

## Proveniência e semântica: uso de ontologias na representação de documentos arquivísticos

Marieta Marks Löw<sup>1</sup>; Rafael Port da Rocha<sup>2</sup>

### RESUMO

O texto apresenta pesquisa de mestrado no contexto de representação semântica da produção documental arquivística. Trata-se de pesquisa exploratória aplicada, de abordagem qualitativa. A pesquisa tem por objetivo analisar as ontologias de documentos (IAO, D-Acts) na representação semântica da produção documental em contrato de partilha de petróleo no domínio de óleo e gás. Apresenta resumidamente pesquisa envolvendo análise das ontologias IAO e D-Acts, derivadas de BFO, e do modelo conceitual RiC-CM para representação semântica de documentos arquivísticos produzidos em razão de contrato de partilha para exploração e produção de petróleo e gás.

**Palavras-chave:** Proveniência Arquivística; Representação Semântica; Ontologias; Basic Formal Ontology; Record in Context Conceptual Model

### INTRODUÇÃO

O estudo aqui apresentado aborda a representação da documentação de contrato de partilha para exploração e produção de petróleo, a partir do olhar arquivístico e com o uso de ontologias como ferramentas de organização do conhecimento para representar semanticamente a informação. A documentação arquivística produzida em ambiente organizacional, quando observada na perspectiva da Proveniência, além de fonte informacional, por seu próprio conteúdo, reflete, no seu conjunto, uma estrutura de funções e atividades, bem como um contexto legal e

---

<sup>1</sup> Mestranda em Ciência da Informação; PPGCIN/UFRGS; marietalow@gmail.com

<sup>2</sup> Professor no Departamento de Ciência da Informação; FABICO/UFRGS; rafael.rocha@ufrgs.br

administrativo, do qual tais documentos se originam (COOK, 2012). A Arquivologia tem desenvolvido instrumentos que descrevem o contexto e o conteúdo de documentos, como as normas de descrição arquivística, por exemplo. Tais instrumentos, entretanto, não tiveram como enfoque a representação semântica da informação. Para sanar isso o Conselho Internacional de Arquivos (ICA) está desenvolvendo uma proposta de modelo conceitual denominada *Records in Context Conceptual Model (RiC-CM)* (ICA, 2019). Uma ontologia, apresentando os conceitos do modelo e seus relacionamentos, também foi desenvolvida. Contudo, esta não foi derivada da ontologia de topo *Basic Formal Ontology (BFO)*, que é usada em muitos setores da indústria para representação semântica de entidades de seu domínio. O uso de ontologias derivadas de BFO no domínio de óleo e gás é bastante interessante, pois permite a integração com outras ontologias do setor, e também facilita eventual reúso. Algumas ontologias derivadas de BFO trazem conceitos associados à produção de documentos, como a *Information Artifact Ontology (IAO)* (SMITH; CEUSTERS, 2015) e a *Ontology of Document Acts (D-Acts)* (BROCHHAUSEN; *et al.*, 2013), ambas originadas em estudos para ontologias na área da saúde.

A questão de pesquisa foi expressa da seguinte forma: como representar semanticamente a produção documental arquivística de um contrato de partilha de petróleo considerando as ontologias IAO e D-Acts e Modelo Conceitual RiC-CM? Partindo dessa questão, o objetivo da pesquisa é analisar a representação semântica da produção documental arquivística em um contrato de partilha de petróleo considerando as ontologias IAO e D-Acts e Modelo Conceitual RiC-CM. Como principais marcos do trabalho, para se alcançar esse objetivo final, estão: (1) analisar as ontologias de documentos IAO e D-Acts quanto à representação da produção documental usando os principais conceitos do RiC-CM como referência para análise; (2) demonstrar a representação da produção documental de um poço de petróleo em regime de partilha de produção a partir da metodologia arquivística; (3) enriquecer semanticamente a representação da produção documental a partir dos resultados encontrados na análise e avaliar esses resultados.

## METODOLOGIA

Trata-se de pesquisa exploratória aplicada, de abordagem qualitativa que tem por objeto a representação semântica da produção documental no domínio de óleo e gás. O método utilizado é bibliográfico documental. A análise das ontologias de documentos IAO e D-acts foi feita considerando sua relação com a BFO e utilizando os conceitos do modelo RiC-CM. A derivação de classes de IAO e D-Acts como subclasses de BFO é um ponto importante para semântica e reuso de conceitos dessas ontologias para representação dos documentos arquivísticos em estudo. Entretanto algumas definições são de cunho genérico e, em alguns casos, as subclasses existentes não são suficientes representar tal documentação.

Utilizando métodos e normas da Arquivologia, como as normas ISDF (*International Standard for Describing Functions*) e ISO 15489 e o modelo DIRKS (*Design and Implementing Recordkeeping System*), foi desenvolvida uma estrutura de representação da produção documental em estudo com base em uma hierarquia de funções e atividades, associadas às tipologias produzidas. Com essa etapa foi possível explicitar as relações orgânicas existentes entre os documentos e as funções e atividades do negócio, bem como relacionamento entre os documentos. Há uma relação entre funções e atividades e o conceito de ocorrente, de BFO, assim como o documento produzido se relaciona com o conceito de entidade material. A última etapa da pesquisa está em desenvolvimento. Nesta etapa os resultados encontrados nas etapas anteriores serão consolidados e analisados quanto a eventual enriquecimento semântico.

## CONCLUSÕES

A representação formal dos relacionamentos gerados pelos documentos é uma forma de organização do conhecimento que permite a criação de diferentes aplicações, tanto para recuperação da informação, como para outros serviços informacionais, como automatização da produção documental e do fluxo

informacional. Desse modo, representar os documentos e seus relacionamentos orgânicos no contexto complexo da indústria do petróleo pode significar uma melhora na capacidade de organização do conhecimento desse domínio, bem como o desenvolvimento de ferramentas mais efetivas na gestão e recuperação da informação.

## REFERÊNCIAS

BROCHHAUSEN, M.; ALMEIDA, M. B.; SLAUGHTER, L. Towards a formal representation of document acts and the resulting legal entities. **Johanssonian Investigations: Essays in Honour of Ingvar Johansson on his Seventieth Birthday**, [S. l.], p. 120–139, 2013. Disponível em: <https://doi.org/10.1515/9783110322507.120>. Acesso em: 14 mai. 2020.

COOK, T. A ciência arquivística e o pós-modernismo: novas formulações para conceitos antigos. **InCID: Revista de Ciência da Informação e Documentação**, Ribeirão Preto, v. 3, n. 2, p. 3–27, 2012. Disponível em: <https://doi.org/10.11606/issn.2178-2075.v3i2p3-27>. Acesso em: 12 mar. 2020.

INTERNATIONAL COUNCIL ON ARCHIVES ICA. **Records in Context: a conceptual modelo for archival description (Consulation Draft v0.2)**. [s. l.], 2019. Disponível em: [https://www.ica.org/sites/default/files/ric-cm-0.2\\_preview.pdf](https://www.ica.org/sites/default/files/ric-cm-0.2_preview.pdf). Acesso em: 12 fev. 2020.

SMITH, B.; CEUSTERS, W. Aboutness: Towards foundations for the information artifact ontology. *In*: 2015, **CEUR Workshop Proceedings**: CEUR-WS, 2015.