

Estudio y aplicabilidad de los metadatos de preservación de documentos en herramientas de software libre

**PLAN DE INVESTIGACIÓN
PROGRAMA DE DOCTORADO EN FORMACIÓN EN LA SOCIEDAD DEL CONOCIMIENTO
UNIVERSIDAD DE SALAMANCA**

Marcos Vinícius Bittencourt de Souza

DIRECTORES

María Manuela Moro Cabero

José Luis Bonal Zazo

07 de mayo de 2019

INTRODUCCIÓN Y JUSTIFICACIÓN

No mundo atual, quando a grande parte das comunicações entre pessoas, órgãos e entidades acontece através de documentos em meio digital, existe a demanda por preservar esses documentos para que possam ser usados como possíveis provas no futuro. Nesse caso, deve-se garantir que os documentos tenham a mesma forma e conteúdo em que foram criados, podendo-se usar metadados (UNE-ISO 23081-1:2018) como forma de se descrever o contexto de criação dos documentos, seu conteúdo, suas características e etapas do ciclo de vida.

Um dos padrões de metadados existentes para preservação é o PREMIS (*Preservation Metadata: Implementation Strategies*), que define quatro tipos de entidades semânticas (Direitos, Objeto, Agente e Evento) e seus devidos relacionamentos entre elas. Como forma de documentar a estrutura de cada entidade, são definidas propriedades que registram as características de cada entidade, formando um Modelo de Entidade-Relacionamento (PREMIS EDITORIAL COMMITTEE, 2015)

. Uma visão mais global e integrada sobre a descrição arquivística de documentos foi criada pelo Conselho Internacional de Arquivos (ICA - *International Council of Archives*) através do padrão *Records in Contexts: a conceptual model for archival description* (RiC-CM), que aborda a descrição dos documentos através de grafos (INTERNATIONAL COUNCIL ON ARCHIVES, 2016) diferentemente da metodologia praticada no padrão ISAD(G) (*General International Standard Archival Description*) que define a descrição arquivística dos documentos de maneira multinível (INTERNATIONAL COUNCIL ON ARCHIVES, 1999).

Na Espanha, o governo criou o Esquema de Metadados para a Gestão do Documento Eletrônico (e-EMGDE) (DIRECCIÓN DE TECNOLOGÍAS DE LA INFORMACIÓN Y LAS COMUNICACIONES, 2016), como complemento de uma norma técnica usada para a interoperabilidade de sistemas, que define um padrão de metadados a ser usado pelos sistemas do governo e por outros sistemas que desejem interagir com o governo. Esse padrão se baseou no padrão adotado pelo governo australiano AGRkMS (*Australian Government Recordkeeping Metadata Standard - Padrão de Metadados para Armazenamento de Registros do Governo Australiano*), com o qual possui uma tabela de correspondências entre os dois padrões de metadados, além de possuir uma tabela de correspondências com os metadados PREMIS. A Espanha, dispõe, igualmente, do *Modelo conceptual de descripción archivística y requisitos de datos Básicos de la descripción de documentos, agentes y funciones*. (COMISIÓN DE NORMAS ESPAÑOLAS DE DESCRIPCIÓN ARCHIVÍSTICA, 2017), de notável versatilidade e projeção para a preservação (LLANES-PADRÓN Y MORO-CABERO, 2017; REQUEJO ZALAMA, 2017). Por outro lado, no Brasil, o governo criou o padrão e-Arq Brasil (Modelo de Requisitos para SIGAD - Sistemas Informatizados de Gestão Arquivística de Documentos) que define um conjunto de requisitos e metadados a serem implementados pelos SIGADs desenvolvidos no país (CÂMARA TÉCNICA DE DOCUMENTOS ELETRÔNICOS, 2011).

Para fazer a preservação de documentos digitais, se faz uso de repositórios arquivísticos que permitem a gestão de documentos que devem ser acessíveis ao longo do tempo, sendo esse tempo um período de alguns poucos anos ou para a eternidade. A principal norma que define o comportamento e arquitetura de um repositório é a ISO-OAIS (*Open Archival Information System*) (UNE-ISO 14721:2015), que documenta as rotinas responsáveis por receber os documentos (processo de ingestão), administração dos documentos sob sua guarda e disponibilização dos documentos para acesso (processo de difusão).

Entre as ferramentas de repositório disponíveis atualmente está o Archivematica que é um sistema de código livre para a preservação digital baseado na norma ISO-OAIS coordenado pela empresa Artefactual Systems (ARTEFACTUAL SYSTEMS INC, 2019). Entre as características de uso do Archivematica está adoção do

Dublin Core ([ISO 15836-1:2017](#)) como o principal padrão de metadados da plataforma. Após a ingestão dos documentos, o repositório possui funcionalidades que permitem a manipulação dos valores correspondentes a esse padrão, no entanto, deve-se verificar qual é o tratamento realizado para os outros padrões de metadados utilizados na descrição dos documentos advindos de países que adotam um padrão próprio como, por exemplo, Espanha e Brasil.

Uma pesquisa aprofundada nos mecanismos internos do Archivematica deve ser realizada a fim de se identificar o tratamento dispensado a documentos com conjuntos de metadados diferentes do seu padrão suportado. Deve-se levar em consideração se os documentos brasileiros ou espanhóis, por exemplo, estão sendo corretamente preservados e não possuem perda de dados, fazendo com os esforços feitos durante a criação e gestão dos documentos se tornem inválidos.

HIPÓTESIS DE TRABAJO Y PRINCIPALES OBJETIVOS A ALCANZAR

O presente trabalho trata sobre a questão da preservação de documentos e seus respectivos metadados. O correto tratamento dessas informações permite que sejam conhecidos os detalhes a respeito dos documentos, seu contexto de criação e demais interações realizadas e, com isso, eles possam autenticidade e possam ser usados como fonte de prova quando necessário. No entanto, diante de diversos modelos conceituais de descrição arquivística, entre os quais RiC-CM e NEDA-CM, como está sendo tratada a preservação dos documentos? Ainda, à medida em que tais modelos de descrição arquivística são mapeados para conjuntos de metadados, as devidas preocupações a respeito de preservação estão sendo adequadas? Os softwares de código-livre usados como repositórios para armazenar o patrimônio documental das instituições estão realizando o devido tratamento para a preservação dos documentos e os seus respectivos metadados obtidos no decorrer do seu ciclo de vida?

A seguir, são descritos os objetivos a serem alcançados neste trabalho:

1. Determinar qual (quais) é (são) o(s) modelo(s) conceituais para preservar os documentos a longo prazo, que há escrito sobre eles, que normas há específicas de requisitos de dados e metadados no mundo, desde a sua evolução;
2. Comparar os metadados de preservação, as suas taxonomias, o seus elementos de metadados que são usados de acordo com as suas obrigatoriedades para as normas internacionais Dublin Core, OAIS, PREMIS e específicas nacionais: e-Arq Brasil; RiC-CM e NEDA-CM;
3. Avaliar quais são as recomendações no modelo OAIS e normas utilizadas para a certificação de repositórios (UNE-ISO 16363:2012) e recomendações para repositórios de terceiros (UNE-ISO 17068:2017);
4. Fazer uma proposta definitiva de metadados para a preservação que deveriam ser objeto de análise e aplicação nas ferramentas.
5. Comparação do tratamento de metadados de preservação entre as ferramentas de código-livre de repositórios disponíveis.
6. Sugerir modificações na arquitetura do Archivematica para permitir a complementação de metadados a respeito dos documentos sob sua guarda. Desenvolvendo-se, para isso, funcionalidades que permitam ao arquivista a escolha de um padrão de metadados e a apresentação de um formulário que possibilite informar os devidos valores correspondentes aos metadados desejados.
7. Aplicação prática do estudos em um centro espanhol e um centro brasileiro.

METODOLOGÍA

Na fase de revisão bibliográfica, serão identificados os modelos conceituais de descrição arquivística existentes atualmente realizando-se a análise das suas preocupações com a preservação dos documentos. Após o estudo sobre os modelos, serão pesquisados os trabalhos-base que definem os conceitos sobre os metadados para preservação, bem como, os últimos trabalhos publicados sobre o tema a fim de conhecer o estado da arte da preservação de documentos digitais com a utilização de metadados de preservação. Durante essa fase, deve-se ainda realizar um levantamento sobre os principais padrões de metadados utilizados internacionalmente destacando-se as suas principais características para a preservação. Com o intuito de se conhecer os casos específicos dos países Espanha e Brasil, deve-se estudar os padrões desenvolvidos para cada caso fazendo-se um comparativo entre as suas características e àquelas adotadas por normas internacionais.

De forma a se conhecer o funcionamento dos repositórios arquivísticos, deve-se pesquisar as principais ferramentas utilizadas pelas instituições que tratam a preservação de documentos digitais, focando na maneira com os metadados são utilizados e cumprem com o objetivo da preservação.

Um estudo aprofundado do repositório Archivematica permitirá compreender o seu funcionamento e seu comportamento com os padrões de metadados, sendo possível realizar um comparativo da sua forma de tratamento dos diferentes padrões disponíveis atualmente.

Através dos estudos desenvolvidos, será possível elaborar uma proposta de modificação na arquitetura do repositório para um melhor atendimento ao tema da preservação de documentos digitais. A validação da proposta deve ser feita com os arquivistas das instituições escolhidas para esse fim, identificando-se o cenário real e pontos a serem modificados e melhorados. Durante a fase de validação, devem ser utilizados documentos que possuam os padrões de metadados já suportados pelo Archivematica assim como os padrões criados especificamente para os países Espanha (e-EMGDE) e Brasil (e-Arq Brasil).

A partir da observação dos testes feitos com os arquivistas, será possível definir a proposta final sobre o tratamento de metadados no repositório Archivematica e criação de um protótipo a ser validado.

MEDIOS Y RECURSOS MATERIALES DISPONIBLES

O presente trabalho será desenvolvido no âmbito do programa de doutorado “Formación en la Sociedad del Conocimiento” (García-Peñalvo, 2014, 2018; García-Peñalvo et al., 2018), sendo o seu portal eletrônico acessível em <https://knowledgesociety.usal.es/> que será usado como principal ferramenta de comunicação e acompanhamento do desenvolvimento do projeto (García-Holgado et al., 2015). Para isso, no portal serão informadas todas as publicações realizadas e participações em cursos e congressos durante o período de realização dos estudos.

Para a realização dos estudos sobre as ferramentas software livre de preservação de documentos, o estudante vai usar os recursos disponíveis na Universidade. Ademais, para realizar o presente trabalho, deve-se dispor de um computador com sistema operacional Linux no qual seja possível instalar a última versão do repositório Archivematica. Assim mesmo, precisa-se de comunicação direta com aquelas instituições objeto-alvo que tenham implementado ferramentas de software livre como Archivematica.

PLANIFICACIÓN TEMPORAL

1. Fase I: revisão bibliográfica.
 - a. Primeiro ano: meses de janeiro a julho.
 - i. Revisão bibliográfica sobre o estado da arte no que se refere a repositórios digitais arquivísticos;
 - ii. Pesquisa sobre os padrões de descrições arquivísticas de documentos nacionais (Espanha e Brasil) e internacionais;
 - b. Primeiro ano: meses setembro-dezembro.
 - i. Estudo das normas de metadados para a preservação e seus usos nos repositórios;
 - ii. Desenvolvimento do estado do arte de modo escrito como capítulo;
 - iii. Participação em jornadas e seminários para apresentação dos resultados obtidos nesta fase.
2. Fase II: coleta de dados de análise
 - a. Segundo ano: meses de janeiro a julho.
 - i. Comparação das taxonomias empregadas: normas e experiências;
 - ii. Sistematização da proposta de metadados para preservação;
 - b. Segundo ano: meses de julho a dezembro.
 - i. Análise da arquitetura de desenvolvimento das aplicações dos repositórios de software livre. Caso específico do Archivematica;
 - ii. Elaboração e publicação das propostas.
3. Fase III: natureza aplicada
 - a. Terceiro ano: meses de janeiro a julho.
 - i. Estudo dos repositórios;
 - ii. Aplicação dos conceitos no Archivematica
 - iii. Prova de funcionamento e parecer dos responsáveis (arquivistas).
 - iv. Elaboração dos capítulos para a tese e publicação das propostas finais
 - b. Terceiro ano: meses maio a setembro.
 - i. Adaptação das propostas atendendo às experiências;
 - ii. Conclusão do desenvolvimento da solução planejada e realização de testes com arquivistas a fim de validar a solução alcançada;

Elaboração do texto final da tese.

REFERENCIAS

AENOR-Asociación Española de Normalización. **UNE-ISO/TR 14721. Sistema de transferencia de datos e información espaciales**. Sistema abierto de información de archivo (OAIS). Modelo de referencia. Madrid, 2015.

AENOR-Asociación Española de Normalización. **UNE-ISO 23081-1. Información y documentación**. Procesos de gestión de documentos. Metadatos para la gestión de documentos. Parte 1: Principios. Madrid, 2018.

ARTEFACTUAL SYSTEMS INC. **Archivemática: open-source digital preservation system**. Disponible em: <<https://www.archivematica.org/en/>>. Acesso em 18 março 2019.

CÂMARA TÉCNICA DE DOCUMENTOS ELETRÔNICOS. **e-ARQ Brasil: Modelo de Requisitos para Sistemas Informatizados de Gestão Arquivística de Documentos**. [s.l.] Arquivo Nacional, 2011.

COMISIÓN DE NORMAS ESPAÑOLAS DE DESCRIPCIÓN ARCHIVÍSTICA (CNEDA). **NEDA-MC: Modelo conceptual de descripción archivística: Entidades, relaciones y atributos**. Disponible em: <<http://www.mecd.gob.es/cultura/areas/archivos/mc/cneda/documentacion/normas/neda-mc.html>> Acesso em 04 de abr de 2019.

DIRECCIÓN DE TECNOLOGÍAS DE INFORMACIÓN Y LAS COMUNICACIONES (DTIC). **Esquema de Metadatos para la Gestión del Documento Electrónico (e-EMGDE)**. Versión 2.0 Documentación complementaria a la Norma Técnica de Política de gestión de documentos electrónicos. Madrid, 2016. Disponible em: <https://administracionelectronica.gob.es/pae/Home/dam/jcr:26d90313-7af0-4f69-9d62-c853d55d7627/Esquema_Metadatos_e-EMGDE_2016.pdf>. Acesso em 04 de abril de 2019.

GARCÍA-HOLGADO, A., GARCÍA-PEÑALVO, F. J., & RODRÍGUEZ-CONDE, M. J. **Definition of a technological ecosystem for scientific knowledge management in a PhD Programme**. In G. R. Alves & M. C. Felgueiras (Eds.), *Proceedings of the Third International Conference on Technological Ecosystems for Enhancing Multiculturality (TEEM'15) (Porto, Portugal, October 7-9, 2015)* (pp. 695-700). New York, NY, USA, 2015: ACM.

GARCÍA-PEÑALVO, F. J. **Formación en la sociedad del conocimiento, un programa de doctorado con una perspectiva interdisciplinar**. *Education in the Knowledge Society*, 15(1), 4-9. 2014.

GARCÍA-PEÑALVO, F. J. **Edición 2018-2019 del Kick-off del Programa de Doctorado "Formación en la Sociedad del Conocimiento"**. Seminarios del Programa de Doctorado en Formación en la Sociedad del Conocimiento (23 de octubre de 2018), Salamanca, España, 2018. <https://goo.gl/TBaxYD>

GARCÍA-PEÑALVO, F. J., GARCÍA- HOLGADO, A., & RAMÍREZ-MONTOYA, M. S. **The PhD Corner: TEEM 2018 Doctoral Consortium**. In F. J. García-Peñalvo (Ed.), *TEEM'18 Proceedings of the Sixth International Conference on Technological Ecosystems for Enhancing Multiculturality (Salamanca, Spain, October 24th-26th, 2018)* (pp. 979-983). New York, NY, USA, 2018: ACM.

INTERNATIONAL COUNCIL ON ARCHIVES. **Records in Context: a conceptual model for archival description**. Consultation draft v0.1, set. 2016. Disponible em: <<https://www.ica.org/sites/default/files/RiC-CM-0.1.pdf>>. Acesso em 04 de abril de 2019.

INTERNATIONAL COUNCIL ON ARCHIVES. **ISAD (G) : general international standard archival description : adopted by the Committee on Descriptive Standards, Stockholm, Sweden, 19-22 September 1999**. Disponible em: <https://www.ica.org/sites/default/files/CBPS_2000_Guidelines_ISAD%28G%29_Second-edition_EN.pdf>. Acesso em 04 abril de 2019.

LLANES-PADRÓN, D. y MORO-CABERO, M. **Record in contexts: un nuevo modelo para la representación de la información archivística en el entorno de la Web Semántica**. El profesional de la información. v.26. n.3 pp. 525-533. 2017. DOI: <https://doi.org/10.3145/epi.2017.may.19>

PREMIS EDITORIAL COMMITTEE. **PREMIS Data Dictionary for Preservation Metadata**. Version 3.0. 2015. Disponible em: <<https://www.loc.gov/standards/premis/v3/premis-3-0-final.pdf>> Acesso em 04 de abril de 2019.

REQUEJO ZALAMA, J. **Documentos en contexto: las nuevas normas de descripción archivística y su aplicación en la administración digital**. En El archivo electrónico en la administración digital: 23 Jornadas de Archivos Universitarios, 21-23 de junio de 2017, A Coruña (pp. 27-41). DOI capítulo: <https://doi.org/10.17979/spudc.9788497496803.027>