [ilegível] Aggravo para o supremo senado. [linhas ilegíveis] Para Vossa Mercê seja servido mandar saber seu aggravo havendo por lá ficado. [rubrica não identificada] Aos onze dias de fevereiro de mil setecentos e cincoenta e oito annos, nesta cidade de São Sebastião do Rio de Janeiro perante o Doutor Juiz de Fora e Pre-sidente do Senado Antônio de Freitas Silva, aqui o parecerão presente os capitães de navios que navegação para Angola: Baltazar dos Reis, Luis Pereira de Farias e Antônio Carvalho e Francisco Tavares França pelos quaes foi dito que elles com o	Folha 819r devido respeito aggravo do dito Ministro e para Tribunal e Senado da Relação em lhe denegar a vista o pedido em suspensão a respeito do Edital que se havia publicado sobre negro de Angolla servido seu requerimento. [ilegível] Ministro lhe mandou escrever a seu aggravo e no re [ilegível] vinha em que houve [ilegível] fiz ter [ilegível] em que assignou o dito Ministro esse Agravante e eu André Martins Britto, escrivão da Câmara que o escrevo. Mater Baltazar dos Reis Luiz Pereira de Farias Francisco Tavares França Antônio Carvalho
Nota do(a) paleógrafo(a)	O documento apresenta dificuldades de leitura pelo seu estado de conservação	
Datas da transcrição	14/12/2020	
Responsável pela transcrição	Ariadne Pires Barbosa	

Fonte: Elaborado pelo Autor (2022).

Quadro 6. Projeção de alguns campos do manuscrito BR.RJ.AGCRJ.CM.ESC.6.1.9.M1 para o RiC

Projeção do manuscrito BR.RJ.AGCRJ.CM.ESC.6.1.9.M1 (descrita pelos campos da Nobrade) para as possíveis entidades, atributos e relações do novo padrão RiC							
Campo	Conteúdo	Entidades no RiC	Atributos no RiC	Possíveis relações no RiC			Tipo de relação
				Sujeito	Propriedade	Objeto	
1.1 – Código de referência	BR.RJ.AGCRJ.CM.ESC.6.1.9.M1	RiC-E03 Record Part	<i>identifier</i> (RiC-A22)	BR.RJ.AGCRJ.CM.ESC.6.1.9.M1	<i>has subject</i> (RiC-R019)	Mercadores de escravos	Subject relations
	Código 6.1.9.M1	RiC-E03 Record Part		Código 6.1.9	<i>has provenance</i> (RiC-R026)	Série Escravidão ESC	Provenance relations
1.2 – Título	Petição ao juiz de fora	RiC-E03 Record Part	<i>name</i> (RiC-A28)				
1.3 – Data(s)	11/2/1758	RiC-E20 Date Range	<i>expressed date</i> (RiC-A19)				
1.4 – Nível de descrição	Item documental (N5)	RiC-E03 Record Part	<i>record set type</i> (RiC-A36)				
1.5 – Dimensão e suporte							
1.5.1 – Gênero	textual	RiC-E06 Instantiation		Gênero textual	<i>is instantiation associated with instantiation</i> (RiC-R034)	Gênero Micrográfico	Instantiation to Instantiation relations
Espécie	manuscrito		<i>carrier type</i> (RiC-A05)				
Quantificação/mensuração	3 folhas de documentos manuscritos		<i>carrier extent</i> (RiC-A04)				
1.5.2 – Gênero	micrográfico	RiC-E06 Instantiation		Gênero micrográfico CM-ESC.001	<i>has derived instantiation</i> (RiC-R014)	Gênero textual BR.RJ.AGCRJ.CM.ESC.6.1.9	Instantiation to Instantiation relations
Espécie	microfilmes (35 mm)		<i>carrier type</i> (RiC-A05)				
Quantificação/mensuração	3 slides de microfilme		<i>carrier extent</i> (RiC-A04)				
1.5.3 – Gênero	eletrônico/digital	RiC-E06 Instantiation		Gênero eletrônico/digital	<i>has derived instantiation</i> (RiC-R014)	Gênero textual	Instantiation to Instantiation relations
Espécie	DVD		<i>carrier type</i> (RiC-A05)				

Quantificação/mensuração	3 arquivos .png		<i>carrier extent</i> (RiC-A04)				
Espécie	transcrição		<i>carrier type</i> (RiC-A05)				
Quantificação/mensuração	1 arquivo .docx		<i>carrier extent</i> (RiC-A04)				
2.1 – Nome(s) do(s) Produtor(es)	Baltazar dos Reis, Luís Pereira de Farias, Antônio Carvalho, Francisco Tavares França e Antônio de Freitas Silva	Person (RiC-E07)	<i>name</i> (RiC-A28)	Câmara Municipal da Cidade do Rio de Janeiro	<i>is owner of</i> (RiC-R037)		Management relations
3.1 – Âmbito e conteúdo	Este manuscrito refere-se à uma petição escrita por Baltazar dos Reis, Luiz Pereira de Farias, Antônio Carvalho e Francisco Tavares França ao juiz de fora e presidente do Senado da Câmara doutor Antônio de Freitas Silva na cidade do Rio de Janeiro e foi redigido pelo escrivão André Martins Britto em 11/2/1758						
4.1 – Condições de acesso	Não descrito						
5.4 – Notas sobre publicação	Não descrito						
6.2 – Notas Gerais	Não descrito						
8.1 – Pontos de acesso e indexação de assuntos	Juiz de Fora, Presidente do Senado da Câmara, Homens de negócio de navios (mercadores de escravos) Reino de Angola, (Reino do Congo), Capitães de Navios, Negro de Angola						

Fonte: Elaborado pelo Autor (2022).

5 LINKED OPEN DATA (LOD), LINKED OPEN ARCHIVES (LOA) E A SÉRIE ESCRAVIDÃO

As entidades, os atributos e as muitas relações projetadas e inferidas no estudo anterior de aplicação, utilizando o modelo conceitual RiC-CM na série Escravidão, formam o primeiro passo para entender como o resultado de uma representação da informação, que é um modelo conceitual, dá origem ou pode ser transformada em um recurso da organização da informação, que, no nosso caso, é a ontologia Records in Contexts, versão (RiC-O.v0.1) e a sua aplicação na mesma série de documentos.

Baseada nas tecnologias da *web* semântica [10] e do Linked Open Data (LOD) [11], a noção de ontologia [12] e sua linguagem OWL [13] foram originalmente propostas por Berners-Lee, Hendler e Lassila (2001), no artigo “The semantic web”, para que as informações com conteúdo semântico pudessem ser processadas por máquinas e interligadas por computadores, através da linguagem de marcação RDF [14], que opera por meio de triplas compostas dos elementos sujeito, propriedade e objeto e, por serem semântica e univocamente

identificadas na *web* através de uma URI (Uniform Resource Identifier) [15], geram grafos que interligam esses elementos.

Da forma como são apresentadas, essas tecnologias prometem representar e organizar as informações de um dado domínio de conhecimento por meio dessa arquitetura de grafos, que permite interconectar todos esses elementos quando eles são representados e organizados em um mesmo vocabulário na *web* (também conhecidos como ontologia), como é o caso da ontologia RiC-O.v0.1.

Essa promessa tecnológica para os arquivos também começa a se tornar realidade, mas, dada a complexidade de aplicação dessas tecnologias, presume-se que o caminho será longo para que os documentos arquivísticos participem desse universo Linked Open Data: todo um processo de tratamento dos próprios documentos de arquivo precisa ser seguido e cumprido. O primeiro passo é repensar o próprio exercício da descrição, ressignificar os antigos padrões de descrição de documentos e se abrir para a abordagem multidimensional trazida pelo modelo conceitual e pela ontologia RiC.

Propomos tal começo em uma pesquisa aplicada, tendo como ponto de partida o esquema de implementação das cinco estrelas para dados abertos proposto por Berners-Lee e o método de desenvolvimento de aplicação semântica (ISOTANI; BITTENCOURT, 2015, p. 147). Teve-se como meta a disponibilização dos dados do conjunto documental série Escravidão do Arquivo Geral da Cidade do Rio de Janeiro (AGCRJ) na *web*, utilizando a ontologia RiC-O para uma futura conexão com os dados de outros conjuntos documentais de Angola e Portugal sobre o tráfico atlântico de escravos (Atlantic Slave Trade) no ambiente Linked Open Data (LOD).

Figura 2. Esquema das cinco estrelas para dados abertos

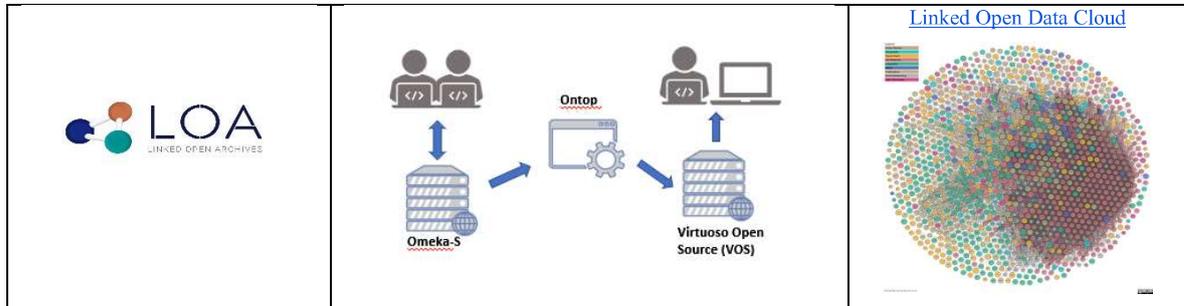


Fonte: <https://5stardata.info/en/>.

Para tanto, criamos experimentalmente a plataforma Linked Open Archives (LOA) (www.linkedopenarchives.com), uma arquitetura tecnológica que abriga o repositório digital Omeka S [16], inserimos nele a ontologia RiC-O 0.1 e o site Atlantic Slave Trade, de forma a testar o conteúdo que foi transcrito da série Escravidão. Essa arquitetura proposta tem o

objetivo de disponibilizar os dados no ambiente Linked Open Data (LOD), sendo composta por três componentes principais: o Omeka S, o Ontop e o Virtuoso Open Source (VOS).

Figura 3. Arquitetura da plataforma LOA para o universo Linked Open Data



Fonte: Elaborado pelo autor (2022).

O Omeka S é um software para o gerenciamento de acervos digitais, que permite ao produtor e ao mantenedor dos dados a inserção, a atualização, a remoção e a publicação dos dados do ecossistema. Entretanto, esse componente armazena as informações em um banco de dados relacional, fornecendo uma interface de programação de aplicações (Application Programming Interface – API) para a disponibilização dos dados apenas no formato JSON-LD [17], que ainda não é amplamente suportado, possuindo assim capacidade limitada de integração com outras fontes de dados. De acordo com Berners-Lee (2006), um dos princípios que devem ser seguidos para que dados sejam publicados como *linked data* é a capacidade de serem acessados através de um formato RDF e, potencialmente, através de um *endpoint* SPARQL. [18]. Para atendê-lo será utilizado o componente Ontop [19], que assumirá a tarefa de exportar os dados do banco do Omeka S para um arquivo RDF/Turtle [20]. Esse arquivo, por sua vez, será empregado como entrada para o Virtuoso Open Source [21], que armazenará essas informações em seu banco de dados de grafos e as publicará em seu próprio *endpoint* SPARQL. Vale lembrar que, para os usuários, essa arquitetura será transparente, dado que todo o processo de comunicação entre os componentes será realizado automaticamente. O arquivista usará como interface o Omeka S, para inserir e manter os dados, e os usuários consultarão as informações armazenadas e utilizarão interfaces amigáveis para seres humanos, por meio do site Atlantic Slave Trade [22], ilustrado na Figura 4.

Figura 4. Imagem de página do site Atlantic Slave Trade



Fonte: <https://linkedopenarchives.com/projects/atlanticslavetrade/s/website/item/47>

6 CONCLUSÃO

Diante do quadro de incerteza que se apresenta no futuro próximo dos arquivos quanto ao impacto do novo padrão RiC nos seus princípios e nas suas práticas, consideramos, nesta prévia conclusão, após a análise do modelo conceitual RiC-CM-0.2, do estudo da sua aplicação na série Escravidão, na implantação inicial da sua ontologia RiC-O-v0.1 na plataforma Linked Open Archives (LOA), que os arquivos, os arquivistas e a arquivologia vão viver mudanças importantes nos seus princípios e na sua prática e que elas serão ampliadas com o advento das novas tecnologias da *web* semântica e dos dados abertos conectados.

Sobre quais mudanças no domínio dos arquivos ficaram refletidas no modelo conceitual RiC-CM-0.2, ficou evidente que a absorção de todos os padrões anteriores (ISAD(G), ISAAR(CPF), ISDF e ISDIAH) e a inclusão da abordagem multidimensional de descrição à tradicional abordagem multinível foram as mais significativas. Abordagem multidimensional esta, que é ao mesmo tempo multicontextual e facetada, por meio das principais entidades (Agent, Record Resource, Event, Space, Time) do RiC, similares àquelas categorias primárias PMEST (Personality, Matter, Energy, Space e Time), propostas pelo matemático e bibliotecário Shiyaly Ramamrita Ranganathan (Ranganathan, 1967), que o autor (Miranda, 1986) aplicou na mesma série de documentos Escravidão do AGCRJ em 1986 [23].

Sobre como a ontologia Records in Contexts (RiC-O-v0.1) pode ampliar e representar os muitos contextos que se apresentam na realidade atual dos documentos de arquivos, ilustramos na sua aplicação experimental na série Escravidão do AGCRJ a potencialidade que o universo Linked Open Data oferece para indicar futuras conexões e contextos para além do

tradicional contexto do produtor de documentos de arquivo, que a abordagem multinível oferece. Presume-se também que essa nova realidade, vista através das tecnologias da ontologia e dos dados abertos conectados, pode conferir aos arquivos um novo *ethos* no mundo contemporâneo, já que, teoricamente, pode conectar todos os documentos de arquivo na *web*, anunciando a própria *web* como um arquivo e como um novo objeto de estudo da arquivologia.

Notas

[1] <https://www.ica.org/en/about-egad>

[2] As tecnologias da web semântica, Linked Open Data (LOD) e a noção de ontologia, originalmente descritas por Berners-Lee, Hender e Lassila (2001) no artigo “The semantic web”, propuseram que as informações com conteúdo semântico pudessem ser processadas por máquinas e interligadas por computadores, através da linguagem de marcação RDF (Resource Date Framework).

[3] A possível crise mencionada aqui não deve ser vista de forma a promover uma tecnofilia contra uma possível tecnofobia. Apenas de forma a anunciar uma possibilidade a ser discutida e respondida neste artigo.

[4] Como os conceitos de Multiproveniência e Macroavaliação, surgido no Canadá e de Records Continuum, surgido na Austrália.

[5] Subentendido implicitamente nas normas anteriores ao RiC-CM.

[6] Como nos formatos EAD e EAC.

[7] Por limitação de espaço neste artigo, apresentamos como exemplo do nosso estudo de transposição apenas o da área de identificação da Nobrade para a série Escravidão e para um de seus manuscritos.

[8] Nomeado pelo autor por inexistir um código de identificação no inventário analítico do AGCRJ.

[9] Cópias gentilmente cedidas pela diretora do AGCRJ para uso acadêmico.

[10] A web semântica interliga significados de palavras e, nesse âmbito, tem como finalidade conseguir atribuir um significado (sentido) aos conteúdos publicados na internet, de modo que este seja perceptível tanto pelo humano como pelo computador.

[11] Linked Open Data (LOD) usa tecnologias da web semântica para publicar dados estruturados na web e estabelecer links entre dados de uma fonte para outras.

[12] Ontologias são utilizadas em inteligência artificial, web semântica, engenharia de software e arquitetura da informação como uma forma de representação de conhecimento sobre o mundo ou alguma parte deste.

[13] OWL (Web Ontology Language) é uma linguagem para definir e instanciar ontologias na web. Uma linguagem OWL pode incluir descrições de classes e suas respectivas propriedades e relacionamentos.

[14] RDF é uma sintaxe padrão para representar um grafo dirigido em XML. Elementos do modelo: recurso (resource), propriedade (propertie), valor (value) e declaração (statement).

[15] Uniform Resource Identifier (URI) é um termo técnico que foi traduzido para a língua portuguesa como um "identificador uniforme de recurso", é uma cadeia de caracteres compacta usada para identificar ou denominar um recurso na internet.

[16] Embora não seja compatível com a tecnologia Open Archive Information System (OAIS), recurso importante para a preservação digital de documentos de arquivo (previsto para a próxima versão do LOA), a opção pelo repositório digital OmekaS (<https://omeka.org/>) deve-se por ele oferecer recursos para inserção de ontologias/vocabulários e publicação dos dados no ambiente Linked Open Data (LOD).

[17] <https://json-ld.org/>.

[18] SPARQL é uma linguagem para consulta a grafos RDF, padronizada pelo Data Access Working Group do World Wide Web Consortium (W3C).

[19] <https://ontop-vkg.org/>.

[20] <https://www.w3.org/TR/turtle/>.

[21] <http://vos.openlinksw.com/owiki/wiki/VOS>.

[22] Site em desenvolvimento pela equipe do projeto Linked Open Archives (LOA).

[23] Aplicação experimental do formato de descrição Delcom (descrição legível por computador) na série Escravidão, custodiada pelo Arquivo Geral da Cidade do Rio de Janeiro (AGCRJ) e apresentada em 1986, no VI Congresso Brasileiro de Arquivologia (BOTTINO, 2014, p. 103).

REFERÊNCIAS

ARQUIVO GERAL DA CIDADE DO RIO DE JANEIRO (AGCRJ). **Inventário analítico:** série Escravidão. Rio de Janeiro: AGCRJ, 2007.

BERNERS-LEE, T. **Linked data.** [S. l.]: w3.org, 2006. Disponível em: <http://www.w3.org/DesignIssues/LinkedData.html>. Acesso em: 2 dez. 2019.

BERNERS-LEE, T; HENDLER, J.; LASSILA, O. The semantic web: a new form of web content that is meaningful to computers will unleash a revolution of new possibilities. **Scientific American**, May 2001. Disponível em: https://www-sop.inria.fr/acacia/cours/essi2006/Scientific%20American_%20Feature%20Article_%20The%20Semantic%20Web_%20May%202001.pdf. Acesso em: 2 dez. 2019.

BOTTINO, M. **O legado dos congressos brasileiros de arquivologia (1972-2000)**. Rio de Janeiro: Editora da FGV, 2014.

BRASIL. Conselho Nacional de Arquivos. **Nobrade**: Norma Brasileira de Descrição Arquivística. Rio de Janeiro: Arquivo Nacional, 2006.

INTERNATIONAL COUNCIL ON ARCHIVES (ICA). **Records in contexts**: a conceptual model for archival description. Consultation Draft v0.2 (preview). Experts Group on Archival Description, 2019. Disponível em: https://www.ica.org/sites/default/files/ric-cm-0.2_preview.pdf. Acesso em: 12 dez. 2019.

INTERNATIONAL COUNCIL ON ARCHIVES (ICA). **Records In Contexts Ontology**. V0.1. ed., 2019. Disponível em: <https://www.ica.org/standards/RiC/ontology.html>. Acesso em: 12 dez. 2019.

INTERNATIONAL COUNCIL ON ARCHIVES (ICA). **Records in contexts**: a conceptual model for archival description. Consultation Draft v0.1 ed. [S. l.] Experts Group on Archival Description, 2016.

INTERNATIONAL COUNCIL ON ARCHIVES (ICA). **ISDIAH**: International Standard for Describing Institutions with Archival Holdings. 1. ed. London, United Kingdom: Committee on Best Practices and Standards, 2008.

INTERNATIONAL COUNCIL ON ARCHIVES (ICA). **ISDF**: International Standard for Describing Functions. 1. ed. Desden, Germany: Committee on Best Practices and Standards, 2007.

INTERNATIONAL COUNCIL ON ARCHIVES (ICA). **ISAAR (CPF)**: International Standard Archival Authority Record for Corporate Bodies, Persons and Families. 2. ed. Canberra, Australia: Committee on Descriptive Standards, 2004.

INTERNATIONAL COUNCIL ON ARCHIVES (ICA). **ISAD(G)**: General International Standard Archival Description: adopted by the Committee on Descriptive Standards, Stockholm, Sweden, 19-22 September 1999. 2nd ed. Ottawa: International Council on Archives, 2000.

ISOTANI, S.; BITTENCOURT, I. I. **Dados abertos conectados**: em busca da web do conhecimento. São Paulo: Novatec, 2015.

LAUFER, C. **Guia de web semântica**. São Paulo: Ceweb, 2015. Disponível em: <https://ceweb.br/guias/web-semantica/>. Acesso em: 9 nov. 2020.

MARCONDES, C. H. Linked data:dados interligados e interoperabilidade entre arquivos, bibliotecas e museus na web. **Encontros Bibli**: Revista Eletrônica de Biblioteconomia e Ciência da Informação, v. 17, n. 34, p. 171-192, 2012. DOI: 10.5007/1518-2924.2012v17n34p171.

MIRANDA, J. M. **Do multinível ao multidimensional**: análise do novo padrão de descrição Records in Contexts (RiC) e do seu impacto no acesso aos arquivos, à luz da web semântica.

2020. Relatório (Pós-doutorado em Ciência da Informação) – Programa de Pós-Graduação em Ciência da Informação, Universidade Federal Fluminense, Niterói, 2020.

MIRANDA, J. M. Tratamento da informação em arquivo: uma metodologia para construção de banco de dados referenciais. *In: CONGRESSO BRASILEIRO DE ARQUIVOLOGIA*, 6., 1986, Rio de Janeiro. **Anais** [...]. Rio de Janeiro: Associação dos Arquivistas Brasileiros, 1986.

RANGANATHAN, S. R. **Prolegomena to library classification**. Bombay: Asia Publishing House, 1967.