

# Proveniência arquivística em um poço de petróleo: uso de ontologias para representação semântica de documentos.

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

<sup>1</sup>Aluna no Programa de Pós-Graduação em Ciência da Informação (PPGCIN)  
Universidade Federal do Rio Grande do Sul (UFRGS) – Porto Alegre – RS – Brasil

<sup>2</sup>Professor do Departamento de Ciência da Informação na Faculdade de Biblioteconomia e Comunicação (FABICO)  
Universidade Federal do Rio Grande do Sul (UFRGS) – Porto Alegre – RS – Brasil.

marieta.low@ufrgs.br, rafael.rocha@ufrgs.br

**Abstract.** *The paper presents a master's degree research in the context of semantic representation of records production. It focuses on applied exploratory research, with a qualitative approach. The research aims at analyzing the document ontologies (IAO, D-Acts) in the semantic representation of records production in an oil production sharing contract in the oil and gas domain. It presents the problem of semantics in the representation of archival record production, theoretical references, and the proposed solution involving analysis of document ontologies, representation of record production based on archival theory, semantic enrichment with the result of analysis and evaluation of results.*

**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 um contrato de partilha de petróleo no domínio de óleo e gás. Apresenta o problema da semântica na representação da produção documental arquivística, o referencial teórico e a proposta de solução envolvendo análise de ontologias de documentos, representação de produção documental a partir da teoria arquivística, enriquecimento semântico da representação com resultado da análise e, ainda, avaliação dos resultados.*

## 1. Apresentação da pesquisa e contexto

A gestão da informação documental arquivística produzida em uma instituição é feita seguindo os princípios arquivísticos, sendo o principal deles o chamado Princípio da Proveniência, que estabelece que os documentos de uma mesma entidade produtora devem ser mantidos reunidos na instituição arquivística e com a organização dada por aquele produtor [Schellenberg, 2006]. Tal organização reflete, através dos documentos produzidos, as funções, os processos e as atividades dessa instituição produtora [Cook, 2012]. Tais princípios se consolidaram com o desenvolvimento da Arquivologia, enquanto campo de conhecimento, e são amplamente aceitos e aplicados em sistemas de gestão documental, tanto no momento de produção, na instituição de proveniência, como na etapa de custódia, na instituição arquivística. Os documentos, observados na perspectiva da Proveniência, além de fonte informacional por seu próprio conteúdo, refletem o fazer da instituição geradora.

Esse fazer, entretanto, não costuma ser expresso nas representações semânticas desenvolvidas em sistemas de informação, fator que dificulta seu uso integrado com ontologias de domínio e na recuperação da informação tendo como base a proveniência. Os sistemas de busca de documentos, por exemplo, são baseados, de maneira geral, na recuperação a partir do conhecimento sobre o domínio de uma determinada área em dado sistema de gestão de informação, e normalmente não incluem o contexto em que os documentos são produzidos. Já os instrumentos arquivísticos que expressam esse conhecimento não foram desenvolvidos para representação semântica, tornando difícil sua integração com outras representações semânticas do domínio em sistemas de informações.

Agregar uma camada semântica à representação da produção documental pode contribuir não somente para uma melhor compreensão do documento como também para o desenvolvimento de ferramentas de gestão e uso da informação mais eficientes e precisas.

Existem ontologias que trazem conceitos associados à produção de documentos. Vale citar, especialmente, a *Information Artifact Ontology* (IAO) [Smith, 2012; Smith & Ceusters, 2015] e a *Ontology of Document Acts* (D-Acts) [Brochhausen et al., 2013; Smith, 2014]. Ambas partem da *Basic Formal Ontology* (BFO) como ontologia de fundamentação, o que favorece o seu reuso, à medida que BFO é base para o desenvolvimento de várias ontologias de domínio, incluindo as desenvolvidas no escopo do projeto em que este estudo se insere. Ambas derivam de estudos de Barry Smith para ontologias na área da saúde. Mesmo não originadas de estudos da Arquivologia, percebemos uma visível convergência de conceitos para seu uso em representação da produção documental arquivística.

Pela via dos estudos arquivísticos, por outro lado, surge em 2016 uma proposta de modelo conceitual, ainda em desenvolvimento, apresentada pelo Conselho Internacional de Arquivos (ICA) denominada Record in Context Conceptual Model (RiC-CM) [International Council on Archives ICA, 2016; Llanes-Padrón & Pastor-Sánchez, 2017]. Esse modelo foi revisado e uma segunda versão rascunho (*draft*) foi apresentada em dezembro de 2019. A partir do modelo conceitual foi derivada uma ontologia chamada RiC-O. Essa ontologia apresenta os conceitos do modelo e uma série de relacionamentos possíveis. Contudo esta, por derivar do modelo RiC-CM, não está fundamentada em BFO, como as anteriormente citadas, dificultando o reuso e sua integração com outras ontologias que utilizam BFO.

Nesta pesquisa propomos um estudo sobre a representação da documentação produzida em razão do poço de petróleo, sob contrato de partilha, a partir do olhar arquivístico e com o uso de ontologias como ferramentas de organização do conhecimento para melhoria semântica dessa representação. Dessa forma, uma ontologia do domínio de produção e exploração do petróleo sob regime de partilha pode englobar também o fazer da instituição, sob o olhar da gestão arquivística de documentos, na qual funções, processos e atividades representam a proveniência dos documentos produzidos ou recebidos.

O objetivo final do estudo é analisar as ontologias de documentos (IAO e D-Acts) na representação semântica da produção documental arquivística em um contrato de partilha de petróleo no domínio de óleo e gás. 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.

Os documentos arquivísticos produzidos em razão da existência de um poço de petróleo permitem compreender as funções e atividades desenvolvidas pela indústria do petróleo, bem como conhecer os relacionamentos entre o produtor e os diferentes agentes da cadeia produtiva, como parceiros comerciais, agências governamentais e outros. 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

P&D e Inovação estão presentes à medida que o estudo: potencializa a organização da informação apresentando recursos que permitem recuperação e uso de documentos a partir de seu contexto de produção; analisa e classifica o documento a partir das funções do negócio que o geraram, a partir de regulamentações que constituem esse negócio; discute e enriquece, com o uso de ontologias, as técnicas e os modelos arquivísticos para representação de funções de negócio, frente à complexidade da representação da produção documental de um poço.

## **2. Metodologia e Proposta de Solução**

Trata-se de pesquisa exploratória aplicada, de abordagem qualitativa. Seu objeto será a representação da produção documental relacionada a funções e atividades desenvolvidas em um poço de petróleo sob contrato de regime de partilha. O método utilizado será bibliográfico documental. O levantamento dos documentos relacionados ao domínio de óleo e gás será feito a partir de literatura da área, documentação disponível em repositórios e acervos públicos e legislação da área. Para a análise de ontologias serão utilizados, além da literatura, a documentação disponível em repositórios.

O desenvolvimento da pesquisa se apoiará em instrumentos e técnicas da Arquivologia para o desenvolvimento da representação da produção documental e em ontologias para a análise de conceitos. Considerando o objetivo final da pesquisa, foram estabelecidas três principais etapas de desenvolvimento da pesquisa, como indicado anteriormente.

Na primeira etapa será feita uma análise das ontologias IAO e D-Acts. Para isso, elas serão analisadas quanto ao seu uso para representação da produção documental. Cada ontologia será analisada no seu conjunto e também quanto às entidades que podem servir à representação da documentação arquivística. Para isso serão usados os princípios arquivísticos e os principais conceitos do modelo RiC-CM como referência de análise. Os conceitos de BFO serão utilizados subsidiariamente para compreensão do contexto. Com esta etapa espera-se identificar estratégias para o enriquecimento semântico da representação documental, bem como apontar possíveis ausências conceituais nas ontologias analisadas.

Na segunda etapa será demonstrada uma representação da produção documental de um poço de petróleo em regime de partilha. Os instrumentos da Arquivologia darão apoio à definição da estrutura de representação da produção documental e subsidiarão a definição de estratégias a serem adotadas para o processo de construção da representação da produção documental de um poço. A estrutura do modelo de representação da produção documental tem como base esquemas de classificação de atividades de negócio propostos pela metodologia DIRKS (*Design and Implementing Recordkeeping System*), pela norma ISO

15489 e pela norma arquivística para a descrição de funções ISDF (International Standard for Describing Functions). Esses esquemas são estruturas que ligam documentos aos contextos de suas criações, contextos estes representados através de funções de negócio e suas subdivisões em atividades e transações. Dessas propostas são também utilizados os aspectos considerados relevantes para a construção de um esquema de classificação de atividades do negócio, quais sejam, os passos nomeados na metodologia como: Investigação Preliminar, Análise das Atividades de Negócio e Avaliação de Sistemas Existentes. Com essa etapa espera-se identificar a representação da produção documental a partir dos instrumentos tradicionais para o desenvolvimento de sistemas arquivísticos de gestão documental.

A terceira etapa tem como objetivo utilizar as estratégias de enriquecimento semântico identificadas na etapa um para melhorar a representação demonstrada na segunda etapa. A partir disso será feita uma avaliação dos resultados obtidos. Como critérios para essa avaliação serão considerados aspectos semânticos representados e não representados na representação com instrumentos tradicionais resultantes da segunda etapa e aspectos relacionados à integração da representação semântica obtida na terceira etapa com representação do domínio.

BFO é usada ao longo das etapas para apoiar a compreensão do contexto da produção documental. Considerando a complexidade que é a constituição e a operação de um poço de petróleo sob o regime de partilha, o que envolve diversificadas relações entre agentes (como consórcio, operador, membro de consórcio, agência regulamentadora), diferentes atividades de negócio (como exploração e produção) e diferentes conceitos (como bacia, área de produção, campo de petróleo, poço de exploração de petróleo, poço de produção de petróleo), considerou-se que uma ontologia de fundamentação como a BFO pode auxiliar na estruturação do modelo e na classificação da natureza dos conceitos envolvidos.

### **3. Revisão de Literatura**

Tradicionalmente, a produção documental de uma determinada instituição é vista pela Arquivologia como o resultado do fazer dessa instituição. O conjunto dos documentos produzidos compõe o arquivo da instituição. São justamente as características próprias de produção dos documentos e do contexto de produção que tornam dado conjunto de documentos único. Aí reside a base teórica da Arquivologia, consubstanciada no chamado Princípio da Proveniência [Bellotto, 2004]. O Dicionário Brasileiro de Terminologia Arquivística define como: “Princípio básico da arquivologia segundo o qual o arquivo(1) produzido por uma entidade coletiva, pessoa ou família não deve ser misturado aos de outras entidades produtoras. Também chamado princípio do respeito aos fundos”. Assim, os documentos de um dado produtor compõem um conjunto que representa o produtor, sua história, suas atividades e funções, ou seja, o chamado contexto de produção.

Nas últimas décadas, com o desenvolvimento tecnológico, a Arquivologia passou a dialogar com outros campos científicos, como a Ciência da Informação, Administração e Tecnologia da Informação. Essa aproximação com a tecnologia e, por consequência, com as questões de organização e recuperação da informação em ambiente digital, fez com que as práticas da área fossem também modernizadas e saíssem dos arquivos para atuar com soluções para as necessidades de gestão nas etapas de produção e uso da documentação. Os estudos da área se voltaram novamente para o Princípio da Proveniência e métodos de classificação e descrição dos conjuntos documentais. A metodologia de análise funcional se

consolida como a principal ferramenta para construção dos instrumentos de gestão e representação arquivística.

A classificação por funções está pacificada na literatura arquivística. Ela ganha força com as padronizações administrativas e o desenvolvimento da burocracia moderna, a partir do final de Segunda Guerra Mundial. Isso proporcionou uma racionalização e, ao mesmo tempo, um aumento exponencial da complexidade de produção e uso dos documentos jurídico-administrativos. A partir de estudos realizados na Austrália foi construída uma metodologia para desenvolvimento e implantação de sistemas de gestão de documentos chamada DIRKS – *Designing and Implementing Recordkeeping Systems*. Para Dias [2010, p. 124], essa metodologia “assegura que a gestão da informação esteja baseada nas necessidades do negócio da organização, e, por isso, possibilita um estudo minucioso das atividades, processos e funções.”

Apesar das normas e do desenvolvimento, nas últimas décadas, de metodologias para classificação e organização da documentação arquivística, os estudos aproximando a documentação de arquivo com a área de Organização do Conhecimento ainda são tímidos. Entretanto, já é possível localizar alguns autores que tratam do tema, tanto epistemologicamente, quanto metodologicamente [Tognoli & Guimarães, 2018].

A representação semântica de documentos em ontologias foi também discutida para desenvolvimento de ontologias para a área médica. A necessidade de entidades que retratassem os atos sociais foi apontada por Smith [2012]. Desses estudos derivaram as duas ontologias aqui estudadas. A IAO é uma ontologia de entidades de informação, tendo sido “criada para servir como um recurso de domínio neutro para a representação de tipos de entidades de conteúdo informacional (ICE) como documentos, bases de dados e imagens digitais” [Smith & Ceusters, 2015, p. 1, tradução nossa]. Nessa ontologia há uma relação entre a entidade informacional e a realidade expressada pelo assunto (*aboutness*). Além disso, essa entidade informacional, para existir, depende de uma entidade material, como um contrato, que é uma entidade informacional que depende de um suporte para existir. Essa dependência é genérica, pois o contrato pode mudar de entidade material, ao passar a existir em um suporte eletrônico, por exemplo.

A D-Acts originou-se na teoria dos atos documentais proposta por Smith [2012]. Está fundamentada na concepção de que os documentos possuem estabilidade e, portanto, podem ser preservados no tempo. Além disso, os documentos são a materialização de ações que ocorrem no mundo real, cujo registro atesta essa ação. Ao mesmo tempo, o fato de a ação ser registrada traz consequências sociais que impactam o mundo real. O mesmo documento pode servir a diferentes atos sociais ao longo de sua existência, e ter diferentes destinatários.

Smith [2014, pp. 4–5] observa a relação entre documentos e atos de fala: os documentos são objetos, portanto, continuantes, pela BFO, enquanto os atos de fala são eventos, ou seja, ocorrentes. Isso significa que os documentos não apenas se mantêm idênticos no tempo, como se separam de seu criador e ganham vida própria. É possível, assim, entender ODA como uma ontologia que estende a IAO pela incorporação da teoria dos atos sociais. Com isso, é possível descrever o documento e o que as pessoas podem fazer com os documentos.

Em estudo sobre a dimensão normativa das organizações, Almeida e outros [2017] verificam a possibilidade de aplicação da ODA em contextos institucionais. São analisadas o que os autores entendem por dimensão descritiva (estrutura física) e dimensão prescritiva

(estrutura de regras). Essa abordagem nos parece bastante interessante e pode contribuir com a compreensão sobre o uso de ontologias de documentos em ambientes corporativos. Já a abordagem arquivística de gestão de documentos utiliza as funções e as atividades das organizações para nortear a classificação da documentação produzida, recebida e acumulada. Podemos inferir que o documento é visto, então, como consequência de atos sociais que ocorrem na realidade. Eles são criados para registrar ações e informações da realidade, e podem produzir efeitos nessa realidade.

## Referências

- Almeida, M. B., Brochhausen, M., Silva, F. B., & Santos, R. B. M. dos. (2017). Ontological approach to the normative dimension of organizations: an application of Documents Acts Ontology. *Ciência Da Informação*, 46(1).
- Bellotto, H. L. (2004). *Arquivos Permanentes, Tratamento documental*. (2ª Ed.). FGV.
- Brochhausen, M., Almeida, M. B., & Slaughter, L. (2013). Towards a formal representation of document acts and the resulting legal entities. *Johanssonian Investigations: Essays in Honour of Ingvar Johansson on His Seventieth Birthday*, 120–139. <https://doi.org/10.1515/9783110322507.120>
- Cook, T. (2012). A ciência arquivística e o pós-modernismo: novas formulações para conceitos antigos. *InCID: Revista de Ciência Da Informação e Documentação*, 3(2), 3–27. <https://doi.org/10.11606/issn.2178-2075.v3i2p3-27>
- Dias, C. C. (2010). Análise de domínio organizacional na perspectiva arquivística: potencialidade no uso da metodologia DIRKS - Designing and Implementing Recordkeeping Systems [Universidade Federal de Minas Gerais]. In *Ciência da Informação* (Vol. 40, Issue 2). <https://repositorio.ufmg.br/handle/1843/ECIC-85CNSE?mode=full>
- International Council on Archives ICA. (2016). *Records in Context: a conceptual model for archival description (Consultation Draft v0.1)*. <https://www.ica.org/sites/default/files/RiC-CM-0.1.pdf>
- Llanes-Padrón, D., & Pastor-Sánchez, J. A. (2017). Records in contexts: the road of archives to semantic interoperability. *Program*, 51(4), 387–405. <https://doi.org/10.1108/PROG-03-2017-0021>
- Schellenberg, T. R. (2006). *Arquivos Modernos, princípios e técnicas* (6ª Ed.). FGV.
- Smith, B. (2012). How to Do Things with Documents. *Rivista Di Estetica*, 50(2012), 179–198.
- Smith, B. (2014). Document Acts. In A. Konzelmann-Ziv & H. B. Schmid (Eds.), *Institutions, Emotions, and Group Agents: Contributions to Social Ontology* (pp. 19–31). Springer. [https://doi.org/https://doi.org/10.1007/978-94-007-6934-2\\_2](https://doi.org/https://doi.org/10.1007/978-94-007-6934-2_2)
- Smith, B., & Ceusters, W. (2015). Aboutness: Towards foundations for the information artifact ontology. *CEUR Workshop Proceedings*, 1515.
- Tognoli, N. B., & Guimarães, J. A. C. (2018). Provenance. In *Encyclopedia of Knowledge Organization*. International Society for Knowledge Organization. <https://www.isko.org/cyclo/provenance>