

Aspectos do ciclo de vida de dados em processos de construção de ontologias biomédicas

Jeanne Louíze Emygdio^a e Eduardo Ribeiro Felipe^b

Resumo: Tendo por foco o estudo das coisas existentes no mundo e das relações que estas estabelecem entre si, as Ontologias vêm ganhando campo especial de investigação científica por serem consideradas alternativas viáveis à organização das informações massivas e ininterruptamente alimentadas pelos seres humanos nos espaços digitais, por meio das tecnologias da informação e comunicação (TICs). As perspectivas centrais da utilização das ontologias repousam sobre as possibilidades de prover interoperabilidade semântica entre os sistemas de informação, pouco alcançada por meio dos modelos, técnicas e ferramentais usualmente aplicados. O objetivo deste artigo é descrever de que forma a gestão do ciclo de vida dos dados pode contribuir para prover maior qualidade à construção de ontologias biomédicas. A metodologia adotada foi a pesquisa qualitativa, de caráter descritivo sobre os processos de construção da Ontologia realista Obstétrica e Neonatal – OntoNEO. Os resultados encontrados comprovaram a eficiência do plano de gestão de dados como requisito de atribuição de qualidade à referida ontologia, o que a habilitou a integrar o projeto Basic Formal Ontology – BFO, padrão de referência na construção de ontologias realistas de alto nível.

Palavras-chave: Ontologias biomédicas. Ciclo de vida de dados. Plano de gestão de dados. Representação do conhecimento.

Aspects of the data life cycle in processes of construction of biomedical ontologies

Abstract: With the focus of the study of the existent things in the world and of the relations that these establish with each other, the Ontologies have been gaining special field of scientific investigation for being considered viable alternatives to the organization of the massive information and uninterruptedly fed by the human beings in the digital spaces, through of information and communication technologies (ICTs). The central perspectives of the use of ontologies rest on the possibilities of providing semantic interoperability among information systems, little achieved through the models, techniques and tools usually applied. The purpose of this article is to describe how the data life cycle management can contribute to provide higher quality to the construction of biomedical ontologies. The methodology adopted was the qualitative research, of a descriptive character on the construction processes of the realistic Ontology Obstetric and Neonatal – OntoNEO. The results obtained proved the efficiency of the data management plan as a requirement of quality attribution to this ontology, which enabled it to integrate the Basic Formal Ontology – BFO project, a reference standard in the construction of high level realist ontologies.

Keyword: Biomedical ontologies. Data life cycle. Data management plan. Knowledge representation.

a Universidade Federal de Minas Gerais (UFMG). E-mail: jeanne.emygdio@gmail.com. ORCID: <https://orcid.org/0000-0002-7329-4447>. Currículo: <http://lattes.cnpq.br/6849175974671037>

b Universidade Federal de Minas Gerais (UFMG). E-mail: erfelipe@hotmail.com. ORCID: <https://orcid.org/0000-0003-1690-2044>. Currículo: <http://lattes.cnpq.br/1010588591399870>

1 Introdução

1.1 Ontologias

De origem filosófica, o estudo das ontologias busca compreender a realidade dos seres, sua essência, princípios e causas (SEARLE, 1998), baseando-se em exaustivas classificações dos elementos do mundo real visando identificar seu tipo (categoria), estrutura, propriedades, eventos, processos e as inter-relações estabelecidas entre estes elementos (SMITH, 2003).

Atualmente, estes estudos têm sido caracterizados pela interdisciplinaridade, envolvendo campos como a Ciência da Informação (VICKERY, 1997; SOERGEL, 1999; ALMEIDA e BAX, 2003; ALMEIDA, 2013) e a Ciência da Computação (WAND e WEBER, 1990; GUARINO, 1998; WAND, STOREY e WEBER, 1999; CANTELE et al, 2004; FONSECA, 2007), com implicações em domínios da medicina (FARINELLI et al, 2016; MENDONÇA e ALMEIDA, 2016, ALMEIDA e FARINELLI, 2017), gestão e organização do conhecimento (TORRES, SIMÕES e ALMEIDA, 2017), biologia (HARRIS e PARKINSON, 2003), engenharia, gestão pública (FARINELLI, SILVA e ALMEIDA, 2013), direito (BREUKERS e HOEKSTRA, 2004; GRIFFO, ALMEIDA e GUIZZARDI, 2015) e geografia (RODRIGUES, CHAVES e SILVA, 2006; MACHADO et al, 2011).

Em função do desenvolvimento da Web Semântica (BERNERS-LEE; HENDLER e LASSILA, 2001) e da necessidade de lidar com a recuperação adequada de massivas informações digitais (ALMEIDA; SOUZA e BARACHO, 2015), as ontologias têm sido utilizadas como recurso alternativo para a representação da realidade em função da formalidade com que seus modelos são concebidos, baseados em lógica descritiva e axiomas formais. Quando convertidos em artefatos de informação proveem vocabulários semânticos consensuais que viabilizam a interoperabilidade semântica entre sistemas computacionais.

Na área da saúde, conforme pesquisas realizadas nas últimas décadas, grande esforço tem sido destinado a integrar e recuperar dados para uso profícuo destes sistemas, especialmente visando a criação do prontuário eletrônico de pacientes – EHR (RECTOR, 2000; SMITH et al, 2007; ANDRADE, ALMEIDA e SCHULZ, 2012; FARINELLI et al, 2016; MARTÍNEZ-COSTA, C., SCHULZ, 2017; BODENREIDER, 2018; HAENDEL et al, 2018).

Em função da complexidade dos dados e informações biomédicas (insumos para a construção de ontologias), compreende-se a relevância das pesquisas no entorno de seu ciclo de vida. Embora a incipiência de tais pesquisas, iniciativas de aspectos complementares podem ser observadas.

Em Ibraim e Ataelfadiel (2017), e Noy et al (2010), o ciclo de vida é visto sob o enfoque principal da Engenharia de Ontologias, subárea da Engenharia do Conhecimento e que aborda processos, métodos e metodologias de desenvolvimento e disseminação de ontologias, bem como plataformas tecnológicas relacionadas (NOY et al, 2010). Embora em Noy et al (2010) as análises ocorram sobre a construção de ontologias biomédicas, apenas em Ibraim e Ataelfadiel (2017) encontram-se aportes teóricos para classificação dos dados em um ciclo de vida de ontologias.

Em VUMC (2005) encontra-se o modelo “Translational Research” que visa a descrição do processo de incorporação de resultados de uma pesquisa científica às práticas reais e, a partir

delas, a retroalimentação de novas pesquisas. O modelo compreende aspectos relacionados aos processos de acesso a dados, sendo composto por quatro ciclos sequenciais que abrangem a coleta, o consumo e a geração de insumos para um novo processo ou pesquisa.

Em Hunkler et al (2011), encontra-se a apresentação de um processo de governança de grandes volumes de dados provenientes da área da saúde, coletados via surveys (Survey of Health, Ageing and Retirement in Europe – SHARE). O processo envolve as tarefas e atores relacionados à governança, além de cuidados especiais para a disseminação pública de dados sensíveis.

Na seção seguinte são apresentados aspectos teóricos acerca dos dados e ciclos de vida que serviram de subsídio para análise de aderência das etapas de construção de uma ontologia biomédica e de identificação dos aspectos de qualidade resultantes.

1.2 Dados e Ciclos de Vida

O conceito de dado tem sua referência histórica na teologia, sua utilização tanto no singular, quanto no plural é encontrada em registros que datam do século XVII (BORGMAN, 2015). É importante ressaltar que em sua percepção, o significado do dado não reside nele mesmo, sendo reflexo de um contexto, um símbolo, um referencial que depende da interpretação do observador.

Embora pareça uma abstração óbvia em nosso dia a dia, o dado não possui uma definição simples, devido sua ampla abrangência semântica. Ele é o insumo base tanto para os sistemas de informação, quanto para os instrumentos informacionais como Listas Controladas, Taxonomias, Tesouros e Ontologias. Constitui-se de uma unidade celular que acrescida de metainformação e semântica, contribui para que a informação seja produzida. Esta visão estruturada em camadas (dado – informação – conhecimento) possui ampla aceitação nos estudos e práticas referentes ao estudo da informação e do conhecimento.

No entanto, sua definição tem sido simplificada e associada a exemplos como números, símbolos, letras e fatos (National Research Council 1999, apud Borgman). O delineamento de suas fronteiras conceituais muitas vezes não são claras e se mesclam com conceituações sobre informação. Neste trabalho um conceito norteador para dado, é o utilizado pela Organisation for Economic co-operation and Development (OECD1) em sua definição sobre “dado de pesquisa”; a organização integrada por 36 nações membro, expressa que: dados de pesquisa são definidos como registros factuais (pontuações numéricas, registros textuais, imagens e sons) usados como fontes primárias para pesquisa científica, e que são comumente aceitos na comunidade científica como necessários para validar os resultados de uma pesquisa. Um conjunto de dados de pesquisa constitui uma representação parcial e sistemática do assunto investigado.

O dado, em comparação com outras estruturas que lidam com informação, como sistemas computacionais por exemplo, também possui “Ciclo de Vida”: Um modelo sequencial de ações que resultam em decisões na utilização/manutenção do dado. Briney (2015) elenca uma comparação entre ciclo de vida antigo e novo na perspectiva do dado. Na Figura 1, do Apêndice A, são representadas as quatro etapas consideradas como ciclo de vida antigo, sendo que na primeira etapa, “Planejamento do Projeto”, é necessário pensar em quais dados devem ser adquiridos. O próximo passo é trabalhar a “Aquisição dos Dados” propriamente dita. Em

seguida é vista a “Análise dos Dados” para que a última etapa, “Publicação” possa ser realizada. A autora destaca que neste modelo composto por quatro estágios, a aquisição e análise são críticos para o sucesso do processo, e sua crítica reside no sentido de que este ciclo não recompensa o uso do dado além da análise, considerando-os como um subproduto da pesquisa em vez de serem produtos com importância em si mesmos, como um artigo por exemplo. Este modelo foi usado por centenas de anos, mas com o advento digital, o ciclo demanda outras etapas, abordadas no Novo Ciclo de Vida.

Este novo conjunto de processos considera “O Compartilhamento de Dados”, a “Preservação” e o “Reuso” como passos no processo de pesquisa, além de incluir “Planejamento de Gerenciamento de Dados e Projeto”, “Aquisição de Dados”, “Análise de Dados”, “Publicação de Artigos” e “Divulgação de Dados”, “Preservação” e “Reuso”. Na Figura 2, do Apêndice A, são apresentados estes processos e seu fluxo. Este ciclo assemelha-se a outras propostas, como a da DataONE(2018), demonstrada na Figura 3 no mesmo apêndice.

Para Mearian (2003), o ciclo de vida informacional também deve abordar sua persistência, o suporte para sua gravação e recuperação. Observou-se nos artigos pesquisados que a preocupação com a questão processual das etapas de tratamento do dado em seu ciclo de vida foi uma etapa posterior, anteriormente havia uma atenção em relação à arquitetura de armazenamento. Período aquele em que muitas informações foram perdidas por falta de hardware compatível com mídias substituídas por outras. Políticas de migração de dados entre sistemas e sua possível interoperabilidade entre aplicações diversas eram importantes para minimizar o impacto da evolução tecnológica.

Diversos autores e instituições propuseram ciclos de vida de dados abordando processos que, em várias etapas se assemelham. Santana (2013) destaca que não é possível abordar um processo de ciclo de vida sem considerar o contexto em que os dados devam ser tratados, bem como sua especificidade, seja administrativa, descritiva, técnica, estrutural ou de preservação. O autor ainda propõe um Ciclo de Vida dos Dados para Ciência da Informação, CVD-CI, representado na Figura 4, do Apêndice A.

Em Lenhardt et al (2014) encontram-se passos comuns em várias propostas de ciclo de vidas observadas e já descritas anteriormente, tais como: Planejar, Coletar, Controlar Qualidade, Documentar, Preservar, Usar.

Justifica-se o investimento na adoção de um ciclo de vida de dados por benefícios como: facilidade na recuperação, uso e análise dos dados, tanto pelo pesquisador como por colaboradores; fornecimento de dados de qualidade para pesquisas translacionais; melhoria na tomada de decisão; incremento do valor investido na pesquisa; expansão da compreensão sobre um determinado problema e/ou domínio, etc. Um gerenciamento eficiente pode permitir que cientistas externos ao projeto possam encontrar, entender e utilizar os dados no futuro. Adicionando a este cenário a documentação por metadados, a possibilidade de citações e créditos futuros é garantida, além de permitir a rastreabilidade autoral.

2 Objetivos

Embora robustas as pesquisas voltadas à aplicação de ontologias na área da saúde, as respectivas práticas de governança sobre os dados tratados nestas pesquisas precisam ser

aprimoradas. Conforme apresentado anteriormente, não foram identificadas práticas de governança do ciclo de vida do dado dentro do ciclo de vida ontológico, sendo este o fator motivador da presente pesquisa.

Desta forma, o objetivo deste estudo será o de compreender o ciclo de vida de dados biomédicos a fim de representá-los em uma ontologia e, por extensão, compreender ainda, de que forma a gestão deste ciclo de vida poderá agregar maior qualidade à representação de tais dados.

Para o estudo foi selecionada a Ontologia Obstétrica e Neonatal – OntoNEO2 (FARINELLI, ALMEIDA, ELKIN e SMITH, 2016), uma ontologia realista que abrange os domínios de cuidados no pré-natal, intraparto e pós-natal. Tal escopo foi estabelecido baseando-se na análise de um conjunto de padrões de EHR (nacionais e internacionais), sistemas de EHR adotados por diferentes hospitais e entrevistas com obstetras americanos e brasileiros visando identificar os fluxos de trabalho das clínicas de saúde e as demandas informacionais dos médicos envolvidos. A ontologia foi construída por membros do grupo de Pesquisa RECOL3 da Universidade Federal de Minas Gerais (UFMG), do qual os autores do presente artigo participam.

3 Procedimentos Metodológicos

A pesquisa de natureza qualitativa e caráter descritivo baseou-se em uma análise dos procedimentos que conduziram a construção da OntoNEO, em comparação com as etapas dos ciclos de vida previamente apresentados neste estudo. Estes procedimentos foram organizados em etapas selecionadas dos estudos de Briney (2015), DataONE (2018) e Dias (2018). Por fim foram analisados aspectos de qualidade agregados à OntoNEO em função das práticas adotadas em sua construção. Os resultados serão apresentados na próxima seção.

4 Resultados

Os resultados das análises realizadas serão apresentados em cinco etapas, a saber: 1) Informações sobre dados e formatos de dados, 2) Conteúdo dos metadados e formatos, 3) Políticas para acesso, compartilhamento e reúso, 4) armazenamento de longo prazo e, 5) questões orçamentárias.

4.1 Informações sobre dados e formatos de dados

A pesquisa foi realizada a partir do consumo e da produção de dados experimentais variados, como:

(i) Dados brutos ou derivados:

- a) consumo de dados provenientes dos padrões adotados para a definição de seu escopo e para prover insumos de dados para a construção da ontologia: Registro de Saúde da Mulher, Registro Anteparto e Formulário Pós-Parto fornecidos pelo Colégio Americano de Obstetras e Ginecologistas (ACOG); o Formato de Registro Eletrônico de Saúde das Crianças fornecido pela Agência de Pesquisa e Qualidade em Assistência à Saúde (AHRQ); Padrões para a Estrutura Clínica e Conteúdo dos Registros de Pacientes fornecidos pelo Centro de Informação de Saúde e Assistência Social do Reino Unido e pela Academia de Faculdades Reais de Medicina; Padrão Nacional para Informações de

Resumo de Alta do Paciente da Autoridade Australiana de Informações e Qualidade em Saúde. Observou-se a preocupação com as especificidades terminológicas, técnicas e administrativas do contexto, o que, em termos de qualidade, contribuiu para a expansão da compreensão sobre o domínio do problema e para a padronização dos insumos de dados para a construção da ontologia. Tais práticas se assemelham aos processos de Planejamento e Coleta encontrados em DataONE (2018);

b) consumo de dados provenientes de pesquisa translacional: um conjunto de 11 ontologias abertas, previamente construídas e publicadas no Bioportal⁴ serviram de base para a construção da OntoNEO. Tais artefatos proveram classes, entidades e relacionamentos para reuso, em aderência aos princípios da OBO Foundry⁵ para a construção de ontologias interoperáveis voltadas à representação de realidades biológicas e biomédicas. Entre elas encontram-se a Ontologia de Artefatos e Informação (IAO), a Ontologia dos Genes (GO) e a Ontologia de Doenças Humanas (DOID). Neste sentido observou-se a preocupação com a qualidade dos dados brutos a serem reutilizados, com a perspectiva de prover, ao término da construção, uma ontologia de qualidade para alimentar novas pesquisas translacionais, conforme visto em VUMC (2005).

(ii) Modelos e saídas

a) Produção de modelo de representação do conhecimento abrangendo coleções físicas de classes, atributos e relacionamentos, distribuídas em três conjuntos de dados: informações básicas sobre os cuidados de saúde pertencentes ao escopo da ontologia; dados de formulários e documentos relacionados e entidades sociais envolvidas no contexto. A OntoNEO encontra-se disponível publicamente no repositório de ontologias biomédicas Bioportal. Neste aspecto observou-se a preocupação com as especificidades técnicas e administrativas para que a ontologia pudesse ser publicada respeitando-se as normas de qualidade da OBO Foundry. Tal cuidado constituiu em um incremento do valor investido na pesquisa uma vez que possibilita à comunidade externa o acesso e reuso de sua produção. As práticas adotadas assemelham-se ao processo de Publicação/Divulgação de Dados descrito no novo modelo de Briney (2015);

b) Realização de Prova de Conceito (PoC) sobre a capacidade da ontologia em prover interoperabilidade semântica no domínio estabelecido. Estes resultados podem ser encontrados em Farinelli, Almeida e Emygdio (2018), tendo por saída a proposição de uma arquitetura de interoperabilidade semântica. Observou-se a preocupação com a validação do dado produzido (OntoNEO) quanto à proposta de se constituir em um artefato de informação realmente capaz de integrar semanticamente sistemas computacionais. A PoC garante a eficácia da ontologia produzida e a qualidade do processo de construção.

(iii) Materiais científicos

a) Publicação de artigos científicos relacionados à pesquisa: no período de 2013 a 2017, relativo à produção da pesquisa de doutorado que culminou na construção da OntoNEO, foram produzidos 11 artigos, direcionados a periódicos científicos e congressos e simpósios de grande reconhecimento na área. Dentre eles destaca-se a publicação no

JASIST (ALMEIDA e FARINELLI, 2017). Neste aspecto observou-se a preocupação quanto a divulgação dos conhecimentos teóricos e empíricos que poderão nortear a condução de novas pesquisas, na comunidade nacional e internacional. Tais atividades relacionam-se às previstas em Briney (2005), e compreendem uma contrapartida relevante às agências financiadoras, além de agregarem valor ao reconhecimento da pesquisadora.

(iv) Sobre processamento de dados

A construção da OntoNEO baseou-se em um modelo de ciclo de vida iterativo e incremental, provendo desenvolvimento gradativo da ontologia, com a definição de novas entidades e relações a cada iteração. Três níveis conceituais foram desenvolvidos antes de sua formalização: a) definição do escopo da iteração; b) avaliação de aderência às ontologias preexistentes no domínio para a definição de vocabulário consensual. Os seguintes critérios foram estabelecidos para a seleção de classes para reuso: i) a ontologia em questão é aderente aos princípios da OBO Foundry?; ii) as classes possuem definição?; iii) quantas vezes o termo foi previamente reutilizado? iv) qual o status da ontologia no repositório da OBO Foundry?; c) construção de novos termos e relações para atender aos requisitos da OntoNEO não cobertos na fase anterior. As ferramentas OntoFox e Protegé foram utilizadas para importação dos termos e axiomas necessários para a geração do arquivo em formato OWL da OntoNEO. O controle de versões da ontologia foi realizado a partir do ambiente colaborativo para construção de ontologias CMAP Tools.6 Cada versão foi submetida para validação por representantes de diferentes comunidades de especialistas, incluindo médicos. Há compromisso do autor principal em manter atualizações desta ontologia, em função dos avanços científicos na área. O controle de qualidade realizado nestes procedimentos repousam sobre os dados de reuso para construção da OntoNEO, integração de versões, e preservação da ontologia, consonância com Briney (2005), DataONE (2018) e Dias (2018).

4.2 Conteúdo dos metadados e formatos

Conforme citado anteriormente, foram utilizados diversos padrões para formalização de conteúdos de metadados e formatos, além dos princípios da OBO Foundry, essenciais para a construção de ontologias realistas de alto nível agregáveis ao repositório em questão. Houve preocupação em relação à qualidade de representação dos dados, aumentando a possibilidade de reuso futuro e realimentação de pesquisas translacionais e independentes de domínio em função do nível da ontologia.

4.3 Políticas para acesso, compartilhamento e reuso

A OntoNEO segue ainda as diretrizes da OBO Foundry para acesso, compartilhamento e reuso. Os direitos autorais são garantidos aos autores e devem ser mantidos pela licença Creative Commons (License: CC-BY). Tais práticas, associadas às mencionadas no item 4.2 contribuem para facilitar a rastreabilidade autoral, garantir maior número de citações e créditos futuros. E ainda, aproximam-se dos processos de controle encontrados em Dias (2018).

4.4 Armazenamento a longo prazo

A ontoNEO é um projeto que se encontra na fase inicial de desenvolvimento, podendo ser acessada pelos sites: <<https://ontoneo.com/>> e <<https://bioportal.bioontology.org/ontologies/>>. O ontoNEO, onde será mantida em caráter permanente. Os aspectos de qualidade envolvidos neste item relacionam-se aos processos de preservação do dado, políticas de arquivamento, padrões da comunidade e atribuição de responsabilidades, encontrados em Briney (2005) e Dias (2018).

4.5 Questões orçamentárias

O projeto foi mantido em parte por meio de financiamento das agências de fomento CAPES e CNPQ ao autor principal. Outra parte do projeto foi financiada pelo NIH NCATS sob o prêmio CTSA Número UL1TR001412. As questões orçamentárias são previstas no plano de gestão de dados proposto por Dias (2018).

5 Considerações finais

O presente artigo apresentou uma revisão teórica acerca dos estudos sobre ontologias e sobre os ciclos de dados de pesquisa buscando traçar um caminho de compreensão da relevância da adoção das práticas de governança de dados como premissa para a atribuição de maior qualidade aos processos de construção de ontologias biomédicas.

Uma análise descritiva foi realizada sobre todas as fases de construção da ontologia ontoNEO, abrangendo desde a aquisição de variados dados brutos, puros ou provenientes de pesquisas translacionais, até a publicação das diversas saídas do projeto. Estas fases, apresentadas na seção de resultados caracterizam um tipo de ciclo de vida dos dados para a construção de ontologias biomédicas de maior qualidade, conforme discutido em cada análise.

Em função do volume e dos valores resultados oriundos da pesquisa, especialmente, por sua aceitação como uma ontologia de alto nível, integrante do projeto OBO Foundry, é possível concluir que a adoção das práticas de governança, incluindo as de versionamento e atualização das ontologias constituem uma iniciativa relevante para agregar maior qualidade a estes artefatos, tornando mais eficaz sua capacidade de prover interoperabilidade semântica entre sistemas computacionais.

Referências

ALMEIDA, M. B. (2013). Revisiting Ontologies: a necessary clarification. **Journal of the American Society of Information Science and Technology**. v. 64, n. 8. p. 1682–1693. DOI: 10.1002/asi.22861

ALMEIDA, M. B.; BAX, M. P. Uma visão geral sobre ontologias: pesquisa sobre definições, tipos, aplicações, métodos de avaliação e de construção. **Ci. Inf.**, Brasília, v. 32, n. 3, p. 7-20, set./dez. 2003.

ALMEIDA, M. B.; FARINELLI, F. (2017). Ontologies for Representation of Electronic Medical Records: the Obstetric and Neonatal Ontology. **Journal of the Association for Information Science and Technology**. doi:10.1002/asi.23900.

ALMEIDA, M. B.; SOUZA, R.; BARACHO, R. M. A. Looking for the identity of Information Science in the age of big data, computing clouds and social networks. In: **Proceedings of the 4th International Symposium of Information Science (ISI 2015)**, Croatia.

ANDRADE, A.Q.; ALMEIDA, M.B.; SCHULZ, S. Revisiting ontological foundations of the OpenEHR Entry Model. **Proceedings of the International Conference of Biomedical Ontologies (ICBO)**, 2012, Austria.

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.**, v.1, p. 34 – 43, 2001.

Best Practices | DataONE. Disponível em: <<https://www.dataone.org/best-practices>>. Acesso em: 26 set. 2018.

BORGMAN, C. L. **Big data, little data, no data: scholarship in the networked world.** The MIT Press: England, 2015.

BREUKERS, J. A. P. J.; HOEKSTRA, R. J. Epistemology and ontology in core ontologies: FOLaw and LRI-Core, two core ontologies for law. **CEUR Workshop Proceedings**, 2004.

BRINEY, K. **Data management for researchers: organize, maintain and share your data for research success.** Exeter, UK: Pelagic Publishing, 2015.

CANTELE, R. C.; ADAMATTI, D. F.; FERREIRA, M. A. G. V.; SICHMAN, J. S. Reengenharia e ontologias: análise e aplicação. **I Workshop de Rede Semântica (WWS2004)**, 1, 2004.

DIAS, G. A. **Guia de Dados de Pesquisa**, 2018. Acesso em: 22 ago. 2018

FARINELLI, F.; ALMEIDA, M. B.; ELKIN, P.; SMITH, B. OntONeo: The Obstetric and Neonatal Ontology. Proceedings of the **International Conference of Biomedical Ontologies**, Corvallis, U.S, 2016.

FARINELLI, F.; ALMEIDA, M.B.; ELKIN, P.; SMITH, B. (2016). Dealing with Social and Legal Entities in the Obstetric and Neonatal Domain. In Proceedings of the International Conference of Biomedical Ontologies, Corvallis, U.S.

FARINELLI, F.; ALMEIDA, M.B. EMYDGIO, J.L. Arquitetura de dados orientada por ontologias: interoperando sistemas de saúde. In: XVI Congresso Brasileiro de Informática em Saúde (CBIS 2018).

FARINELLI, F.; SILVA, S. de M.; ALMEIDA, M. B. Iniciativas governamentais para interoperabilidade semântica. **Revista Fonte**, ano 10, v. 13, p.66-74, dez. 2013.

FONSECA, F. The double role of ontologies in information science research. **Journal of the American Society for Information Science and Technology**, v. 58, p. 786-793, 2007.

GRIFFO, C.; ALMEIDA J. P. A.; GUIZZARDI, G. Towards a legal core ontology based on alexy's theory of fundamental rights. In: **15th International Conference on Artificial Intelligence and Law (ICAIL)**, 2015.

GUARINO, N. Formal ontology and information systems. In N. Guarino (Ed.). **Formal ontology in Information Systems** (1998, pp. 3–15). Amsterdam: IOS Press.

HAENDEL, M. A.; MCMURRY, J. A.; RELEVO, R.; MUNGALL, C. J.; ROBINSON, P. N.; CHUTE, C. G.. A Census of Disease Ontologies. **Annual Review of Biomedical Data Science**, v.1, p. 305-331, 2018.

HARRIS, M.; PARKINSON, H.. Standards and Ontologies for Functional Genomics:Towards Unified Ontologies for Biology and Biomedicine. **Comp. Funct. Genom.** 2003; 4:116–120.

IBRAHIM, E. A. A.; ATAELFADIEL, M. A. M. Ontology Life Cycle: A Survey on the Ontology and its Development Steps. *International Journal of Science and Research (IJSR)*, v. 6, n. 8, 2017, p. 1352-1357.

LENHARDT, W. C.; AHALT, S.; BLANTON, B.; CHRISTOPHERSON, L.; IDASZAK, R. Data Management Lifecycle and Software Lifecycle Management in the Context of Conducting Science. **Journal of Open Research Software**, v. 2, n. 1, p. e15, 9 jul. 2014.

MACHADO, I. M. R.; ALENCAR, R. O. de J.; CAMPOS, R. de O.; JUNIOR, C.I A. D.. An ontological gazetteer and its application for place name disambiguation in text. **Journal of the Brazilian Computer Society**, 17(4):267-279, 2011.

MARTÍNEZ-COSTA, C.; SCHULZ, S. Validating EHR clinical models using ontology patterns. **Journal of Biomedical Informatics**, 76 (2017) 124–137.

MEARIAN, L. The Data Life Cycle. **Computerworld**, Knowledge Center Storage. v. November 17, 2003, p. 39, 40, 2003.

NOY, N.; TUDORACHE, T.; NYULAS, C.; MUSEN, M. The ontology life cycle: Integrated tools for editing, publishing, peer review, and evolution of ontologies. **AMIA Annual Symposium Proceedings**, v. 2010, p. 552–556, 2010.

RECTOR, A. L. **Clinical terminology: why is it so hard?** *Methods Inf Med.* 1999 Dec;38(4-5):239-52.

RODRIGUES, C.; CHAVES, M. S.; SILVA, M. J.. Uma Representação Ontológica da Geografia Física de Portugal. **IX Encontro de Utilizadores de Informação Geográfica**, ESIG-Oeiras, Portugal, 2006.

SANTANA, R. C. G.. Ciclo de Vida dos Dados e o Papel da Ciência da Informação. **XIV Encontro Nacional de Pesquisa em Ciência da Informação (Enancib 2013)**, p. 21, [s.d.], 2013.

SMITH, B.; ASHBURNER M.; ROSSE, C.; BARD, J.; BUG, W.; CEUSTERS, W.; GOLDBERG, L. J.; EILBECK, K.; IRELAND, A.; MUNGALL, C. J.; The OBI Consortium; LEONTIS, N.; ROCCA-SERRA, P.; RUTTENBERG, A.; SANSONE, S.; SCHEUERMANN, R. H.; SHAH, N.; WHETZEL, P.; LEWIS, S.. The OBO Foundry: coordinated evolution of ontologies to support biomedical data integration. **Nat Biotechnol**. 2007 November; 25(11): 1251. doi:10.1038/nbt1346.

SOERGUEL, D.. **The rise of ontologies or the reinvention of classification**.1999. Disponível em: <http://www.dsoergel.com/cv/B70.pdf>. Acesso em: out. 2018.

STRASSER, C.; COOK, R.; MICHENER, W.; BUDDEN, A. Primer on Data Management: What you always wanted to know. **DataONE**, p. 11, [s.d.].

TORRES, S.; SIMÕES, M.G.; ALMEIDA, M.B. (2017). Princípios para modelagem de domínios em sistemas de organização do conhecimento In: **Anais do III Congresso ISKO Espanha-Portugal**.

VICKERY, B. C. Ontologies. **Journal of Information Science**, London, v. 23, n. 4, p. 227-286, 1997.

VUMC – Vanderbilt University Medical Center **Strategic Plan for VUMC Informatics**. 2005. Disponível em: <https://ncs.mc.vanderbilt.edu/Data/NonSecure/IC_Strategic_Plan_9-12-05.pdf>.

WAND, Y.; WEBER, R. An Ontological Model of an Information System. **IEEE Transactions on Software Engineering**, Vol. 16, No. 11, November 1990.

_____. An Ontological Analysis of the Relationship Construct in Conceptual Modeling. **ACM Transactions on Database Systems**, Vol. 24, No. 4, December 1999, Pages 494–528.

WAND, Y.; STOREY, V. C.; WEBER, R.. An Ontological Analysis of the Relationship Construct in Conceptual Modeling. **ACM Transactions on Database Systems**, Vol. 24, No. 4, December 1999.

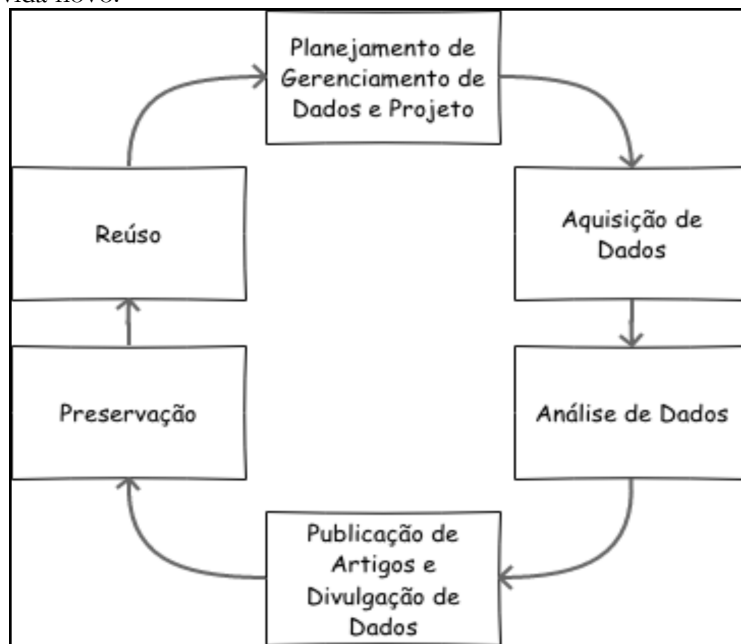
Apêndice A

Figura 1 – Ciclo de vida antigo.



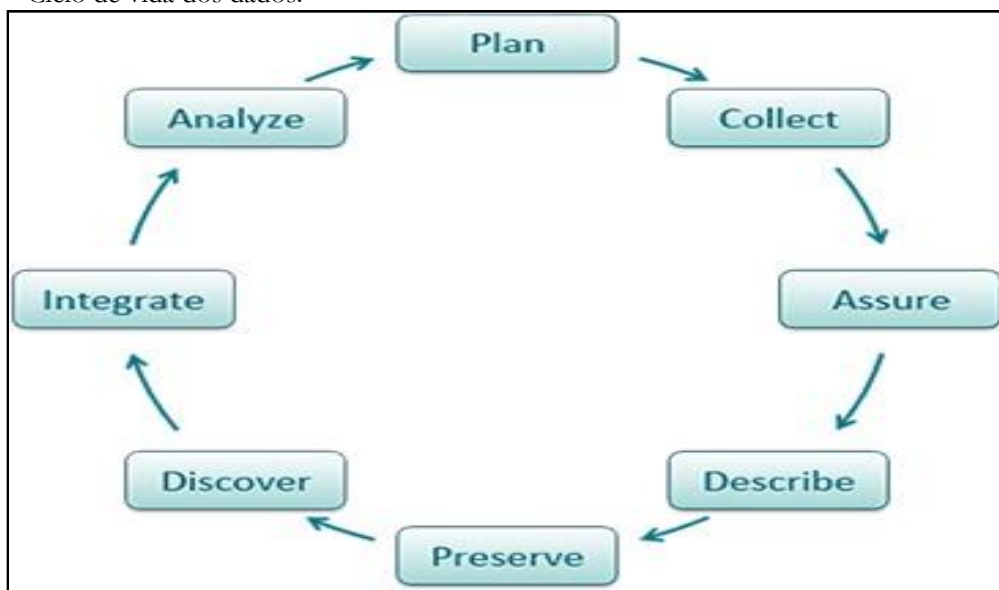
Fonte: Adaptado de Briney (2015).

Figura 2 – Ciclo de vida novo.



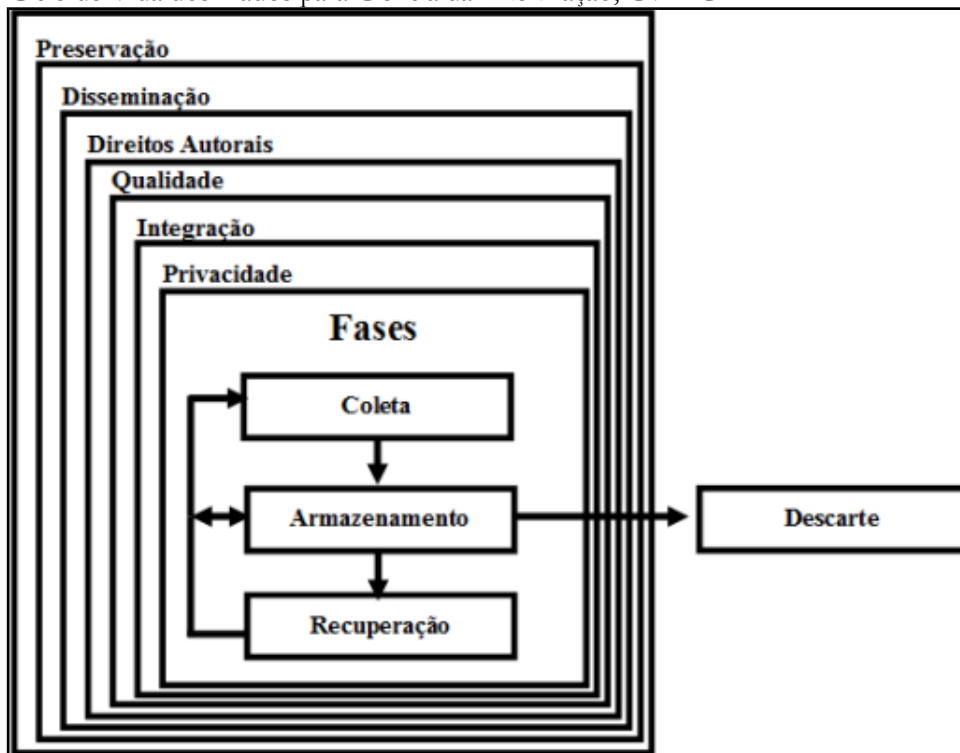
Fonte: Adaptado de Briney (2015).

Figura 3 – Ciclo de vida dos dados.



Fonte: DataONE (2018).

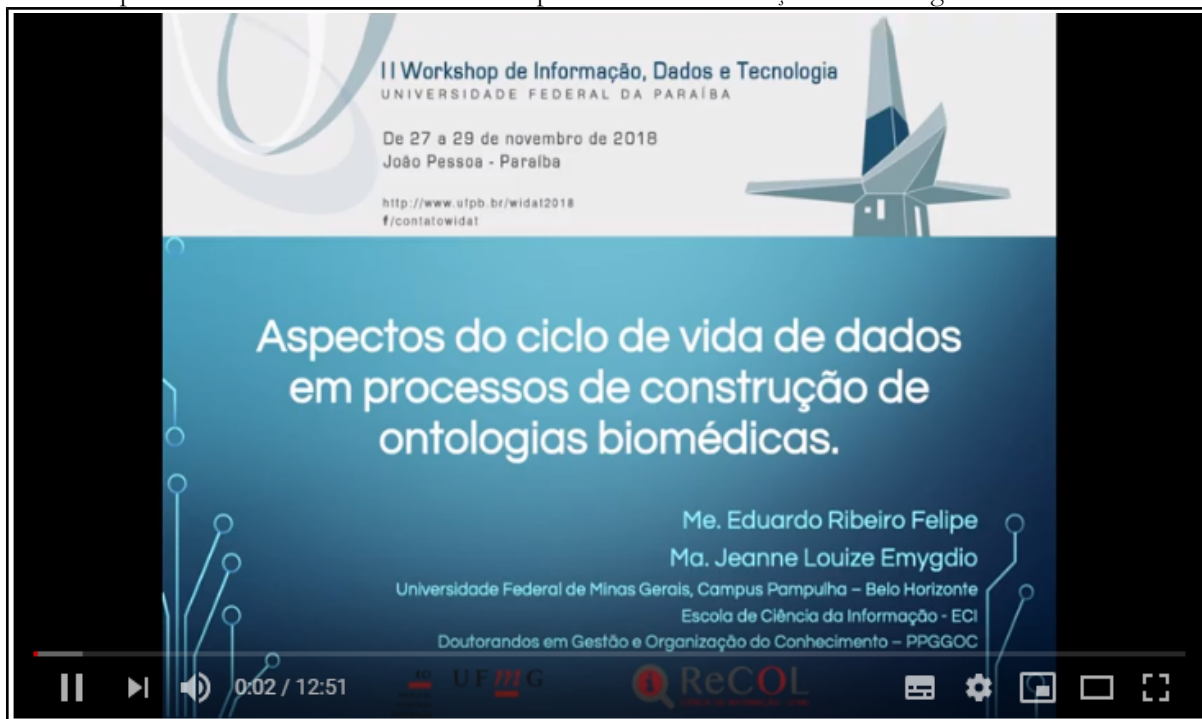
Figura 4 – Ciclo de Vida dos Dados para Ciência da Informação, CVD-CI.



Fonte: Santana (2013).

Vídeo da apresentação

Título: Aspectos do ciclo de vida de dados em processos de construção de ontologias biomédicas.



Disponível em: http://dadosabertos.info/enhanced_publications/idt/video.php?id=22

Transcrição da apresentação

Olá a todos meu nome é Eduardo sou aluno de doutorado na Universidade Federal de Minas Gerais sob a orientação do professor doutor Maurício Barcelo de Almeida eu e minha colega Jeane gostaríamos de apresentar o nosso artigo intitulado Aspectos do ciclo de vida de dados em processo de construção de ontologias biomédicas no segundo o Workshop de Informação, Dados e Tecnologia dos nossos colegas da Universidade Federal da Paraíba.

Estará apresentando a primeira parte com a introdução, referencial teórico, objetivos e logo em seguida minha colega dará sequência a essa apresentação. Nessa primeira etapa eu dividi em quatro grandes grupos, onde o artigo na introdução apresenta conceitos ontologias em suas origens na filosofia, a sua vocação na representação na realidade dos seres, os seus aspectos intimamente ligados às classificações e principalmente a sua capacidade de estabelecer relações entre os elementos.

Também o artigo retrata interdisciplinaridade como um ponto forte das ontologias e contribuições na área médica na questão da web semântica não podemos deixar de citar a capacidade da ontologia de retratar a realidade em um ambiente virtual na questão dos dados, também colocamos a origem das primeiras representações de conceitos de dados na teologia, o dado enquanto reflexo do contexto, o dado dependente da interpretação do observador e também a sua questão como insumo dos sistemas de informação. Insumo é uma unidade básica dos instrumentos informacionais utilizamos a definição da OECD com uma definição norteadora dos conceitos estabelecidos no artigo e com relação ao ciclo de vida citamos nesses slides aqui quatro boas referências que nos ajudaram na construção do trabalho falando mais especificamente sobre ciclo de vida a identificação inicial foi que ciclo de vida muitas vezes é citado em desenvolvimento de software em construção de ontologias mas foi um pouco mais difícil encontrar essa definição relacionada aos dados é importante ressaltar que encontramos definições de ciclo de vida em dados, porém é direcionado unicamente à persistência dos dados em um período onde a preocupação era o armazenamento e recuperação das informações visto que muitas informações foram perdidas quando os suportes informacionais tecnológicos, as fitas magnéticas nos disquetes eram substituídos com uma velocidade muito rápida impossibilitavam muitas vezes a recuperação dos dados, a autora Briney ela apresenta dois ciclos de vida interessantes, um ciclo de vida que ela denomina de antigo com quatro processos básicos como é visto no slide e um ciclo de vida que ela denomina novo onde ela acrescenta os processos de preservação e reúso.

Também é importante notar nos dois outros ciclos de vidas importantes no artigo o da DataONE e de Santana especificamente direcionam ciclo de vida para a ciência da informação e nos remete muitas vezes e nos lembra muitas vezes no modelo interativo de desenvolvimento de software, pois bem em relação aos objetivos nós procuramos combinar os interesses de nossas teses do grupo de pesquisa Recol e da disciplina do professor Guilherme Dias, encontramos no trabalho da professora Fernanda Farinelli a ontologia OntoNEO e procuramos então compreender o ciclo de vida dos dados biomédicos em processo de construção ontológica.

Pelo tempo já procurando concluir a primeira parte eu quero destacar então que o ciclo de vida de dados ele traz diversos benefícios, ele pode tratar facilidade na recuperação dos dados da

informação, ele pode permitir o uso em análise dos dados na pesquisa pelo próprio pesquisador como colaboradores ele é o ciclo de vida de dados, ele pode melhorar a tomada de decisão e principalmente fornecer dados de qualidade para pesquisas translacionais, onde um instrumento é utilizado como origem como fonte para outros instrumentos, pode permitir que cientistas externos ao projeto encontrem, entendam e utilizem esses dados no futuro e dentro desse cenário acrescentando a documentação por metadados é possível então estabelecer situações e créditos futuros de uma maneira mais ampliada além de permitir a rastreabilidade autoral.

Eu passo então para a segunda parte da apresentação do artigo agradeço a atenção de todos muito obrigado.

Saudações a todos meu nome Jeanne eu sou aluna de doutorado do programa de pós-graduação em Gestão e Organização do Conhecimento da Universidade Federal de Minas Gerais e orientanda do Professor Doutor Luis Barcelos já aumenta nessa segunda parte vou apresentar a vocês os procedimentos metodológicos os resultados e as considerações finais do nosso estudo.

O estudo se caracterizou por uma natureza qualitativa caráter descritivo tendo sido realizado em três fases: a primeira delas compreendeu uma análise dos procedimentos de construção do OntoNEO em comparação com os processos de governança de dados previamente estudados representados pelo colega Eduardo utilizei aqui os termos processo de governança para compreender as fases do ciclo de vida de dados e processos dos planos de gestão de dados com vistas ao simplificar a minha apresentação.

A fase 2 nós organizamos os procedimentos de construção da OntoNEO em processos de governança de dados selecionados de Briney (2005), DataONE (2018) e Dias (2018). Esse último trabalho se baseia em padrões da Universidade Virginia para governança de dados.

Na fase 3 fizemos uma análise dos aspectos de qualidade e agregados ao OntoNEO em função das práticas de governança que foram adotadas em sua construção, encontramos dois resultados principais ao término do nosso estudo. O primeiro deles repousa sobre uma percepção clara dos aspectos de qualidade que a governança de dados pode proporcionar aos processos de construção de ontologias biomédicas e o segundo reforço de identificação de cinco macros processos relacionados à governança de dados e que podem ser aplicáveis a construção de ontologias biomédicas.

Então detalhando um pouquinho desses dois resultados nós temos entre os aspectos de qualidade identificados o primeiro deles repousando sobre uma premissa de que a ontologia na ser construída ela seria integrante de uma iniciativa denominada OBO Foundry é uma iniciativa internacional de construção de ontologias de alto valor de qualidade para integração de dados biomédicos.

Então pra que uma ontologia se agregue a esta iniciativa ela precisa também ser construída com altos padrões de qualidade seguindo as premissas definidas por esse projeto.

Então para o alcance desse nível de qualidade nós percebemos que a professora Fernanda ela adotou padrões tecnológicos internacionais especificidades técnicas e administrativas do contexto real médico que eles vão depressivo reuso de dados de outras ontologias já integrantes a iniciativa OBO Foundry. Portanto com qualidade já previamente certificada pela iniciativa ou redução de modelos de ciclo de vida para a construção de sistemas de informação da área de engenharia de

software da ciência da computação para a condução das práticas de engenharia de ontologias no campo da ciência da informação o que também contribuiu para o aumento da qualidade dos artefatos final gerando.

Foi realizada uma prova de conceito que garantiu a eficiência do artefato informacional gerado no atendimento aos propósitos que motivaram sua construção. Esse modelo portanto ele é um modelo de representação do conhecimento de alta qualidade para a representação de dados biomédicos e foi disponibilizado em formato aberto em caráter permanente tendo um site próprio e ainda sendo possível o seu acesso no repositório de acesso público bio portal que também integra a iniciativa OBO Foundry paralelamente percebemos uma ampla divulgação e compartilhamento do conhecimento científico produzido pela pesquisadora.

A pesquisa também recebeu financiamentos de agências de fomento nacionais e internacionais porque ela foi realizada com uma pesquisa do tipo sanduíche, ela se realizou em parte do departamento de filosofia informática biomédica da Universidade de Buffalo e em parte aqui na escola sem essa informação da Universidade Federal de Minas Gerais.

Na sequência nós passamos ao detalhamento dos cinco macros processos que foram percebidos e que compreendem a gestão de informações sobre dados e formatos a gestão de conteúdo dos metadados e formatos a gestão de políticas para acesso, compartilhamento erros de dados a gestão de armazenamento de longo prazo e gestão de questões orçamentárias, então esse 5 macros processos eles nos permitiram é uma visão de aderência ao ciclo de vida estudado em cada um daqueles autores que já foram citados nos dando uma percepção de que ainda não tenha sido esse o propósito inicial da pesquisa realizada pela nossa colega o cuidado a meculosidade dela durante todas as fases de construção da ontologia ela permitiu uma aderência clara a ciclos de vida relevantes para a governança de dados e uma visualização de macroprocesso candidatos a estruturação de planos de gestão de dados específicos para esse tipo de pesquisa sendo assim nas nossas considerações finais temos retornar aos objetivos do nosso estudo afirmarmos que em função do volume e do valor dos resultados oriundos da pesquisa realizada pela professora doutora Fernanda Farinelli e que foi analisado em nosso estudo concluímos que a adoção de práticas de governança de dados em processos de construção de ontologias elas agregam maior qualidade aos artefatos informacionais gerados possibilita a formalização de melhores práticas para os processos de engenharia de ontologias e podem ainda fomentar pesquisas para formalização de ciclo de vida de dados ou planos de gestão de dados aplicáveis aos processos de engenharia de ontologias promovendo também avanços nas pesquisas voltadas à governança de dados.

Deixamos aqui os nossos contatos pra elucidação de possíveis dúvidas e agradecemos a atenção de todos e pela grata oportunidade participarmos do 2º WIDAT um forte abraço.

Slides da apresentação

Título: Aspectos do ciclo de vida de dados em processos de construção de ontologias biomédicas.



II Workshop de Informação, Dados e Tecnologia
UNIVERSIDADE FEDERAL DA PARAÍBA

De 27 a 29 de novembro de 2018
João Pessoa - Paraíba

<http://www.ufpb.br/widat2018>
f/contatowidat

**Aspectos do ciclo de vida de dados
em processos de construção de
ontologias biomédicas.**

Ma. Jeanne Louize Emygdio
Me. Eduardo Ribeiro Felipe

Universidade Federal de Minas Gerais, Campus Pampulha – Belo Horizonte
Escola de Ciência da Informação - ECI
Doutorandos em Gestão e Organização do Conhecimento – PPGOC

ECI ESCOLA DE CIÊNCIA DA INFORMAÇÃO - UFMG
UFMG
ReCOL CENTRO DE INFORMAÇÃO - UFMG

Disponível em: http://dadosabertos.info/enhanced_publications/idt/presentation.php?id=22