



**UNIVERSIDADE ESTADUAL PAULISTA “JÚLIO DE MESQUITA FILHO”
FACULDADE DE FILOSOFIA E CIÊNCIAS
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA INFORMAÇÃO**

PALOMA MARÍN ARRAIZA

 <https://orcid.org/0000-0001-7460-7794>

TESE DE DOUTORADO

**MULTIMODALIDADE NA PUBLICAÇÃO CIENTÍFICA AMPLIADA:
CONSIDERAÇÕES SEMIÓTICAS E MODELO DE REPRESENTAÇÃO**

**MARÍLIA
2019**



Esta obra está licenciada com uma Licença [Creative Commons
Atribuição 4.0 Internacional](https://creativecommons.org/licenses/by/4.0/).

PALOMA MARÍN ARRAIZA

**MULTIMODALIDADE NA PUBLICAÇÃO CIENTÍFICA AMPLIADA:
CONSIDERAÇÕES SEMIÓTICAS E MODELO DE REPRESENTAÇÃO**

Tese de doutorado apresentada ao Programa de Pós-graduação em Ciência da Informação (PPGCI) da Universidade Estadual Paulista “Júlio de Mesquita Filho”, como requisito para a obtenção do título de doutor em Ciência da Informação.

Área de concentração: Informação, tecnologia e conhecimento.

Linha de pesquisa: Informação e tecnologia

Orientadora: Profa. Dra. Silvana Aparecida Borsetti Gregorio Vidotti

Financiamento: CNPq (Período abril 2018 – setembro 2018).

**MARÍLIA
2019**

M337m

Marín-Arraiza, Paloma

Multimodalidade da publicação científica ampliada:
considerações semióticas e modelo de representação / Paloma
Marín-Arraiza. -- Marília, 2019
189 f.

Tese (doutorado) - Universidade Estadual Paulista (Unesp),
Faculdade de Filosofia e Ciências, Marília

Orientadora: Silvana Aparecida Borsetti Gregorio Vidotti

1. Publicações eletrônicas acadêmicas. 2. Semiótica. 3.
Modelo entidade-relacionamento. 4. Tecnologia da informação.
I. Título.



UNIVERSIDADE ESTADUAL PAULISTA

Câmpus de Marília

CERTIFICADO DE APROVAÇÃO

TÍTULO DA TESE: MULTIMODALIDADE NA PUBLICAÇÃO CIENTÍFICA AMPLIADA: CONSIDERAÇÕES SEMIÓTICAS E MODELO DE REPRESENTAÇÃO

AUTORA: PALOMA MARIN ARRAIZA

ORIENTADORA: SILVANA APARECIDA BORSETTI GREGORIO VIDOTTI

Aprovada como parte das exigências para obtenção do Título de Doutora em CIÊNCIA DA INFORMAÇÃO, área: Informação, Tecnologia e Conhecimento pela Comissão Examinadora:

Profa. Dra. SILVANA APARECIDA BORSETTI GREGORIO VIDOTTI

Departamento de Ciência da Informação e Programa de Pós-Graduação em Ciência da Informação / Unesp, Faculdade de Filosofia e Ciências, Marília

Profa. Dra. SILVANA DRUMOND MONTEIRO

Departamento de Ciência da Informação e Programa de Pós-Graduação em Ciência da Informação / Universidade Estadual de Londrina

Prof. Dr. LUIS FERNANDO SAYÃO

Centro de Informações Nucleares e Pós-Graduação em Ciência da Informação / Comissão Nacional de Energia Nuclear

Prof. Dr. HENRY PONCIO CRUZ DE OLIVEIRA

Departamento de Ciência da Informação / Universidade Federal da Paraíba

Prof. Dr. FERNANDO DE ASSIS RODRIGUES

Instituto de Ciências Sociais Aplicadas / Universidade Federal do Pará

Marília, 20 de dezembro de 2019


Prof. Dra. Marta Ligia Ponzim Valentim
Coordenadora do PPG em
Ciência da Informação

Faculdade de Filosofia e Ciências - Câmpus de Marília -
Av. Hygino Muzzi Filho, 737, 17525-900
<http://www.marilia.unesp.br/poso/CHPJ> 48.031.916/0008-09

“

A PhD is so much more than a degree. It can break you down into your most vulnerable form, but has the potential to build you back together to become a resilient, determined, humble, and knowledgeable researcher. This process takes time and patience. Please don't give up on yourself. It's not just about getting the degree. It's about becoming who you are meant to be.

– PhD SOS

Fonte: PhD_SOS <https://www.instagram.com/p/BuHSkJFBEe5/>

Tradução

O doutorado é muito mais que um diploma. Ele pode fazer com que você se sinta mais vulnerável, mas tem o potencial de fortalecê-lo para se tornar um(a) pesquisador(a) resiliente, determinado(a), humilde e conhecedor(a). Esse processo leva tempo e paciência. Por favor, não desista de si mesmo(a). Não se trata apenas de obter o diploma. É sobre se tornar quem você está destinado(a) a ser.

AGRADECIMENTOS

“Gratitude is the sign of noble souls”

Muitas são as pessoas que me auxiliaram no desenvolvimento desta tese de uma forma ou outra, direta ou indiretamente. A todas elas gostaria de transmitir aqui minha mais sincera gratidão.

Especialmente:

Aos meus pais, Carmen e Apolinar que, mesmo discordando de muitas das minhas decisões, sempre as aceitaram e me apoiaram para continuar.

Às minhas irmãs, Patricia e Leyre, que mesmo estando cada uma em uma ponta do planeta, literalmente, sempre estiveram presentes no meu dia a dia.

Às minhas colegas de estudos, amigas e confidentes Larissa de Mello Lima e Mirelys Puerta Díaz, sem vocês esta etapa não teria sido possível. Também ao Ilídio Lobato Manique, com quem é possível conversar sobre Semiótica, métricas alternativas, história do mundo... Muito obrigada aos três, tenho aprendido muito com vocês.

Aos colegas do grupo de pesquisa “Novas Tecnologias em Informação”, sobretudo à Larissa Pavarini da Luz, pela sua amabilidade, boa vontade e disponibilidade sempre para ajudar. Aos demais colegas do PPGCI com que pude compartilhar aulas e debates, em especial à Selma Leticia Capinzaiki Ottonicar pelas parcerias e convergências entre linhas.

Ao meu companheiro, Bruno Ferreira Batista, quem bem poderia defender esta tese também de tantas vezes que me ouviu falar sobre ela, pelo apoio e por ter aguentado todos os desabafos.

À minha orientadora, a Profa. Dra. Silvana Aparecida Borsetti Gregorio Vidotti, pelas orientações e o acompanhamento ao longo deste trabalho.

À Profa. Dr. Maria José Jorente por todas as conversas sobre pesquisa e política e por me abrir as portas da casa dela.

Ao Prof. Dr. Carlos Cândido de Almeida pelos ensinamentos ao longo da sua disciplina e pelas valiosas contribuições como membro da banca do meu exame de qualificação. Quem teria me dito em 2017 que a Semiótica tornar-se-ia tão importante na minha tese!

Ao Prof. Dr. Fernando de Assis Rodrigues por me apresentar o modelo entidade-relacionamento, outro pilar desta tese, e pela frutífera troca de conhecimento na sala de aula.

Ao Prof. Dr. Luiz Fernando Sayão pela disponibilidade em participar do meu exame de qualificação e da minha defesa, pelas discussões construtivas e pelas contribuições e apontamentos sobre a Ciência Aberta e a gestão de dados de pesquisa.

À Profa. Dr. Silvana Drumond Monteiro e ao Prof. Dr. Henry Poncio de Oliveira, pela sua disponibilidade e por ter aceito o convite para fazer parte da minha banca de defesa.

Por último, mas não menos importante aos integrantes do Centro de Gestão de Dados de Pesquisa da Universidade Técnica de Viena pela maravilhosa colaboração e trabalho diário e estar fazendo minha vida em Viena muito mais agradável.

Muito obrigada, muchas gracias e vielen Dank!

MARÍN-ARRAIZA, PALOMA. **Multimodalidade da publicação científica ampliada: considerações semióticas e modelo de representação**. 2019. 189 f. Tese de doutorado (Doutorado em Ciência da Informação) – Faculdade de Filosofia e Ciências, Universidade Estadual Paulista (UNESP), Marília.

RESUMO

A publicação de resultados de pesquisa com base no documento textual vem sendo considerada desde o século XVII como a principal via de comunicação científica. A chegada na Ciência do paradigma baseado nos dados e do movimento da Ciência Aberta inicia a demanda de publicação de outros ativos/assets de pesquisa de diferente natureza. Contrastamos a necessidade de novas vias e abordagens à publicação de forma que os ativos/assets possuam uma adequada descrição para garantir a encontrabilidade, o acesso, o reuso e a reprodutibilidade. Uma das vias para essa publicação é a publicação científica ampliada. Com base nas ferramentas fornecidas pelo método quadripolar, esta pesquisa analisa as transformações epistemológicas, teóricas e técnicas do processo de comunicação científica com o objetivo de formalizar um modelo para as publicações ampliadas, que considere os aspectos semióticos derivados da hibridação de linguagens neste tipo de publicação e dos relacionamentos entre as diferentes entidades que compõem a publicação. A análise parte da conceitualização da publicação ampliada como um espaço semiótico de topografia tanto homogênea quanto heterogênea, devido a sua composição por objetos de diferente natureza. Realiza-se um mapeamento de diferentes ambientes de publicação para extrair as entidades que podem compor uma publicação ampliada. Essas entidades, junto com a dimensão executável que representa e estabelece o processo de criação e lógica entre as entidades, conformam a base da modelagem conceitual, produto ou objeto morfológico desta tese. Para o desenvolvimento da modelagem conceitual utiliza-se a análise semântica que permite a descrição de cada uma das entidades, bem como as dependências ontológicas entre elas. Concluimos apresentando teoricamente tanto uma definição Semiótica para as publicações ampliadas quanto uma proposta de modelo conceitual para esse tipo de publicações. Esse modelo oferece um mecanismo de publicação de ativos/assets de pesquisa que garanta a integridade de cada um deles e sua contextualização no ciclo de publicação científica.

PALAVRAS-CHAVE: publicação científica ampliada, linguagem híbrida, modelo de publicação, modelagem entidade-relacionamento, ciência aberta.

MARÍN-ARRAIZA, PALOMA. **Multimodality in enhanced scientific publications: semiotic considerations and representation model**. 2019. 189 p. Doctoral thesis (Doctorate in Information Science) – Faculdade de Filosofia e Ciências, Universidade Estadual Paulista (UNESP), Marília.

ABSTRACT

The publication of research results based on the textual document has been considered since the 17th century as the primary means of scientific communication. The arrival in Science of the paradigm based on data and the Open Science movement initiates the demand for publication of other research assets/assets of different nature. We contrast the need for new ways and approaches to the publication of assets to have an adequate description to ensure the findability, access, reuse and reproducibility. One of the avenues for this publication is the expanded scientific publication. Based on the tools provided by the quadripolar method, this research analyses the epistemological, theoretical and technical transformations of the scientific communication process. The main aim is to formalize a model for enhanced publications, which considers the semiotic aspects derived from the hybridization of languages in this type of publication and the relationships between the different entities that comprise the publication. The analysis starts from the conceptualization of enhanced publications as semiotic spaces of homogeneous and heterogeneous topography, due to their composition by objects of different nature. A mapping of different publication environments is performed to extract the entities that can compose an enhanced publication. These entities, together with the executable dimension that represents and establishes the creation process and logic between the entities, form the basis of the conceptual modelling, product or morphological object of this thesis. For the conceptual modelling development, we used the semantic analysis, allowing the description of each entity, as well as the ontological dependencies between them. We conclude by theoretically presenting both a semiotic definition for enhanced publications and a proposal for a conceptual model for this type of publications. This model offers a publication mechanism of research assets that ensures the integrity of each asset and their contextualization in the scientific publication cycle.

KEYWORDS: enhanced scientific publication, hybrid language, publication model, entity-relation model, open science.

MARÍN-ARRAIZA, PALOMA. **Multimodalidad en la publicación científica ampliada: consideraciones semióticas y modelo de representación**. 2019. 189 p. Tesis de doctorado (Doctorado en Ciencia de la Información) – Faculdade de Filosofia e Ciências, Universidade Estadual Paulista (UNESP), Marília.

RESUMEN

La publicación de los resultados de la investigación a partir del documento textual ha sido considerada desde el siglo XVII como el principal medio de comunicación científica. La llegada a la Ciencia del paradigma basado en datos y el movimiento Open Science inicia la demanda de publicación de otros activos/assets de investigación de diferente naturaleza. Contrastamos la necesidad de nuevas formas y enfoques en la publicación de activos/assets para tener una descripción adecuada que asegure la encontrabilidad, el acceso, la reutilización y la reproducibilidad. Una de las vías para esta publicación es la publicación científica ampliada. Basada en las herramientas proporcionadas por el método cuadripolar, esta investigación analiza las transformaciones epistemológicas, teóricas y técnicas del proceso de comunicación científica. El objetivo principal es formalizar un modelo de publicaciones ampliadas, que tenga en cuenta los aspectos semióticos derivados de la hibridación de las lenguas en ese tipo de publicaciones y las relaciones entre las diferentes entidades que componen la publicación. El análisis parte de la conceptualización de las publicaciones ampliadas como espacios semióticos de topografía homogénea y heterogénea, debido a su composición por objetos de distinta naturaleza. Se realiza un mapeo de los diferentes entornos de publicación para extraer las entidades que pueden componer una publicación ampliada. Estas entidades, junto con la dimensión ejecutable que representa y establece el proceso de creación y la lógica entre las entidades, constituyen la base del modelado conceptual, producto u objeto morfológico de esta tesis. Para el desarrollo del modelado conceptual, se utilizó el análisis semántico, permitiendo la descripción de cada entidad, así como las dependencias ontológicas entre ellas. Concluimos presentando teóricamente tanto una definición Semiótica de las publicaciones ampliadas como una propuesta de modelo conceptual para ese tipo de publicaciones. Ese modelo ofrece un mecanismo de publicación de los activos de investigación que garantiza la integridad de cada activo y su contextualización en el ciclo de publicación científica.

PALABRAS CLAVE: publicación científica ampliada, lenguaje híbrida, modelo de publicación, modelado entidad-relación, ciencia abierta

DECLARAÇÃO DE AUTORIA

Esta tese está composta pelo meu trabalho original e não contém material escrito por outra pessoa, a exceção das citações diretas e indiretas referenciadas no texto. Além disso, esclareço a seguir a contribuição de outros nos trabalhos de autoria conjunta inclusos parcialmente nesta tese.

Declaro a contribuição de outros na minha tese como um todo, incluindo desenho da pesquisa, procedimentos técnicos e metodológicos significativos, estruturação e qualquer outro trabalho de pesquisa original utilizado ou relatado nesta tese. As figuras 1 e 9 utilizadas nesta tese são da minha elaboração e autoria conceitual, os créditos gráficos são do arquiteto e designer Bruno Ferreira Batista. O conteúdo da minha tese é o resultado do trabalho realizado desde o início da minha candidatura com o processo seletivo em agosto de 2016, incluindo as disciplinas do Programa de Pós-Graduação em Ciência da Informação da UNESP, até a finalização da redação da tese. O conteúdo não inclui uma parte substancial que tenha sido apresentada para qualificar para a atribuição de qualquer outro grau ou diploma em qualquer universidade ou outra instituição terciária. Eu declaro quais partes da minha tese, se houver, foram apresentadas para se qualificar para outro fim.

Declaro que os direitos autorais de todo o material contido na minha tese residem com o(s) detentor(es) de direitos autorais desse material.

PUBLICAÇÕES DURANTE O PROCESSO DE DOUTORADO

Declaro que todas as publicações realizadas durante o doutorado estão disponíveis em acesso aberto. Aquelas publicações realizadas em periódicos contêm também um identificador persistente. Não todas as publicações listadas a seguir tem relação direta com o conteúdo da tese e sim com pesquisas desenvolvidas anteriormente ou de forma paralela aos estudos doutorais.

Artigos em periódicos:

- MARÍN-ARRAIZA, P.; VIDOTTI, S. Implementação de serviços institucionais de administração de dados. **Liinc em Revista**, v.15, n.2, p. 259-274, 2019. DOI: <http://doi.org/10.18617/liinc.v15i2.4819>
- MARÍN-ARRAIZA, P.; PUERTA-DÍAZ, M.; VIDOTTI, S. Gestión de datos de investigación y bibliotecas: preservando los nuevos bienes científicos. **Hipertext.net**, n.19, p. 13-31, 2019. DOI: <http://doi.org/10.31009/hipertext.net.2019.i19.02>
- MARÍN-ARRAIZA, P. ORCID in the Open Science Scenario: Opportunities for academic libraries. **Communications of the Association of Austrian Librarians**, v.72, n.2, 2019. DOI: <https://doi.org/10.31263/voebm.v72i2.2811>
- MARÍN-ARRAIZA, P.; GONÇALEZ, P.R.V.A.; VIDOTTI, S. Recomendações para a integração de publicações ampliadas em repositórios digitais confiáveis. **Encontros Bibli**, v.24, n.55, p.1-23, 2019. DOI: <https://doi.org/10.5007/1518-2924.2019.e58556>
- MELLO-LIMA, L.; MARÍN-ARRAIZA, P. A Análise de Discurso de Matriz Francesa enquanto polo técnico na pesquisa em Ciência da Informação. **Investigación Bibliotecológica**, v.33, n.79, p. 67-81, 2019. DOI: <http://dx.doi.org/10.22201/iibi.24488321xe.2019.79.57991>
- MARÍN-ARRAIZA, P.; VIDOTTI, S. El vídeo como ampliación de publicaciones científicas: Descripción y modelos de datos. **Informação e Tecnologia**, v.5, n.1, p. 52-64, 2018. DOI: <https://doi.org/10.22478/ufpb.2358-3908.2018v5n1.38381>
- BERRÍO-ZAPATA, C.; MARÍN-ARRAIZA, P.; SILVA, E. F.; SOARES, E. C. Desafíos de la Inclusión Digital: antecedentes, problemáticas y medición de la Brecha Digital de Género. **Psicología, Conocimiento y Sociedad**, v.7, p.162-198, 2017. DOI: <http://dx.doi.org/10.26864/pcs.v7.n2.8>

Capítulos de livros:

- MARÍN-ARRAIZA, PALOMA; MOLNÁR, ATTILA DÁVID. **Autocreación de video abstracts como parte de la investigación multimodal**. In: Esteban Romero Frías; Lidia Bocanegra Barbecho. (Org.). Ciencias

sociales y Humanidades Digitales aplicadas. Casos de estudio y perspectivas críticas. Nova lorque: Downhill Publishing, 2018, p. 66-84. DOI: <https://doi.org/10.5281/zenodo.1469337>

Trabalhos em anais de conferências:

- PUERTA-DÍAZ, M.; MARÍN-ARRAIZA, P.; VIDOTTI, S. Bibliotecas y datos: preservando los bienes científicos del futuro. In: IX Encuentro Ibérico EDICIC, 2019. Barcelona (Espanha). <http://doi.org/10.5281/zenodo.2820732>
- MARÍN-ARRAIZA, P.; LIBERATORE, G.; VIDOTTI, S. Publicação científica ampliada: desafios desde as Ciências Sociais e Humanidades. In: XIX Encontro Nacional de Pesquisa em Ciência da Informação (XIX ENANCIB), 2018. Londrina (PR, Brasil). *Anais... XIX ENANCIB*. <http://enancib.marilia.unesp.br/index.php/XIXENANCIB/xixenancib/paper/view/1260>
- PLANK, M.; MARÍN-ARRAIZA, P.; MOLNÁR, A.D. Video abstracts for scientific education. In: European Distance and E-Leaning Network 2018 Annual Conference: Exploring the Micro, Meso and Macro, 2018. Genova (Italia). *Anais... EDEN 2018*. p. 123-130. http://www.eden-online.org/wp-content/uploads/2018/06/Annual_2018_Genova_Proceedings.pdf
- MARÍN-ARRAIZA, P. Multimodalidade na pesquisa científica: Desafios para a curadoria digital. In: II Encontro de Pesquisa em Curadoria Digital, 2018. Marília (SP, Brasil). Disponível em: <http://doi.org/10.5281/zenodo.3369743>
- PLANK, M.; MOLNÁR, A.D.; MARÍN-ARRAIZA, P. Extending Media Literacy Education: The Popular Science Video Workshop. In: IFLA World Library and Information Congress (IFLA WLIC 2017), 2017. Wroclaw (Polônia). *Anais... IFLA WLIC 2017*. <http://library.ifla.org/1776/1/242-plank-en.pdf>
- MARÍN-ARRAIZA, P.; BOLAÑOS-CARMONA, J.M., VIDOTTI, S. As formas da informação: Um olhar aos conceitos de informação e fluxo de informação. In: XVIII Encontro Nacional de Pesquisa em Ciência da Informação (XVIII ENANCIB), 2017. Marília (SP, Brasil). *Anais... XVIII ENANCIB*. <http://enancib.marilia.unesp.br/index.php/xviiienancib/ENANCIB/paper/viewFile/167/1076>

PUBLICAÇÕES PARCIALMENTE INCLUÍDAS NESTA TESE

- MELLO-LIMA, L.; MARÍN-ARRAIZA, P. A Análise de Discurso de Matriz Francesa enquanto polo técnico na pesquisa em Ciência da Informação. **Investigación Bibliotecológica**, v.33, n.79, p. 67-81, 2019. DOI: <http://dx.doi.org/10.22201/iibi.24488321xe.2019.79.57991>

Parte do conteúdo desta publicação, com as adaptações correspondentes está incluído no Capítulo 2 “*Aspectos metodológicos*” na parte de apresentação do Método Quadripolar e na parte de desenho do polo técnico.

Autoras	Declaração de contribuição
Larissa Mello-Lima	Desenho do texto e da ideia (60 %) Escrita do texto (50 %)
Paloma Marín-Arraiza	Desenho do texto e da ideia (40 %) Escrita do texto (50 %) Autora de correspondência (*)

(*) A escolha da autora de correspondência foi feita devido a língua da revista ser o espanhol.

- MARÍN-ARRAIZA, P.; GONÇALEZ, P.R.V.A.; VIDOTTI, S. Recomendações para a integração de publicações ampliadas em repositórios digitais confiáveis. **Encontros Bibli**, v.24, n.55, p.1-23, 2019. DOI: <https://doi.org/10.5007/1518-2924.2019.e58556>

Parte do conteúdo desta publicação, com as adaptações correspondentes está incluído no Capítulo 5 “*A publicação científica ampliada: abordagens e estrutura*” na subseção 5.1.

Autoras	Declaração de contribuição
Paloma Marín-Arraiza	Desenho do texto e da ideia (90 %) Escrita do texto (85 %) Autora de correspondência
Paula Regina Ventura Amorim Gonzalez	Desenho do texto e da ideia (10 %) Escrita do texto (10 %)
Silvana Aparecida Borsetti Gregorio Vidotti	Escrita do texto (5 %)

As porcentagens foram calculadas aproximadamente segundo o trabalho desenvolvido e tempo investido por cada uma das autoras. Não foi seguida nenhuma regra bibliométrica específica.

LISTA DE FIGURAS

Figura 1 Aplicação do método quadripolar nesta pesquisa	33
Figura 2 Aplicação dos princípios da <i>openness</i> nos diferentes estágios do processo de pesquisa	46
Figura 3 Publicação de dados? Os principais aspectos legais	49
Figura 4 Taxonomia da ciência aberta	50
Figura 5 101 Inovações em comunicação científica	61
Figura 6 Esboço das publicações científicas ampliadas e as possíveis conexões entre elas	84
Figura 7 Exemplo publicação no limite baixo com núcleo textual	89
Figura 8 Exemplo de publicação no limite baixo com núcleo audiovisual	92
Figura 9 Desenvolvimento temporal da publicação científica ampliada ..	104
Figura 10 Fluxos de comunicação entre autores, avaliadores e leitores na máquina virtual <i>Paper Mâché</i>	108
Figura 11 Modelo do FAIR Data Object	117
Figura 12 Compartilhamento livre, descoberta e reutilização dos ativos/assets de pesquisa identificados por PIDs	120
Figura 13 Filtro “PID systems” no registro re3data.org	122
Figura 14 Diagrama A0: pesquise, comunique e implemente os resultados	131
Figura 15 Diagrama A231: facilite a recuperação dos resultados	132
Figura 16 Dependência entre as entidades “Autor”, “Texto” e “Texto – resumo”	140
Figura 17 Modelo ER para publicações ampliadas	164

LISTA DE VÍDEOS

Vídeo 1 Exemplo de publicação no limite baixo com matriz verbo-visual central	91
Vídeo 2 Exemplo de publicação no meio do espectro com diversos ativos/assets adicionais	93
Vídeo 3 Exemplo de publicação no meio do espectro com base do conhecimento integrada	94
Vídeo 4 Exemplo de publicação no meio do espectro com integrações audiovisuais	95
Vídeo 5 Exemplo de projeto de publicação no limite superior	96

LISTA DE QUADROS

Quadro 1 As transformações técnicas para a integração da publicação ampliada	38
Quadro 2 Termos utilizados nas buscas nas fontes de informação e número de documentos recuperados	39
Quadro 3 Matriz de comparação entre as perspectivas da ciência aberta e as escolas de pensamento da ciência aberta	54
Quadro 4 Abordagens da ciência aberta por área de publicação	55
Quadro 5 Definições de ciência aberta em documentos oficiais	57
Quadro 6 Escolas de pensamento da ciência aberta e as implicações para publicações ampliadas	64
Quadro 7 Classificação dos signos	70
Quadro 8 O marco de trabalho semiótico	72
Quadro 9 Exemplos das matrizes da linguagem segundo as categorias semióticas	76
Quadro 10 Características de cada nível de interação com a publicação e exemplos	88
Quadro 11 As entidades na publicação científica e sua classificação segundo as matrizes	97
Quadro 12 Combinação dos elementos do artigo modular para a estrutura da publicação ampliada	114
Quadro 13 Alguns identificadores persistentes para objetos digitais	120
Quadro 14 Exemplo de anatomia do DOI	123
Quadro 15 Granularidade na atribuição de DOI num artigo da PLOS	125
Quadro 16 Campos de metadados no padrão DataCite	126
Quadro 17 Exemplos de formatos recomendados para arquivamento a longo prazo para algumas das entidades identificadas	127
Quadro 18 Classificação das entidades	139
Quadro 19 Padrões para a descrição das entidades e relacionamentos	143
Quadro 20 Propriedades da entidade “Autor”	145
Quadro 21 Propriedades da entidade “Código” e “Software”	146
Quadro 22 Propriedades da entidade “Conjunto de dados brutos”	147
Quadro 23 Propriedades da entidade “Arquivo multimídia”	149

Quadro 24 Propriedades das entidades “Áudio”, “Figura”, “Imagem”, “Vídeo” e “ <i>Workflow</i> ”	151
Quadro 25 Propriedades da entidade “Quadro”	152
Quadro 26 Propriedades da entidade “Texto”	154
Quadro 27 Propriedades da entidade “Texto - referências”	156
Quadro 28 Propriedades da entidade “Metadados”	158
Quadro 29 Descrição dos relacionamentos	159

LISTA DE SIGLAS

ABNT	Associação Brasileira das Normas Técnicas
ACRL	<i>Association of College and Research Libraries</i> - Associação de Bibliotecas Universitárias e de Pesquisa
ALI	<i>Access and License Indicator</i>
CMS	<i>Content Management System</i> – Sistema de Gestão de Conteúdo
COAR	<i>Confederation of Open Access Repositories</i>
CSMD	<i>Core Scientific Metadata Model</i>
DCAT	<i>Data Catalog Vocabulary</i>
DCMI	<i>Dublic Core Metadata Initiative</i>
DILAM	<i>Digital Images in Libraries, Archives and Museums</i>
DOI	<i>Digital Object Identifier</i> – Identificador de objeto digital
EOSC	<i>European Open Science Cloud</i>
FAIR	<i>Findable, Accessible, Interoperable, Reusable</i> – Encontrável, acessível, interoperável, reutilizável
FAPESP	Fundação de Amparo à Pesquisa do Estado de São Paulo
FOAF	<i>Friend of a friend</i>
GND	<i>Gemeinsame Normdatei</i>
HTTP	<i>Hypertext Transfer Protocol</i>
LD	<i>Linked Data</i>
LERU	<i>League of European Research Universities</i> - Liga de Universidades de Investigação Europeias
LIBER	<i>Ligue des Bibliothèques Européennes de Recherche</i> – Associação de bibliotecas de pesquisa europeias
NSF	<i>National Science Foundation</i>
OA	<i>Open Access</i> – Acesso Aberto
OCO	<i>Open Citation Ontology</i>
ODIN	<i>ORCID and DataCite Interoperability Network</i>
OJS	<i>Open Journal System</i>
OPAC	<i>Open Public Access Catalog</i> – Catálogo em linha
OWL	<i>Ontology Web Language</i>
PID	<i>Persistent identifier</i> – identificador persistente
PLOS	<i>Public Library of Science</i>
RDF	<i>Resource Description Framework</i>
RIP	<i>Rich Internet Publication</i>
ROR	<i>Research Organisation Registry</i>
SKO	<i>Scientific Knowledge Object</i>
SPP	<i>Scientific Publication Package</i>
STAR	<i>Structured, Transparent, Accessible Reporting Method</i>
Method	
THOR	<i>Technical and Human Infrastructure for Open Research</i>

TI	Tecnologia da Informação
UNESP	Universidade Estadual Paulista “Júlio de Mesquita Filho”
URI	<i>Uniform resource identifier</i>
URL	<i>Uniform resource locator</i>

SUMÁRIO

1 INTRODUÇÃO.....	24
2 ASPECTOS METODOLÓGICOS.....	31
2.1 O POLO EPISTEMOLÓGICO	34
2.2 O POLO TEÓRICO	36
2.3 O POLO TÉCNICO	37
2.4 O POLO MORFOLÓGICO.....	42
3 A CIÊNCIA ABERTA E GESTÃO DE DADOS DE PESQUISA COMO CONTEXTO.....	43
3.1 A CIÊNCIA ABERTA E A GESTÃO DE DADOS DE PESQUISA: CONCEITOS, CONTRASTES E PERSPECTIVAS	44
3.2 AS ESCOLAS DE PENSAMENTO DA CIÊNCIA ABERTA E A PUBLICAÇÃO AMPLIADA	58
4 SEMIÓTICA PEIRCEANA E MULTIMODALIDADE NA PESQUISA CIENTÍFICA.....	67
4.1 NOTAS SOBRE SEMIÓTICA PEIRCEANA.....	68
4.1.1 As matrizes da linguagem	76
4.2 A MULTIDIMENSIONALIDADE E O HIPERTEXTO	77
4.3 A MULTIMODALIDADE NA PESQUISA CIENTÍFICA.....	80
4.4 NÍVEIS DE INTERAÇÃO NAS PLATAFORMAS DE PUBLICAÇÃO, CATEGORIAS FENOMENOLÓGICAS E HIBRIDAÇÃO DAS LINGUAGENS	86
4.5 AS ENTIDADES NA PUBLICAÇÃO AMPLIADA E A DEFINIÇÃO SEMIÓTICA DA PUBLICAÇÃO AMPLIADA	97
5 A PUBLICAÇÃO CIENTÍFICA AMPLIADA: ABORDAGENS E ESTRUTURA	100
5.1 EVOLUÇÃO E ABORDAGENS DA PUBLICAÇÃO CIENTÍFICA AMPLIADA	103
5.1.1. Repercussão da publicação ampliada no contexto de pesquisa brasileiro ...	112
5.2. PROPOSTA DE ESTRUTURA DE UMA PUBLICAÇÃO CIENTÍFICA AMPLIADA.....	113
5.3 IDENTIFICADORES PERSISTENTES	115
6 MODELAGEM CONCEITUAL PARA A PUBLICAÇÃO CIENTÍFICA AMPLIADA.....	129
6.1 ANÁLISE SEMÂNTICA.....	134
6.1.1 Definição do problema e problema de estudo	136
6.1.2 Identificação das unidades semânticas.....	136
6.1.3 Classificação das unidades semânticas.....	137
6.1.4 Identificação das dependências ontológicas	140
6.1.5 Finalização do modelo semântico	141
6.2. PROPOSTA DE MODELAGEM CONCEITUAL PARA PUBLICAÇÕES CIENTÍFICA AMPLIADAS	143
6.2.1 Entidades	145
6.2.2 Relacionamentos	159

6.2.3 Diagrama do modelo	162
6.3 CONSIDERAÇÕES SOBRE O MODELO.....	165
7 CONSIDERAÇÕES FINAIS	167
REFERÊNCIAS.....	175

1 INTRODUÇÃO

Encontrar vias para a disseminação e troca de conhecimento científico tem sido uma das preocupações centrais da prática científica. Historicamente, precisavam-se de espaços para a publicação de novas teorias e a posterior discussão das mesmas. Assim, destacam-se no século XVII dois acontecimentos históricos: a fundação em 1660 na Inglaterra da *Royal Oscite of London for the Improvement of Natural Knowledge* (ou simplesmente *Royal Oscite*), bem como a criação em 1665 dos primeiros periódicos científicos *Le Journal des Sçavans* na França e *Philosophical Transactions of the Royal Oscite* na Inglaterra.

O surgimento destes dois periódicos científicos contribuiu enormemente ao compartilhamento de conhecimento científico durante os séculos XVII e XVIII. Por primeira vez era possível “[...] identificar erros, apoiar, rejeitar ou refinar teorias e reutilizar dados para o entendimento futuro” (ROYAL SOCIETY, 2012, p. 13, tradução própria). Além disso, as revistas científicas constituíram um meio para o fomento da dialética na ciência, entendendo a dialética como uma forma de evitar contradições na ciência e permitindo alterar a observação ou afirmações feitas sobre a observação (WAGENSBERG, 1998).

Esta troca de conhecimento dentro da comunidade científica é conhecida como “comunicação científica”. A Associação de Bibliotecas Universitárias e de Pesquisa (Association of College & Research Libraries - ACRL) dos Estados Unidos define a comunicação científica como:

O sistema por meio do qual a pesquisa e outros escritos acadêmicos são criados, avaliados segundo a qualidade, disseminados entre a comunidade científica e preservados para uso futuro. O sistema inclui tanto canais formais de comunicação, tais como publicações em periódicos revisados por pares, quanto canais informais, tais como listas de distribuição eletrônicas. (ACRL, 2003, tradução própria¹).

Meadows e Lemos (1999) destacam que os artigos de periódicos revisados por pares são “[...] considerados como as publicações definitivas dos resultados de pesquisa.” (p. 166). Os artigos são os ativos/assets² que mais

¹ ACLR, Scholarly Communication: <https://www.arl.org/focus-areas/scholarly-communication#.XIPfBNKjBI>

² Utilizaremos o termo “ativo/asset” para definir os objetos digitais com os que os usuários podem interagir segundo a definição de Abrams (2015) exposta no Capítulo 4.

facilmente chegam à comunidade científica da área tratada e ao redor dos quais podem surgir interações, citações e reflexões. Os autores afirmam que a importância da publicação em um periódico pode mudar no tempo, dependendo das condições. É precisamente na atualidade com a utilização dos ambientes web onde percebe-se esta mudança, pelo menos no formato do artigo científico.

Atualmente, na pesquisa existe a demanda de oferecer mais ativos/assets de pesquisa além do artigo textual tradicional (como conjuntos de dados ou *software*) que devem ser reconhecidos como contribuições intelectuais dos pesquisadores (PIWOWAR, 2013). Portanto, os dados de pesquisa adquirem uma grande relevância na pesquisa científica.

A ciência atravessou os paradigmas empírico, baseado na descrição de fenômenos naturais; teórico, baseado no uso de modelos e generalizações; e computacional, baseado na simulação de fenômenos complexos. Porém, a prática científica atual atravessa uma mudança de paradigma que envolve câmbios na forma de entender a ciência e desenvolvê-la caracteriza-se pelo uso de informação e conhecimento armazenado em computadores, bem como pelo uso intensivo de redes de computadores para a gestão da explosão de dados, e constitui o denominado *Quarto Paradigma Científico* (HEY; TANSLEY; TOLLE, 2009).

A gestão de dados, a estatística e os novos *softwares* de processamento desempenham um papel essencial neste paradigma, que é conhecido pelo termo em inglês *e-Science*. Geralmente, o termo *e-Science* é abordado desde a perspectiva da transformação e melhora das infraestruturas para a transmissão, processamento e arquivamento de dados digitais e informação (HEY; TREFETHEN, 2005; STEIN, 2008).

Porém, a prática científica baseia-se em processos que vão além da ciberinfraestrutura. Aparece assim o termo ciência aberta (*Open Science*) como um termo guarda-chuva definido no projeto FOSTER Open Science como:

A prática da ciência de tal maneira que outros possam colaborar e contribuir, onde os dados de pesquisa, as notas de laboratório e outros processos de pesquisa estejam livremente disponíveis sob termos que permitam o reuso, a redistribuição e a reprodução da pesquisa e de seus dados e métodos subjacentes. (FOSTER OPEN SCIENCE, 2018).

Mais especificamente para a comunicação científica, a ciência aberta envolve a publicação dos resultados de pesquisa em um formato digital e com

as menores restrições de acesso possíveis (OECD, 2015). Aliás, envolve a publicação tanto do artigo textual quanto dos outros ativos/assets produzidos durante e após a pesquisa —ou pelo menos dos metadados descritivos dos ativos/assets —, tais como: bancos de dados, vídeos, questionários, apresentações, planilhas de dados ou conjuntos de dados brutos. Como será aprofundado no Capítulo 4, ao longo deste trabalho utilizaremos o termo ativo/asset para fazer referência aos resultados de pesquisa (tanto publicados quanto não publicados). O ativo/asset é um objeto útil que expõe características apropriadas para os sentidos humanos. Portanto, é o tipo de objeto com o qual interatua o leitor ou o pesquisador.

Um aspecto importante da publicação de ativos/assets oriundos da mesma pesquisa é a necessidade de contextualizá-los e vinculá-los entre si. A contextualização e vinculação esclarecem a proveniência do ativo/asset e garantem sua confiabilidade. Para realizar estas duas ações —a contextualização e a vinculação—, um dos mecanismos existentes é a publicação ampliada ou melhorada (do inglês *enhanced publicação*), que consiste na publicação conjunta dos diferentes ativos/assets oriundos da mesma pesquisa. Embora esse tipo de publicação costuma ser caracterizado por uma parte narrativa textual acompanhada de subpartes (BARDI; MANGHI, 2014), a parte central da publicação ampliada pode não ser textual e incluir ativos/assets de natureza não textual.

Nesse sentido, considera-se a publicação científica ampliada como um objeto de investigação de caráter científico e social. Ao longo do texto esta publicação será caracterizada como um espaço semiótico de topografia tanto heterogênea quanto homogênea. Abordar as publicações ampliadas com esta perspectiva permite analisar a sua inclusão dentro dos processos de comunicação e manter uma estrutura para a agrupação dos ativos/assets. Justifica-se esta abordagem na atual crise de reprodutibilidade, destacada por autores como Chen *et al.* (2019). Um dos problemas desta crise é precisamente a falta de descrição, escolha de formatos e interoperabilidade dos ativos/assets de pesquisa o que impede o posterior reuso desses produtos morfológicos das pesquisas.

Um processo de comunicação científica que inclua as publicações científicas ampliadas desafia em vários aspectos o ciclo de

comunicação/publicação científica estabelecido e requer mudanças desde um ponto de vista epistemológico, teórico e técnico.

Epistemologicamente, requer-se uma nova abordagem à publicação científica distinta ao artigo textual tradicional. Esta abordagem resulta em outro tipo de publicação entendida como uma linguagem híbrida que permita a organização reticular da informação. Teoricamente, é necessário estudar os modelos de dados existentes e verificar como os ativos/assets de pesquisa podem estar integrados neles. Tecnicamente, é preciso ter novas infraestruturas informacionais e novos ciclos de publicação que abranjam todas as necessidades de avaliação dos ativos/assets (textuais e não textuais), tanto na concepção conceitual quanto nas diretrizes para a implementação técnica e a política da infraestrutura.

Diante do exposto, traçamos como objetivo geral desta pesquisa **propor, considerando os aspectos semióticos, um modelo para publicações ampliadas que sirva como mecanismo para a integração dos diferentes ativos/assets científicos na comunicação científica.**

Para alcançar o objetivo geral da pesquisa, traçamos os seguintes objetivos específicos:

- a. Contextualizar as mudanças que a ciência aberta (como marco epistemológico) leva ao contexto da comunicação científica (ambientes informacionais científicos e ciclo de publicação);
- b. Definir a publicação ampliada a partir da ótica da Semiótica Peirceana, bem como considerando conceitos como a semiosfera;
- c. Identificar as tecnologias da *web* semântica existentes para modelos de dados de publicação ampliada mediante uma análise da evolução destas publicações com o fim de definir a estrutura de representação da publicação científica ampliada;
- d. Analisar o ciclo de pesquisa científica para identificar as entidades necessárias na modelagem conceitual para publicações ampliadas de forma que todos os elementos estejam interconectados entre si e respeitem uma lógica de vinculação.

Nesta tese, partimos das seguintes duas hipóteses ou premissas:

A publicação ampliada pode ser conceitualizada como um espaço semiótico composto por ativos/assets heterogêneos e que abrangem as interações dentro do contexto da comunicação científica.

A completa descrição dos ativos/assets dentro de um modelo de publicação ampliada serve como mecanismo de publicação dos diferentes ativos/assets científico-acadêmico.

Do ponto de vista científico a presente pesquisa fomenta a inter e transdisciplinaridade, já que estabelece um diálogo entre a Ciência da Informação, as abordagens semióticas e a modelagem conceitual de entidade-relacionamento própria da Ciência da Computação. Além disso, a publicação ampliada é objeto de análise de outras comunidades discursivas ou sujeitos informacionais procedentes de outros contextos culturais, externos ou não à comunidade científica.

Portanto, do ponto de vista social, a publicação ampliada aporta transparência ao processo de pesquisa e serve também como mecanismo para devolver à sociedade o conhecimento científico criado nas universidades e centros de pesquisa financiados com fundos públicos. Não obstante, esta disponibilização deve ser realizada sempre sob licenças apropriadas que garantam o respeito da propriedade intelectual ao mesmo que tempo que contribuam ao conhecimento coletivo.

Academicamente, esta pesquisa insere-se na linha de pesquisa Informação e Tecnologia do Programa de Pós-graduação em Ciência da Informação da Universidade Estadual Paulista “Júlio de Mesquita Filho” (UNESP), já que incorpora elementos tecnológicos e aborda a inserção dos dados na comunicação científica dentro do contexto de investigação da Ciência da Informação. Ainda, possibilita outras pesquisas em nível de mestrado e doutorado que abordem temas como a modelagem de bancos de dados, os dados no percurso do processo de pesquisa científica, o ciclo de vida dos dados, a publicação de dados, os aspectos semióticos dos processos de curadoria digital e preservação, o *design* de ambientes informacionais e a avaliação e medição do impacto dos resultados de pesquisa, entre outros.

Além disso, acreditamos que o resultado desta pesquisa poderá subsidiar as atividades de cientistas da informação no seu papel de bibliotecários de dados, gestores da informação, desenvolvedores de repositórios e designers, entre outros, na construção de ambientes para a inserção de publicações científicas ampliadas.

Esta tese está estruturada em sete capítulos.

Esse primeiro capítulo introduz o tema e o objeto de pesquisa. Além disso, detalham-se os objetivos e como será estruturada a tese para atingir estes objetivos.

No segundo capítulo, são tratados os aspectos metodológicos mediante a apresentação do método quadripolar. O método quadripolar entende-se como uma aproximação geral a todos os aspectos do processo de investigação. Especificamente, há uma incidência em como o objeto de pesquisa, a publicação ampliada, pode desafiar a publicação científica epistemológica, teórica e tecnicamente. Por isso, descrevem-se as transformações técnicas que supõem as publicações ampliadas.

O terceiro capítulo aborda a ciência aberta como marco epistemológico da pesquisa. Tratam-se as características da ciência aberta como mudança de paradigma que representa um novo *modus operandi* na forma de fazer ciência, bem como a interseção entre ciência aberta e gestão de dados de pesquisa (GDP³). Descrevem-se também as escolas de pensamento nas que se divide o movimento da ciência aberta —escolas pública, pragmática, democrática, da infraestrutura e das métricas—, contrastando-as com as implicações na GDP, e como a publicação ampliada pode ser abordada desde cada uma delas.

No quarto capítulo, incide-se sobre os aspectos semióticos dentro da pesquisa científica. Estes aspectos semióticos são considerados devido à combinação de ativos/assets de pesquisa de diferente natureza que convivem em uma publicação ampliada. A análise Semiótica inclui a revisão de ambientes de publicação dos quais se extraem as entidades que compõem uma publicação ampliada. Ainda, propõe-se uma definição Semiótica deste tipo de publicação.

³ Consideramos oportuno utilizar a sigla em português GDP (Gestão de dados de pesquisa) neste texto, embora alguns documentos encontrados na literatura mantenham a sigla em inglês RDM (*Research Data Management*) independentemente da língua do texto.

O quinto capítulo trata diretamente da publicação ampliada. Apresenta um levantamento da literatura sobre a publicação digital e a publicação ampliada para poder contextualizar o objeto desta pesquisa em um marco temporal e dentro dos ciclos de publicação. Esse capítulo também apresenta uma estrutura de publicação ampliada que servirá de base para o proposto nesta tese. Além disso, introduz-se um apartado sobre os identificadores persistentes e sua relevância como identificadores nos entornos de publicação para o enriquecimento semântico da representação das publicações.

O sexto capítulo apresenta o modelo e sua construção. Para isso, descrevem-se as etapas da análise semântica que permitem delimitar as unidades semânticas e suas dependências ontológicas. Depois aprofunda-se na descrição dos atributos das entidades e seus relacionamentos. Para exemplificar a modelagem das entidades, utilizam-se valores desta tese, quando possível. Finalmente, mostra-se o diagrama do modelo.

O sétimo capítulo conclui o trabalho e delinea futuras linhas de pesquisa.

Finalmente apresenta-se a bibliografia utilizada no desenvolvimento da presente tese. Caso a referência não estivesse disponível em acesso aberto, procurou-se uma versão aberta utilizando a extensão para o navegador da ferramenta Unpaywall⁴. Não obstante, nem sempre foi possível encontrar uma versão aberta.

⁴ Unpaywall: <https://unpaywall.org/>

2 ASPECTOS METODOLÓGICOS

Para abordar esta pesquisa utiliza-se o método quadripolar, pois, como indicado na introdução, a análise das publicações científicas ampliadas envolve considerações epistemológicas, teóricas e técnicas.

O método quadripolar foi proposto em 1974 pelos pesquisadores belgas De Bruyne, Herman e De Schoutheete da Universidade de Lovaina (Bélgica), com o objetivo de criar um instrumento de investigação para a área de Ciências Sociais e Humanas (DE BRUYNE; HERMAN; DE SCHOUTHEETE, 1982).

Apresenta-se como um método dinâmico e adaptado aos requerimentos de complexidade e análise global de uma pesquisa em Ciência da Informação. Constitui uma alternativa à dicotomia entre metodologias “quantitativas” e “qualitativas” (SILVA, 2006). Assim, uma das suas principais características é o tratamento da pesquisa de forma não linear nem sequencial, entendendo a pesquisa científica como um conjunto de quatro polos que dialogam entre eles e se articulam.

O método quadripolar possui uma plasticidade que permite ajustá-lo às características dos fenômenos da pesquisa. Não há um único objeto ou fenômeno no qual esse método possa ser aplicado, o que permite utilizar vários aparatos teóricos e epistemológicos, diversas técnicas e produzir diferentes resultados ou morfologias. A plasticidade consegue que o método se aproxime às pesquisas qualitativas, características das Ciências Sociais (SILVA; RIBERO, 2008).

A denominação “quadripolar” deve-se à sua divisão em quatro polos que abrangem as necessidades da pesquisa: polo epistemológico, polo teórico, polo técnico e polo morfológico. Estes polos, como dito anteriormente, não representam momentos isolados da pesquisa, senão que fomentam a conectividade entre as ações e práticas desenvolvidas.

O polo epistemológico serve como base para a construção do objeto científico definido dentro de uma problemática de pesquisa. Reflete-se também sobre os paradigmas nos quais a pesquisa será baseada. De acordo com De Bruyne, Herman, De Schoutheete (1982):

O polo epistemológico exerce uma função de vigilância crítica. Ao longo de toda a pesquisa ele é a garantia da objetivação – isto é, da produção – do objeto científico, da explicitação das problemáticas da pesquisa. Encarrega-se de

renovar continuamente a ruptura dos objetos científicos com os do senso comum. Decide, em última instância, das regras de produção e de explicitação dos fatos, da compreensão e da validade das teorias. Explicita as regras de transformação do objeto científico, critica seus fundamentos. (DE BRUYNE; HERMAN; DE SCHOUTHEETE, 1982, p. 35).

Em algumas ocasiões, além do polo epistemológico, podem ser identificadas outras duas dimensões: política e ética (BUFREM, 2013). Estas referem-se ao estabelecimento de prioridades, autonomia do pesquisador e objetividade e fidedignidade dos resultados. Dessa maneira, para o desenvolvimento da presente pesquisa, estas dimensões serão consideradas parte do polo epistemológico e inclusas na construção do objeto científico e na liberdade de definir a problemática.

O polo teórico é o lugar de abordagem do objeto anteriormente construído, da formulação das hipóteses, teorias e conceitos. Constitui em si o marco referencial e teórico da pesquisa desenvolvida, como definem De Bruyne, Herman, De Schoutheete (1982):

O polo teórico guia a elaboração das hipóteses e a construção dos conceitos. É o lugar da formulação sistemática dos objetos científicos. Propõe regras de interpretação dos fatos, da especificação e de definição das soluções provisoriamente dadas às problemáticas. É o lugar de elaboração das linguagens científicas, determina o movimento da conceitualização. (DE BRUYNE; HERMAN; DE SCHOUTHEETE, 1982, p. 35).

O polo técnico lida com a instrumentalização da pesquisa e estabelece o contato e confrontação entre a teoria e os objetos com o objetivo de validar os processos e dispositivos metodológicos. Neste polo podem ser desenvolvidas operações como “[...] observação de casos e de variáveis, a avaliação, retrospectiva e prospectiva, a informetria e até a experimentação mitigada ou ajustada ao campo de estudo de fenomenalidades humanas e sociais [...]” (SILVA, 2006, p. 155). Também, de acordo com De Bruyne, Herman, De Schoutheete:

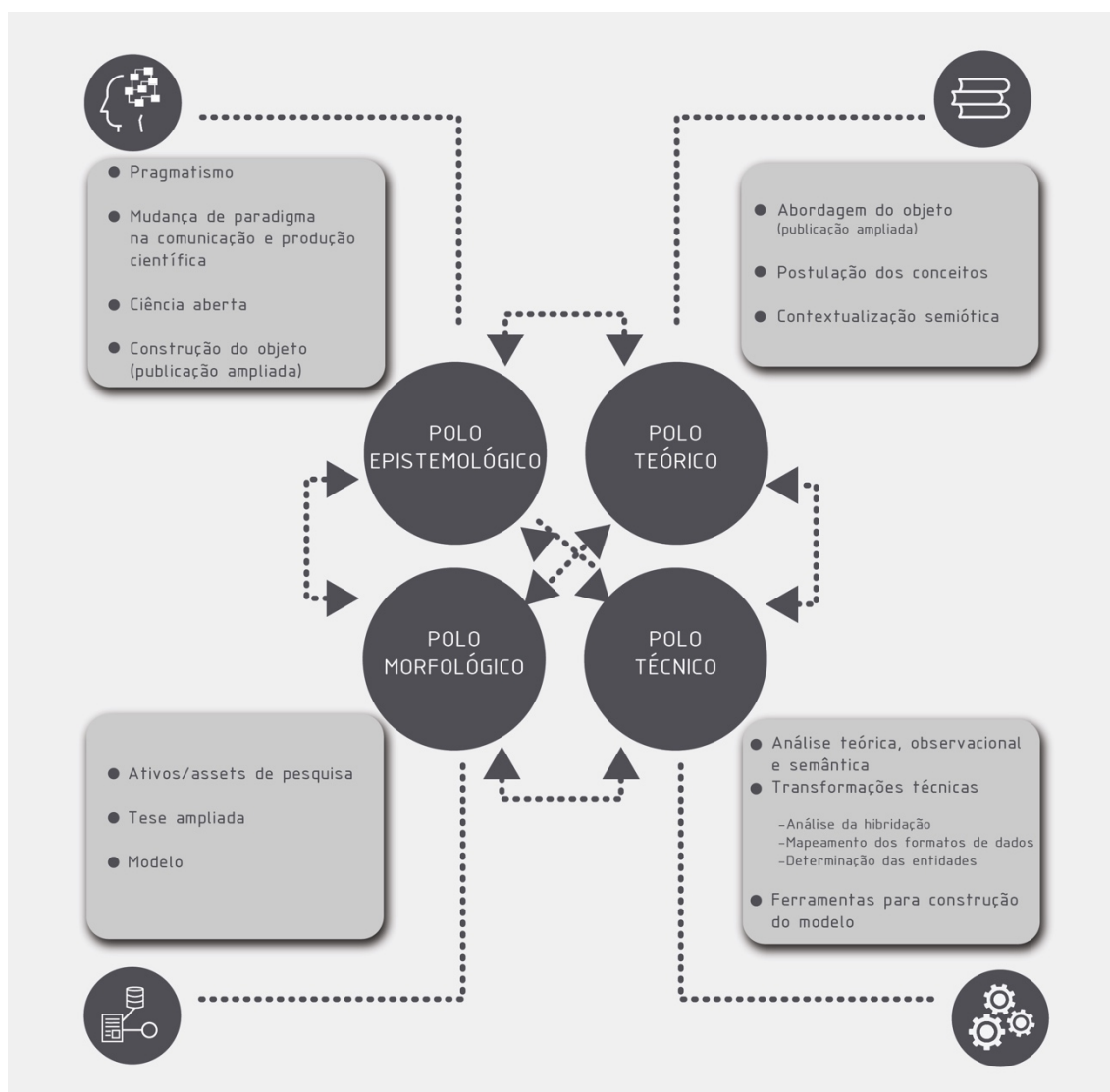
O polo técnico avizinha-se dos ‘quadros de referência’ que lhe fornecem inspirações e problemáticas provenientes das contribuições teórico-práticas das disciplinas e dos ‘hábitos’ adquiridos. Esses quadros de referência desempenham um papel paradigmático implícito. São principalmente os quadros de referência ‘positivista’, ‘compreensivo’, ‘funcionalista’ e ‘estruturalista’. (DE BRUYNE; HERMAN; DE SCHOUTHEETE, 1982, p. 35).

O polo morfológico trata a formalização dos resultados de pesquisa. Para De Bruyne, Herman, De Schoutheete:

[...] é a instância que enuncia as regras de estruturação, de formação de objeto científico, impõe-lhe uma certa figura, uma certa ordem entre seus elementos. Permite colocar um espaço de causação em rede onde se constroem os objetos científicos, sejam como modelos/cópias, seja como simulacros de problemáticas reais. (DE BRYNE; HERMAN; DE SCHOUTHEETE, 1982, p. 35-36).

Ao considerar estes quatro polos, o método quadripolar se apresenta como não redutor e adaptável a diferentes objetos científicos. Assim, a seguir, aprofunda-se a construção de cada polo para a presente pesquisa.

Figura 1 Aplicação do método quadripolar nesta pesquisa.



Fonte: Elaboração própria.

2.1 O POLO EPISTEMOLÓGICO

Quando existe uma revolução na ciência, os cientistas, mesmo empregando os instrumentos do momento, veem o mundo desde uma nova perspectiva que lhes permite enxergar aspectos anteriormente ocultos (KUHN, 1962). Assim, “[...] os grandes problemas epistemológicos nascem a partir de crises nas ciências, do questionamento, não de seus resultados, mas dos seus fundamentos.” (BRUYNE; HERMAN; DE SCHOUTHEETE, 1982, p.42).

A publicação científica possui desde há mais de quatro séculos um instrumento de comunicação principal: o artigo científico. Porém, no contexto atual, existe um interesse por conhecer mais ativos/assets de pesquisa para atingir uma melhora nos princípios da dialética dentro e fora da comunidade científica e oferecer uma pesquisa mais transparente.

Nesse contexto, configuram-se duas mudanças de paradigma: a primeira, relacionada ao conceito de *e-Science* ou paradigma dos dados e a segunda referida à abertura do processo de pesquisa como um todo, conhecida como ciência aberta. Cabe destacar que a *e-Science* e a ciência aberta têm objetos centrais diferentes, porém complementares. O objeto da *e-Science* são os dados e o trabalho com eles, já a ciência aberta versa trabalhar com o processo de pesquisa completo e o estabelecimento das menores barreiras possível no acesso aos elementos deste processo (FOSTER, 2018). Sem dúvida, os dados desempenham um papel importante na ciência aberta, mas não são o único interesse dela.

Para a explicitação do objeto de estudo precisamos entendê-lo como “[...] um sistema de relações construídas expressa e explicitamente.” (DE BRUYNE; HERMAN; DE SCHOUTHEETE, 1982, p.51) que nasce após um processo de objetivação, “[...] de produção dos fatos e dos objetos científicos que é a progressão da formação, da estruturação e do recorte dos fatos até os procedimentos de coleta de informação.”

Assim, partimos para a *ruptura epistemológica* que envolve um distanciamento do objeto científico das pré-noções existentes sobre ele (DE BRUYNE; HERMAN; DE SCHOUTHEETE, 1982). O paradigma dos dados e o movimento de ciência aberta abandeiram esta ruptura que considera que o artigo científico não é o único e principal meio de comunicação científica; as

comunidades científicas deixam de ser grupos fechados; existem meios de expressão de diferente natureza. Esta ruptura não é ainda completa e não abandona totalmente as ideias previamente concebidas sobre a prática científica, mas insere uma série de mudanças e transições nos afazeres científicos. Tanto em contextos de ciência aberta como de GDP, é comum falar em “Mudança cultural”. Surgem assim conceitos como *FAIR⁵ culture* ou *Open Science Culture*.

Além disso, a Liga de Universidades de Investigação Europeias (LERU – *League of European Research Universities*) propõe um roteiro para atingir a mudança cultural necessário para tornar real a ciência aberta.

Assim, enuncia oito dimensões da ciência aberta: (1) o futuro da publicação acadêmica/científica, (2) dados FAIR, (3) infraestrutura (p.ex. as arquiteturas federadas da *The European Open Science Cloud* – EOSC), (4) formação e habilidades, (5) recompensas e incentivos, (6) próxima geração de métricas, (7) integridade na pesquisa, (8) ciência cidadã (LERU, 2018).

A presente pesquisa gira ao redor do futuro da publicação científica e é aí onde centra a ruptura epistemológica.

Depois, temos que conceitualizar o *objeto percebido* “[...]aquele que se dá aos sentidos sob forma de imagens.” (DE BRUYNE; HERMAN; DE SCHOUTHEETE, 1982, p.51). Esse objeto não é toda a realidade e, portanto, não é tudo o que acontece na pesquisa, senão o que o sujeito percebe e entende como informativo. Desde um ponto de vista semiótico seria o “[...] enraizamento da significação, significante e significado, no referente.” (DE BRUYNE; HERMAN; DE SCHOUTHEETE, 1982, p.51). Daí a importância de delimitar as questões semióticas da multimodalidade na pesquisa científica e análise das linguagens híbridas na publicação científica/acadêmica.

Esta análise relaciona-se diretamente com a Fenomenologia Peirceana, encarregada de descrever e analisar os fenômenos naturais e mentais, quer dizer aquelas experiências abertas para todo ser humano. A Fenomenologia, mediante observação direta dos fenômenos, permite descrever as diferenças

⁵ Conceito aplicado aos dados de pesquisa que devem ser Encontráveis (Findable), Acessíveis (Accessible), Interoperáveis (Interoperable) e Reutilizáveis (Reusable). Esse conceito será abordado no Capítulo 3.

entre eles e generalizar as observações. Isso torna-se especialmente relevante na análise dos ambientes de publicação desenvolvida no Capítulo 4.

Finalmente é necessário falar do *objeto de conhecimento* como “[...] objeto que responde a leis de composição autônomas, construído por métodos explícitos.” (DE BRUYNE; HERMAN; DE SCHOUTHEETE, 1982, p. 51-52).

O objeto científico se forma após esse processo de conceptualização e formalização. Desta maneira, o nosso objeto científico será a publicação ampliada percebida na literatura às vezes como um mecanismo para a publicação de dados (PAMPEL et al., 2013) e outras, um objeto dinâmico e composto do conhecimento (MUCHERONI; SILVA; PALETTA, 2015). Não obstante, ao longo desta tese, contextualiza-se como um espaço semiótico.

2.2 O POLO TEÓRICO

No polo teórico se elaboram as hipóteses e se constroem os conceitos a partir da análise das pré-noções sobre o tema (DE BRUYNE; HERMAN; DE SCHOUTHEETE, 1982). Silva (2006), afirma que no polo teórico:

[...] centra-se a racionalidade do sujeito que conhece e aborda o objeto, bem como a postulação de leis, a formulação de hipóteses, teorias e conceitos operatórios e consequente confirmação ou informação do «contexto teórico» elaborado. (SILVA, 2006, p. 154).

O polo teórico fornecerá o referencial necessário para fundamentar a pesquisa desenvolvida. No polo epistemológico definimos a publicação ampliada como o objeto de pesquisa inserido na ruptura epistemológica propiciada pelo movimento da ciência aberta.

Portanto, as bases teóricas da pesquisa partem do processo de comunicação científica, sobre as características semióticas presentes neste tipo de publicação e do conceito de publicação ampliada e modelos de dados para ela. Esta publicação é principalmente viável devido a existência do ambiente da *web* e das relações que se estabelecem neste ambiente caracterizado como ecologia informacional complexa.

Entende-se uma ecologia informacional complexa como “[...] uma conjunção sinérgica de ambientes analógicos, digitais e híbridos, tecnologias analógicas e digitais, utilizadas de maneira holística, em multicanais pelos sujeitos em determinado contexto cultural.” (OLIVEIRA; LIMA, 2016, p. 53).

O processo científico que leva a publicação de uma publicação ampliada, bem como a publicação em si, reúne espaços de diferentes naturezas (analógicos, digitais e híbridos), no processo existem diversos aparatos tecnológicos e na criação e publicação final da publicação ampliada existe uma grande variedade de formatos de dados.

Os estudos teóricos desenvolvidos neste polo permitiram o delineamento da pesquisa e esclarecimento de cada um dos conceitos, dando sustentabilidade para a condução da coleta de dados nos ambientes de publicação, bem como o delineamento de modelos que possam ser aplicados no contexto da publicação ampliada. Considera-se também a Semiótica, tanto Peirceana quanto da cultura, como apoio teórico para a elaboração da definição semiótica de publicação ampliada (ver Capítulo 4). Define-se dessa forma um conceito operatório que posteriormente será um dos objetos morfológicos do trabalho.

2.3 O POLO TÉCNICO

O polo técnico abrange e descreve os processos de coleta, organização e tratamento de dados, possuindo esses dados uma natureza diversa. Assim, no polo técnico:

[...] consuma-se, por via instrumental, o contacto com a realidade objectivada, aferindo-se a capacidade de validação do dispositivo metodológico, sendo aqui que se desenvolvem operações cruciais como a observação de casos e de variáveis e a avaliação retrospectiva e prospectiva, sempre tendo em vista a confirmação ou refutação das leis postuladas, das teorias elaboradas e dos conceitos operatórios formulados. (SILVA, 2006, p. 154).

De Bruyne, Herman, De Schoutheete (1982) definem três campos na pesquisa técnica: o campo doxológico, o campo epistêmico e o campo teórico. O primeiro é a realidade diária que contém as informações; o segundo é a transformação dessa realidade devido aos processos de coleta de dados orientados pelas hipóteses de trabalho; e o terceiro é a redução dos dados em fatos quando têm sido confrontados com as pré-noções da pesquisa teórica. Esse percurso entre os três campos permite a objetivação das informações que se tornam parte de um referencial teórico. Reforça-se assim a ideia de plasticidade do método quadripolar e a conexão e conversação entre os polos da pesquisa. Portanto, é importante destacar as diferenças entre o polo técnico e o polo teórico. “O polo técnico é o momento da observação, do relatório dos

fatos, enquanto o polo teórico é o momento da interpretação e da explicação desses fatos.” (DE BRUYNE; HERMAN; DE SCHOUTHEETE, 1982, p. 204).

O polo técnico traz consigo uma série de transformações técnicas (como definidas pelos autores De Bruyne, Herman e De Schoutheete) associadas aos processos de observação das informações, seleção dos dados e redução a fatos. Para a presente pesquisa, as transformações técnicas necessárias para a consecução do objetivo geral esquematizam-se no Quadro 1.

Quadro 1 As transformações técnicas para a integração da publicação ampliada.

	Transformações técnicas
Observação Informações	• Necessidade de novas abordagens na publicação científica/acadêmica; • Presença de hibridação de linguagens na multimodalidade da pesquisa científica/acadêmica.
Seleção Dados	• Mapeamento teórico dos elementos nas publicações científica/acadêmica inseridas nos ambientes <i>web</i>; • Mapeamento dos formatos de dados nos ambientes <i>web</i>; • Análise dos elementos da publicação segundo a hibridação da linguagem e sua presença dentro da ecologia informacional.
Redução Fatos	• Determinação das entidades da publicação ampliada e sua descrição.

Fonte: Elaboração própria a partir de De Bruyne, Herman e De Schoutheete (1982, p. 207).

Para a consecução das transformações técnicas, utilizam-se duas técnicas de coleta de dados. Por um lado, a análise bibliográfica e documental e, por outro, a análise observacional dos ambientes *web* de publicação científica/acadêmica.

Por meio da análise bibliográfica e documental, analisamos o “estado da arte” do nosso objeto de pesquisa —a publicação ampliada—, bem como os demais elementos presentes na pesquisa —o movimento da ciência aberta, as questões semióticas na pesquisa científica e as tecnologias semânticas para a descrição de entidades—. Esta revisão da literatura foi realizada principalmente a partir dos itens selecionados em grandes bases de dados como *Scopus*, *SciELO* e *Dimensions* e ampliada com resultados obtidos em *Google Scholar* e informações recuperadas de livros e teses doutorais. Os principais critérios de seleção dos conteúdos foram títulos, resumo e palavras-chave, desde um ponto de vista descritivo, bem como área do conhecimento, favorecendo a área de Ciência da Informação para poder abordar o objeto da pesquisa com o olhar das Ciência da Informação. A intenção inicial era incluir apenas aqueles documentos (artigos, livros, teses doutorais, relatórios e roteiros) disponibilizados em acesso

aberto; porém, alguns documentos que foram considerados de interesse não ofereciam essa opção e foram também incluídos.

As buscas realizaram-se essencialmente seguindo as expressões de busca detalhadas no Quadro 2.

Quadro 2 Termos utilizados nas principais buscas nas fontes de informação e número de documentos considerados.

Termo	Material	Quantidade
“Enhanced publication”, “publicación ampliada”, “publicação ampliada”,	Artigo	62
	Livro	2
	Tese	4
	Relatório e roteiro	1
“Open Science”, “Ciencia Abierta”, “ciência aberta”,	Artigo	64
	Livro	3
	Tese	-
	Relatório e roteiro	6
“multimodal scholarship”, “investigación multimodal”, “pesquisa multimodal”	Artigo	1
	Livro	-
	Tese	-
	Relatório e roteiro	-
“scholarly communication lifecycle”, “ciclo de comunicação científica”, “ciclo de comunicación científica”.	Artigo	30
	Livro	-
	Tese	-
	Relatório e roteiro	1
“Semiotics” AND “Information systems”, “Semiótica” AND “sistemas de la información”, “Semiótica” AND “sistemas da informação”	Artigo	34
	Livro	6
	Tese	-
	Relatório e roteiro	-

Fonte: Elaboração própria.

É importante ressaltar que algumas expressões de busca não ofereceram resultados relevantes para os temas. Por exemplo, não há literatura específica em espanhol sobre *publicaciones ampliadas*. O 75 % dos textos recuperados foram em língua inglesa; o 16,2 %, em língua portuguesa; o 7,3 % em língua espanhola e um 1,5 %, em outras línguas. Embora não foram utilizados termos de busca em outras línguas, as fontes de informação utilizadas implementam *cross-lingual retrieval*, o que permitiu obter textos também em outros idiomas.

A alta porcentagem de texto em inglês permite descrever o caráter internacional das pesquisas nesta temática, já que tanto autores anglófonos quanto não anglófonos optam pela publicação neste idioma. A respeito da faixa temporal, para o caso das publicações ampliadas não se estabeleceram limites temporais, pois o interesse era conhecer um “estado da arte” e uma perspectiva da evolução do conceito. Para os demais termos consideraram-se principalmente os documentos dos últimos sete anos. O principal levantamento

bibliográfico realizou-se desde o início do doutorado em março de 2017 até dezembro de 2018. Posteriormente incluíram-se outros documentos; porém, sem seguir uma metodologia sistemática.

Por sua vez os termos “semiosphere”, “persistent identifiers”, “digital curation” e “information systems” (e suas correspondentes traduções em espanhol e português) foram utilizados de forma acessória. Seu uso foi diretamente em inglês, dada a projeção internacional anteriormente mencionada.

Como segunda técnica de coleta de dados, utiliza-se a análise observacional dos ambientes *web* de publicação científica/acadêmica que permite delinear como se configuram as publicações atualmente no espaço *web*. Para isso, definem-se os objetos e unidades a serem observados, sendo que os modelos de dados seguem as publicações e se existe ou não alguma hibridação nas linguagens utilizadas. Esta técnica de coleta possui um caráter mais sistemático e a intervenção como pesquisadora é mínima. Os ambientes analisados foram os seguintes:

- A editora Copernicus Publications⁶, cujas publicações estão ampliadas de maneira modular com dados de pesquisa e vídeos. Os vídeos são arquivados no TIB AV-Portal;
- a plataforma ScienceDirect⁷ da editora Elsevier, que incluem resumos em vídeo em várias das suas publicações, bem como bases de conhecimento e outros ativos/assets;
- a editora CellPress⁸ (propriedade de Elsevier) que implementou um novo formato para a metodologia da pesquisa chamado *STAR (Structured, Transparent, Accessible Reporting) Methods format*;
- a editora Public Library Of Science⁹ (PLOS) que oferece conteúdo relacionado com a publicação textual identificado mediante identificador de objeto digital (*Digital Object Identifier – DOI*). A granularidade da atribuição do DOI atinge cada parte do artigo (figuras, tabelas, quadros, vídeos, materiais complementares, etc.);

⁶ Copernicus Publications: <https://publications.copernicus.org/>

⁷ Science Direct: <https://www.sciencedirect.com/>

⁸ CellPress: <https://www.cell.com/>

⁹ PLOS: <https://www.plos.org/>

- o projeto *Vectors, Journal of Culture and Technology in a Dynamic Vernacular*¹⁰, que oferece um ambiente de publicação dinâmico e baseado em diferentes formatos;
- o periódico *The Comics Grid, Journal of comics scholarship*¹¹, da Open Library of Humanities, que introduz o formato do visual do quadrinho como meio de comunicação científica.

O último passo dentro das transformações técnicas próprias deste polo é a delimitação e extração das entidades que passaram a formar parte da modelagem conceitual para publicação ampliada. A modelagem se realiza seguindo o modelo de Entidade-Relacionamento (ER). São estabelecidas também as relações entre cada uma das entidades e a descrição de cada uma dessas entidades (seu identificador único e seus atributos mínimos para um objeto digital). Para a modelagem ER será utilizado o *software* de modelagem StarUML¹² apoiado na linguagem de modelagem unificada (*Unified Modeling Language* - UML). A escolha do *software* foi baseada nos seguintes critérios:

- Disponibilidade total ou parcial sob licença de código aberta. As primeiras versões do StarUML possuíam uma licença GNU GPL (General Public License)¹³, atualmente a licença é proprietária. Não obstante, a versão não registrada permite o uso sem restrições de *software* para fins não comerciais;
- aspectos de usabilidade. A interface do usuário está estruturada de forma simples; uma vez selecionado o tipo de diagrama desejado, a introdução das entidades e atributos é intuitiva. Além disso, é possível trabalhar facilmente com atalhos do teclado para introduzir o tipo de atributo;
- documentação disponível. Há disponível uma ampla documentação sobre o uso do *software*, tanto diretamente na página oficial¹⁴, quanto no correspondente repositório de GitHub¹⁵. A disponibilidade em GitHub também permite a edição colaborativa da documentação; e
- compatibilidade com o sistema operacional MacOS.

¹⁰ Vectors: <http://vectors.usc.edu/journal/index.php?page=Introduction>

¹¹ The Comics Grid: <https://www.comicsgrid.com/>

¹² StarUML: <http://staruml.io/>

¹³ GNU GPL: <https://www.gnu.org/licenses/gpl-3.0.en.html>

¹⁴ Documentação StarUML: <https://docs.staruml.io/>

¹⁵ StarUML gitbook: <https://github.com/staruml/staruml-gitbook>

Não obstante, StarUML não é o único *software* para esse tipo de modelagem. Para quem deseja executar os *scripts* de linguagem de consulta estruturada (SQL), MySQL Workbench¹⁶ é outra opção. Outra ferramenta interessante, e útil para fins didáticos, é brModelo¹⁷, desenvolvida pela Universidade Federal de Santa Catarina e o Centro Universitário de Várzea Grande. O uso do brModelo não foi considerado nesta tese, dado que não é compatível com MacOS.

2.4 O POLO MORFOLÓGICO

Após percorrer todos os polos, a pesquisa atinge o polo morfológico, onde se realiza a formalização dos resultados. Isto implica apresentar o objeto de estudo, os resultados e o processo de pesquisa (SILVA, 2006).

No que tange a uma tese doutoral, o polo morfológico abrange os resultados parciais que foram sendo publicados ao longo do processo em periódicos científicos ou anais de congressos. Também, incluem-se no polo morfológico os relatórios de pesquisa tanto de qualificação como a tese final a ser defendida.

Esta pesquisa em particular colocou como objetivo principal a apresentação de uma definição de publicação ampliada com uma abordagem Semiótica e a elaboração de uma modelagem conceitual para publicações ampliadas onde consta cada entidade com as suas relações, os seus identificadores e os seus atributos. Como indicam De Bruyne, Herman e De Schoutheete (1982), “O espaço morfológico pode se desdobrar em vários planos para acolher a construção de ‘modelos’ teóricos que se referem a modelos materiais que descrevem estados de coisas.” (p.163).

O “modelo teórico” desta tese descreve o possível estado ou configuração de uma publicação ampliada em um ambiente digital. Tanto a modelagem quanto a definição com abordagem Semiótica são, portanto, as contribuições principais desta pesquisa e o núcleo principal do polo morfológico.

¹⁶ MySQL Workbench: <https://www.mysql.com/products/workbench/>

¹⁷ brModelo: <http://www.sis4.com/brModelo/>

3 A CIÊNCIA ABERTA E GESTÃO DE DADOS DE PESQUISA COMO CONTEXTO

Nos últimos anos, o crescimento do conhecimento científico e da produção tem sido notório. De fato, cada nove anos a produção científica se duplica (BORNMANN; MUTZ, 2014). Isso deve-se em parte às práticas científicas abertas e ao uso e produção intensivos de dados. Como assinala a Royal Society (2012) no relatório *Science as an open enterprise* (A ciência como uma empresa aberta), a prática científica atual gira ao redor da *openness*¹⁸. O conceito de *openness* faz referência à eliminação de restrições e ao fomento do acesso (para pessoas e máquinas) e pode ser aplicado em uma ampla gama de contextos.

Dentro da comunidade científica termos como *open source* (código aberto), *open access* (acesso aberto), *open knowledge* (conhecimento aberto), *open government* (governo aberto) e *open science* (ciência aberta) têm ganhado muita popularidade e são amplamente utilizados. Porém, seu uso estendido também os torna mais difíceis de desambiguar e de associar com um conceito ou prática específico. Nesse sentido, também podemos entender que o termo *open* (aberto) está relacionado com os direitos, o acesso, o uso, a transparência e a participação (POMERANTZ; PEEK, 2016).

Abre-se uma janela de oportunidade para colocar o conceito de *openness* no debate e analisar as implicações que terá dentro de cada comunidade científica, bem como as áreas de trabalho que podem ser desenvolvidas desde outras partes interessadas (p.ex. agências financiadoras de pesquisa, desenvolvedores de políticas públicas, jornalistas científicos ou cientistas cidadãos). Isto envolve também uma análise das estratégias de GDP a serem desenvolvidas.

Nesse contexto, esse capítulo explora o conceito de ciência aberta, o contrasta com a GDP e contextualiza as consequentes mudanças de paradigma na prática científica marco epistemológico dentro da Ciência da Informação para o desenvolvimento de publicações científicas ampliadas.

¹⁸ O termo *openness* será mantido em inglês ao longo do texto, pois não há uma tradução clara e direta ao português. *Openness* entende-se como uma mistura entre abertura e transparência.

3.1 A CIÊNCIA ABERTA E A GESTÃO DE DADOS DE PESQUISA: CONCEITOS, CONTRASTES E PERSPECTIVAS

Uma forma de abordar a ciência aberta é a partir do discurso de Thomas Kuhn no livro *A estrutura das revoluções científicas* (KUHN, 1962). Para o autor, as revoluções científicas estão caracterizadas por mudanças de paradigma, que trazem consigo alterações nos conceitos básicos e nas práticas experimentais. Os paradigmas definem-se como: “[...] realizações científicas universalmente reconhecidas que, durante certo tempo, proporcionam modelos de problemas e soluções a uma comunidade científica.” (KUHN, 1962, p. 13). A mudança entre paradigmas costuma ser conflituosa, pois há incerteza frente ao abandono de práticas enraizadas para começar a se apropriar das novas práticas.

Bartling e Friesike (2014) entendem dois paradigmas na história da prática científica contemporânea que chamam de revoluções científicas (BARTLING; FRIESIKE, 2014): a primeira e a segunda revolução. Aqui o termo “revolução” não têm o mesmo significado que para Kuhn, senão que se refere diretamente à mudança.

A primeira revolução baseou-se na profissionalização da criação do conhecimento científico e caracterizou-se pelo compartilhamento de resultados científicos com os pares. Esse processo permitiu identificar erros para polir teorias ou experimentos, o que “[...] tornou a ciência um processo autocorretivo.” (ROYAL SOCIETY, 2012, p. 13). Foi nesse ponto quando o artigo científico se legitimou como principal meio para disseminar resultados.

Por sua vez, a segunda revolução está caracterizada pelo uso da Internet como ferramenta para publicar resultados ou partes da pesquisa em qualquer ponto do ciclo de pesquisa (ideia de pesquisa, resultados intermediários, resultados conclusivos, revisões pós-publicação).

Seguindo na distinção de paradigmas, Hey, Tansley e Tolle (2009) trazem a divisão de Tim Gray dos quatro paradigmas da ciência: o empírico, o teórico, o computacional e o orientado por dados. A ciência começa descrevendo fenômenos naturais por meio da percepção sensorial. A introdução de modelos, como são as leis de Newton, para explicar estes fenômenos (a gravidade no caso das leis de Newton) supôs a primeira mudança de paradigma. Estes modelos permitiam tanto a descrição empírica quanto a predição de novos

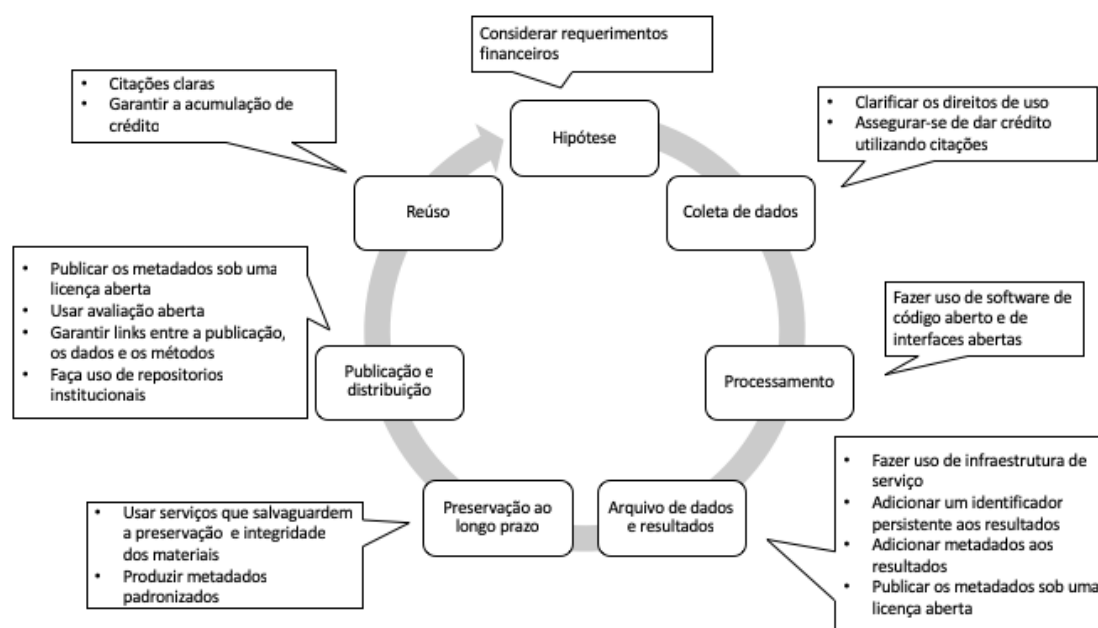
fenômenos, o que seria o paradigma teórico. Posteriormente, no paradigma computacional, a simulação de fenômenos complexos tornou-se a chave do desenvolvimento científico. A rigorosidade das simulações impulsionou o desenvolvimento da computação.

Finalmente, encontramos o processo científico baseado na exploração de dados e na gestão de grandes quantidades de dados. Esse último degrau da evolução dos paradigmas é conhecido como paradigma orientado por dados ou Quarto Paradigma. Considerando que esse paradigma não substitui completamente os anteriores, não está pensado como uma mudança completa. Porém, sim emergem novas práticas e ferramentas e existe uma incerteza para a adaptação ao novo fazer científico.

Consequente à ideia do Quarto Paradigma, “[...] os dados são a infraestrutura da ciência.” (TENOPIR et al., 2011, p. 1) surge o termo *e-Science*. Na *e-Science*, a forma de abordar os dados é completamente nova. As pesquisas podem começar “[...] pelos dados em vez de pelas hipóteses que levam à coleta dos dados” (DE ROURE, 2014, p. 234). Devido também ao avanço das ferramentas *web* os usuários “[...] podem acessar, minerar, explorar, reproduzir e disseminar sem custo os dados de pesquisa que estão livremente acessíveis.” (COMISSÃO EUROPEIA, 2016, p.4).

Nesse contexto, a ciência aberta constitui uma nova abordagem ao processo científico com o objetivo de extrapolar o conceito de *openness* a todos os estágios do ciclo de pesquisa (ver Figura 2).

Figura 2 Aplicação do conceito de *openness* nos diferentes estágios do processo de pesquisa.



Fonte: Open Science and Research Handbook¹⁹ (2014, p. 8, tradução própria)

- **Hipótese:** na hora de planejar a pesquisa e as hipóteses do projeto, é necessário considerar os custos derivados da mesma. Por exemplo, em projetos europeus é possível considerar até 5 % do total para gastos de gestão de dados de pesquisa. Na fase do planejamento da pesquisa e da elaboração das hipóteses, é necessário em muitas ocasiões se adaptar às chamadas de projetos temáticos disponíveis. Daí a necessidade de pensar na hipótese segundo o tipo de projeto e aporte financeiro disponível.
- **Coleta de dados:** a coleta de dados envolve o esclarecimento dos direitos de uso, quais dados podem ser coletados, de qual forma e qual será o uso posterior desses dados. Além disso, no caso de coleta de conjuntos já existentes, citar as fontes para garantir o conhecimento da proveniência dos dados.
- **Processamento:** para garantir a aplicação dos princípios da *openness* o processamento dos dados deve ser baseado em *software* de código aberto. Por exemplo, utilizar R Studio em vez de SPSS para análises estatísticas.
- **Arquivo de dados e resultados:** no arquivo de dados e resultados, é importante considerar repositórios institucionais ou temáticos que,

¹⁹ Open Science and Reserch Handbook:

<https://avointiede.fi/sites/avointiede.fi/files/openscience%20handbook.pdf>

idealmente, possuam alguma certificação de confiabilidade (p.ex. CoreTrustSeal²⁰). Além disso, atribuir um identificador persistente ao conjunto de dados para garantir, entre outros aspectos, sua encontrabilidade e citabilidade e com uma licença aberta para que possam ser reutilizados. O mesmo é válido também para os metadados. No caso específico dos metadados, muitas vezes recomenda-se o uso de uma licença de domínio público (p.ex. Creative Commons 0).

- **Preservação ao longo prazo:** o ponto anterior é válido também para a preservação ao longo prazo. Aqui, a ideia de *openness* traduz-se principalmente na padronização de metadados seguindo padrões gerais ou da área (*common standards*). Ainda, devem ser considerados os formatos dos arquivos e o armazenamento em um ambiente confiável.
- **Publicação e distribuição:** na publicação e distribuição de resultados de pesquisa é especialmente importante a identificação de todos os ativos/assets e o estabelecimento de enlaces entre a partes (aliás, a criação de uma publicação ampliada que é objeto desta tese). A *openness* faz referência também a como esses ativos/assets são avaliados, destacando a ideia de avaliação aberta e debate com a comunidade sobre os resultados alcançados.
- **Reuso:** descrever com os metadados e a documentação apropriados e atribuir um identificador persistente aos ativos/assets garantem não só a possibilidade de reprodutibilidade dos resultados, senão também servem como base para evitar práticas científicas inadequadas, para mostrar uma base sólida para a pesquisa e para dar crédito aos autores dos ativos/assets.

Aplicar o conceito da *openness* a cada um dos estágios do processo de pesquisa está diretamente relacionado com a estratégia de GDP planejada para a pesquisa. A GDP é entendida como “[...] um conjunto de práticas para lidar com a informação coletada e criada durante a pesquisa.” (HIGMAN; BANGERT; JONES, 2019, p. 2). A GDP está presente em todos os estágios da vida dos dados e cobre todos os processos de planificação, recopilação, seleção, transformação e arquivo de dados de pesquisa, sempre com o objetivo de

²⁰ CoreTrustSeal, Core Trustworthy Data Repositories: <https://www.coretrustseal.org/>

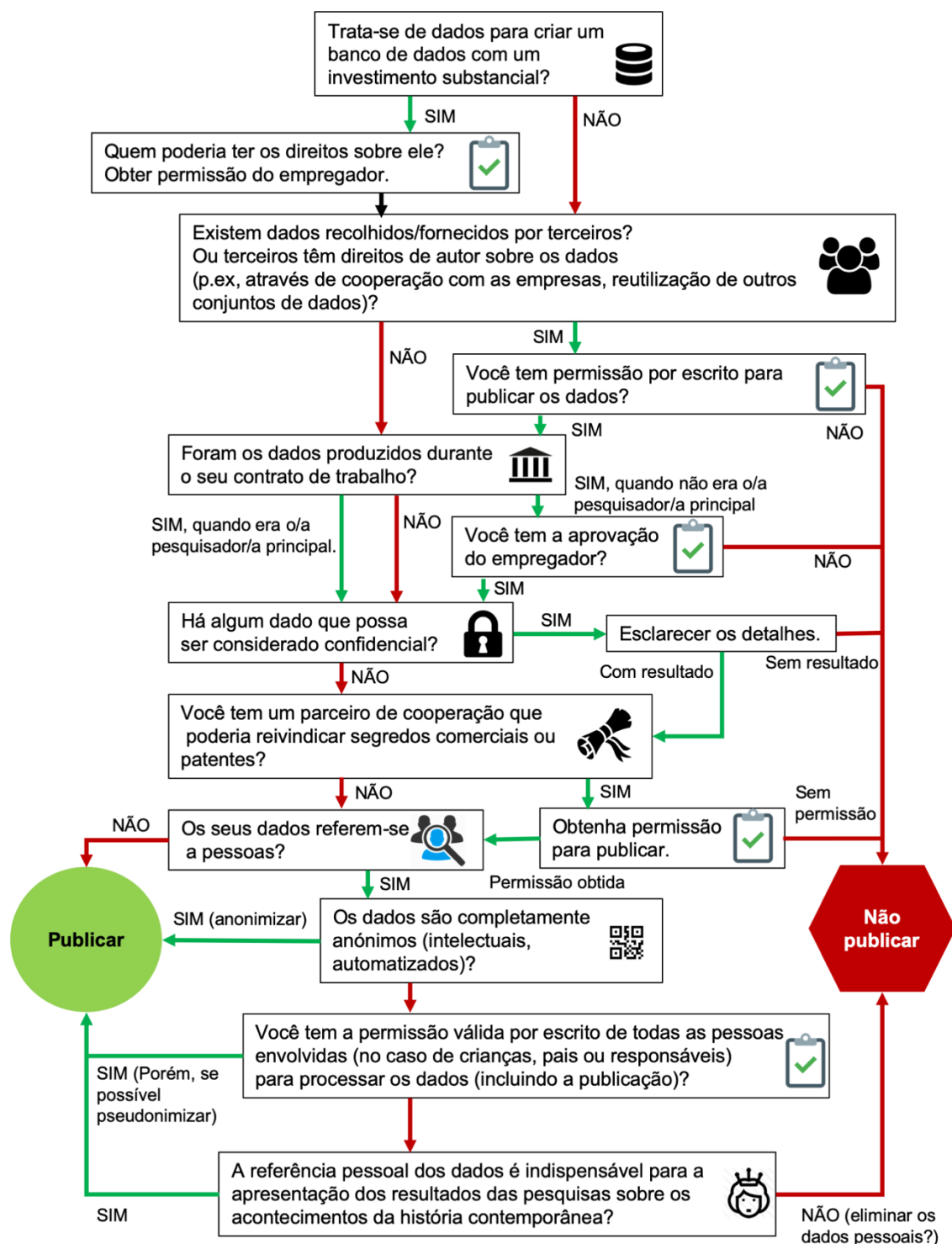
armazená-los no longo prazo e de forma independente ao criador dos dados, aliás, em uma infraestrutura externa (WHYTE; TEDDS, 2011).

As demandas relacionadas com a GDP variam muito de uma área de pesquisa para outra e também entre os atores envolvidos no processo. Os pesquisadores com frequência têm necessidades e ideias diferente às de uma agência de financiamento ou um parceiro de cooperação na indústria, dependendo da fase de pesquisa e do tipo de dados gerados.

Consequentemente, nem sempre é possível publicar de forma aberta os ativos/assets gerados em uma pesquisa. Às vezes outras formas de acesso são requeridas (p.ex., acesso restrito ou acesso após um determinado período de embargo no caso de pesquisas patenteáveis) ou a anonimização dos dados (p.ex., dados clínicos ou pessoais ou aqueles que possam ser inferidas atividades humanas). Nos casos em que os dados não possam ser disponibilizados de nenhuma maneira devido a diversas restrições recomenda-se publicar os metadados descritivos dos dados, incluindo a licença aplicada aos dados (uma licença não aberta).

Para facilitar o processo de tomada de decisão frente à publicação de dados, Schleußinger e Rex (2019) apresentam uma árvore de decisão que inclui perguntas tanto de aspectos legais quanto de aspecto éticos (ver Figura 3).

Figura 3 Publicação de dados? Os principais aspectos legais.



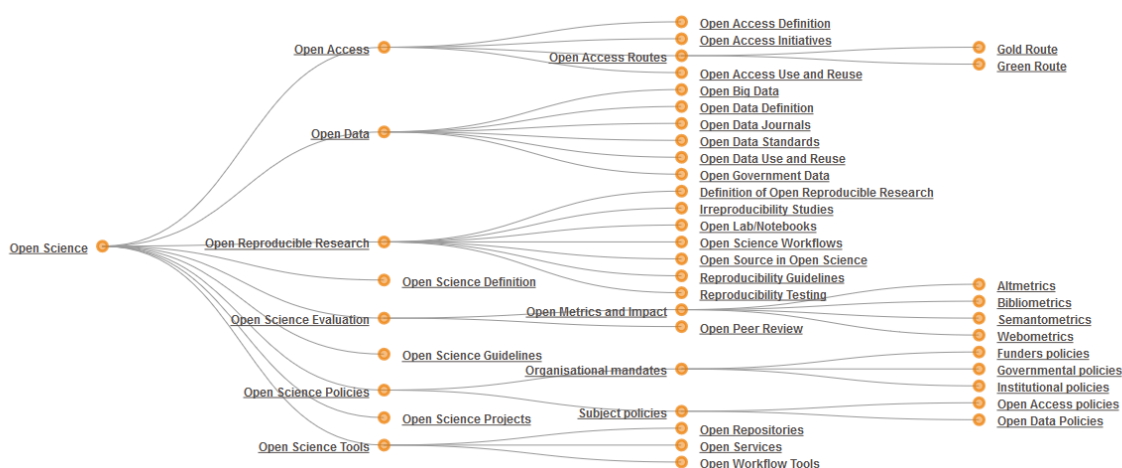
Fonte: Schleußinger e Rex (2019, tradução própria)²¹.

²¹ No original em alemão incluem-se referências ao direito alemão de autor e ao Regulamento Geral sobre a Proteção de Dados da União Europeia. Essas referências têm sido eliminadas na tradução já que as leis aplicáveis variam entre territórios e não se aplicam da mesma maneira.

Como apontam Higman *et al.* (2019), as estratégias de GDP coexistem com as práticas de ciência aberta e com o conceito dos princípios FAIR. Estes princípios advogam por um maior acesso aos ativos/assets de pesquisa e seus respectivos metadados tanto por humanos quanto por máquinas, pela utilização de padrões e formatos abertos e independentes (não proprietários) e pela documentação dos ativos/assets.

Por sua vez, segundo o projeto FOSTER Open Science (2018) as práticas de ciência aberta classificam-se em seis pilares fundamentais: o acesso aberto, os dados abertos, a pesquisa aberta reproduzível, a avaliação aberta da ciência, as políticas de ciência aberta e as ferramentas de ciência aberta (ver Figura 4). Parte destas práticas coincidem com as estratégias para uma adequada GDP como detalha-se a seguir.

Figura 4 Taxonomia da ciência aberta.



Fonte: FOSTER Open Science (2018).

- **Acesso aberto (*Open Access*):** refere-se ao acesso *online*, sem custos e, de preferência, sob licença aberta (p.ex. Creative Commons) a conteúdo científico revisado por pares. O acesso aberto tem seu foco principal nas publicações textuais e pode ser desenvolvido seguindo três vias principais: a via verde, a via dourada e a via diamante. A via verde implica a publicação em repositórios de um *pós-print* do texto. A via dourada consiste em publicação em periódicos de acesso aberto, mas, geralmente, envolve custos de publicação, conhecidos como *Article Processing Charges* (APC). A via diamante também implica publicação em periódicos de acesso aberto,

porém, sem custos de publicação, pois os periódicos estão subvencionados por uma universidade ou associação científica.

- **Dados abertos (*Open Data*):** dados disponíveis online, sem custos e de forma acessível que podem ser utilizados, reutilizados e distribuídos, sempre seguindo as indicações da fonte dos dados. Os dados abertos requerem a ampla documentação para garantir a adequada reutilização e podem ser publicados em repositórios de dados, em periódicos de dados (*data journals*) ou seguindo publicações como os artigos de dados (*data papers*). Em muitos casos, os dados devem passar por um processo de anonimização, pois podem conter conteúdos sensíveis (p.ex. histórias clínicas de pacientes). Atualmente, projetos como o Horizon 2020 seguem a seguinte premissa para os dados: “Tão abertos como seja possível, tão fechados quando necessário.” (*As open as possible, as closed as necessary*). Ainda, paralelamente, tem surgido a iniciativa de dados FAIR (*Findable – Encontráveis, Accessible – Acessíveis, Interoperable – Interoperáveis, Reusable – Reutilizáveis*) para garantir uma série de descrições dos dados e dos ambientes de publicação (WILKINSON *et al.*, 2016). Desta forma, cria-se uma categoria no contexto da Ciência Aberta aplicável também aos dados fechados.
- **Pesquisa reproduzível aberta (*Open Reproducible Research*):** implica a disponibilização não só dos resultados da pesquisa, senão também dos elementos experimentais para que a pesquisa possa ser reproduzida. Exemplos são os fluxos de trabalho que levaram à obtenção dos resultados ou os cadernos de laboratório. Em termos de ferramentas, nas Ciências Exatas e Computação têm se popularizado os documentos produzidos Jupyter notebooks, que contêm código (p.ex. Python ou R) e elementos de texto enriquecido (p.ex. parágrafos, equações e figuras).
- **Avaliação aberta da ciência (*Open Science Evaluation*):** consiste na avaliação dos resultados não baseada unicamente nas revisões por pares, mas também na contribuição da comunidade. Nesse sentido, também se propõem avaliações por pares abertas, onde tanto os nomes dos autores do texto quanto os nomes dos avaliadores são conhecidos. Além disso, identificam-se as falhas dos indicadores bibliométricos tradicionais na hora de avaliar o impacto de um texto e aparecem alternativas que avaliam o

impacto ao nível do artigo e não da revista. Estes indicadores alternativos também são aplicáveis à produção não textual.

- **Políticas de ciência aberta (*Open Science Policies*):** refere-se às diretrizes de boas práticas para aplicar a Ciência Aberta e atingir seus objetivos fundamentais. No contexto europeu, criou-se a Plataforma de Políticas para a Ciência Aberta (*Open Science Policy Platform* - OSPP) e já existem quatro (4) relatórios e oitenta e sete (87) recomendações sobre esse movimento. Com frequência, o que se identifica não é a falta de políticas senão a falta de ação na aplicação destas políticas (MÉNDEZ RODRÍGUEZ, 2019), devido a falta de comunicação e a diferenças de interesses entre pesquisadores e desenvolvedores de políticas públicas.
- **Ferramentas de ciência aberta (*Open Science Tools*):** abrange todas as ferramentas que ajudam o processo e as práticas da Ciências Aberta, desde as ferramentas para criar fluxos de trabalho até os repositórios abertos para depositar os ativos/assets de pesquisa.

A partir desta taxonomia percebemos que o movimento de ciência aberta atua como guarda-chuva e abrange muitas das práticas no panorama científico atual. Assim, identificam-se quatro perspectivas principais na ciência aberta (FRIESIKE *et al.*, 2015):

1. **Perspectiva filantrópica (*philanthropic perspective*):** centra-se na ideia da democratização da ciência e da pesquisa e de abrir a possibilidades de compartilhamento da produção. Exemplos desta perspectiva são os cursos massivos online abertos (MOOC) como forma de compartilhar publicamente o conhecimento universitário ou os periódicos científicos em acesso aberto, como aqueles indexado no Diretório de Periódicos em Acesso Aberto (DOAJ);
2. **Perspectiva reflacionária (*reflationary perspective*):** envolve o compartilhamento de conhecimento (ou de ativos/assets de pesquisa) desde os primeiros estágios do processo de pesquisa. O motivo desta ação é incrementar a criação de ideias dentro da comunidade científica, a recepção de comentários e sugestões (*feedback*) e o fomento dos ambientes colaborativos. Assim, surgem ambientes de publicação de *preprints*, como o

e-Prints in library & information science (e-LiS)²² ou o jornal *Research Ideas and Outcomes* (RIO)²³ que fornece um ambiente para publicação de resultados ao longo do processo de pesquisa (propostas de projetos, ideias de pesquisa, marcos de trabalho, dados, *software*, etc.);

3. **Perspectiva construtivista (*constructivistic perspective*)**: mantém a ideia de que se a ciência se abre, aparecem novas oportunidades para a criação de conhecimento tanto dentro das universidades e dos centros de pesquisa quanto nas empresas e nos novos modelos de negócio. Um exemplo é a utilização do *crowdsourcing* para a resolução de problemas, como apresenta Sarasua²⁴ (2014) para a interligação de dados em bibliotecas semânticas;
4. **Perspectiva exploratória (*exploitative perspective*)**: possui um foco na transformação do conhecimento científico teórico em conhecimento com uma orientação prática. Assim, podem ser estabelecidas relações entre indústria e universidade, como acontece nas incubadoras empresariais ou nos centros de gestão do conhecimento.

No âmbito acadêmico, principalmente se observam as três primeiras perspectivas e a quarta fica um pouco mais reservada para os contextos de inovação aberta. Dentro da própria comunidade científica, há diferentes correntes na hora de analisar o futuro da criação e disseminação de conhecimento científico no contexto da ciência aberta. Em uma revisão da literatura, Fecher e Friesike (2014) agrupam estas correntes em “escolas de pensamento” e identificam cinco delas: a escola pública, a escola democrática, a escola pragmática, a escola da infraestrutura e a escola das métricas. Estas escolas e suas implicações para as publicações ampliadas serão discutidas na seção 3.2. Porém antes é importante estabelecer as conexões entre as escolas de Fecher e Friesike (2014) e as perspectivas de Friesike *et al.* (2015) para observar os pontos de encontro e contextualizar mais o movimento da ciência aberta (ver Quadro 3).

²² Plataforma e-prints in library & information science: <http://eprints.rclis.org/>

²³ Research Ideas and Outcomes (RIO) Journal: <https://riojournal.com/>

²⁴ Apresentação de trabalho na conferencia Semantic Web in Libraries 2014: http://swib.org/swib14/slides/sarasua_swib14_14.pdf

Quadro 3 Matriz de comparação entre as perspectivas e as escolas de pensamento da ciência aberta.

		Escolas do pensamento da ciência aberta				
		Pública	Democrática	Pragmática	Da infraestrutura	Das métricas
Perspectivas na ciência aberta	Filantrópica	Aproximar a ciência e a pesquisa da sociedade.	Distribuição de conteúdo de forma não restritiva.			
	Refracionária			Abertura do processo de criação do conhecimento científico. Incremento dos debates sobre conteúdos.		Novas métricas para os resultados. Garantia do crédito e da autoria.
	Construtivista			Maior eficiência do trabalho científico.	Plataformas, ferramentas e novas vias para a criação colaborativa de conhecimento científico. Fusões tecnológicas e soluções inovadoras.	
	Exploratória			Descobertas científicas orientadas à aplicação na vida real.		

Fonte: Elaboração própria a partir de Friesike *et al.* (2015) e Fecher e Friesike (2014).

Nem todas as áreas do conhecimento abordam a ciência aberta e as suas implicações da mesma forma. Portanto, considera-se importante analisar as abordagens ao termo presentes na literatura segundo área de publicação e identificando a perspectiva que seguem (ver Quadro 4).

Quadro 4 Abordagens de ciência aberta por área de publicação.

Autor	Abordagem	Área de publicação	Perspectiva principal
(ALBAGLI; CLINIO; RAYCHTOCK, 2014, p. 435)	“Ciência aberta é hoje um termo guarda-chuva, que engloba diferentes significados, tipos de práticas e iniciativas, bem como envolve distintas perspectivas, pressupostos e implicações.”	Ciência da Informação	Todas
(ALI-KHAN <i>et al.</i> , 2018)	“A implementação da ciência aberta será o reconhecimento de uma gama maior de produções como material publicável por periódicos, agências de financiamento e instituições.”	Ciências da vida	Reflacionária
(ANGLADA; ABADAL, 2018, p. 293)	A ciência aberta é isto: “[...] uma mudança de paradigma na maneira de fazer ciência. A mudança não está no que é feito, senão em como é feito.”	Ciência da Informação	Todas
(DELFANTI, 2013, p. 27)	“A ciência aberta é um método para produzir conhecimento científico, disseminando os resultados e abrindo-os para a revisão da comunidade científica inteira, maximizando a circulação e o compartilhamento de informação e conhecimento.”	Biologia e Políticas Públicas	Reflacionária
(DESTRO BISOL <i>et al.</i> , 2014, p. 2)	“[...] um movimento internacional crescente em prol da ‘ciência aberta’, que significa tornar a publicação de conceitos científicos e dos dados em que se baseiam facilmente acessível a todos, bem como os procedimentos de compartilhamento de conjuntos de dados importantes.”	Antropologia	Filantrópica
(CAULFIELD; HARMON; JOLY, 2012, p. 11)	“[...] os pesquisadores são incentivados a compartilhar dados e disseminar conhecimento rapidamente (ou seja, a adotar um modelo de ciência aberta), a fim de promover o progresso científico, atingir metas humanitárias e (novamente) maximizar o impacto da pesquisa.”	Medicina	Filantrópica e reflacionária
(CHESBROUGH, 2015, p. 6)	“[...] o conceito aqui é que um acesso mais amplo, mais rápido e mais barato aos novos conhecimentos promoverá uma compreensão e utilização mais rápidas da ciência.”	Inovação e Gestão do Conhecimento	Filantrópica e reflacionária
(DE ROURE <i>et al.</i> , 2010, p. 2338)	“O movimento da ciência aberta, embora atualmente seja um nicho, defende vocalmente a colaboração aberta e distribuída em larga escala que é possibilitada pela disponibilização gratuita de dados, métodos e resultados na Web.”	Ciência da Computação	Filantrópica e reflacionária
(MASUZZO; MARTENS, 2017, p. 2)	“A ‘ciência aberta’ evoca muitos conceitos diferentes e abrange muitas frentes diferentes, desde o direito de ter livre acesso a publicações acadêmicas (denominado de ‘acesso aberto’), passando pela demanda por um		Filantrópica, reflacionária e construtivista

Autor	Abordagem	Área de publicação	Perspectiva principal
	engajamento público mais amplo (comumente chamado de ciência cidadã), até o desenvolvimento de ferramentas livres para colaboração e revisão aberta por pares (conforme implementado nas mídias sociais orientadas à ciência) plataformas).”		
(PONTIKA <i>et al.</i> , 2015, p. 1)	“[...]uma tendência universal que exige uma mudança no comportamento dos investigadores para conteúdos abertos e a adoção de um vasto leque de práticas e estratégias abertas relacionadas com todo o ciclo de vida da investigação.”	Tecnologias do conhecimento	Reflacionária
(RENTIER, 2016, p. 159)	“A plena abertura, pesquisabilidade, reprodutibilidade e controle pelos pares da investigação (‘ciência aberta’) só podem ser alcançados se o <i>software</i> de investigação utilizado for livre, aberto e totalmente transparente.”	Ciência da Informação	Filantrópica
(SCHMIDT <i>et al.</i> , 2018, p. 1)	“[...] A ciência aberta como paradigma para tirar partido da tecnologia digital para os processos de investigação e colaboração e para os tornar transparentes e compreensíveis, seguiu estas abordagens iniciais para o acesso aberto aos resultados da investigação.”	Ciência da Informação	Filantrópica
(VICENTE-SAEZ; MARTINEZ-FUENTES, 2018, p. 434)	“A ciência aberta é um conhecimento transparente e acessível que é partilhado e desenvolvido por meio de redes de colaboração.”	Negócios	Filantrópica e reflacionária

Fonte: Elaboração própria.

Principalmente, a ciência aberta é abordada em publicações em áreas que envolvem desafios globais, como Biologia ou Ciências da vida. Porém, também existe um interesse na análise desta mudança cultural na Ciência da Informação, a Comunicação e a Ciência da Computação. A ciência da computação costuma basear as práticas de ciência aberta no compartilhamento e na análise de grandes quantidades de dados. Por isso, em muitas ocasiões se estabelecem os conceitos de *e-Science* e ciência aberta como sinônimos ou se destacam apenas as práticas da primeira (OLIVEIRA; SILVA, 2016; ROSS; KRUMHOLZ, 2013; CORDEIRO *et al.*, 2013). Embora os termos não sejam estritamente sinônimos, é importante assinalar que as práticas da abertura e de compartilhamento de dados têm sido mais aprofundadas na literatura.

Observa-se também que as práticas de ciência aberta ainda se consideram extras e não estão normalizadas e inseridas na prática científica

comum (WATSON, 2015). Por isso, os organismos oficiais também têm procurado uma definição de ciência aberta (ver Quadro 4), que seja a mais abrangente possível e que sirva de base para o desenvolvimento de políticas públicas e de roteiros de atuação para governos e entidades (como, por exemplo, bibliotecas de pesquisa).

Quadro 5 Definições de ciência aberta em documentos oficiais.

Organismo ou projeto	Definição (originais em inglês, tradução própria)
(FOSTER OPEN SCIENCE, 2018)	“A prática da ciência de tal forma que outros possam colaborar e contribuir, onde dados de pesquisa, notas de laboratório e outros processos de pesquisa estejam livremente disponíveis, em termos que permitam a reutilização, redistribuição e reprodução da pesquisa e seus dados e métodos subjacentes.”
LIBER (2018, p. 6)	“A ciência aberta torna os resultados científicos publicamente disponíveis para que qualquer pessoa possa ler, compartilhar e desenvolver os resultados. Isso, por sua vez, impulsiona a colaboração e a inovação, e maximiza o potencial para resolver desafios globais, como mudanças climáticas e doenças.”
LERU (2018, p. 2-3)	“A ciência aberta, talvez mais propriamente designada por Investigação Aberta em inglês, representa uma mudança cultural na forma como os intervenientes nas comunidades de investigação, educação e intercâmbio de conhecimentos criam, armazenam e fornecem os resultados da sua atividade.” “A ciência aberta não é um dogma; é uma questão de maior eficiência e produtividade, mais transparência e uma melhor resposta às necessidades da investigação interdisciplinar.”
(COMISSÃO EUROPEIA, 2016, p. 35)	“A ciência aberta é para a ciência o que a web 2.0 era para as transações sociais e económicas: permitir que os utilizadores finais sejam produtores de ideias, relações e serviços e, ao fazê-lo, permitir novos modelos de trabalho, novas relações sociais e conduzir a um novo <i>modus operandi</i> para a ciência.”
(OECD, 2015, p. 7)	“O termo refere-se aos esforços de pesquisadores, governos, agências de financiamento de pesquisa ou da própria comunidade científica para tornar os resultados primários dos resultados de pesquisas financiadas publicamente - publicações e dados de pesquisa acessíveis ao público em formato digital, sem restrições ou com restrições mínimas como meio de acelerar a pesquisa.”

Fonte: Elaboração própria.

Estas definições focam, de forma clara, na mudança cultural, no novo *modus operandi* e no compartilhamento de ativos/assets de pesquisa. Governamentalmente, a Comissão Europeia vem trabalhando na definição de políticas públicas para a ciência aberta no nível europeu, mas com a visão de globalização científica e expansão dos diálogos com outros países. Também se observam movimentos em outros continentes, como o africano, por meio da

plataforma African Open Science Platform²⁵, para favorecer o intercâmbio e acesso à informação, especialmente em áreas mais desfavorecidas.

Por sua vez, LERU e LIBER têm apresentado roteiros de trabalho para as universidades e bibliotecas de pesquisa, respectivamente, nos próximos anos. Nos roteiros destacam-se as seguintes áreas:

- a comunicação científica (LERU, 2018; LIBER, 2018);
- infraestrutura para a pesquisa e a iniciativa EOSC (LERU, 2018; LIBER, 2018);
- dados FAIR (encontráveis, acessíveis, interoperáveis e reutilizáveis) (LERU, 2018; LIBER, 2018);
- competências para o trabalho com ciência aberta (LERU, 2018).
- integridade de pesquisa (LERU, 2018; LIBER, 2018);
- recompensas, reconhecimento e métricas alternativas (LERU, 2018; LIBER, 2018); e
- ciência cidadã (LERU, 2018; LIBER, 2018).

Tanto novos modelos para a comunicação científica como a infraestrutura para a pesquisa e a adequação dos princípios dos dados FAIR são pontos centrais de ambos roteiros de trabalho. As publicações ampliadas podem ser consideradas um desafio para cada um desses pontos, mas ao mesmo tempo podem servir como proposta para novos modelos de publicação. Além disso, como indicam Bardi e Manghi (2015) e Woutersen-Windhouver (2009), a publicação ampliada é um veículo de publicação de dados.

3.2 AS ESCOLAS DE PENSAMENTO DA CIÊNCIA ABERTA E A PUBLICAÇÃO AMPLIADA

Nos últimos anos, a ciência aberta tem sido muito abordada na literatura científica e tudo assinala que seguirá sendo nos próximos anos devido, principalmente, à incorporação dela na agenda política internacional. Como mencionado anteriormente, há diferentes movimentos dentro da comunidade científica na hora de analisar o futuro da criação e disseminação de

²⁵ African Open Science Platform: <http://africanopenscience.org.za/>

conhecimento científico, denominado “escolas de pensamento” por Fecher e Friesike (2014):

- **Escola pública:** parte do pressuposto de que a ciência deve estar disponível para o público. Esta escola abrange principalmente projetos de ciência cidadã. Em certas ocasiões, os projetos de ciência cidadã despertam esteticismo entre a comunidade científica por causa da qualidade dos resultados que possam ser obtidos; por isso, muitos projetos utilizam abordagens *top-down* onde um cientista profissional coordena as atividades. Se bem, também há casos de sucesso no trabalho com cientistas cidadãos (ou amadores). Por exemplo, em um artigo recentemente publicado no *The Astronomical Journal* se reconhece a função dos cientistas cidadãos na pesquisa: “O papel dos cientistas cidadãos se tornará provavelmente até mais crucial na detecção de exoplanetas (ou planetas extrassolares) de trânsito interessantes[...]” e “[...] incentivamos a prática de equipes científicas que se aproximem dos cientistas cidadãos para todos os futuros artigos de descoberta.” (FEINSTEIN *et al.*, 2019, p. 12). Além disso, também estão surgindo laboratórios cidadãos (ou *makerspaces* científicos) para o trabalho com propostas dos cidadãos para a ciência. Também, esta escola tenta ampliar a divulgação científica como prática a partir de *blogs* de pesquisa ou palestras de divulgação que atinjam comunidades fora do contexto científico.
- **Escola democrática:** parte do suposto de que o acesso ao conhecimento está distribuído de forma desigual. Principalmente, esta escola está preocupada com o acesso aberto a todos os ativos/assets de pesquisa, especialmente aqueles que têm sido produzidos com financiamento público. Desta forma se facilita o reuso em estudos da mesma área ou de áreas diferentes. Por exemplo, dentro do movimento de dados abertos, estabelece-se a necessidade de documentar os conjuntos de dados que se publicam, detalhando a coleta, o tipo de dados, a leitura dos dados feita para o estudo e outras possíveis áreas de aplicação.

Esta escola, também, foca no acesso aberto a publicações científicas como uma forma de contribuição ao espaço global de conhecimento, evitando barreiras temporais e financeiras para o acesso. Nesse sentido, tem surgido

iniciativas como o PlanS²⁶ para a mudança completa de publicações europeias ao acesso aberto em um prazo de um ano. Uma das maiores críticas que esse tipo de iniciativas recebe é a perda de liberdade acadêmica e o risco de perder reputação. Porém, como contra-argumento está pensar que os pesquisadores agora ficam forçados a escolher o periódico dependendo do fator de impacto e do quartil que o periódico possua nos *rankings* internacionais (LÓPEZ-BORRULL, 2019).

- **Escola pragmática:** parte do pressuposto de que a criação do conhecimento poderia ser mais eficiente se os cientistas trabalhassem em equipe. Dentro desta escola, considera-se a ciência “[...] um processo que pode ser otimizado, por exemplo, pela modularização do processo de criação de conhecimento, abrindo a cadeia de valor científica, incluindo o conhecimento externo e permitindo a colaboração por meio de ferramentas *online*.” (FECHER; FRIESIKE, 2014, p. 32). Em parte, uma das portas que a ciência aberta abre é a possibilidade de colaboração, sendo esta colaboração uma das bases do desenvolvimento científico (ADAMS, 2013). A colaboração permite agregar experiência, conhecimento e produtividade de diferentes fontes e gerar assim pesquisas mais sólidas e com menos pontos fracos. Para esta escola, as ferramentas da web 2.0 são cruciais, já que permitem o compartilhamento mais eficiente da informação.
- **Escola da infraestrutura:** parte do pressuposto de que a pesquisa eficiente depende das ferramentas e aplicativos disponíveis. O desenvolvimento de infraestrutura que permita desenvolver práticas de pesquisa na rede é o foco principal desta escola. Para poder compartilhar dados em formato aberto é necessário ter um repositório de dados (se possível confiável), para gerenciar citações em aberto²⁷ é necessário um gestor bibliográfico e para escrever de forma colaborativa é necessário um editor de texto web que possa ser utilizado por vários usuários simultaneamente. Estas ferramentas requerem a colaboração de bibliotecários e pesquisadores (HEY; HEY, 2006). Nesse

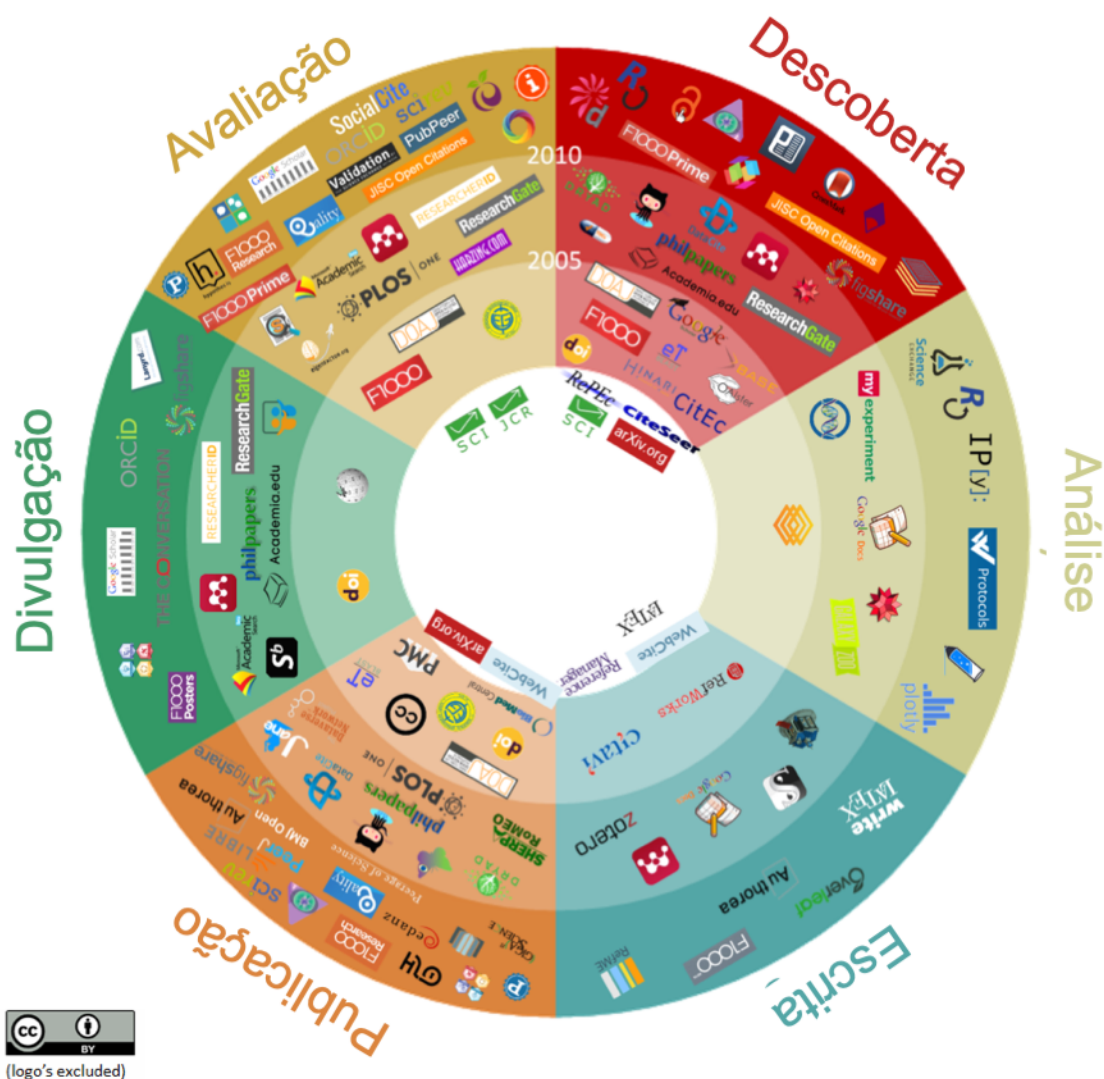
²⁶ Iniciativa iniciada por Science Europe para garantir o acesso aberto a todas as publicações que resultem de projetos de pesquisa financiados com fundos públicos.

²⁷ I4OC – Initiative for Open Citations: <https://i4oc.org/>

sentido, na biblioteca da Universidade de Utrecht (Países Baixos) foram analisadas 101 ferramentas utilizadas pelos pesquisadores ao longo do processo de pesquisa (ver Figura 5).

A Figura 5 agrupa estas ferramentas em seis grupos: descoberta (*discovery*), análise (*analysis*), escrita (*writing*), publicação (*publication*), divulgação (*outreach*) e avaliação (*assessment*). Estes grupos se ajustam as fases principais do ciclo de pesquisa científica.

Figura 5 101 Inovações em comunicação científica



Fonte: Kramer e Bosman (2015). Disponível em: <https://101innovations.wordpress.com/>.

Na fase de descoberta (Discovery) há ferramentas como Google Scholar e base de dados indexadoras tradicionais como Scopus e Web of Science; porém, também aparecem os fornecedores de DOI, como DataCite e Crossref. A expectativa principal nesta fase é o incremento da descoberta de dados. Na fase da análise (Analysis), utilizam-se ferramentas como SPSS e R e também aparecem algumas como ROpenSci na fase mais experimental. A análise deixa também ver um dos desafios do tratamento de dados de pesquisa: a reprodutibilidade. Na fase da escrita (Writing), junto com os processadores de textos como Word e OpenWriter, aparecem as ferramentas de escrita colaborativa, como GoogleDocs, Overleaf (para utilizadores de LaTeX) e Docear (que inclui uma ferramenta de gestão de referências). Na fase de publicação (Publication), aparecem vias mais tradicionais como o artigo em revistas indexadas; porém também novas revistas com revisão por pares abertas e avaliação coletiva como PeerJ, repositórios de *pre-prints* como arXiv. Nesta fase, existe um foco também na publicação de dados e no fomento do acesso aberto como padrão de publicação global. Na divulgação de resultados (Outreach), há um foco no uso de redes sociais acadêmicas e melhorar a conexão entre os diferentes perfis acadêmicos. Uma das formas mais impulsadas é o uso do identificador único para autores ORCID iD e as possibilidades de conexão entre DOI e ORCID iD para a visibilidade dos resultados. Além disso, está o uso de outras ferramentas colaborativas como Wikipedia e Wikidata para os dados. Finalmente, na fase de avaliação (Assessment), começa uma tentativa de ruptura com os indicadores bibliométricos como o fator de impacto para introduzir indicadores de métricas alternativas (p.ex. *altmetric score* da altmetric.com e métricas de nível de artigo (*article-level-metrics*) da PLOS).

Além das ferramentas para trabalho colaborativo, armazenamento e publicação durante o ciclo de pesquisa, estão as infraestruturas para a interação entre pesquisadores mais distantes geograficamente, como é o caso dos ambientes virtuais de pesquisa (*Virtual Research Environments* – VRE). Esta infraestrutura facilita a gestão e o compartilhamento de objetos de pesquisa (DE ROURE; GOBLE; STEVENS, 2009), incentiva os pesquisadores a trabalhar com os objetos existentes, é aberta e extensível,

e fornece uma plataforma de pesquisa-ação onde os objetos de pesquisa são a base dos trabalhos.

- **Escola das métricas:** parte do pressuposto de que as contribuições científicas hoje precisam de métricas de impacto alternativas. Devido ao uso da *web* como base do intercâmbio de conhecimento científico, ao surgimento das redes sociais *online* e ao uso de diferentes ativos/assets de pesquisa são necessárias formas alternativas de medição do impacto. Principalmente, métricas que estejam ligadas diretamente com o ativo/asset e não com o periódico ou com a plataforma onde foram publicados. Assim, obtém-se as métricas de nível de artigo (*article-level-metrics*) (FENNER, 2014).

Para o desenvolvimento destas novas métricas, a empresa Crossref utiliza o *software* Lagotto²⁸ que consegue rastrear o uso virtual e engajamento com os ativos/assets que possuam um DOI. Por sua vez, a empresa Altmetric.com gera o *altmetric score* para a medição do impacto alternativo.

Outro argumento neste sentido são as métricas alternativas que podem ser aplicadas a todos os ativos/assets de pesquisa e não só aos artigos publicados em revistas como acontece com os indicadores bibliométricos tradicionais (PIWOWAR, 2013; PRIEM; GROTH; TARABORELLI, 2012).

A possibilidade de uso destas métricas para a avaliação da produção científica ainda cria muitas dúvidas, por uma parte devido à falta de transparência a respeito da origem dos dados e, por outra parte, pela dificuldade de ter uma interpretação adequada dos indicadores. Em todo caso, as métricas alternativas estão presentes para avaliação nas agendas políticas, por exemplo no marco de trabalho Horizon Europe, a iniciativa que dará continuidade ao Horizon 2020 na União Europeia.

Estas cinco escolas fazem parte do movimento da ciência aberta, cujas mudanças na forma como os acadêmicos e os cidadãos interagem com a prática científica conduzem a uma mudança inevitável na comunicação acadêmica. Estas alterações não envolvem o desaparecimento dos artigos tradicionais, mas sim a sua reestruturação. Uma possível reestruturação é a publicação ampliada

²⁸ Lagotto: <http://www.lagotto.io/docs/crossref/>

e pode ser abordada por cada uma das escolas identificadas por Fecher e Friesike (2014) (ver Quadro 5).

Quadro 6 Escolas de pensamento da ciência aberta, gestão de dados e as implicações para publicações ampliadas.

Escola de pensamento	Objetivo central	Aspecto da GDP	Implicações para as publicações ampliadas
Pública	Tornar a ciência acessível para os cidadãos	Planejamento da coleta dos dados e do compartilhamento futuro segundo a(s) comunidade(s) envolvida(s).	Orientação dos ativos/assets de pesquisa a diferentes comunidades discursivas. Engajamento da comunidade científica profissional e amadora mediante o conteúdo audiovisual.
Democrática	Tornar o conhecimento científico disponível de forma livre para todos	Análise das licenças a serem aplicadas. Documentação adequada dos ativos/assets finais. Escolha dos repositórios apropriados para facilitar o acesso. Considerar ações para a melhorar a qualidade dos ativos/assets disponibilizados.	Publicação em plataformas abertas baseadas em <i>software</i> abertos. Transparência na publicação (partes narrativas, não textuais e <i>workflows</i> publicados simultaneamente). Necessidade de aplicação dos princípios FAIR em cada ativo/asset da publicação.
Pragmática	Abrir o processo de criação de conhecimento	Seleção de plataformas para o compartilhamento e trabalho com dados durante a pesquisa (p.ex. plataformas na nuvem ou laboratórios virtuais de pesquisa).	Entendimento da publicação ampliada como um objeto dinâmico do conhecimento que incluía comentários de outros agentes pré- e pós-publicação
Da infraestrutura	Criar plataformas, ferramentas e serviços para cientistas que estejam disponíveis de forma	Desenvolvimento e escolha de repositórios preferivelmente confiáveis. Análise do tipo de ativos/assets gerados e reutilizados para definir as estratégias de arquivo e curadoria.	Preparação de plataformas para a inserção das publicações ampliadas (p.ex. repositórios confiáveis). Modelagem das publicações para sua inserção.
Das métricas	Desenvolver um sistema de métricas alternativas para o impacto científico	Não se aplica	Medição do impacto de cada ativo/asset. Avaliação da publicação como um conjunto e análise da forma de dar crédito aos autores.

Fonte: Elaboração própria.

As publicações ampliadas, como objetos digitais compostos e dinâmicos do conhecimento, podem ser analisadas desde as correntes discursivas de cada uma das escolas de pensamento da ciência aberta.

A escola democrática estará presente ao sugerir a publicação ampliada como um mecanismo para a publicação de ativos/assets de pesquisa e seus metadados seguindo os princípios FAIR para garantir a encontrabilidade, o acesso, a interoperabilidade e a reutilização. Além disso, será estudado o papel dos identificadores persistentes na caracterização da ecologia complexa que as publicações ampliadas configuram.

A escola pragmática servirá para trazer a ideia de objeto dinâmico do conhecimento que considera as diversas interações discursivas ao longo do ciclo de pesquisa (p.ex. os comentários pré- e pós-publicação e os *workflows* de trabalho).

A escola de infraestrutura será considerada para a inserção das publicações ampliadas no ciclo de publicação científica. Neste ponto, a modelagem conceitual da publicação para seu tratamento dentro dos ambientes de publicação digital definirá quais requerimentos precisam-se para as entidades que conformam a publicação.

A escola pública —cujo principal objeto é ciência cidadã— fica fora do escopo desta tese. Porém, uma das implicações sociais das publicações ampliadas é garantir a transparência na pesquisa financiada com fundos públicos. Esta transparência é um primeiro passo para despertar o interesse pesquisador no público geral. Além disso, diversas pesquisas podem se beneficiar da participação da cidadania. Por exemplo, pesquisas que requeiram de anotações de imagens, pesquisas para a melhora de algoritmos de reconhecimento de caracteres ou de reconhecimento de sentimentos, entre outras. No contexto desta tese, o modelo apresentado Capítulo 6 integra a taxonomia dos roles de contribuição (CredIT) como um dos atributos da entidade autor. Estes roles podem ser adaptados também para reconhecer a contribuição da cidadania (ou de grupos de trabalho externos) na pesquisa.

A escola das métricas tampouco entra no do escopo desta tese. Não obstante, consideramos que futuras pesquisas sobre publicação ampliada podem versar a proposta de um marco de avaliação para esse tipo de publicações, bem como para aqueles ativos/assets não textuais e que não se

encaixam dentro dos indicadores bibliométricos tradicionais. Esse marco de avaliação deveria estar alinhado com propostas internacionais como a Declaration on Research Assessment²⁹ (DORA) e as iniciativas de citação de dados³⁰ e *software*³¹ da Future Research Communications and e-Scholarship (FORCE11).

²⁹ DORA: <https://sfdora.org/>. Declaração que reconhece a necessidade de melhorar as formas de avaliação dos resultados da investigação académica.

³⁰ FORCE 11. Declaração dos princípios de citação de dados: <https://www.force11.org/datacitationprinciples>

³¹ FORCE 11. Princípios da citação de software: <https://www.force11.org/software-citation-principles>

4 SEMIÓTICA PEIRCEANA E MULTIMODALIDADE NA PESQUISA CIENTÍFICA

A palavra “Semiótica” vem do grego para “sintoma” (*sēmeiōtikós* ou *sēmeiōtikḗ*). O matemático, lógico e filósofo Charles Sanders Peirce (1839-1914) fundou a Semiótica como “doutrina formal dos signos”. Seu contemporâneo Ferdinand de Saussure (1857-1913) fundou a semiologia como escola europeia da Semiótica com o objetivo de entender a constituição dos signos e as leis que os governam. “A Semiótica cobre o ciclo completo de um signo desde a sua criação, através do seu processamento, até seu uso, com maior ênfase no efeito dos signos.” (LIU, 2000, p. 13).

Nas dimensões da linguagem, distinguem-se três ramos principais: a sintática, a semântica e a pragmática. Segundo Morris (1985), a pragmática lida com a origem, os usos e efeitos dos signos no ambiente no qual acontecem; a semântica trata a significação dos signos em todos os modos de significação e a sintática trata a combinação dos signos com independência da sua significação ou sua relação com o ambiente no qual acontecem.

A Semiótica serve como base para ampliar a compreensão dos fenômenos complexos que acontecem dentro das ciências e da produção de conhecimento científico. Diferentes autores da Ciência da Informação têm analisado os aspectos semióticos em processos de busca da informação e da topografia do ciberespaço (MONTEIRO, 2007), de preservação e curadoria digital (BECKER, 2018; ABRAMS, 2015; ABRAMS; CRUSE; KUNZE, 2009). Inclusive, a partir da afirmação de Liu (2000), é possível estabelecer uma analogia com o ciclo de vida dos dados de pesquisa e a GDP.

Entendemos que os aspectos semióticos devem ser considerados ao estudar as publicações científicas ampliadas, dada a combinação de formatos e a consequente hibridação de linguagens presentes nelas. Essa combinação é motivada pelo fomento da publicação dos ativos/assets de pesquisa e os procedimentos nas diversas etapas do processo de pesquisa.

4.1 NOTAS SOBRE SEMIÓTICA PEIRCEANA

A noção de signo é essencial na Semiótica. Peirce parte da ideia de que qualquer coisa pode ser um signo, pois qualquer coisa observável remete a outro fenômeno (mental ou observável). Um signo “[...] é algo que representa para alguém outra coisa em algum aspecto ou capacidade.” (LIU, 2000, p. 13).

Para Peirce, um signo pode ser “[...] qualquer pintura, diagrama, grito natural, dedo apontando, piscadela, mancha em nosso lenço, memória, sonho, imaginação, conceito, indicação, ocorrência, sintoma, letra, numeral, palavra, sentença, capítulo, livro, biblioteca.” (SANTAELLA, 2001, p. 39). Também, Buckland (1991) define qualquer coisa que possa ser informativa como “informação como coisa”.

O signo para Peirce pode ter naturezas diversas e podemos considerá-lo uma estrutura complexa constituída por três elementos: fundamento, objeto e interpretante. Cada elemento têm uma qualidade: o fundamento possui a qualidade da possibilidade; o objeto, a da existência; e o interpretante, a da lei do pensamento.

O fundamento é uma propriedade ou caráter ou aspecto do signo que o habilita como tal. **O objeto** é algo diferente do signo, algo que está fora do signo, um ausente que se torna imediatamente presente a um possível intérprete graças à mediação do signo. **O interpretante** é um signo adicional, resultado do efeito que o signo produz em uma mente interpretativa, não necessariamente humana, uma máquina, por exemplo, ou uma célula interpretam sinais. (SANTAELLA, 2001, p. 43).

A relação entre estes três elementos varia segundo o contexto, a cultura ou a língua. Por exemplo, a palavra “rato”, embora tenha a mesma ortografia em espanhol e em português e praticamente a mesma pronúncia, significa “espaço de tempo” em espanhol e “animal roedor” em português. Consequentemente, dizer “tengo un rato” (espanhol) e “tenho um rato” (português) gerarão interpretações completamente distintas.

Ainda, Peirce considera que em todos os fenômenos existem três categorias presentes e os eventos semióticos podem ser estudados seguindo as categorias primeridade, secundidade e terceridade, inicialmente denominadas qualidade, relação e representação.

Peirce concluiu que tudo o que aparece à consciência assim o faz por meio da gradação de três elementos: 1) qualidade ou sentimento (primeiridade); 2) reação (secundidade); 3) mediação (terceiridade). Resumidamente, todo

fenômeno apresenta três elementos: a qualidade, o fato atual e a abstração. (MONTEIRO; FIDENCIO, 2013, p. 44).

Na primeridade encontram-se fenômenos de difícil análise. Peirce refere-se à primeridade como pura liberdade, espontaneidade, possibilidade de que aconteça algo novo. Na primeridade incluem-se as casualidades sem conexão com o redor. Representa a possibilidade futura de formar parte de uma classificação determinada e estabelecer uma interrelação com outros signos possíveis.

A secundidade explica a experiência, ou seja, todos os tipos de fenômenos em experiências diretas. “É a compulsão, a absoluta coação sobre nós de alguma coisa que irrompe o fluxo de nossa quietude, obrigando-nos a pensar de modo diferente daquilo que estivemos pensando, que constitui a experiência.” (SANTAELLA, 1983, p. 49). Na secundidade, abordamos os fenômenos da primeridade como um fato. “É para nós um simples ‘fato bruto’, como parte do nosso mundo físico, o é uma imaginação ou em pensamento na mente.” (MERRELL, 2001, p. 3).

A terceridade reúne os fenômenos das duas categorias anteriores para que sejam inteligíveis.

Finalmente, terceridade, que aproxima um primeiro e um segundo numa síntese intelectual, corresponde à camada de inteligibilidade, ou pensamento em signos, através da qual representamos e interpretamos o mundo. Por exemplo: o azul, simples e positivo azul, é um primeiro. O céu, como lugar e tempo, aqui e agora, onde se encarna o azul, é um segundo. A síntese intelectual, elaboração cognitiva —o azul no céu, ou o azul do céu—, é um terceiro. (SANTAELLA, 1983, p. 51).

Esta categoria define-se mediante a mediação, a transformação e a evolução ou crescimento vital. A mediação envolve a interrelação de duas entidades por meio de uma terceira entidade mediadora. Essa entidade mediadora atua como uma substância catalizadora em certas reações químicas. “Um signo media entre um objeto e um interpretante e incorpora também a quem esteja interpretando o signo.” (MERRELL, 2001, p. 3). A terceridade serve como tradução entre uma entidade Semiótica em outra, é uma transformação. Consequentemente, a terceridade determina o desenvolvimento vital dos signos. É um processo de criação que inclui a passagem da variedade para a uniformidade (do caos à ordem). “Com a passagem da variedade para a uniformidade, quero dizer que a variedade ao ser multiplicada quase em todos os departamentos da experiência

mostra uma tendência a formar hábitos.”³² (PEIRCE, 1958 CP 6.97, 1903, tradução própria).

Seguindo as categorias fenomenológicas, Santaella (2005) propõe três matrizes que irão sustentar a variedade e multiplicidade das formas de linguagens puras: sonora (ou virtual), visual e verbal. Fazendo uma analogia entre a Semiótica peirceana e a teoria de Santaella, a linguagem sonora pertence à primeridade e envolve ao ícone, a linguagem visual encontra-se na secundidade e envolve ao índice e a linguagem verbal está na terceridade e envolve ao símbolo. Assim, a classificação peirceana dos signos resume-se no seguinte quadro.

Quadro 7 Classificação dos signos³³.

	Relação entre signo e meio. Quando o signo serve como meio para se referir a algo.	Relação entre signo e objeto. O signo é introduzido como meio para se referir a um objeto.	Relação entre signo e interpretante. Único elemento para o qual o signo tem um significado.
	Fundamento (1º)	Objeto (2º)	Interpretante (3º)
Possibilidade (1º)	Quali-signo	Ícone	Rema
Existência (2º)	Sin-signo	Índice	Dicente
Lei (3º)	Legi-signo	Símbolo	Argumento

Fonte: Ampliação a partir de (SANTAELLA, 1983, p. 62).

- **Quali-signo:** o signo é em si uma qualidade. Refere-se a uma qualidade intrínseca do signo. Por exemplo: a alegria ou a luz.
- **Sin-signo:** é uma coisa ou evento real, existente e concreto. Por exemplo: uma palavra (essa palavra em particular e especificamente).
- **Legi-signo:** é uma lei em um signo, resultante de uma convenção social. Por exemplo: a cor preta que determina o luto.
- **Ícone:** um signo que imita ao objeto. O importante são os caracteres que possui, independentemente de se o objeto existe ou não existe. Por exemplo: um desenho de um unicórnio.

³² Tradução própria do original em inglês: “By the passage of variety into uniformity, I mean that variety upon being multiplied almost in every department of experience shows a tendency to form habits”.

³³ A indicação dos numerais no quadro serve para indicar a relação com a primeridade, secundidade e terceridade.

- **Índice:** um signo que remete ao objeto que denota porque está afetado diretamente por esse objeto. Por exemplo: a fumaça como indício de um churrasco.
- **Símbolo:** um signo que se refere ao objeto que denota em virtude de uma lei. Por exemplo: os sinais de trânsito.
- **Rema:** um rema é um signo que talvez proporcione uma informação para seu interpretante; porém, não se interpreta a informação que proporciona. Por exemplo: um conceito.
- **Dicente:** é um signo apto para afirmar algo, valorar ou para tomar decisões ou ações dos interpretantes a partir do objeto. Peirce os considera como uma proposição ou quase-proposição. Por exemplo: uma pintura (o tipo de traço, as cores escolhidas, a temática, etc.) que permite julgar ao pintor ou à pintora.
- **Argumento:** é a explicação total e racional de tudo o que conforma o signo. O argumento é a conexão completa dos signos, necessariamente verdadeira. Com frequência expõe-se como exemplo o argumento silogístico.

Ainda, Peirce define a semiose como a conversão de signos em signos em um processo de conversão infinita. Especificamente, “[...] uma ação, ou influência, que é, ou envolve, uma cooperação de três sujeitos, tais como um signo, seu objeto e seu interpretante, esta influência tri-relativa não é resolúvel em ações entre pares.”³⁴ (PEIRCE, 1932 CP 5.484, tradução própria). O interpretante pode ser tanto o terceiro elemento da relação triádica quanto o primeiro da seguinte triada.

Um signo, portanto, é um objeto que está relacionado com seu objeto por um lado, e com um interpretante, por um outro, de modo que leva ao interpretante a uma relação com o objeto que corresponde a sua própria relação com o objeto. (PEIRCE, 1932 CP 8.332, tradução própria³⁵).

Liu (2000) aponta que existem quatro características para descrever a semiose: (1) é universal e aplicável a qualquer tipo de atividade de processamento de signos; (2) é um processo capaz de identificar qualquer coisa presente seguindo um critério específico ou uma norma; (3) é um processo recursivo (um signo pode ser visto como um referente em outro processo de signos, bem como um interpretante ou um referente podem ser um signo); (4) é

³⁴ Tradução própria do original em inglês: “[...]an action, or influence, which is, or involves, a cooperation of three subjects, such as a sign, its object, and its interpretant, this tri-relative influence not being in any way resolvable into actions between pairs”.

³⁵ Tradução própria do original em inglês: “A sign therefore is an object which is in relation to its object on the one hand and to an interpretant on the other, in such a way as to bring the interpretant into a relation to the object, corresponding to its own relation to the object.”

possível fazer qualquer coisa não presente identificável. Podemos dizer assim que a semiose é um processo dependente do sujeito. Está relacionada com o interpretante que pode ser um indivíduo, um grupo social, uma comunidade científica ou qualquer comunidade que compartilhe um determinado conhecimento e esteja regulada por umas determinadas normas.

Como indicado no início do capítulo, as dimensões da Semiótica têm sido tradicionalmente a sintática, a semântica e a pragmática. A primeira trata as estruturas; a segunda, os significados; e a terceira, o uso dos signos e a relação desses signos com outros. Stamper (1973) adiciona outras três dimensões: a física, a empírica e o mundo social. A física preocupa-se com os aspectos físicos dos signos no nível de sinais e propriedades físicas como forma, tamanho, velocidade, aceleração ou fonte. A empírica estuda as propriedades estatísticas dos signos, tais como padrões, ruído ou entropia, quando são utilizados diferentes meios ou aparelhos. Finalmente, a dimensão do mundo social serve para o estudo dos efeitos do uso dos signos nos assuntos humanos.

Seguindo estas seis dimensões, Liu (2000) propõe um marco de trabalho semiótico para abordar a Semiótica nos sistemas da informação (ver Quadro 8).

Quadro 8 O marco de trabalho semiótico.

Funções de informação humana	MUNDO SOCIAL: crenças, expectativas, funções, compromisso, contratos, lei, cultura...
	PRAGMÁTICA: intenções, comunicações, conversações, negociações...
	SEMÂNTICA: significados, proposições, validade, verdade, significação, denotação...
Plataforma de Tecnologias da Informação	SINTÁTICA: estrutura formal, linguagem, lógica, dados, dedução de registros, <i>software</i> , arquivos...
	EMPÍRICA: padrão, variedade, ruído, entropia, capacidade do canal, redundância, eficiência, código...
	MUNDO FÍSICO: sinais, rastro, distinções físicas, hardware, densidade, velocidade, economia...

Fonte: Liu (2000, p.27, tradução própria).

- **Física (ou mundo físico):** “Um signo em uma forma física é um fenômeno.”³⁶ (LIU, 2000, p. 27). Pode ser tanto um signo em movimento (sinal) quanto

³⁶ Tradução própria do original em inglês: “A sign in a physical form is a phenomenon”

estático (marca). Um signo possui diferentes propriedades físicas tais como forma, tamanho, contraste, intensidade, amplitude, fonte, destino, entre outras. Estas propriedades são medíveis mediante métodos da física e da engenharia. No mundo físico, um banco de dados é simplesmente uma combinação de *tokens* que “[...] podem ser armazenados, movimentados, para entrada (*input*), saída (*output*) e exibição (*display*)”³⁷ (LIU, 2000, p. 28).

- **Empírica:** a dimensão empírica estuda as propriedades estatísticas dos signos, sendo o objeto de estudo uma coleção de sinais ou marcas. A informação a partir da perspectiva empírica considera-se “[...] um fluxo de sinais que devem ser transportados de um local para outro, independentemente do seu significado.”³⁸ (LIU, 2000, p. 28). Essa perspectiva está presente, entre outras, na teoria matemática da comunicação proposta por Shannon e Weaver (1949). A codificação da informação deve incluir o menor número possível de sinais portadores para atingir uma eficiência na transmissão.
- **Sintática:** a dimensão sintática foca nas regras para compor signos complexos a partir de signos simples. A codificação da informação segue uma estrutura determinada, “[...] um signo complexo, uma palavra, uma expressão matemática ou uma frase podem estar compostos de algumas partes mais básicas seguindo as regras.”³⁹ (LIU, 2000, p. 29). Estas regras são denominadas “gramática”.
- **Semântica:** a dimensão semântica analisa a relação entre signos e aquilo ao que se referem. O significado de um signo está estritamente relacionado com o contexto social. A semântica está presente em todos os estágios do processamento do signo: o estado de ser um signo, o interpretante e o fato de denotar o significante (MORRIS, 1985). Em termos funcionais, ao empregar um signo, o primeiro objetivo é exprimir um significado; o segundo objetivo é transmitir uma determinada intenção pragmática; o objetivo final é produzir um efeito no nível social (LIU, 2000, p. 31). Por exemplo, se

³⁷ Tradução própria do original em inglês: “[...] which can be stored, moved around, for input, output and display”.

³⁸ Tradução própria do original em inglês: “[...] stream of signals which must be transported from one location to another, regardless of what they mean”.

³⁹ Tradução própria do original em inglês: “[...] a complex sign, a word, a mathematical expression., or a sentence can be composed of some basic parts according to the rules”.

recebemos um *e-mail* do(a) nosso(a) orientador(a) dizendo que “há uma chamada de artigos aberta no periódico XY”, entendemos facilmente o significado da frase e a intenção de transmitir a informação. Porém, o efeito da frase não acaba aqui, o(a) nosso(a) orientador(a) quer que escrevamos um texto e submetamos para essa chamada.

- **Pragmática:** os signos que possuem um significado são utilizados com uma intenção determinada, por exemplo, para a comunicação. “A pragmática, no caso do uso intencional dos sinais, é um ramo da Semiótica preocupado com as relações entre sinais e agentes de comportamento.”⁴⁰ (LIU, 2000, p. 31). Porém, estas relações podem mudar segundo a comunidade discursiva. Assim, fatores como o grupo social, a experiência pessoal, as expectativas ou o contexto no qual acontece comunicação afetam (e as vezes dificultam) a interpretação de um signo. Consequentemente, a pragmática deve considerar fatores como o emissor, o receptor, e seus estados psicológicos (desejos, crenças, etc.), a intenção, o propósito, o tema, a data ou a localização.
- **Mundo social:** cada conversação entre humanos supõe mudanças no nível social (LIU, 2000). Em um ato de fala há inicialmente uma intenção, depois direciona-se o discurso para alguém e existe uma expectativa de que esse alguém responda independentemente do contexto da conversação. Esses processos de interação estão regidos por normas, sejam essas normas perceptivas, cognitivas, avaliativas ou comportamentais (LIU, 2000).

Seguindo estas dimensões, a Semiótica serve de guia na compreensão do funcionamento dos sistemas de informação. Desde o ponto de vista operacional, serve de guia na análise de processos comunicativos simples.

Para exemplificar, analisamos, seguindo estas dimensões e de forma simplificada, um processo de coleta, análise e publicação de dados.

No nível físico, os aparelhos de medição devem estar conectados e corretamente configurados.

⁴⁰ Tradução própria do original em inglês: “Pragmatics, in such a case of the purposeful use of signs, is a branch of semiotics concerned with the relationships between signs and behaviour agents.”.

No nível empírico, os dados coletados devem ser transformados em elementos binários que possam ser arquivados em um banco de dados e transmitidos entre computadores.

No nível sintático, o formato dos dados coletados e o formato aceito pelo *software* de análise deve ser o mesmo, ou seja, seguem as mesmas regras.

No nível semântico, os vocabulários escolhidos para a descrição e enriquecimentos dos dados, bem como a documentação realizada sobre o processo de coleta, devem possuir um significado dentro da comunidade de pesquisa que lide com esse tipo de dados. Também, devem fomentar a conexão com outras entidades que sigam os mesmos padrões.

No nível pragmático, há uma intenção na publicação dos dados (p. ex. depositando os dados em um repositório). O criador dos dados (emissor) espera que o leitor dos dados (receptor) os observe, analise e, se descobre algum erro, melhore ou aplicação em outra área do conhecimento, reutilize os dados.

No nível social, se o leitor dos dados os reutiliza para outra pesquisa, deve citar os dados apropriadamente para atribuir e dar crédito ao trabalho desenvolvido pelo criador dos dados.

Desde o ponto de vista do criador dos dados, entende-se que o propósito final ao publicar os dados, além de passar credibilidade sobre o estudo realizado e compartilhar os dados por questões de transparência, é receber uma atribuição pela criação dos dados e que os dados sejam reutilizados (e citados) em futuras pesquisas.

4.1.1 AS MATRIZES DA LINGUAGEM

A teoria de Peirce deve interagir com outras teorias mais específicas, principalmente para adaptá-la à era tecnológica, baseada na *web* e nos meios digitais. Nesse contexto, Santaella (2001) define a matriz como o lugar onde algo se gera e é preciso objetivar a origem das múltiplas linguagens existentes, cuja base é a percepção humana, e trasladar essa classificação aos ambientes da hipermídia.

A linguagem sonora utiliza a percepção auditiva e combina os elementos para formar unidades mais complexas: sua propriedade é a sintaxe. As linguagens sonoras podem acontecer por sintaxes naturais, por aquelas produzidas por corpos sonoros ou pelo resultado de convenções musicais e rítmicas. A linguagem visual utiliza a percepção visual e caracteriza-se pelo aspecto exterior dos corpos: sua propriedade é a forma. A forma pode ser abstrata, figurativa ou simbólicas. A linguagem verbal utiliza a faculdade de verbalização como forma de mediação. Organiza sequencialmente a capacidade discursiva: sua propriedade é o discurso. O discurso pode ser descritivo, narrativo e dissertativo.

Quadro 9 Exemplos resumidos das matrizes da linguagem segundo as categorias semióticas.

	Primeiridade	Secundidade	Terceridade
Sonora	Vento, chuva, tosse	Buzina, apito	Ritmo, melodia
Visual	Um quadro de Pollock	Uma imagem de um objeto	Uma expressão matemática ou um alfabeto
Verbal	Descrições em um livro de instruções	Narração de uma história	Argumentação em um texto científico

Fonte: Elaboração própria a partir de Santaella (2001).

Porém, a existência e isolamento de uma matriz da linguagem carece de sentido. As linguagens são híbridas, sendo a hipermídia a mais híbrida das linguagens. Dessa hibridização, tão necessária para as novas formas de comunicação, surge a seguinte classificação:

- **Sonoro-verbais (orais):** canção, lição universitária, *speech*;
- **Sonoro-visuais:** música instalação, *performance* sonora, teatro instrumental;
- **Visuais-sonoras:** arquitetura, computação gráfica, dança;

- **Visuais-verbais:** escritura, pictografia, publicidade impressa, quadrinhos, jornal;
- **Verbo-sonoras:** fala, literatura oral, poesia sonora;
- **Verbo-visuais:** gestualidade, mímica, performance e happening; e
- **Verbo-visuais-sonoras:** cinema, televisão, vídeo, materiais audiovisuais.

A virtualização e digitalização de recursos na Internet vão unidas a uma multiplicidade de linguagens no ciberespaço. Portanto, a digitalização envolve uma combinação das linguagens, mas também “[...] permite a organização reticular dos fluxos informacionais em arquiteturas hipertextuais.” (SANTAELLA, 2001, p.393). Não estaremos mais frente à estrutura linear e sequencial e sim frente à uma estrutura não sequencial e interativa com uns nexos (ou nós) que permitem ao leitor escolher entre diferentes itinerários.

Nesse contexto, esse capítulo pretende apresentar uma abordagem Semiótica às publicações científicas baseadas na multimodalidade da pesquisa científica, em especial no conjunto de signos distribuídos no entorno *web* e sua representação. Inicialmente, descrevemos o hipertexto e multidimensionalidade que regem na *web*. Seguidamente, apontamos as características da multimodalidade próprias da publicação ampliada e finalizamos com uma descrição das entidades extraídas dos ambientes para compor uma publicação ampliada.

4.2 A MULTIDIMENSIONALIDADE E O HIPERTEXTO

As estruturas hipertextuais iniciam-se antes do surgimento do ciberespaço. Porém, é na era digital, como base das conexões *web*, quando elas atingem sua maior importância. Antes do ciberespaço, escritores como Julio Cortázar forneceram opções de leitura não lineares para suas obras. É o caso de *O Jogo da Amarelinha*, onde podemos optar entre duas leituras ao início do livro:

O primeiro livro pode ser lido na forma corrente e termina no capítulo 56, ao término do qual aparecem três vistosas estrelinhas que equivalem à palavra ‘Fim’. Assim, o leitor prescindirá sem remorsos do que virá depois.

O segundo livro pode ser lido começando pelo capítulo 73 e continua, depois, de acordo com a ordem indicada no final de cada capítulo. (CORTÁZAR, 1963).

Utilizando como base as descrições que anos antes havia realizado Vannevar Bush (1945) no seu trabalho *As we may think?*⁴¹, o termo hipertexto foi introduzido por Theodor Nelson em 1965 (DE SALVADOR AGRA, 2016). Nelson define hipertexto como:

Uma escritura não sequencial, um texto que bifurca, que permite que o leitor escolha e que se leia melhor em uma tela interativa. De acordo com a noção popular, trata-se de uma série de blocos de texto conectados entre si por nexos que forma diferentes itinerários para o usuário. (NELSON, 1987, p. 2).

Ainda, Nelson apresenta a ideia visionária de um enorme sistema de hipertexto conhecido como “Projeto Xanadú”. Com ele, concebia o desenvolvimento de uma *interface* para computadores e o documento global. O modelo de documento baseia-se em uma presunção de mudança e reutilização contínua, já que os documentos incluem automaticamente as fontes das que proveem. Ainda, os nexos seriam estáveis e independentes das versões e o autor poderia ver o original de cada citação feita em um documento. Nelson inaugurou elementos das tecnologias digitais existentes na atualidade, os nexos estáveis não seriam diferentes dos atuais identificadores persistentes do tipo DOI e com os originais de cada citação estaríamos frente a implementações das tecnologias da *web* semântica para um documento (p.ex. páginas *web* atuais implementadas utilizando a base do conhecimento da BBC⁴² —ontologias e LD— como BBC Education⁴³).

Naranjo (2010) aponta que a partir da Semiótica é possível classificar o hipertexto como texto, mas supõe uma atualização comunicativa de códigos ou estruturas de significação. As condições para a existência de um texto são “[...] instância emissora, intencionalidade comunicativa, proposta de significado, situação ou contexto comunicativo e materialidade significante.” (NARANJO, 2010, p. 100). Embora seja considerado um texto, existem novas vias para enriquecer a interpretação de um hipertexto. O hipertexto deve criar uma proposta própria de significado. Os *links* externos e internos fortalecem “[...] a competência interpretativa do usuário.” (NARANJO, 2010, p.101). Nos ambientes digitais, os documentos são escritos na linguagem de hipertexto. “O

⁴¹ As we may think?: <https://www.theatlantic.com/magazine/archive/1945/07/as-we-may-think/303881/>

⁴² BBC Ontologies: <https://www.bbc.co.uk/ontologies>

⁴³ BBC Education: <https://www.bbc.com/bitesize>

hipertexto permite a interatividade, a navegação contígua sem limites de espaço físico e de margens.” (MONTEIRO, 2000, p. 28).

Para Lévy (1993, p. 15), o modelo do hipertexto é caracterizado mediante seis princípios abstratos:

1. **Princípio de metamorfose:** A rede hipertextual está em constante construção e renegociação. Ela pode permanecer estável durante um certo tempo, mas esta estabilidade é em si mesma fruto de um trabalho. Sua extensão, sua composição e seu desenho estão permanentemente em jogo para os atores envolvidos, sejam eles humanos, palavras, imagens, traços de imagens ou de contexto, objetos técnicos, componentes destes objetos, etc.
2. **Princípio de heterogeneidade:** Os nós e as conexões de uma rede hipertextual são heterogêneos. Na memória serão encontrados imagens, sons, palavras, diversas sensações, modelos, etc., e as conexões serão lógicas, afetivas, etc. Na comunicação, as mensagens serão multimídias, multimodais; analógicas, digitais, etc. O processo sociotécnico colocará em jogo pessoas, grupos, artefatos, forças naturais de todos os tamanhos, com todos os tipos de associações que pudermos imaginar entre estes elementos.
3. **Princípio de multiplicidade e de encaixe das escalas:** O hipertexto se organiza em um modo "fractal", ou seja, qualquer nó ou conexão, quando analisado, pode revelar-se como sendo composto por toda uma rede, e assim por diante, indefinidamente, ao longa da escala dos graus de precisão. Em algumas circunstâncias críticas, há efeitos que podem propagar-se de uma escala a outra: a interpretação de uma vírgula em um texto (elemento de uma microrrede de documentos), caso se trate de um tratado internacional, pode repercutir na vida de milhões de pessoas (na escala da macrorrede social).
4. **Princípio de exterioridade:** A rede não possui unidade orgânica, nem motor interno. Seu crescimento e sua diminuição, sua composição e sua recomposição permanente dependem de um exterior indeterminado: adição de novos elementos, conexões com outras redes, excitação de elementos terminais (captadores), etc. Por exemplo, para a rede semântica de uma pessoa escutando um discurso, a dinâmica dos estados de ativação resulta de uma fonte externa de palavras e imagens. Na constituição da rede sociotécnica intervêm o tempo todo elementos novos que não lhe pertenciam no instante anterior: elétrons, micróbios, raios X, macromoléculas, etc.
5. **Princípio de topologia:** Nos hipertextos, tudo funciona por proximidade, por vizinhança. Neles, o curso dos acontecimentos é uma questão de topologia, de caminhos. Não há espaço universal homogêneo onde haja forças de ligação e separação, onde as mensagens poderiam circular livremente. Tudo que se desloca deve utilizar-se da rede hipertextual tal como ela se encontra, ou então será obrigado a modificá-la. A rede não está no espaço, ela é o espaço.
6. **Princípio de mobilidade dos centros:** A rede não tem centro, ou melhor, possui permanentemente diversos centros que são como pontas luminosas perpetuamente móveis, saltando de um nó a outro, trazendo ao redor de si uma ramificação infinita de pequenas raízes, de rizomas, finas linhas brancas esboçando por um instante um mapa qualquer com detalhes delicados, e depois correndo para desenhar mais à frente outras paisagens do sentido.

Segundo Santaella (2005), as unidades básicas de informação em um hipertexto são os nós. Os documentos hipertextuais estão compostos por recursos de diferente natureza (ou multimídia), não só texto senão gráficos, desenhos, vídeos, *clips* de áudio, etc. Esses recursos multimídia formam parte dos nós e, para Santaella (2005), “[...] os nós de informação podem aparecer na forma de texto, gráficos, sequências de vídeos ou áudios, janelas ou de misturas entre eles.” (p.394). Com isso, a autora apresenta a ideia de estrutura reticular recorrente na *web*. Estamos frente à hipermídia, uma combinação de hipertexto com multimídias e multilinguagens.

Lemke (2002) afirma que “[...] a evolução de qualquer meio procura primeiro replicar gêneros familiares[...]” e depois “[...] cria novos gêneros por si só.” (LEMKE, 2002, p. 30). Para o autor, uma página *web* pode ser entendida como uma evolução da página impressa e “[...] divergirá provavelmente em muitos gêneros novos que se adaptem aos nichos especializados e funcionais.” (LEMKE, 2002, p.30). Nesse sentido, e dependendo do tipo de recurso, as matrizes da linguagem e pensamento enunciadas por Santaella predominaram de forma diferente no ciberespaço. A autora destaca o predomínio da matriz sonora em narrações *voice-over*, o predomínio da matriz visual nos *sítes* que descobrem imagens com *mouse-over*, o predomínio da matriz verbal em registros enciclopédicos e o equilíbrio entre as três matrizes, por exemplo, em narrações *web* que combinem textos, áudios, vídeos e outros recursos da hipermídia. A hipermídia, como apresentada por Bairon (2011) não é simplesmente a reunião de meios existentes de diferente natureza senão a fusão desses meios por meio de elementos não lineares. É precisamente nesse contexto de fusão onde deve se encaixar a publicação ampliada.

4.3 A MULTIMODALIDADE NA PESQUISA CIENTÍFICA

A publicação digital permite combinar diferentes matrizes da linguagem em uma publicação científica. A adição de conteúdo não se realiza de maneira arbitrária e desorganizada senão que se pretende tornar a publicação um objeto dinâmico e estruturado de informação e conhecimento mediante a combinação de diferentes ativos/assets resultantes da pesquisa em formatos diversos (áudio,

vídeo, *slide shows*, conjuntos de dados ou *software*), bem como comentários dos leitores e outras contribuições pós-publicação.

Entre os ativos/assets resultantes da pesquisa devem se estabelecer relações e conexões e, com isto, transformam-se em nós contextualizados. O nó, apresentado na descrição do hipertexto, é entendido por Santaella (2001) como modular, funcional e integrado dentro de um conjunto maior. “Um capítulo, uma secção, uma tabela, uma nota rodapé, [...], ou qualquer outra subestrutura do documento.” (SANTAELLA, 2001, p.394) podem ser um nó. Na hora de selecionar o tipo de informação desejada, o leitor pode pular os módulos (nós)⁴⁴ que não sejam do seu interesse. Portanto, um artigo modular está composto por módulos (nós) e as ligações entre eles (KIRCZ, 1998). Assim, é constituída “[...] uma unidade coerente para fins de comunicação.” (MUCHERONI, DA SILVA E PALETTA, 2015, p.5).

Também nesse sentido, Spicer (2014) define a investigação multimodal como:

[...] o uso de múltiplos suportes digitais (p.ex. imagens fixas e em movimento, objetos digitais interativos, áudio, conjuntos de dados, dados geoespaciais ou texto), com frequência compostos, exibidos, interligados e disseminados por uma série de plataformas de publicação digital (p.ex. sites web, blogs, aplicativos móveis e redes sociais), para comunicar a pesquisa. (SPICER, 2014, p. 2).

O resultado de um processo de investigação multimodal é uma série de ativos/assets que configuram todos os aspectos da pesquisa científica. Estes ativos podem ser estruturados como uma publicação ampliada. Por sua vez, uma publicação ampliada entra no contexto da ecologia informacional complexa, sendo esta “[...] a conjunção sinérgica de ambientes analógicos, digitais e híbridos, tecnologias analógicas e digitais, utilizadas de maneira holística, em multicanais pelos sujeitos em determinado contexto cultural.” (OLIVEIRA; LIMA, 2016, p.53). Ao mesmo tempo, esta publicação ampliada, como ecologia informacional complexa, está inserida em um ambiente —seja um periódico, uma plataforma de publicação, um repositório, etc.— que também é caracterizado como uma ecologia informacional complexa. É possível distinguir aqui a articulação entre signos dentro de diferentes estruturas (a publicação ampliada, o periódico, o repositório, etc.).

⁴⁴ Diferentes autores utilizam tanto a palavra “modulo” quando a palavra “nó” para se referir aos núcleos na rede.

Além disso, a perspectiva ecológica envolve que a publicação científica ampliada seja um processo hipermidiático, quer dizer, que na sua estrutura haja espaço para “[...] textos escritos, imagens estáticas, animações, animações interativas, áudios, vídeos, entre outras possibilidades midiáticas.” (p.59).

Nesse contexto de agregação de conteúdo e objetos dinâmicos, é importante o estudo da multimodalidade como interação entre diversos signos com diferentes origens semióticas (O’HALLORAN *et al.*, 2010). As três matrizes tradicionais combinam-se e essa multimodalidade resulta em interações entre diferentes ativos/assets (texto escrito, imagens, vídeos, áudios, etc.). Indubitavelmente, esta combinação é um desafio para a articulação entre componentes a partir da perspectiva da pesquisa multimodal.

A articulação entre signos, discursos e culturas (cultura científica para as publicações ampliadas) pode ser analisada desde o conceito da semiosfera, definida por Lotman (1996) como a esfera de semiose (transformação de signos em signos) onde os processos operam em um conjunto de ambientes interconectados. Lotman (2005) entende o universo semiótico como a totalidade textos individuais e linguagens isoladas desde que se relacionem entre si. Ainda, o autor entende que para definir a semiosfera é necessária a existência de uma fronteira.

Essa fronteira é entendida como uma membrana celular: define a divisão entre estruturas internas e externas, mas permite a filtração elementos externos. Esta filtração envolve uma semiotização desses elementos e a sua transformação em informação, já que tudo o que existe fora da fronteira considera-se caótico ou sem estrutura. Consequentemente, dentro da fronteira, aceleram-se os processos semióticos. A fronteira integra também a semiosfera e constitui “[...] um mecanismo bilíngue que traduz as comunicações externas para a linguagem interna da semiosfera e vice-versa.”⁴⁵(LOTMAN, 2005, p. 210 tradução própria).

A topografia da semiosfera é descontínua e heterogênea (NÖTH, 2015). A descontinuidade entende-se como a existência de objetos separados dentro das fronteiras da semiosfera. Porém, nela existem tanto características de homogeneidade e heterogeneidade.

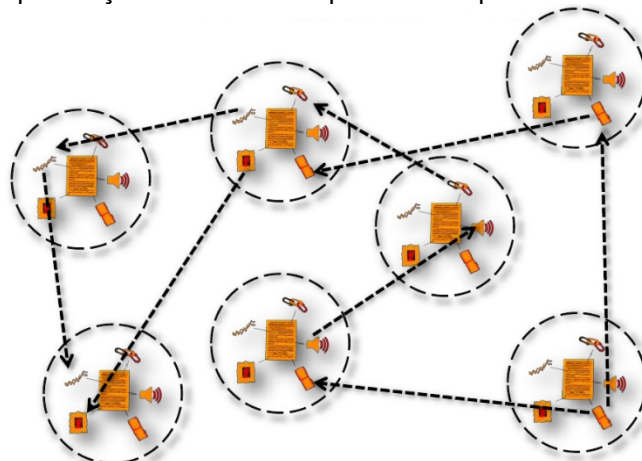
⁴⁵ Tradução própria do original em inglês: “[...]a bilingual mechanism, translating external communications into the internal language of the semiosphere and vice versa”.

A homogeneidade está relacionada como o isolamento dentro de uma fronteira que separa a semiosfera de “[...] esperadas situadas fora ou além das suas fronteiras.” (NÖTH, 2015, p. 18). Simultaneamente, a semiosfera apresenta irregularidades semióticas que a caracterizam como heterogênea. “O espaço semiótico caracteriza-se pela presença de estruturas nucleares (frequentemente múltiplas) e por um mundo semiótico mais amorfo e visivelmente organizado, gravitando para a periferia, no qual as estruturas nucleares estão imersas”⁴⁶ (LOTMAN, 2005, p. 213 tradução própria). Ainda, os elementos dentro da semiosfera diferem em natureza e função.

Estas características da topografia da semiosfera podem ser trasladadas à descrição das publicações ampliadas. Cada publicação ampliada é uma das estruturas nucleares da semiosfera. Esta publicação é heterogênea, já que está composta por ativos/assets conectados de distinta natureza conectados (ver Figura 6); porém, ao mesmo tempo homogênea, já que todos esses ativos/assets são produtos do mesmo processo de pesquisa. Cada publicação ampliada está em contato e intercâmbio com outras publicações ampliadas ou com ativos/assets pertencentes à outras publicações ampliadas, bem como com outros elementos dos ambientes virtuais de publicação. Esse intercâmbio faz com que as publicações ampliadas estejam em um processo constante de transformação e possam criar novas versões os ativos/assets, adicionar novos ou ser enriquecidas com comentários pós-publicação.

⁴⁶ Tradução própria do original em inglês: “Semiotic space is characterised by the presence of nuclear structures (frequently multiple) and a visibly organised more amorphous semiotic world gravitating towards the periphery, in which nuclear structures are immersed”.

Figura 6 Esboço das publicações científicas ampliadas e as possíveis conexões entre elas.



Fonte: SURF Foundation (2011).

Para aprofundar na estrutura da publicação ampliada é necessário abordar as características mínimas de cada um dos seus elementos (elementos da semiosfera). Cada elemento é objeto digital, que se tornará um ativo/asset após a interação com o receptor. Um objeto digital deve ser entendido como a soma da integridade e a identidade, características que garantem a autenticidade do objeto (BUDRONI, 2017)⁴⁷. A identidade do objeto digital é garantida mediante a consistência do objeto, a informação contida nele, a alegação do contexto do objeto e a informação da proveniência do objeto (principalmente, o criador ou criadora).

O objeto digital abrange tanto dados digitais quanto interpretações, anotações, interações ou representações de domínio. Por isso, Abrams (2015) considera seis dimensões características para o conteúdo digital: sintática, semântica, empírica, pragmática, diplomática e dinâmica. Estas dimensões levam a uma nova conceitualização do objeto digital. Para cada dimensão o autor considera a tipologia dos objetos como⁴⁸: bolha (*blob*), artefato (*artifact*), exemplar (*exemplar*), produto (*product*), ativo (*asset*), registro (*record*), herança (*heirloom*).

- **Dimensão sintática:** considera a relação entre conteúdo e expressão simbólica. Os objetos são de tipo *exemplar*, sendo um *exemplar* um documento com codificações simbólicas da sua expressão e estrutura

⁴⁷ Paolo Budroni, LIBER Workshop, julho 2017 Workshop: Management of Born Digital Materials at Research Libraries.

⁴⁸ No texto serão utilizados os termos originais em inglês para não perder o significado inicial na tradução.

interna, como, por exemplo, uma imagem com metadados técnicos, mas ausente de qualquer conhecimento do que a imagem representa.

- **Dimensão semântica:** considera as relações de conteúdo com significado ou influência abstratas. Os objetos são de tipo *product*, sendo um *product* um objeto interpretável dentro de um domínio, mas sem possibilidades de exploração da mensagem contida, como, por exemplo, um arquivo JPEG com descrições associadas.
- **Dimensão empírica:** considera a relação entre conteúdo e representação física. Os objetos são de tipo *blob* e *artifact*, sendo um *blob* uma sequência indefinida de *bits* e um *artifact* uma sequência fixa de *bits* no espaço-tempo, como, por exemplo, um arquivo com um tamanho e *timestamps* específicos, mas sem expressão de conteúdo.
- **Dimensão pragmática:** considera a relação entre conteúdo e consumidor. Os objetos são de tipo *asset*, sendo um *asset* um objeto útil que expõe características apropriadas para os sentidos humanos. É o caso da renderização da imagem JPEG.
- **Dimensão diplomática:** considera as relações entre conteúdo e autenticidade da sua expressão, representação, gestão e transmissão. Os objetos são de tipo *record*, sendo um *record* um objeto confiável cujas propriedades servem para a verificação e conservação da sua autenticidade e verificação, por exemplo, uma imagem avaliada e considerada confiável.
- **Dimensão dinâmica:** considera as relações entre os diferentes estados dos conteúdos que persistem e evoluem no tempo e no espaço. Os objetos são de tipo *heirloom*, sendo um *heirloom* um objeto flexível resultante de uma intervenção proativa e uma comprovação constante da sua usabilidade, como, por exemplo, uma imagem pronta para o engajamento futuro com o consumidor.

Estas dimensões podem ser trasladadas ao processo de pesquisa científica, desde os dados observacionais (*blob*), o conjunto de dados brutos (*product*), a descrição dos conjuntos (*exemplar*), a relação entre descrição e conjunto (*artifact*), o compartilhamento desse conjunto de dados descrito (*asset*), a curadoria do conjunto como objeto confiável (*record*) e a apresentação do conjunto junto com a publicação final como parte de um objeto dinâmico do

conhecimento para reuso de outros pesquisadores (*heirloom*). O pesquisador “[...] não interatua com um *artifact*, exemplar ou *product*, senão com um *asset*: o desempenho realizado pelo *product*.” (BECKER, 2018, p.32) e essa interação é possível devido à presença dos objetos no ambiente digital.

Não obstante, para o objeto desta pesquisa, é necessário expandir a ideia de dimensão dinâmica. A classificação dos objetos de Abrams (2015) constitui uma classificação nominal, ou seja, uma vez que a ideia ou o dado é “congelado” em um elemento arquivável. Porém, esse modelo esquece a dimensão da ação onde não é substantivo o protagonista, senão o verbo. Partindo para um pensamento computacional, esta ação poderia ser traduzida em um arquivo executável (p.ex. “.exe”). Portanto, incluiremos uma nova dimensão, a **dimensão executável**, que nos permitirá definir a lógica dos relacionamentos entre cada uma das entidades. Esta dimensão traz consigo a abordagem linguagem-ação necessária para repensar o *design* de sistemas.

4.4 NÍVEIS DE INTERAÇÃO NAS PLATAFORMAS DE PUBLICAÇÃO, CATEGORIAS FENOMENOLÓGICAS E HIBRIDAÇÃO DAS LINGUAGENS

As novas formas de publicação e comunicação científica precisam ser abordadas desde óticas diferentes. O dinamismo e a hipertextualidade são dois aspectos a serem tidos em consideração nos estudos sobre evolução da web e das publicações na web. Na atualidade, as publicações (e a própria web) vão além de uma representação estática, são objetos compostos por módulos bem definidos cognitivamente (KIRCZ, 1998) e formas dinâmicas de transmitir informação e conhecimento. A partir da mesma perspectiva devemos abordar os ambientes informacionais nos que elas estejam inseridas. Esses ambientes permitem trabalhar não só com o armazenamento, a encontrabilidade ou a recuperação, senão também com novas apresentações das pesquisas e o enriquecimento delas (BREURE; VOORBIJ; HOOGERWERF, 2011) e interagir com os diferentes ativos/assets de pesquisa (BECKER, 2018).

Breure, Voorbij e Hoogerwerf (2011) classificam as publicações e seus ambientes em três níveis segundo a sua interatividade e conceitualização

hipermidiática: o limite inferior, o meio do espectro e o limite superior. O **limite inferior** caracteriza-se pela translação do artigo tradicional ao meio digital e, como muito, a adição de um *link* para oferecer dados complementares. O **meio do espectro** envolve publicações com um grau maior de interatividade, já que se introduzem elementos como galerias de imagens, mapas, enlaces entre conteúdos e a bases de conhecimento, entre outros. Nesse nível intermédio, ainda existe uma dominação textual, mas a hierarquia linear já não é tão pré-definida. O limite superior se caracteriza pela narrativa visual altamente baseada em multimídia interativa. A estrutura de leitura linear desaparece. Porém, é difícil achar publicações científica periódicas neste nível, já que são altamente demandantes temporal e economicamente falando. Além disso, precisam de uma programação específica e não estão baseadas em um *software* genérico de gestão de publicações como o *Open Journal System* (OJS) ou *Janeaway*. Portanto, chamaremos projetos de publicação às publicações pertencentes ao limite superior.

Esses três níveis relacionam-se diretamente com as categorias fenomenológicas de Peirce. Como descrito anteriormente, as categorias fenomenológicas (primeridade, secundidade e terceiridade) visam descrever as características nos fenômenos naturais e mentais. Essas características contemplam as características da comunicação, que inclui a comunicação científica e a interatividade nas plataformas de publicação.

No Quadro 10, representam-se as características dos níveis de interação relacionados com as categorias fenomenológicas a partir da semiose de um estágio da terceiridade e mostram-se exemplos de plataforma de publicação que seguem esses níveis. Parte-se da terceiridade e exemplifica-se o crescimento da complexidade do signo nas plataformas.

Quadro 10 Características de cada nível de interação com a publicação e exemplos.

Categoria fenomenológica	Características da categoria para a interação	Nível	Características do nível (BREURE; VOORBIJ; HOOGERWERF, 2011)	Exemplos (inserção própria)
3.1	Presentidade Apresentação dos fatos Não existe participação ou interação Comunicação linear Independência do receptor	Limite baixo	Baseado no texto Alguns elementos visuais Leitura linear Enlaces internos Metadados semânticos Partes separadas Camada de dados	Plataforma da editora Copernicus Publications Periódico científico <i>The Comics Grid</i> Plataformas baseadas no sistema OJS.
3.2	Ação e reação Inclusão da experiência Construção da apresentação	Meio do espectro	Baseado no texto Mais elementos visuais Principalmente leitura linear Enlaces internos Mais metadados semânticos providenciados Partes integradas Camada de dados Ferramentas	Plataforma PLOS Plataforma ScienceDirect Plataforma CellPress
3.3	Interatividade (suma da atração e da participação). Ação mútua entre apresentação do conteúdo, conteúdo e receptor. Mensagem alterável Alta presença de multilinguagens	Limite alto	Aplicações web autônomas Baseado na imagem Leitura linear e não linear Enlaces internos Mais metadados semânticos providenciados Partes integradas Camada de dados Ferramentas	Projeto <i>Vectors Journal</i>

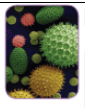
Fonte: Elaboração própria.

A seguir, vemos como estes níveis estruturam-se nos ambientes de publicação selecionados. Esta análise dos ambientes entra no polo técnico como parte das transformações técnicas. Estas transformações são entendidas como a redução aos fatos (entidades da publicação) após a observação do entorno (as plataformas de publicação).

O limite inferior

No limite inferior observa-se uma predominância da linguagem verbal e visual-verbal. O texto constitui a base da publicação; porém, aparecem outros ativos/assets de pesquisa que permitem a interação. A Figura 7, a seguir, apresenta um exemplo deste limite obtido da editora de acesso aberto Copernicus Publications. Os artigos publicados nesta editora adicionam alguns elementos visuais e possuem enlaces internos e externos. Os enlaces externos ficam reservados para a conexão com *assets* disponíveis em outras plataformas, como plataformas de arquivo de dados ou vídeos (ver Figura 7).

Figura 7 Exemplo publicação no limite baixo com núcleo textual



Climate of the Past
An interactive open-access journal of the European Geosciences Union

[EGU.eu](#) | [EGU Publications](#) | [EGU Highlight Articles](#) | [Contact](#) | [Imprint](#) | [Data protection](#)

Clim. Past, 15, 153–168, 2019
<https://doi.org/10.5194/cp-15-153-2019>
 © Author(s) 2019. This work is distributed under the Creative Commons Attribution 4.0 License.

Volume 15, issue 1



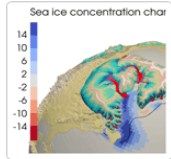
[Article](#) | [Assets](#) | [Peer review](#) | [Metrics](#) | [Related articles](#)

Research article 24 Jan 2019

Heinrich events show two-stage climate response in transient glacial simulations

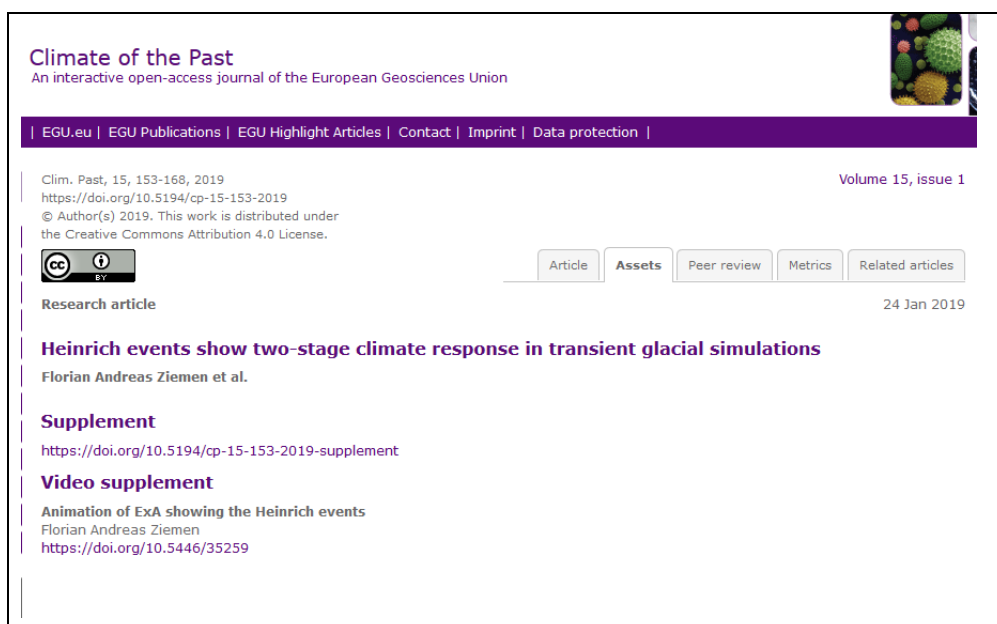
Florian Andreas Ziemer[✉], Marie-Luise Kapsch, Marlene Klockmann[✉], and Uwe Mikolajewicz
 Max Planck Institute for Meteorology, Bundesstraße 53, 20146 Hamburg, Germany

Received: 23 Feb 2018 – Discussion started: 21 Mar 2018 – Revised: 14 Dec 2018 – Accepted: 22 Dec 2018
 – Published: 24 Jan 2019



Abstract. Heinrich events are among the dominant modes of glacial climate variability. During these events, massive iceberg armadas were released by the Laurentide Ice Sheet and sailed across the Atlantic where they melted and released freshwater, as well as detritus, that formed characteristic layers on the seafloor. Heinrich events are known for cold climates in the North Atlantic region and global climate changes. We study these events in a fully coupled complex ice sheet–climate model with synchronous coupling between ice sheets and oceans. The ice discharges occur as an internal variability of the model with a recurrence period of 5 kyr, an event duration of 1–1.5 kyr, and a peak discharge rate of about 50 mSv, roughly consistent with reconstructions. The climate response shows a two-stage behavior, with freshwater release effects dominating the surge phase and ice sheet elevation effects dominating the post-surge phase. As a direct response to the freshwater discharge during the surge phase, deepwater formation in the North Atlantic decreases and the North Atlantic deepwater cell weakens by 3.5 Sv. With the reduced oceanic heat transport, the surface temperatures across the North Atlantic decrease, and the associated reduction in evaporation causes a drying in Europe. The ice discharge lowers the surface elevation in the Hudson Bay area and thus leads to increased precipitation and accelerated ice sheet regrowth in the post-surge phase. Furthermore, the jet stream widens to the north, which contributes to a weakening of the subpolar gyre and a continued cooling over Europe even after the ice discharge. This two-stage behavior can explain previously contradicting model results and understandings of Heinrich events.

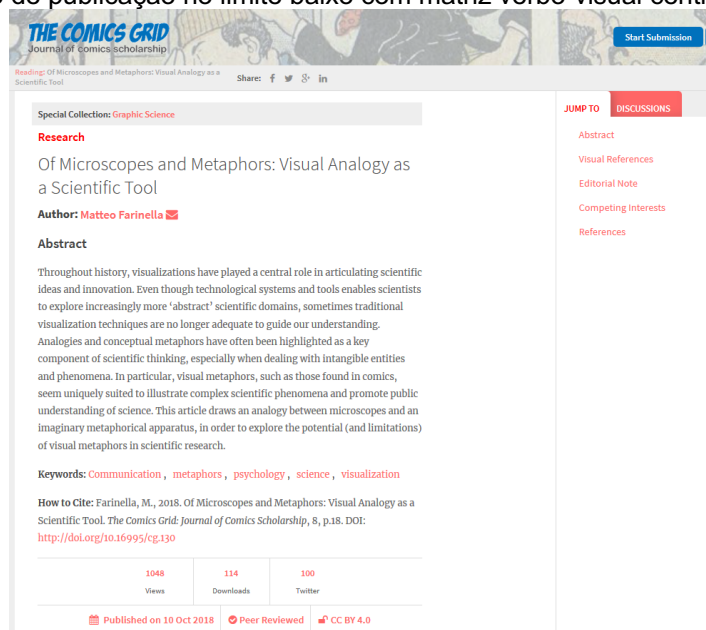
How to cite: Ziemer, F. A., Kapsch, M.-L., Klockmann, M., and Mikolajewicz, U.: Heinrich events show two-stage climate response in transient glacial simulations, *Clim. Past*, 15, 153–168, <https://doi.org/10.5194/cp-15-153-2019>, 2019.



Fonte: Elaboração própria a partir de <https://doi.org/10.5194/cp-15-153-2019>.

Além das estruturas centradas na matriz verbal, observam-se também iniciativas baseadas na matriz verbo-visual (ver Vídeo 1). Iniciativas como a da revista *The Comics Grid* ainda não suportam uma interatividade ou semantização tão alta quanto para pertencerem ao meio do espectro, mas abrem a possibilidade de inserção de outras formas de expressão na pesquisa científica. Se o contexto, o tipo de pesquisa e a temática assim o permitirem, nada impede elaborar os resultados de pesquisa em forma de quadrinho.

Vídeo 1 Exemplo de publicação no limite baixo com matriz verbo-visual central.



Fonte: Elaboração própria a partir de <http://doi.org/10.16995/cg.130>. Acesso em: <https://doi.org/10.6084/m9.figshare.11663592>

A linguagem visual-verbal costuma ocupar o lugar principal neste limite; porém também é possível encontrar núcleos audiovisuais. Esta situação se observa nos portais de vídeos científicos, como TIB|AV-Portal (ver Figura 8), ou em periódicos científicos com um conteúdo audiovisual maior, como JOVE. Apesar da integração de mais conteúdo audiovisual, mantém-se uma estrutura linear, com metadados semânticos, mas com pouca interação. Todas as ações podem ser realizadas mediante *links* externos ou internos.

Não obstante, a existência deste tipo de portais ou publicações mostra a relevância que está adquirindo das linguagens visual-sonora (simulação 3D ou computação gráfica) e verbo-visual-sonora (vídeos no artigo) no contexto acadêmico.

O meio do espectro

No meio do espectro tem lugar uma maior hibridação das linguagens e uma interação maior. Há navegações que permitem acessar a parte da publicação que resulte de interesse e o enriquecimento semântico dos metadados associados também aumenta. Um claro exemplo desta categoria são as revistas da PLOS (ver Vídeo 2).

Vídeo 2 Exemplo de publicação no meio do espectro com diversos ativos/assets adicionais.

The screenshot shows the PLOS Biology website interface. At the top, there's a navigation bar with 'BROWSE', 'PUBLISH', and 'ABOUT' links, a search bar, and a 'sign in' button. Below the navigation bar, the article title 'Evolution of correlated complexity in the radically different courtship signals of birds-of-paradise' is prominently displayed. The authors listed are Russell A. Ligon, Christopher D. Diaz, Janelle L. Morano, Jolyon Troscianko, Martin Stevens, Annalyse Moskeland, Timothy G. Laman, and Edwin Scholes III. The article is dated November 20, 2018. On the right side, there's a statistics box showing 0 saves, 1 citation, 6,366 views, and 42 shares. Below the article title, there's a tabbed interface with 'Article' selected. The abstract is visible, discussing the evolutionary dynamics of ornamental traits in birds-of-paradise. On the right, there are buttons for 'Download PDF', 'Print', and 'Share', along with a 'Check for updates' button. At the bottom right, there's a 'Related PLOS Articles' section with a link to 'How do complex animal signals evolve?' and a 'Subject Areas' section with tags for 'Animal behavior' and 'Acoustics'.

Fonte: Elaboração própria a partir de <https://doi.org/10.1371/journal.pbio.2006962>. Acesso: <https://doi.org/10.6084/m9.figshare.11663721>

Nestas revistas, cada artigo torna-se uma coleção de elementos onde cada um deles possui uma identificação persistente pertencente ao artigo central.

Outro exemplo é o fornecido no ambiente ScienceDirect da editora Elsevier. Esse ambiente tem surgido após muitas implementações para obter um “artigo do futuro”, como proposto na publicação *Article of the future* (AALBERSBERG *et al.*, 2012). Um aspecto inovador deste ambiente é a integração semântica com uma base de conhecimento que possa ampliar certos conceitos ao longo do artigo. Dessa forma, cria-se um vínculo entre um módulo textual dentro do artigo e um módulo de informação complementar dentro da base do conhecimento.

Vídeo 3 Exemplo de publicação no meio do espectro com base do conhecimento integrada.

The screenshot shows the ScienceDirect interface for the article "Aluminium in brain tissue in autism" by Matthew Mold, Dorcas Umar, Andrew King, and Christopher Exley. The article is published in the Journal of Trace Elements in Medicine and Biology, Volume 46, March 2018, Pages 76-82. The page includes an abstract, keywords, and a list of figures (5). The abstract states: "Autism spectrum disorder is a neurodevelopmental disorder of unknown aetiology. It is suggested to involve both genetic susceptibility and environmental factors including in the latter environmental toxins. Human exposure to the environmental toxin aluminium has been linked, if tentatively, to autism spectrum disorder. Herein we have used transversely heated graphite furnace atomic absorption spectrometry to measure, for the first time, the aluminium content of brain tissue from donors with a diagnosis of autism. We have also used an aluminium-selective fluor to identify aluminium in brain tissue using fluorescence microscopy. The aluminium content of brain tissue in autism was consistently high. The mean (standard deviation) aluminium content across all 5 individuals for each lobe were 3.82(5.42), 2.30(2.00), 2.79(4.05) and 3.82(5.17) µg/g dry wt. for the occipital, frontal, temporal and parietal lobes respectively. These are some of the highest values for aluminium in human brain tissue yet recorded and one has to question why, for example, the aluminium content of the occipital lobe of a 15 year old boy would be 8.74 (11.59) µg/g". The page also features a sidebar with recommended articles and a table of contents.

Fonte: Elaboração própria desde <https://doi.org/10.1016/j.jtemb.2017.11.012>. Acesso: <https://doi.org/10.6084/m9.figshare.11663685>

De forma semelhante, a plataforma CellPress (da editora Elsevier) apresenta os conteúdos com um maior nível de interação e com a possibilidade de vincular termos chave em diferentes textos da plataforma. Adicionalmente, integram-se tanto o resumo em vídeo quanto o resumo gráfico. É importante sinalar que no resumo vídeo acontecem uma série de mudanças no registro, começando pelo título e seguindo pela forma de apresentar os conteúdos. Esse tipo de “tradução” do conteúdo textual ao audiovisual mostra claramente que cada ativo/asset está direcionado a uma audiência diferente.

Também, na plataforma CellPress introduz-se o método STAR (*Structured, Transparent, Accessible Reporting*) para apresentar os métodos do artigo de

forma clara e concisa e garantir assim a transparência e possível futura reprodutibilidade de métodos e resultados.

Vídeo 4 Exemplo de publicação no meio do espectro com integrações audiovisuais.

The screenshot shows the top half of a Cell journal article page. The header includes the Cell logo, navigation links (Log in, Register, Subscribe, Claim), and a search icon. Below the header, the article title is 'orco Mutagenesis Causes Loss of Antennal Lobe Glomeruli and Impaired Social Behavior in Ants'. The authors listed are Waring Trible, Leonora Olivos-Cisneros, Sean K. McKenzie, Benjamin J. Matthews, Peter R. Oxley, and Daniel J.C. Kronauer. A DOI link is provided: <https://doi.org/10.1016/j.cell.2017.07.001>. On the right, there are icons for PDF (2 MB), Figures, Save, Share, Reprints, and Request. Below the title, there is a 'Highlights' section with four bullet points:

- We generated odorant receptor co-receptor (*orco*) mutants in an ant
- *orco* mutants did not follow pheromone trails or cluster with other ants
- Mutant ants also had greatly reduced antennal lobes in the brain
- Odorant receptors are essential for the complex organization of ant societies

 To the left of the highlights is a vertical menu with links to Summary, Graphical Abstract, Keywords, Introduction, Results, Discussion, STAR Methods, and Author Contributions. At the bottom right of the highlights section is a 'Feedback' button.

Fonte: Elaboração própria desde <https://doi.org/10.1016/j.cell.2017.07.001>. Acesso em: <https://doi.org/10.6084/m9.figshare.11663559>

A partir destes ambientes de publicação é possível identificar mais entidades que compõem uma publicação científica. Entre as entidades de conteúdo estão o texto central, as figuras, as imagens, as tabelas, os áudios, as planilhas de dados, os conjuntos de dados brutos, os vídeos, os códigos ou algoritmos, o *software* e as bases de conhecimento. Além disso, há uma entidade descritiva: a entidade autor. Esta entidade é a que possui a característica de criador do conteúdo, entidade que pode gerar o resto e tem a capacidade intelectual para tal fim. Esta ideia de criação reflete em vocabulários controlados como o Dublin Core Terms na propriedade “creator” (dc:creator).

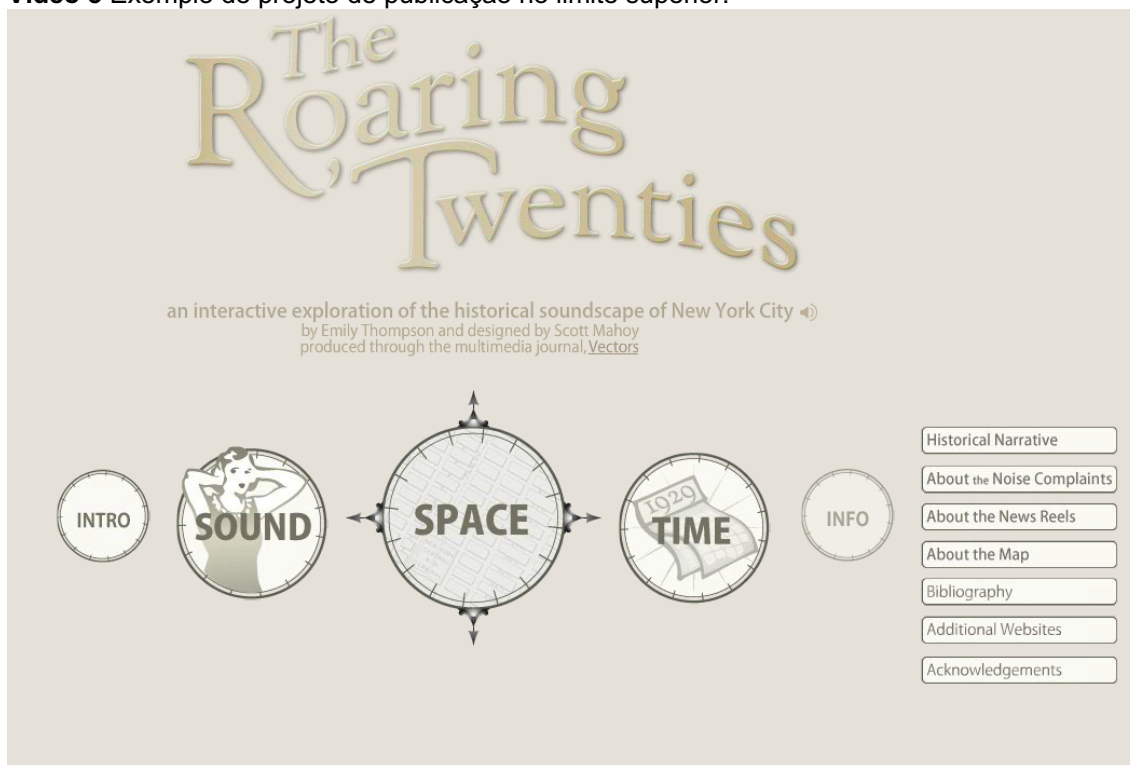
O limite superior

Neste limite a interação é muito elevada, praticamente permitindo a navegação livre e criação de uma própria história dentro da publicação. Como comentado anteriormente, no limite superior existem poucas publicações periódicas, dada a dificuldade de programação deste tipo de publicações e os custos temporais e econômicos. Em muitos casos estes projetos estariam

relegados a divulgação científica e não seriam considerados como resultados de pesquisa.

Um exemplo dos projetos do limite superior é o periódico *Vectors Journal*. Esse periódico identifica as publicações baseadas no texto com escalares e os projetos contidos nele como vetores⁴⁹. Utiliza assim o conceito matemático de vetor como magnitude física que possui um módulo, uma direção e uma orientação em contraste com um escalar que apenas define o fenômeno físico com magnitude. *The Roaring Twenties* é um dos projetos contidos no *Vectors Journal* (ver vídeo 5).

Vídeo 5 Exemplo de projeto de publicação no limite superior.



Fonte: Elaboração própria desde <http://vectors.usc.edu/projects/index.php?project=98>. Acesso: <https://doi.org/10.6084/m9.figshare.11663691>

Esse projeto permite a navegação entre diferentes momentos da história do primeiro terço do século XX em Nova Iorque. Possui uma grande quantidade de documentação visual, auditiva e audiovisual de forma individual, o que compõe um projeto nativo da linguagem híbrida.

⁴⁹ Vectors Journal: <http://vectors.usc.edu/issues/index.php?issue=7>

4.5 AS ENTIDADES NA PUBLICAÇÃO AMPLIADA E A DEFINIÇÃO SEMIÓTICA DA PUBLICAÇÃO AMPLIADA

Considerando o exposto anteriormente, entendemos a publicação científica ampliada, objeto deste trabalho, não só como uma forma de agregação de dados em uma publicação baseada no texto, como defendem muitos autores (PAMPEL *et al.*, 2013; BREURE; VOORBIJ; HOOGERWERF, 2011; WOUTERSEN-WINDHOUWER, 2009). A ideia defendida neste trabalho aproxima-se mais ao conceito de “publicação enriquecida na internet” proposto por Breure, Voorbij e Hoogerwerf (2011).

Em cada nó da rede aparece um ativo/asset da pesquisa científica (ou entidades como p.ex. texto, planilha de dados, vídeo, imagem ou software) identificado por um identificador persistente (p.ex. DOI) que garanta a sua descrição como signo individual dentro do sistema semiótico. Assim, cada entidade pode ser classificada dentro de uma matriz da linguagem distinta.

O Quadro 11 descreve a ligação entre as entidades identificadas e a matriz a que pertencem.

Quadro 11 As entidades na publicação científica e sua classificação segundo as matrizes.

Entidade	Descrição	Matriz da linguagem
Áudio	Gravação de áudio (p.ex. em entrevistas) ou arquivo musical (p.ex. de produção artística do autor)	Sonoro-verbal
Base de conhecimento	Base para acessar conhecimento adicional para complementar as informações fornecidas no artigo. Podem ser legíveis por máquinas e por humanos e podem usar como base uma ontologia para a semantização.	Verbo-visual-sonora
Código	Algoritmo fonte (quando elaborados ou modificados pelo autor) utilizados para a coleta e/ou análise dos dados.	Verbo - visual
Conjunto de dados brutos	Dados brutos coletados e que sustentam os resultados apresentados na pesquisa e que não possuem intervenção intelectual nem direitos de autor	Verbo - visual
Arquivo multimídia	Arquivos com combinação de formatos, principalmente com caráter interativo.	Verbo-visual-sonora
Figura	Visualização dos dados	Visual
Imagem	Elemento imagético inserido na pesquisa	Visual
Planilha de dados	Dados analisados e trabalhados que levam aos resultados da pesquisa	Verbal
Quadro	Informação textual apresentada de forma tabulada	Verbal
Software	Programas informáticos (quando elaborados ou modificados pelo autor) utilizados para a coleta e/ou análise dos dados mediante uma interface interativa	Visual-verbal
Tabela	Apresentação tabulada final dos dados analisados	Verbal
Texto	Parte narrativa escrita resultante da pesquisa realizada. Pode ser dividido em módulos (p.ex. resumo, métodos ou discussão)	Verbal
Vídeo	Recurso audiovisual para o artigo. Podem ser desde figuras dinâmicas até resumos em vídeo.	Verbo-visual-sonora

Entidade	Descrição	Matriz da linguagem
<i>Workflow</i>	Série estruturada de passos executados para produzir um resultado final. Permite especificar o trabalho de forma mais detalhada e garantir a possível reprodutibilidade.	Verbo-visual-sonora, verbo-visual ou verbal

Fonte: Elaboração própria.

O quadro anterior serve de base para a elaboração da estrutura e modelo propostos para a publicação ampliada no contexto deste trabalho.

Além das entidades expostas anteriormente, consideram-se o dinamismo e a hipertextualidade como dois aspectos fundamentais na execução de uma publicação ampliada. Por isso, é necessário considerar a dimensão executável que represente e estabeleça o processo de criação e a lógica entre as entidades.

Diante do exposto, a publicação ampliada define-se como um espaço semiótico de topografia tanto heterogênea quanto homogênea. Nele, a heterogeneidade é representada pelos ativos/assets de distinta natureza e pertencentes a qualquer matriz da linguagem, cuja identidade é garantida por meio de informação descritiva, identificação persistente, descrição da proveniência e da utilização. A homogeneidade resulta ao entender os ativos/assets como integrantes do mesmo processo de pesquisa. As fronteiras da publicação ampliada são permeáveis, já que cada ativo/asset está em contato e intercâmbio com outros ativos/assets pertencentes a outras publicações ampliadas, bem como com outros elementos da *web*. Esse intercâmbio ativa os processos de semiose e transforma em signos informações externas ao espaço semiótico inicial. O dinamismo da publicação ampliada envolve estas interações, enriquecimentos e consequentes evoluções, resultantes da comunicação dentro da comunidade científica e com outras comunidades discursivas. Essas evoluções conectam-se com o interpretante dinâmico peirceano (PEIRCE, 1958, CP 8.315, 1909), o signo adicional que resulta da interpretação de uma mente (humana ou computacional) do signo.

Dessa forma, abarca-se a perspectiva da linguagem-ação que pode auxiliar como base teórica a modelagem conceitual da publicação ampliada dentro de um sistema da informação.

Esta definição é construída após uma análise teórica de diferentes referências da Semiótica e da publicação ampliada. A definição constitui em si

um objeto morfológico da tese. Não obstante, devido à permeabilidade do método quadripolar, esta definição é candidata a fazer parte do polo teórico de pesquisa futuras.

No capítulo seguinte, analisa-se a evolução do conceito de publicação ampliada para chegar à proposta de estrutura que depois será modelada tendo em consideração o exposto neste capítulo.

5 A PUBLICAÇÃO CIENTÍFICA AMPLIADA: ABORDAGENS E ESTRUTURA

O artigo de periódico científico revisado por pares —seja mediante revisão cega ou aberta— segue sendo a forma principal de comunicação de resultados de pesquisa e, portanto, a base da comunicação científica.

Porém, os canais formais de comunicação vêm mudando suas formas nas últimas décadas, principalmente devido às possibilidades e funcionalidades da publicação digital. Segundo Lancaster (1995) a publicação digital permite:

- A publicação mais rápida dos resultados de pesquisa mediante submissão eletrônica dos artigos, redes de comunicação entre autores, editores e avaliadores, e publicação em uma base de dados como publicação ‘aceita’ em vez de ter que aguardar o próximo número do periódico;
- A disseminação mais eficiente da publicação pela correspondência com outros artigos presentes na base de dados que possam atrair o interesse de leitores potenciais;
- Formas inovadoras de apresentar os resultados de pesquisa e outras formas de dados e informação;
- A facilitação da revisão por pares pública mediante a possibilidade de enlaçar os comentários dos leitores e as avaliações aos artigos publicados;
- Menores custos para a correspondência exitosa entre artigo e leitor;
- A publicação mais rápida e a facilitação de comunicação que conduzem a uma revista mais interativa, na qual uma contribuição pode gerar respostas rápidas de outros investigadores. (LANCASTER, 1995, p. 523)⁵⁰.

Um avanço muito relevante da publicação digital é, como indica Lancaster (1995) no ponto 3, é a possibilidade de representar os resultados de pesquisa de formas inovadoras. De fato, cada vez mais existe a necessidade de publicação de dados de pesquisa junto com o artigo textual e melhorar as possibilidades a reutilização desses dados (BORGMAN, 2008). É importante destacar que ao falar em dados de pesquisa não há uma referência exclusiva a dados resultantes de medições ou dados observacionais, senão que o conceito abrange a totalidade de ativos/assets de pesquisa.

⁵⁰ Tradução e adaptação própria do original em inglês. Não é uma tradução exata e literal do original.

Nesse sentido, a Fundação Nacional de Ciência dos EUA (National Science Foundation - NSF) destaca a natureza diversa dos dados que incluem “[...] números, imagens, vídeos ou áudios, *software* e informação de versão de *software*, algoritmos, equações, animações ou modelos/simulações” (NSF, 2005, p.18). A Comissão Europeia também destaca a heterogeneidade dos dados no contexto da pesquisa tais como “[...] estatísticas, resultados de experimentos, medições, observações resultantes de trabalhos de campo, resultados de questionários, gravações de entrevistas e imagens.” (COMISSÃO EUROPEIA, 2016, p.16). Para a Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP), os dados produzidos em um projeto podem ser “[..] amostras, registros de coleta, formulários, modelos, resultados experimentais, *software*, gráficos, mapas, vídeos, planilhas, gravações de áudio, bancos de dados, material didático e outros.” (FAPESP, 2017).

Ainda, como destaca Borgman (2008), os dados de pesquisa possuem um forte caráter disciplinar. Assim, para a Ecologia, os dados de pesquisa incluem medições do tempo, leituras de sensores ou registros históricos; para a Medicina, raios-x, resultados de exames, relatórios de patologias; para a Química, estruturas de proteínas; para a astronomia, estudos espectrais; para a biologia, espécimes; e para a Física, eventos e objetos. Nas Ciências Sociais, há sondagens de opinião, pesquisas públicas, entrevistas, experiências de laboratório ou de campo, registros demográficos, registros de censo, registros de voto e indicadores econômicos. Nas humanidades, os dados são os registros de atividades humana como jornais, fotografias, cartas, diários, livros, certidões de nascimento, óbito ou casamento, mapas, entre outros.

Além da identificação disciplinar, Sales e Sayão (2019) propõem uma taxonomia para os dados de pesquisa, na qual eles seriam classificados quanto à origem, grau de processamento, natureza, abordagem da pesquisa, nível de sensibilidade, materialidade, perenidade e abertura. Assim, podemos classificar os dados segundo a forma de coleta e uso para a pesquisa (origem); o grau de processamento (brutos, pré-processados ou processados); a abordagem (qualitativos ou quantitativos); a natureza ou o formato dos dados (números, textos, artefatos), o nível de sensibilidade (não sensíveis, confidenciais ou pessoais), a materialidade (digitais ou físicos), a perenidade (dados canônicos ou episódicos) e a abertura (grau de abertura ou licença).

Dessa taxonomia, consideramos importante destacar o grau de abertura e processamento, já que os dados de pesquisa se estruturam ao longo da investigação de forma diferente e não todos os dados são publicados finalmente, como se indica na Figura 3 (Capítulo 3). Uma possível abordagem para a publicação de dados é publicar os denominados *underlying data* (dados subjacentes), aliás, aqueles dados que servem de base direta para os resultados apresentados no texto.

Cada conjunto de dados, e segundo disciplinas, terá um processo diferente de gestão, diferentes requerimentos para a infraestrutura de arquivamento (DALLMEIER-TIESEN *et al.*, 2014) e diferentes modelos de citação (CALLAGHAN *et al.*, 2012). Além disso, dada a relevância da publicação de dados como ativos/assets de pesquisa, tem surgido a iniciativa FAIR (*Findable, Accessible, Interoperable, Reusable*) (WILKINSON *et al.*, 2016) que apresenta quinze princípios⁵¹ para a adequada publicação dos dados.

Destaca-se a necessidade de que todo conjunto de dados que chegue a ser publicado tenha um identificador persistente e metadados associados (ou seja, informação descritiva sobre esses dados). Os metadados não possuem direitos autorais, como acontece também com os dados brutos resultantes de processos de medição (p.ex. dados coletados por satélites). O modelo para publicação ampliada objeto desta tese versa incluir esses princípios para que a publicação científica ampliada tenha também as características FAIR e possa ser uma ferramenta para a publicação de dados procedentes de qualquer camada.

Diante do exposto, esse capítulo explora a literatura científica sobre publicação ampliada e suas abordagens e propõe uma estrutura para a mesma. Além disso, analisa-se o papel dos identificadores persistentes (PIDs) como identificação dos ativos/assets de pesquisa e sua função no modelo de publicação ampliada que constitui o objetivo deste trabalho.

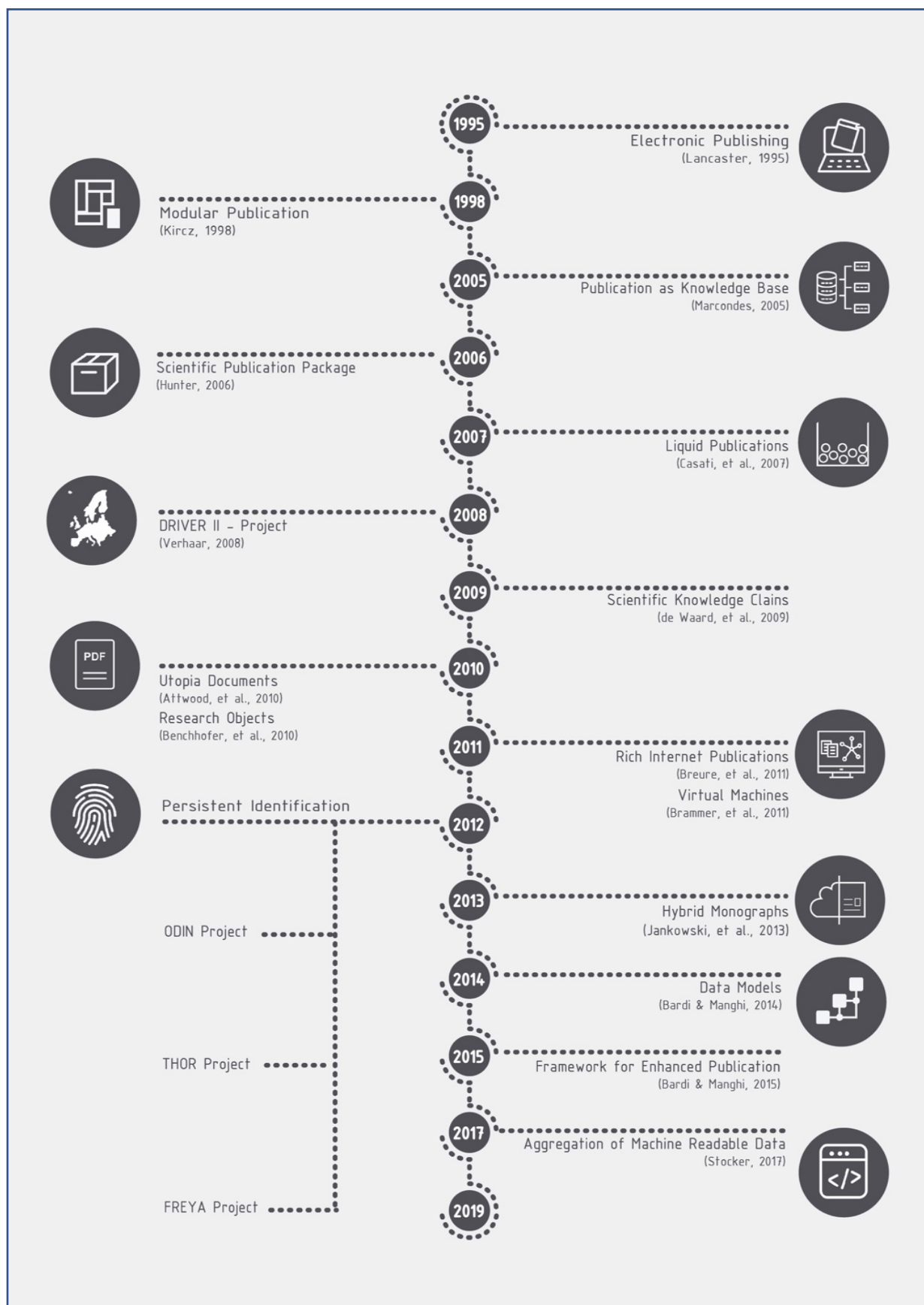
⁵¹ Princípios FAIR: <https://www.go-fair.org/fair-principles/>

5.1 EVOLUÇÃO E ABORDAGENS DA PUBLICAÇÃO CIENTÍFICA AMPLIADA

Devido a necessidade de adicionar mais conteúdo à publicação digital, surge uma nova forma de entender as publicações científicas. As publicações deixam de estar compostas por uma única parte narrativa textual central e começam a ser entendidas como um conglomerado de ativos/assets de pesquisa que tornam visível tanto os resultados quanto o processo para lográ-los.

O conceito de publicação ampliada surge da evolução da publicação digital e de diferentes abordagens à comunicação acadêmica (ver Figura 9). Estas abordagens vão sendo inseridas segundo os avanços tecnológicos e a inserção de elementos como as tecnologias da *web* semântica.

Figura 9 Desenvolvimento temporal da publicação científica ampliada.



Fonte: Elaboração própria

Além das possibilidades introduzidas por Lancaster (1995) anteriormente mencionadas, existem na literatura duas abordagens da publicação ampliada: a publicação modular e a publicação semântica. Não necessariamente estas duas abordagens devem ser consideradas opostas, senão uma evolução a uma da outra. Ambas abordagens oferecem formas de aumentar a interação entre leitores e informação científica e apresentam formas de incrementar a compreensão e a consequente reutilização da informação.

A ideia de apresentar um artigo modular surge da ruptura com a tradicional estrutura linear textual (KIRCZ, 1998). Kircz (1998, 2002) considera o artigo científico como um conjunto de módulos, cada um deles entendido como “[...] uma representação textual, pictórica ou de outro tipo de uma quantidade de informação que é por si mesma suficientemente compreensível para transmitir significado ao leitor.” (KIRCZ, 2002, p. 29).

Essa proposta retoma o conceito de “informação como coisa” de Buckland (1991), onde qualquer elemento pode ser informativo desde que transmita um significado ao leitor, aliás, desde que ative o processo cognitivo dele. Os módulos podem ser elementares (abstract, imagem, corpo do texto, etc.) ou complexos. Nos complexos há dois tipos: compostos (uma agregação de módulos de independentes) e aglomerados (centrados em um conceito que se expande nos elementos restantes).

De fato, uma das características oferecidas pelo ambiente *web* é a possibilidade de compor ativos/assets científicos que depois possam ser partes de outras publicações, como afirmam Casati, Giunchiglia e Marchese, (2007) na proposta de *Scientific Knowledge Objects* (SKOs).

Kircz (2002) exemplifica o módulo composto com uma casa que possui módulos independentes (porta, parede e teto). Por sua vez, o módulo aglomerado seria um conjunto de portas. Cada módulo deve possuir uma série de metadados que identifiquem o autor, a data de criação e a integração modular correspondente.

Nesse sentido, a unidade de comunicação científica pode ser repensada sem que exista uma discriminação entre formatos (VAN DE SOMPEL *et al.*, 2004). Cada unidade deve possuir uma descrição por meio de normas e padrões e contar com cinco componentes estruturais (tipo semântico, tipo de mídia,

formato de mídia, localização na rede, acesso por diferentes repositórios). Devido à dispersão das unidades na rede, sugere-se uma descrição mediante um PID (p. ex. DOI) que garanta a conexão do recurso e dos metadados. No entanto, percebe-se em comunicação acadêmica uma preferência pela localização do ativo/asset mediante um localizador uniforme de recurso (URL) (VAN DE SOMPEL; KLEIN; JONES, 2016), especialmente quando se fala de conjuntos de dados (PARK; WOLFRAM, 2017).

Posteriormente, começam a ser utilizados elementos das tecnologias da *Web Semântica* na elaboração das publicações científicas. Marcondes (2005, 2011) propõe um modelo semântico para publicações digitais composto por duas fases: a primeira baseada em uma ontologia na qual o autor expressaria as relações entre as entidades do texto; a segunda baseada na interface do sistema de submissão para adicionar as principais afirmações do texto junto com os metadados bibliográficos.

Nessa mesma linha, define-se o “Pacote de Publicação Científica (*Scientific Publication Package* - SPP)”, que encapsularia dados brutos e derivados, algoritmos, *software*, fórmulas matemáticas (p. ex., anotadas em MathML⁵²), publicações textuais e todos os metadados associados, precisando de uma nova arquitetura da informação científica (HUNTER, 2006).

O SPP é identificado como um objeto digital composto e é representado como um pacote de *Resource Description Framework* (RDF). Cada tipo de informação contida nele é considerado como um átomo do pacote. A identificação mediante metadados é realizada tanto para a informação atômica quanto para o SPP.

Esse conceito de agregação também é discutido por Bechhofer *et al.* (2010) que definem o *Research Object* como uma unidade de conhecimento científico agregado. O objetivo não é apenas agrupar a informação essencial da investigação, seus resultados e métodos de análise de dados, senão também incluir as pessoas envolvidas na investigação. Isto serve como destaque da componente social dos estudos científicos atuais.

Esse componente social permite elaborar o discurso científico e fortalecer a transmissão do conhecimento. Por isso, De Waard (2010) destaca a

⁵²MathML: <https://www.w3.org/Math/>

necessidade de estruturação desse discurso e sua conexão e integração com dados nas publicações para conhecer as origens da pesquisa. Uma forma de realizar esta estruturação do conhecimento é mediante anotação semântica, por exemplo em textos escritos em LaTeX, que permite a estruturação das diferentes retóricas na pesquisa (DE WAARD *et al.*, 2009).

Attwood *et al.* (2009) destacam como as entidades na área de Biologia podem ser anotadas baseando-se em ontologias (como a *RNA ontology* ou a *GEN ontology*). PLoS também utiliza anotações como parte da ampliação semântica das publicações, assim as publicações podem ser ampliadas com:

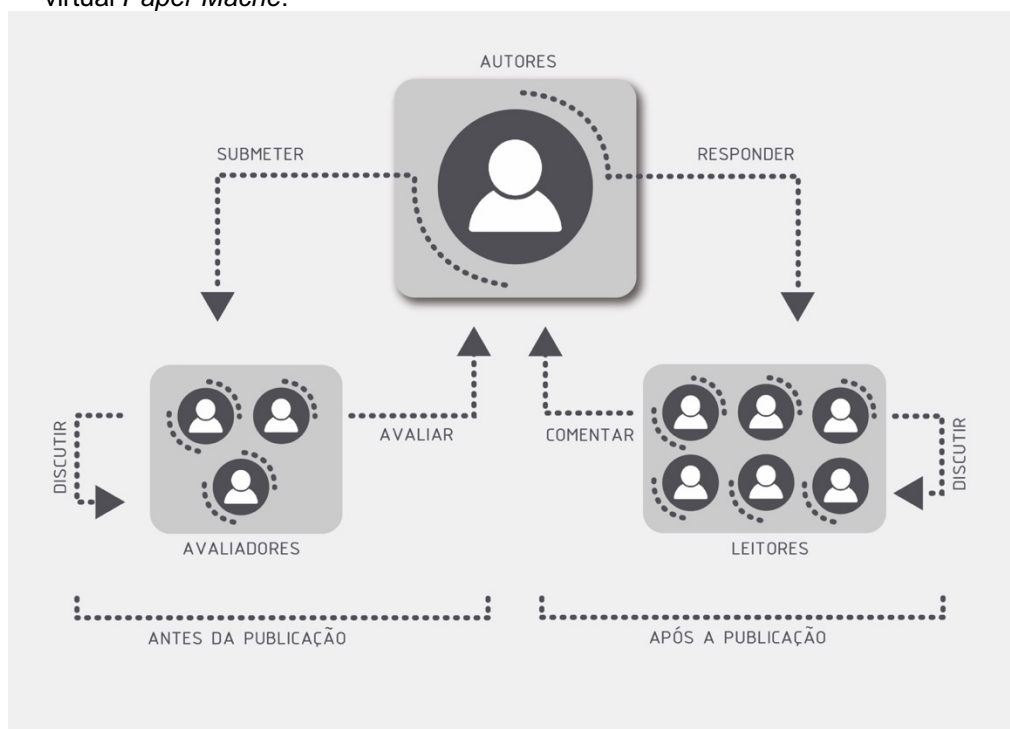
[...] DOIs e hyperlinks; anotações de termos textuais (doença, habitat, organismo, proteína, taxon, etc) com links a fontes externas de informação; figuras interativas, lista de referências reorganizável; resumo de documentos com um resumo do estudo; nuvem de *tags* e análise de citação; janelas *mouse-over* para exibir os principais argumentos de uma fonte citada; e uma árvore de *tags* para traçar termos relacionados semanticamente. (ATTWOOD *et al.*, 2009, p. 323).

Para a leitura destas ampliações os autores propõem o uso de Utopia Documents⁵³, um leitor de PDF que permite ver as anotações semânticas do documento.

Brammer *et al.* (2011) apresentam um sistema para criar artigos dinâmicos e executáveis chamado *Paper Mâché*. Baseiam-se no uso estendido de máquinas virtuais em artigos de pesquisa, que permitirá aos revisores, autores e leitores interagir com o documento e os suplementos que esse contenha. Esses processos de comunicação, ilustrados na Figura 10, fomentam a melhora tanto do documento final quanto dos ativos/assets associados, pois permitem os comentários sobre o documento. Além disso, facilitam acrescentar interações (comentário-resposta) entre o autor e o leitor, outorgando dinamismo pós-publicação.

⁵³ Utopia Documents: <http://utopiadocs.com/>

Figura 10 Fluxos de comunicação entre autores, avaliadores e leitores na máquina virtual *Paper Mâché*.



Fonte: Adaptação própria a partir de Brammer *et al.* (2011, p.660).

Também, a proposta *Article of the Future* (HUNTER, 2006) pretende melhorar a apresentação dos artigos na editora Elsevier, enriquecer seu conteúdo do artigo e adicionar contexto. Essa adição de contexto é realizada mediante bases do conhecimento, como exemplificado nos vídeos do capítulo anterior. A adição de contexto para um artigo pode estar centrada na literatura cinzenta, ou seja, aqueles elementos textuais (atas dos eventos, folhetos ou pôsteres) ou não textuais que ficaram fora do ciclo de publicação científica, mas que tenham sido apoio da publicação central (FARACE *et al.*, 2013).

Nas Humanidades, Heuvel *et al.* (2009) põem em prática as publicações ampliadas no projeto Veteran Tapes Project⁵⁴ ⁵⁵. Para esse projeto, realizam-se uma série de entrevistas com veteranos das guerras que os Países Baixos têm participado. A ideia do projeto foi fornecer os resultados não só como documento textual, mas também apresentando os fragmentos de áudio como forma de aportar humanismo e conexão entre o leitor e os testemunhos dos veteranos.

⁵⁴ Veteran Tapes Project: <http://www.watveteranenvertellen.nl/> (site disponível apenas em neerlandês).

⁵⁵ Veteran Tapes Enhanced Publication: https://youtu.be/Z_XJXDjxAj4 (vídeo em neerlandês com legenda em inglês).

Nesse sentido, Jankowski *et al.* (2012) destacam a necessidade de que os acadêmicos nas áreas das Ciências Sociais e Humanidades aproveitem os recursos da *Web 2.0* nas suas publicações. Por exemplo, as visualizações em livros são recursos tradicionais das Humanidades que não se traspassaram à publicação editorial na era digital. Portanto, os autores propõem um projeto baseado em Wordpress para ampliar quatro monografias nessas áreas que consideram os seguintes elementos adicionais: “(1) recursos suplementares (p.ex., *links*, blogs, apêndices de capítulos, perfis de autor); (2) elementos visuais em cor (animações, figuras, tabelas); (3) *hiperlinks* internos e externos aos textos dos monográficos; (4) atualização de materiais; (5) recursos de pesquisa.” (JANKOWSKI *et al.*, 2012, p. 8, tradução própria).

Breure, Voorbij e Hoogerwerf (2011) aprofundam mais no conceito de publicação ampliada, estendendo-o até o conceito de *Rich Internet Publication* (RIP), pois entendem que a publicação no ambiente *web* deve ir além da união dado-texto, como analisado no capítulo anterior.

Breure (2014) retoma o conceito de RIP e apresenta um modelo conceitual para transformar uma publicação tradicional em uma RIP. Para isso, estabelece quatro passos: a criação do foco (seleção da informação mais importante), a visualização do foco (seleção do formato ótimo para essa informação), a reestruturação do conteúdo (ruptura da estrutura linear do artigo tradicional), e a adição de informação suplementar e dados (suplementos como adição ou integrados no texto). Ainda, o autor apresenta a publicação ampliada como pacote e sua conexão com serviços na nuvem, o artigo criado em um sistema de gestão de conteúdo (CMS) e o e-book em formato EPUB3 como possibilidades reais para a criação de uma publicação ampliada.

No contexto dos e-books, Carbonell-Carrera *et al.* (2016) exploram a incorporação de objetos 3D em livros digitais, por exemplo, mediante códigos QR, abrindo outra porta ao enriquecimento da publicação e à interação leitor-conteúdo.

Seguindo na ideia da publicação ampliada, Bardi e Manghi (2014) identificam cinco tipos de modelos de dados recorrentes na estruturação de publicações ampliadas, a saber:

- **Partes embutidas:** descrevem a adição de arquivos suplementares que carecem de metadados descritivos. Consequentemente, não podem ser recuperados ou compartilhados individualmente e carecem de conteúdo semântico. Um exemplo em sistemas de informação é a agregação de materiais a um elemento em repositórios DSpace ou em OJS;
- **Partes de texto estruturado:** descrevem publicações cujo parte narrativa tem sido estruturada em seções interconectadas. Um exemplo é a proposta de artigo modular de Kircz (1998) ou a construção de publicações da editora PLOS;
- **Partes de referência:** descrevem as publicações ampliadas, cuja adição de conteúdo depende de links a objetos externos. Normalmente, esse *link* é realizado mediante uma URL o que pode ocasionar erros de “*broken link*”, por isso é recomendável utilizar PIDs para conectar os objetos descritos. Um exemplo é a inserção de resumos em vídeo em publicações científica, para o que a maioria das editoras se apoia em plataformas externas como YouTube, Vimeo ou Figshare, que nem sempre oferecem uma identificação persistente (MARÍN ARRAIZA; VIDOTTI, 2018);
- **Partes executáveis:** descrevem a ampliação que requer um *software* e dados para executar uma experiência;
- **Partes geradas:** descrevem a ampliação que, mediante a utilização de um *input* e uma aplicação para a geração de resultados por meio de um *workflow*. Um exemplo seriam as tabelas de dados que podem mudar dinamicamente, dependendo das atualizações de entradas de dados de pesquisa.

No que respeita à integração das publicações ampliadas em sistemas de informação, no relatório DRIVER – II, Verhaar (2008) identifica dez requerimentos e especificações estruturais para publicações ampliadas e para seu acesso em contextos institucionais:

1. Deve ser possível em qualquer momento especificar as partes componentes de uma publicação ampliada;
2. Tanto a publicação ampliada quanto seus componentes devem estar disponíveis como recursos *web* que possam ser referenciados via identificadores uniformes de recursos (URIs);
3. Deve ser possível agregar objetos digitais compostos à publicação ampliada;
4. Deve ser possível acompanhar as diferentes versões das publicações ampliadas, tanto da publicação ampliada como um todo quanto das suas partes constituintes;

5. Deve ser possível registrar as propriedades básicas da publicação e dos outros recursos que estão adicionados a ela;
6. Deve ser possível registrar a autoria da publicação ampliada e das suas partes constituintes;
7. Deve ser possível assegurar a preservação de longo prazo das publicações ampliadas;
8. Deve ser possível registrar as relações entre os recursos *web* que fazem parte da publicação ampliada;
9. Instituições que oferecem acesso a publicações ampliadas devem assegurar que elas possam ser descobertas;
10. Instituições que oferecem acesso a publicações ampliadas devem assegurar que estas estejam disponíveis como documentos baseados na norma *Open Archives Initiative – Object Exchange and Reuse* (OAI-ORE). (VERHAAR, 2008, p. 15–22 grifos e tradução próprios).

A norma OAI-ORE permite descrever as agregações efetuadas no entorno *web* as quais criam objetos digitais compostos que “[...] podem combinar recursos distribuídos com vários tipos de mídia, incluindo texto, imagens, dados e vídeo.” (OAI-ORE), com o objetivo de expor o conteúdo enriquecido de forma padronizada.

Todavia, Bardi e Manghi (2015) apresentam um marco de trabalho para publicações ampliadas com nove requerimentos. Esses requerimentos dividem-se em quatro gerais para publicações científicas (suportar diferentes *back-ends* para armazenamento de dados; oferecer definição de dados, manipulação e acesso linguístico; possibilitar o compartilhamento de dados; e suportar a portabilidade dos dados) e cinco específicos para publicações ampliadas (suportar a integração de fontes de dados heterogêneos; suportar o gerenciamento de fontes dinâmicas de dados; suportar a integração de conteúdo; permitir a customização do modelo de dados para publicações ampliadas; e suportar o enriquecimento e curadoria de conteúdo).

5.1.1. REPERCUSSÃO DA PUBLICAÇÃO AMPLIADA NO CONTEXTO DE PESQUISA BRASILEIRO

No contexto brasileiro, a temática das publicações ampliadas é ainda incipiente com só duas teses defendidas sobre esta temática na área da Ciência da Informação até o ano 2018⁵⁶. Araya (2014) apresenta a temática da agregação de informação como parte da comunicação científica. Sales (2014) retoma a ideia de integração semântica das partes de uma publicação, destacando as dificuldades existentes nos diferentes entornos de publicação, principalmente nos catálogos *online* (*Online Public Access Catalog* - OPAC) de bibliotecas.

Sales e Sayão (2015) apresentam um modelo de publicação científica baseado nos relatórios do projeto DRIVER-II. No modelo, específico para a área de Ciência Nuclear, os dados de pesquisa e as publicações estão ligadas mediante relações semânticas seguindo uma taxonomia própria.

Mucheroni, Silva e Paletta (2015) reconhecem a publicação ampliada como um objeto dinâmico de informação e conhecimento e destacam a multimodalidade da mesma, já que pode conter agregação de diferentes elementos multimídia e desde diversos canais de publicação. Para os autores,

⁵⁶ O levantamento bibliográfico para esta tese foi realizado até dezembro 2018. Há teses e dissertações publicadas ao redor da temática da publicação de dados, políticas de dados, gestão de dados e Biblioteconomia de dados. Porém, não com incidência direta na temática da publicação ampliadas. Alguns exemplos dessas teses e dissertação são:

- JORGE, Vanessa de Arruda. **Abertura e compartilhamento de dados para pesquisa nas situações de emergência em saúde pública: o caso do vírus Zika**. Rio de Janeiro, 2018. 263 f. Orientadora: Sarita Albagli. Tese (Doutorado em Ciência da Informação) – Escola de Comunicação, Universidade Federal do Rio de Janeiro, Instituto Brasileiro de Informação em Ciência e Tecnologia, Rio de Janeiro, 2018.
- MONTEIRO, Elizabete Cristina de Souza de Aguiar. **Direitos autorais nos repositórios de dados científicos: análise sobre os Planos de Gerenciamento dos Dados**. 115 f. Dissertação (Mestrado) - Faculdade de Filosofia e Ciências, Universidade Estadual Paulista, Marília, 2017.
- SEMELER, Alexandre Ribas. **Ciência da informação em contextos de e-science: bibliotecários de dados em tempos de Data Science**. 168 f. Tese (doutorado) - Universidade Federal de Santa Catarina, Centro de Ciências da Educação, Programa de Pós-Graduação em Ciência da Informação, Florianópolis, 2017.
- COSTA, Máira Murrieta. **Diretrizes para uma política de gestão de dados científicos no Brasil**. 2017. 288 f., il. Tese (Doutorado em Ciência da Informação)—Universidade de Brasília, Brasília, 2017.
- MACHADO, Denise Ramires. **Dados de pesquisa em repositório institucional: o caso do Edinburgh DataShare**. Dissertação—Porto Alegre: Universidade Federal do Rio Grande do Sul, 2015.

uma publicação ampliada não pode prescindir dos *Linked Data* como tecnologia da *web* semântica “[...] para viabilizar o relacionamento de conteúdos ou dados integrados.” (p.13).

Rodrigues e Sant’Ana (2016) incidem no volume de produtos de pesquisa de diferentes naturezas que compõem o desenvolvimento de uma pesquisa de pós-graduação. Os autores realizam um estudo de caso de uma tese que possui 14 tipos de ativos/assets (texto da tese, arquivo de texto para a Qualificação, arquivo de Backup, captura de tela para coleta de dados, planilha eletrônica para coleta de dados, planilha eletrônica para tabelas e quadros, desenho vetorial, recursos visuais, *script*, modelagem entidade-relacionamento e dicionário de dados).

Se pensarmos no polo morfológico de qualquer pesquisa e consideramos os dados de pesquisa de cada área, como mencionado no início deste capítulo, chegamos à conclusão de que qualquer projeto de pesquisa vai ter uma grande produção de ativos/assets. A publicação de todos estes ativos/assets requer, porém, de uma estruturação e da descrição de cada um dos elementos. Estes dois pontos serão analisados a seguir.

5.2. PROPOSTA DE ESTRUTURA DE UMA PUBLICAÇÃO CIENTÍFICA AMPLIADA

A partir do anterior, consideramos importante definir uma estrutura genérica da publicação ampliada. Esta estrutura servirá como base para o modelo final. Para a construção desta estrutura, foram considerados os módulos e elementos definidos por Kircz (1998) na proposta de artigo modular.

Os módulos são contrastados com o esquema para artigo científico da Associação Brasileira das Normas Técnicas da NBR 6022 para determinar a posição do elemento no texto. Embora o transfundo de uma publicação ampliada seja a ruptura da linearidade, esta ruptura afeta apenas a leitura do texto e não a sua estruturação. Na ABNT (2003) contemplam-se três possíveis posições: pré-textual, pós-textual e textual. É importante assinalar que a posição “textual” se refere à integração no corpo do texto e não à natureza do elemento.

O Quadro 12 apresenta a combinação da estrutura modular de Kircz (1998) contrastado com a da (ABNT) (2003) e inclui uma proposta de localização

das entidades extraídas no capítulo anterior após a análise dos ambientes de publicação.

Quadro 12 Combinação dos elementos do artigo modular para a estrutura da publicação ampliada.

Módulo (KIRCZ, 1998)	Posição (ABNT, 2003)	Elemento (KIRCZ, 1998)	Entidades adicionais
Meta-informação	Pré-textual	Informação bibliográfica	Autor
	Pré-textual	Conteúdo (Sumário)	Texto
	Pré-textual	Termos de indexação	Texto – palavras chave
	Pós-textual	Referências bibliográficas	Texto - referências
	Pós-textual	Agradecimentos	Texto - agradecimentos
	Pré-textual	Resumo	Texto - resumo
Objetivos e configurações	Textual	Definição do problema	Texto - objetivos
		Contextualização	Base de conhecimento
Métodos	Textual	Modelos	Texto - métodos Base do conhecimento Quadro
		Cálculos	Código <i>Workflows</i>
Medições	Textual	Descrição da ferramenta	Texto - medições
		Uso da ferramenta	<i>Software</i>
Resultados	Textual	Dados brutos	Texto - resultados Conjunto de dados brutos
		Dados processados	Texto - resultados Planilhas de dados Tabela Áudio Arquivo multimídia Figura Imagem Vídeo Quadro <i>Workflow</i>
Discussões	Textual	Dos resultados objetivos	Texto – Discussões
		Das interpretações subjetivas	Quadro Tabela Imagem Figura
Conclusões	Textual	Respostas ao módulo “Objetivos e configurações”	Texto - Conclusões
		Sugestões para futuras pesquisas	
Informação complementar (inserção própria)	Pós-textual (inserção própria)		Todas as anteriores à exceção das sub-entidades do texto

Fonte: Elaboração própria a partir de diversas fontes.

Esta estrutura contempla um total de quatorze entidades, uma das quais, a entidade “texto” divide-se em 10 sub-entidades. Além disso, a informação

bibliográfica é definida pela entidade autor. Portanto, o modelo final contará com 25 entidades.

Ainda, é relevante assinalar a padronização da entidade “Texto-referências” seguindo o OpenCitations Data Model⁵⁷, baseado na aplicação de tecnologias semânticas para a publicação de dados bibliográficos e de citação. Esse modelo integra as recomendações da LERU (2018) para a abertura dos processos de publicação como rota para a ciência aberta nas universidades:

Na publicação e abertura dos dados de apoio à investigação, é altamente desejável que um certo número de normas identificadores/processos sejam usados para ajudar na descoberta e reutilização de resultados abertos - ORCID para identificar os autores; FundRef5, uma taxonomia comum de nomes de financiadores de investigação; DOIs para identificar e localizar publicações; DataCite para identificar e localizar conjuntos de dados; OpenCitations, um movimento para promover a disponibilidade irrestrita de citações acadêmicas dados, e para disponibilizar esses dados. (LERU, 2018, p.5).

Considerando tanto esta recomendação quanto à necessidade de um identificador para a estruturação do modelo de entidade-relacionamento, passaremos agora a analisar os PIDs no âmbito da publicação.

5.3 IDENTIFICADORES PERSISTENTES

Nos últimos anos, os sistemas de PIDs evoluíram e a sua utilização aumentou para atenuar o problema dos *hiperlinks* quebrados (*link rot*), principalmente quando se trata de informações científicas e culturais acessíveis na *web*. Portanto, a ideia por trás de um sistema PID é oferecer uma referência duradoura a uma entidade (física, digital ou abstrata), por exemplo, um documento digital, página web, pessoa ou instituição. Alguns sistemas PID bem conhecidos são Archival Resource Key⁵⁸ (ARK), Digital Object Identifier (DOI), Handle system⁵⁹, Persistent Uniform Resource Locator (PURL), Uniform Resource Name (URN) e Open Researcher and Contributor ID (ORCID), sendo esse último exclusivamente para pessoas.

⁵⁷ OpenCitation Data Model:

https://figshare.com/articles/Metadata_for_the_OpenCitations_Corpus/3443876

⁵⁸ ARK: https://n2t.net/e/ark_ids.html

⁵⁹ Handle System: <https://www.handle.net/>

Um PID possui uma série de metadados associados que são legíveis por máquinas, portanto, identificam o objeto e não a localização dele, como acontece com os URL (DAPPERT *et al.*, 2017). Um PID pode ser implementado seguindo o protocolo HTTP o que o torna acionável e permite dirigir ao leitor à página onde o recurso pode ser encontrado (*landing page*) (LÓPEZ-PELLICER *et al.*, 2016; VAN DE SOMPEL *et al.*, 2014).

No entanto, é importante ressaltar que a persistência está relacionada ao serviço oferecido pelo sistema e não ao identificador em si. Isto significa que uma entidade se compromete a manter o identificador resolúvel. O identificador leva os utilizadores aos serviços que garantem a referência (KUNZE, 2013). Por exemplo, os ARKs podem ser mantidos e resolvidos através do serviço EZID⁶⁰ (Universidade da Califórnia); as DOIs são geridas pela International DOI Foundation⁶¹ e pelas suas agências de registo correspondentes, tais como Crossref e DataCite e centros de dados; os Handles são gerenciados pela Corporation for National Research Initiatives⁶² (CNRI); e o sistema PURL foi desenvolvido pelo Online Computer Library Center⁶³ (OCLC).

O uso de PIDs em arquivos e sistemas de informação de pesquisa está hoje em dia generalizado, e os PIDs são considerados uma parte crucial do processo de preservação. Por isso, várias instituições de pesquisa criaram centros de dados para registrar as PIDs, a fim de preservar seus conteúdos e torná-los internacionalmente encontráveis e editáveis. O centro de dados (*data centre*) encarregado da emissão de um PID—por exemplo uma biblioteca de pesquisa— deve também realizar as tarefas de curadoria digital para garantir a manutenção dos metadados do ativo/asset (JOHNSTON *et al.*, 2018).

De fato, todas as diretrizes atuais indicam o uso de PIDs, como é o caso do primeiro princípio FAIR⁶⁴: “Os (meta)dados são atribuídos com identificadores globalmente únicos e persistentes”. O relatório “Turning FAIR into a reality” propõe um modelo de FAIR Data Objects (HODSON *et al.*, 2018), cujas camadas consistem em metadados, padrões, identificadores e dados (ver Figura 11).

⁶⁰ EZID: <https://ezid.cdlib.org/>

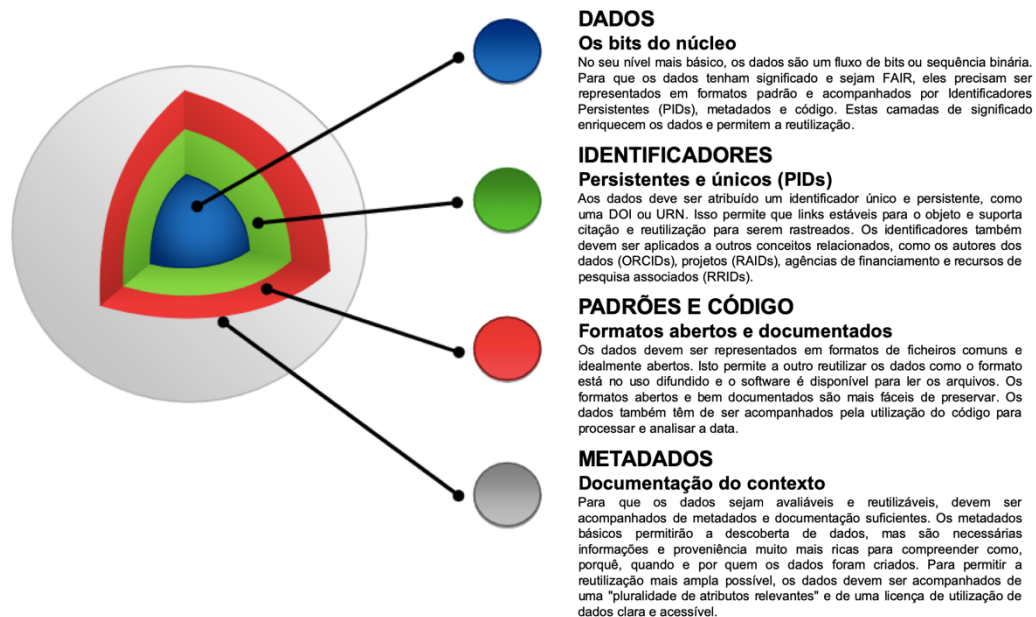
⁶¹ International DOI Foundation: <https://www.doi.org/>

⁶² CNRI: <https://www.cnri.reston.va.us/>

⁶³ OCLC: <https://www.oclc.org>

⁶⁴ Princípio FAIR F1: <https://www.go-fair.org/fair-principles/f1-meta-data-assigned-globally-unique-persistent-identifiers/>

Figura 11 Modelo do FAIR Data Object



Fonte: Hodson *et al.* (2018, p.38, tradução própria).

Para compreender o FAIR Data Object os autores expõem que:

Os dados precisam ser acompanhados por Identificadores Persistentes (PIDs) e metadados básicos de descoberta para que possam ser encontrados, usados e citados de forma confiável. Além disso, os dados devem ser representados em formatos padronizados - e idealmente abertos - e ser ricamente documentados utilizando normas e vocabulários de metadados adotados pelas comunidades de investigação para permitir a interoperabilidade e a reutilização. O compartilhamento de código também é fundamental e deve incluir não apenas a fonte em si, mas também a documentação apropriada, incluindo declarações legíveis por máquinas sobre dependências e licenças.⁶⁵ (HODSON *et al.*, 2018, p.39, tradução própria).

Além da identificação, os PIDs são utilizados para agregar ativos/assets. Os resultados da investigação com um PID são mais fáceis de rastrear, o que facilita as atividades de monitorização da investigação. No entanto, como já foi

⁶⁵ Tradução própria do original em inglês: Data need to be accompanied by Persistent Identifiers (PIDs) and basic discovery metadata to enable them to be reliably found, used and cited. In addition, the data should be represented in common – and ideally open – file formats and be richly documented using metadata standards and vocabularies adopted by the given research communities to enable interoperability and reuse. Sharing code is also fundamental and should include not just the source itself but also appropriate documentation, including machine-actionable statements about dependencies and licencing.

mentionado, a persistência não é uma característica intrínseca de um PID, mas está relacionada com o serviço subjacente.

Nesse sentido, pode-se falar de "identificadores confiáveis" que são —além de persistentes— únicos, descritivos, interoperáveis e governados. O consórcio ODIN (ORCID e DataCite Interoperability Network) propôs as seguintes características para os identificadores confiáveis:

1. São únicos à escala mundial.
2. Resolvem como URI's HTTP persistentes com suporte para negociação de conteúdo.
3. Eles vêm com metadados que descrevem suas propriedades mais relevantes, incluindo um conjunto mínimo de elementos de metadados comuns.
4. São interligáveis.
5. São interoperáveis com outros identificadores através de elementos de metadados que descrevem a sua relação.
6. São geridos através de uma organização empenhada e sustentada e de um processo de governação. (ODIN CONSORTIUM, 2013, p. 19)

Além disso, os PIDs servem como mecanismos de crédito e atribuição, ao citar os resultados da investigação (MCMURRY *et al.*, 2017).

As infraestruturas científicas —por exemplo, repositórios, supercomputadores ou equipamentos físicos— também podem receber um PID (WILKINSON *et al.*, 2016). Atribuir um PID a esse tipo de infraestrutura garante a sustentabilidade da mesma e a reprodutibilidade da pesquisa (CHEN *et al.*, 2019). Portanto, vemos que existem diversas entidades dentro do processo de pesquisa que podem ser identificadas mediante um PID. Dappert *et al.* (2017) as agrupam em cinco categorias:

1. **Agentes:** indivíduos (como investigadores ou curadores); organizações (como financiadores, instituições de investigação, centros de dados, editoras e instituições de arquivo); e outras entidades jurídicas (como consórcios criados para colaborar na investigação);
2. **Recursos**⁶⁶: publicações, dados e outros artefatos de investigação, tais como cadernos de laboratório, *software*, equipamento ou espécimes físicos;
3. **Declarações de direitos:** contratos de subvenção, licenças, patentes;
4. **Eventos:** processos relevantes para a proveniência dos recursos, tais como criação, curadoria, acesso, reivindicação, atualização, citação, revisão;
5. **Entidades derivadas:** tais como projetos, que podem ser vistos como um agregado da entidade jurídica envolvida, da organização que a financia, dos resultados produzidos e das declarações de direitos que a regem.⁶⁷ (DAPPERT *et al.*, 2017, p. 6, tradução própria).

⁶⁶ A palavra "recurso" aqui é sinônima de ativo/asset utilizada ao longo do texto. Manteve-se a palavra recurso por ser uma tradução direta.

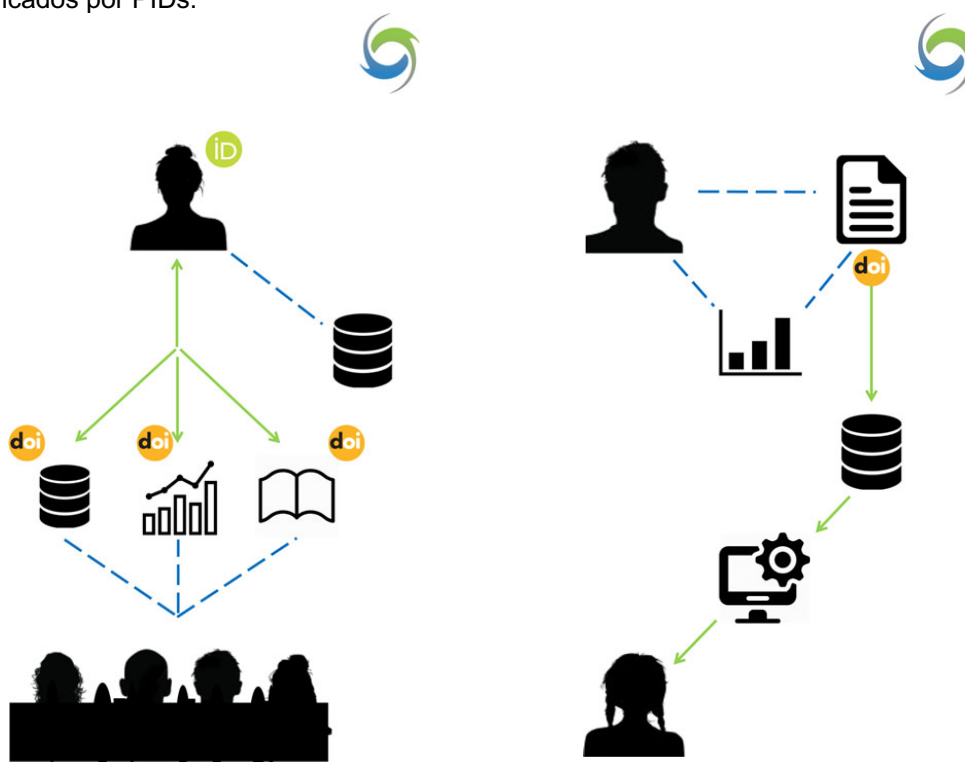
⁶⁷ Tradução própria do original em inglês: (1) Agents: individuals (such as researchers or curators); organisations (such as funders, research institutions, data centres, publishers and archival institutions); and other legal entities (such as consortia that are created to collaborate on research). (2) Resources:

Para efeitos desta tese, serão consideradas a primeira (agentes), a segunda (recursos, especialmente os ativos/assets) e a quarta (eventos) categorias. A categoria “Agentes” abrange a entidade “Autor”, cujo PID principal será o Open Researcher and Contributor iD (ORCID iD); a categoria “recursos” abrange o resto de entidades de ativos/assets de pesquisa e a categoria “eventos” representa as ações entre entidades.

A utilização de PIDs também é uma forma de conectar aos pesquisadores com os seus resultados de pesquisa e serve para identificar ativos/assets “perdidos na rede” (KLEIN; VAN DE SOMPEL, 2017). Esta conexão (ver Figura 12) identifica a autoria do ativo/asset, sua procedência e favorece a citação; consequentemente, os pesquisadores se mostram mais confiantes na hora de compartilhar e publicar seus dados de pesquisa.

publications, data, and other research artefacts, such as lab notebooks, software, equipment, or physical specimens. (3) Rights statements: grant agreements, licenses, patents. (4) Events: processes that are relevant to the provenance of resources, such as creation, curation, access, claiming, updates, citation, review. (4) Derived entities: such as projects, that can be seen as an aggregation of the legal entity involved, the organisation that funds it, the outputs produced and the rights statements that govern it.

Figura 12 Compartilhamento livre, descoberta e reutilização dos ativos/assets de pesquisa identificados por PIDs.



Fonte: Dappert *et al.* (2017, p. 3)

Porém, a identificação de ativos/assets de pesquisa com PIDs não é um movimento tão recente. Alguns centros de dados, como a Biblioteca Nacional Alemã de Ciência e Tecnologia (TIB), oferecem serviços de registro de PIDs há mais de catorze anos (KRAFT *et al.*, 2017). Existem diferentes sistemas de PID para objetos digitais, e o quadro a seguir (ver Quadro 13) apresenta alguns dos mais utilizados, suas características principais e um exemplo da sua anatomia.

Quadro 13 Alguns identificadores persistentes para objetos digitais.

Identificador	Características
Archival Resource Key - ARK	Mantido pela California University Library; Identificador persistente de propósito múltiplo; Pode ser resolvido usando <i>software</i> de código aberto; Conecta o objeto, um breve registro de metadados e um servidor de manutenção. Exemplo de anatomia: https://texashistory.unt.edu/ark:/67531/metaph346793/
Persistent uniform resource locator - PURL	Administrado por Online Computer Library Center (OCLC); Baseado no protocolo HTTP, garante a resolução, mas não tem metadados; Separa o nome do documento da sua ubiquação; Aponta para um serviço de resolução intermediária entre a URL e PURL;

Identificador	Características
	Utilizado na descrição de ontologias e vocabulários controlados. Exemplo de anatomia: http://purl.org/dc/terms/
Handle	Desenvolvido pela Corporation for National Research Initiatives (CNRI); Identifica objetos digitais (vídeos, imagens, revistas, etc.); Os handles são únicos e persistentes; As operações no sistema Handle devem ser autorizadas; Base de outros sistemas, como o DOI; Interoperabilidade completa com o DOI. Exemplo de anatomia: http://hdl.handle.net/11449/168868
Uniform Resource Name - URN	Parte da arquitetura da informação URL + URC + padrão de metadados; Identifica o objeto, mas não garante a disponibilidade do objeto. Exemplo de anatomia: https://resolver.obvsg.at/urn:nbn:at:at-ubtuw:3-5866
Digital Object Identifier - DOI	Nome digital unívoco e permanente de um objeto; Mostra a ubiquação do objeto e armazena os metadados descritivos principais; Possui um modelo de dados associado que garante a interoperabilidade e a descrição através de metadados (DOI Kernel Metadata); Segue a norma ISO 26324:2012. Exemplo de anatomia: https://doi.org/10.1371/journal.pone.0211907

Fonte: Elaboração própria.

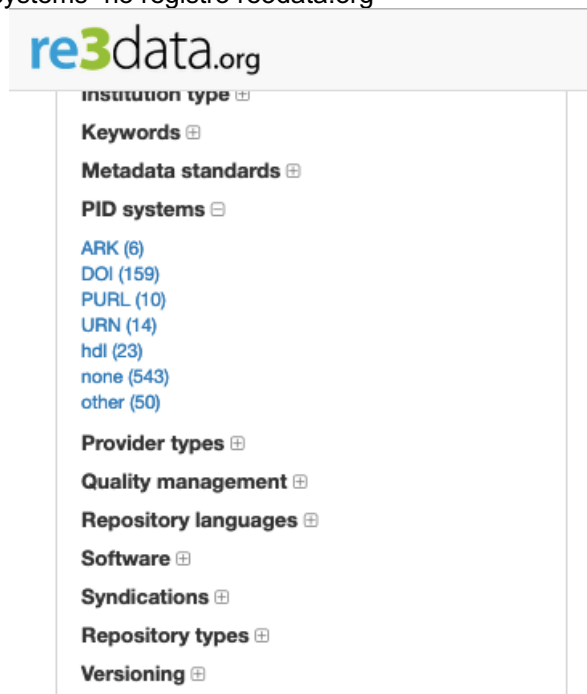
A escolha de um sistema de PID ou outro sistema depende, em parte, do grau de abertura do ativo/asset e da infraestrutura utilizada.

Como indicado no Capítulo 3, os ativos/assets de pesquisa nem sempre podem ser publicados em aberto. Para ativos/assets fechados, cujos metadados também são fechados, recomenda-se o uso de identificadores Handle ou ARK. Nesse caso, um DOI não é apropriado, já que não existe uma *landing page* que possa ser acessada. No caso dos metadados estarem disponíveis em aberto, embora o ativo/asset seja fechado ou de acesso restrito, pode ser usado um DOI.

Segundo a infraestrutura, pode ser utilizado um sistema de PIDs ou outro. Por exemplo, o registro de repositórios de dados de pesquisa re3data permite na busca filtrada escolher o sistema de PID desejado. A seguinte figura (Figura 13) mostra esse filtro para os resultados com a expressão de busca “biology”.

Podemos observar que a maioria dos repositórios ainda carecem de um sistema de identificação persistente.

Figura 13 Filtro “PID systems” no registro re3data.org



Fonte: Registro re3data.org. Acesso em: 10 março 2019

O Serviço de Dados Nacional Australiano (Australian National Data Service – ANDS) sugere o uso do DOI apenas para aqueles ativos/assets citáveis e sugere o uso de outros PIDs, por exemplo Handle, para teses, dissertações ou literatura cinzenta. De qualquer forma, como apresentado no quadro anterior, o sistema DOI está construído sobre o sistema Handle e, portanto, ambos PIDs são perfeitamente interoperáveis.

A recomendação do uso do DOI está contida em diversos documentos oficiais, como Open Science and its role in universities da LERU (2018), e nos resultados de projetos financiados com fundos da União Europeia como ORCID and DataCite Interoperability Network (ODIN) (2012-2014), Technical and Human Infrastructure for Open Research (THOR) (2015-2017) e FREYA (2017-2020). Além disso, o DOI está baseado em um modelo de dados que serve para garantir a persistência se o material for removido, reorganizado ou mercado; a interoperabilidade com outros dados de outras fontes; a extensibilidade através da adição de novos recursos por meio do gerenciamento de grupos de DOIs; a gestão única de dados para múltiplos formatos de saída; a gestão de classes de

aplicativos e serviços; e a atualização dinâmica de metadados, aplicativos e serviços (INTERNATIONAL DOI FOUNDATION, 2018).

Na Norma ISO 26324:2012, o DOI apresenta-se como uma combinação única de caracteres formada por um prefixo e um sufixo que pode estar conectada a um sistema de resolução para transformá-la em um endereço. Dentro da estrutura do DOI está permitido qualquer caráter do Conjunto Universal de Caracteres (Universal Character Set – UCS-2) da norma ISO/IEC 10646. Após o registro pela agência de registro de DOIs (p.ex. DataCite), é atribuído um prefixo individual. A configuração do sufixo é feita pela instituição científica, pela editora ou pelo repositório.

Como indica a International DOI Foundation, um DOI deve ser uma cadeia de caracteres opaca. Isto é, não podem ser inferidas informações partindo dessa cadeia de caracteres. Por isso, recomenda-se evitar a semântica nela. Informação sobre o ativo/asset tais como responsabilidade de gestão, direitos autorais o de propriedade intelectual devem ser descritos nos metadados.

Além disso, destaca-se que um DOI não substitui outros tipos de identificadores como International Standard Book Name (ISBN), International Standard Serial Number (ISSN) ou International Standard Name Identifier (ISNI). Nos padrões de metadados associados ao DOI, outros identificadores podem ser inseridos como “identificadores alternativos”, o que permite enriquecer as funcionalidades dos diferentes sistemas.

O Quadro 14 descreve um exemplo de anatomia de um DOI.

Quadro 14 Exemplo de anatomia do DOI

http://doi.org	10.1371	/journal.pone	.0211907
Serviço de resolver	Prefixo que identifica a instituição de registro (PLOS neste caso)	Sufixo 1 Identificador da instituição ou revista (PLOS One neste caso)	Sufixo 2 Recurso identificado (Artigo do periódico neste caso)

Fonte: Elaboração própria.

Seguindo as recomendações do projeto THOR (2013) e os critérios da Network of Expertise in Long-Term Storage of Digital Resources (NESTOR) (2013), a arquitetura de um sistema PID confiável deve garantir que os PIDs sejam:

- um nome e não um endereço (como acontece com o URL);
- unívocos;
- persistentes e destinados a ter uma vida útil superior à de qualquer sistema ou organização (geralmente com uma garantia de persistência de pelo menos 10 anos);
- resolúveis como um URI apoiado no protocolo HTTP;
- gerenciados por meio de um comitê organizador sustentável e com um processo de governança estipulado;
- identificados com um mínimo de metadados descritivos comuns a todos os elementos do sistema e que mostrem as propriedades mais relevantes;
- interconectáveis;
- interoperáveis com outros identificadores por meio de elementos que descrevam seus relacionamentos;
- indexados e pesquisáveis por seus elementos de metadados, juntamente com todos os outros identificadores confiáveis.

Com isto, para a manutenção do PID, especialmente do DOI, de um ativo/asset de pesquisa, é necessário que o sistema de atribuição cumpra com os seguintes requerimentos:

Garantir a existência e curadoria da página de alojamento (*landing page*): como indicado anteriormente, o DOI deve se referir a uma página onde o recurso esteja descrito e haja informação sobre como acessá-lo. A localização é resolúvel mediante o protocolo HTTP. O sistema DOI não aponta ao ativo/asset diretamente. Porém, existem outros sistemas que oferecem esta resolução direta, como é o ePIC⁶⁸.

Definir o tipo de ativo/asset: no registro de um DOI, deve ser reconhecido o interesse científico (ou artístico) de longo prazo do ativo/asset. Esse ativo/asset pode ser uma das entidades integradoras de uma publicação ampliada ou ativos/assets textuais e não-textuais, por exemplo, literatura

⁶⁸ ePIC: https://www.pidconsortium.eu/?page_id=74

cinzenta, objetos do patrimônio cultural, materiais de ensino e aprendizagem, etc. Dentro dos metadados obrigatórios deve ser incluído o tipo de recurso.

Determinar a granularidade: o DOI pode ser concedido em qualquer nível de granularidade (livro, capítulo, gráfico individual, etc.), dependendo da finalidade e da natureza técnica da plataforma. Para a estrutura de publicação ampliada apresentada nesta tese, a identificação será tanto da publicação completa quanto de cada uma das entidades integradoras.

Quadro 15 Granularidade na atribuição de DOI num artigo da PLOS.

Disposição	Elemento	Identificador persistente em http
Corpo do artigo	Artigo completo	https://doi.org/10.1371/journal.pone.0211907
	Figura 1	https://doi.org/10.1371/journal.pone.0211907.g001
	Figura 2	https://doi.org/10.1371/journal.pone.0211907.g002
	Figura 3	https://doi.org/10.1371/journal.pone.0211907.g003
	Figura 4	https://doi.org/10.1371/journal.pone.0211907.g004
	Figura 5	https://doi.org/10.1371/journal.pone.0211907.g005
Informação de suporte	Vídeo (formato .mp4)	https://doi.org/10.1371/journal.pone.0211907.s001
	Imagens (formato .zip)	https://doi.org/10.1371/journal.pone.0211907.s002
	Dados (formato .txt)	https://doi.org/10.1371/journal.pone.0211907.s003

Fonte: Elaboração própria.

O quadro anterior (ver Quadro 14) mostra um exemplo da granularidade de atribuição de DOI necessária para uma publicação ampliada.

Garantir a disponibilidade: o sistema deve garantir que o ativo/asset identificado com um DOI está acessível *online* em todos os momentos por meio de uma URL e do protocolo HTTP. Seguindo os princípios da *openness*, as restrições de acesso aos ativos/assets devem ser evitadas na medida do possível, e deve ser garantida o acesso da página de destino.

Garantir a qualidade: a fim de garantir a usabilidade a longo prazo, o registo DOI requer a aplicação de normas profissionais na criação de ativos/assets e a existência de metadados. Os ativos/assets de pesquisa devem

possuir uma qualidade (avaliada ou não por pares), ser citáveis e possuir um mínimo de metadados descritivos.

O seguinte quadro (ver Quadro 16) mostra os campos de metadados obrigatórios, recomendados e opcionais do padrão de metadados da DataCite, uma das agências de registro de DOIs.

Quadro 16 Campos de metadados no padrão DataCite

Campo	Propriedade
Obrigatório	• Identificador / URL • Criador • Título • Editor • Ano de publicação • Tipo de recurso
Recomendado	• Área • Contribuinte • Data • Identificador relacionado • Descrição • GeoLocalização
Opcional	• Idioma • Identificador alternativo • Tamanho • Formato • Versão • Licença • Referência de financiamento

Fonte: Elaboração própria a partir do DataCite Metadata Schema 4.3⁶⁹.

Para o modelo proposto nesta tese, os metadados da licença passam a ser considerados obrigatórios, pois só mediante esse atributo podemos garantir os termos de reuso do ativo/asset. Além disso, o criador do ativo/asset atribuído com um DOI deve assegurar que os conteúdos dos ativos/assets cumpram as regras gerais de boa prática científica.

Escolher um formato para o arquivo a longo prazo: o formato de arquivo dos ativos/assets é geralmente aberto. No entanto, se possível, devem ser escolhidos formatos cujo arquivamento a longo prazo possa ser assegurado de acordo com o estado da técnica no momento do registro. No Quadro 17, mostram-se alguns dos formatos recomendados para as entidades identificadas na estrutura de publicação ampliada.

⁶⁹ DataCite Metadata Schema 4.1.: <https://schema.datacite.org/meta/kernel-4.3/>

Quadro 17 Exemplos de formatos recomendados para arquivamento a longo prazo para algumas das entidades identificadas

Entidade	Formato recomendado
Texto	• PDF/A (*.pdf) • Texto sem formato (*.txt, *.asc, *.c, *.h, *.cpp, *.m, *.py, *.r usw) codificado como ASCII, UTF-8 o UTF-16 • XML (inclusive XSD/XSL/XHTML)
Conjuntos e planilhas de dados	• Comma-separated values (*.csv)
Áudio	• WAV (*.wav)
Figura ou imagem	Gráfico rasterizado • TIFF (*.tif) • Portable Network Graphics (*.png) • JPEG2000 • Digital Negative Format (*.dng) Gráfico vectorial • SVG sem JavaScript (*.svg) Gráfico CAD • AutoCAD Drawing (*.dwg) • Drawing Interchange Format, AutoCAD (*.dxf) • Extensible 3D, X3D (*.x3d, *.x3dv, *.x3db)
Vídeo	• FFV1 Codec in Matroska Container (*.mkv) • Motion JPEG 2000 (ISSO/IEC15444-4) (*.mkv) • AVI (*.avi)

Fonte: Elaboração própria com base na informação do Grupo “Preservação de dados digitais” da biblioteca da universidade ETH-Zürich⁷⁰.

Além destas considerações gerais sobre o formato, devem ser considerados aqueles formatos estabelecidos na comunidade científica do autor (por exemplo, *.mat para as comunidades que utilizem Matlab como *software* de cálculo). É importante também utilizar formatos abertos e claramente documentados, para que os ativos/assets possam processados novamente com outros *softwares*, como requerido para o modelo de dados de partes executáveis da publicação ampliada. Ainda, é recomendável escolher formatos que sejam interoperáveis e legíveis por múltiplos produtos de *software* e que não possuam restrições legais ou técnicas.

Definir a versão: de forma geral, um ativo/asset com um DOI não pode ser modificado, pois senão se perderia a persistência. Os ativos/assets alterados e atualizados devem ser salvos como novas versões e registrados com outro DOI.

⁷⁰ Formatos de dados par preservação:

<https://documentation.library.ethz.ch/display/DD/Archivtaugliche+Dateiformate>

Não obstante, existe uma categoria dentro dos ativos/assets de pesquisa que são os dados dinâmicos, dados que estão sujeitos a mudar. O ANDS expõe algumas formas de dados dinâmicos⁷¹:

- Existe um anexo regular e sistemático ao longo do tempo de dados a um conjunto existente de dados. Por exemplo em satélites (Landsat ou MODIS) ou medições meteorológicas;
- Há uma atualização ou modificação de dados pré-existentes em um grande conjunto de dados. Por exemplo, quando são achados erros nos dados pré-existentes, ou nas técnicas analíticas e/ou no processamento que afetam alguns atributos do conjunto de dados existente.

Em circunstâncias desse tipo, pode ser requerida alguma ação especial ou definição de descrições no DOI. Por exemplo, pode ser considerada a agregação de um sufixo que defina a versão ou o período temporal para o acesso aos dados.

Idealmente, para a descrição dos metadados associados ao ativo/asset e ao DOI se utiliza o padrão de metadados fornecido pela agência de registro de DOI ou um padrão comum dentro da comunidade científica ou da infraestrutura.

No contexto de um banco de dados, cada entidade está associada a um identificador primário que a distingue dentro do banco. Para o modelo proposto nesta tese, e descrito no seguinte capítulo, o PID considera-se o identificador primário da entidade. Como mencionado anteriormente, esse não é obrigatoriamente um DOI. A escolha do identificador dependerá do tipo de ativo, da abertura do mesmo e do meio de publicação ou disponibilização.

⁷¹ ANDS – Citing dynamic data: <https://www.ands.org.au/working-with-data/citation-and-identifiers/data-citation/citing-dynamic-data>

6 MODELAGEM CONCEITUAL PARA A PUBLICAÇÃO CIENTÍFICA AMPLIADA

Atualmente, o relacionamento de dados está vinculado diretamente com a descrição de recursos informacionais. A descrição dos recursos e do seu relacionamento também contribui às novas formas de acesso e recuperação.

Um dos processos para a representação dos recursos informacionais é o modelo conceitual. Esse define-se como “[...] o processo de abstração de um sistema real ou proposto para um modelo. É quase exato que a modelagem conceitual é o aspecto mais importante de um projeto de simulação.”⁷² (ROBINSON, 2011, p. 3, tradução própria).

Na construção dos bancos de dados utiliza-se o modelo Entidade-Relacionamento (ER) e o modelo Orientado a Objetos. Com ambos modelos se cria o esquema conceitual, produto da modelagem conceitual.

O modelo ER foi proposto e revisado pelo engenheiro Peter Chen (CHEN, 2002, 1976) para apresentar uma nova visão na modelagem de dados. O modelo cumpre simultaneamente com as necessidades dos fornecedores de dados (p.ex. pesquisadores ou agências de dados governamentais) e organizações (p.ex. bibliotecas acadêmicas ou repositórios) devido a sua flexibilidade e adaptabilidade a diferentes contextos. Esta flexibilidade é possível porque se estabelecem relações simples e triádicas (sujeito – predicado – objeto) baseadas em entidades com atributos e os relacionamentos entre elas. Esse modelo serve de base para a estrutura dos bancos de dados relacionais e para a exportação a RDF-stores (BORNEA *et al.*, 2013), o que permite o trabalho em entornos de *Linked Data*.

Na área da Ciência da Informação existem diversas iniciativas baseadas na modelagem conceitual, entre outras o *Modelo Conceptual de Descripción Archivística y Requisitos de Datos Básicos de las Descipciones de Documentos de Archivo, Agentes y Funciones*⁷³ (CNEDA) da Espanha, o *Modular*

⁷² Tradução própria do original em inglês: “Conceptual modeling is the process of abstracting a model from a real or prososed system. It is almost certainly the most important aspect of a simulation project”.

⁷³ CNEDA: <http://www.culturaydeporte.gob.es/dam/jcr:0313e2cc-85ae-496e-9df3-0dc69a981586/neda-mcda-p1-p2-20120618.pdf>

*Requirements for Records Systems*⁷⁴ (MoReq) da União Europeia ou o Modelo de Entidades e Relacionamentos do e-ARQ Brasil⁷⁵ baseado no MoReq.

Na comunicação científica também têm sido utilizados outros modelos, como o IDEF0 (*Integration Definition for Function Modeling*). O IDEF0 versa modela decisões, ações e atividades de uma organização ou sistema e tem sido tradicionalmente utilizado na engenharia de sistemas e em processos empresariais e industriais, por exemplo, na concepção e fabricação de produtos industriais. Björk (2005) adapta esse modelo aos processos de comunicação científica. Neste modelo, os conceitos principais são a atividade e o fluxo. “As atividades são mostradas como retângulos e seus nomes começam com verbos. Os fluxos são representados mediante setas e seus nomes são nomes”⁷⁶ (BJÖRK, 2005, p. 167, tradução própria). Esse modelo permite representar entradas (*inputs*), saídas (*outputs*), controles (*controls*) e mecanismos (*mechanisms*). Desta forma, através de diversos diagramas, o autor oferece uma visualização do processo de comunicação científica (ver Figura 14).

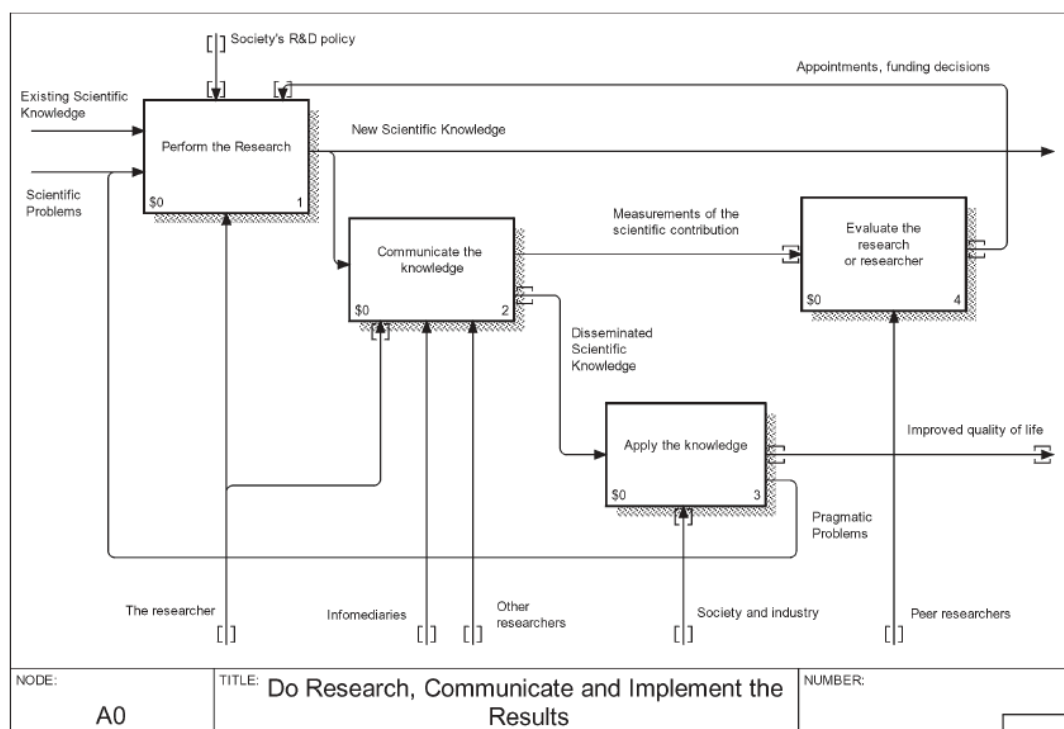
⁷⁴ MoReq: <https://www.moreq.info/>

⁷⁵ Modelo ER do e-ARQ Brasil:

http://www.conarq.gov.br/images/ctde/Orientacoes/Orientacao_tecnica_2.pdf

⁷⁶ Tradução própria do original em inglês: “Activities are shown as rectangles and their names start with verbs. Flows are represented by arrows and the names are nouns”.

Figura 14 Diagrama A0: pesquise, comunique e implemente os resultados.

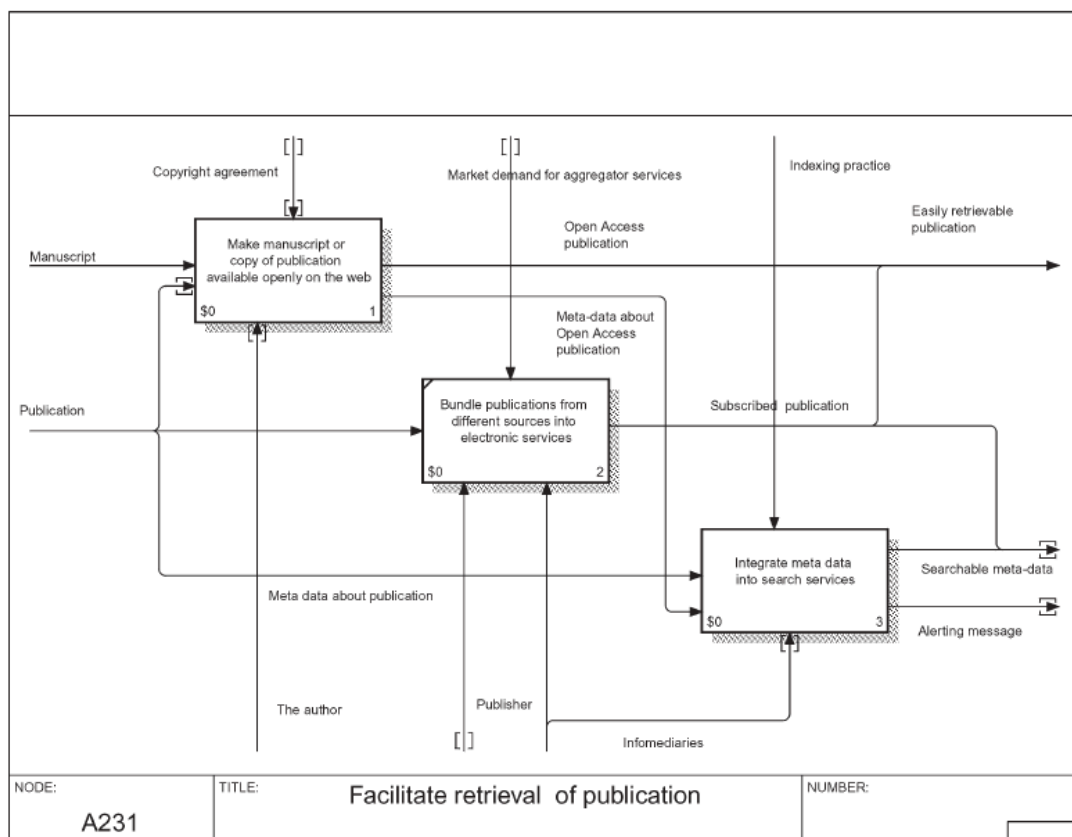


Fonte: Björk (2005, p.170).

Além do diagrama anterior, o autor fornece mais 6 diagramas. Cada um deles foca em um aspecto diferente do processo ou em um sub-ramo dele: Communicate the knowledge (Comunique o conhecimento), Process articles (Processe o artigo), Facilitate retrieval of publication (Facilite a recuperação da publicação) (ver Figura 15), Integrate metadata into search services (Integre metadados nos serviços de busca), Facilitate retrieval inside reader's organization (Facilite a recuperação dentro da organização do leitor) e Study the results (Estude os resultados).

Estes diagramas permitem ver também as diferentes partes interessadas (*stakeholders*) do processo. Por exemplo, na Figura 15, visualizam-se o autor, o editor e os intermediários como partes interessadas. As atividades a serem realizadas são a disponibilização do manuscrito ou uma cópia da publicação abertamente na *web*, unir publicações de fontes diferentes em serviços eletrônicos e integrar metadados nos serviços de busca.

Figura 15 Diagrama A231: facilite a recuperação dos resultados.



Fonte: Björk (2005, p.172).

Segundo Björk (2005), a utilização do modelo IDEF0 para os processos de comunicação científica distingue-se de outros modelos anteriores nos seguintes pontos:

- Estrutura hierárquica do modelo.
- Mais construções de modelagem, ou seja, controles e mecanismos.
- Modelagem muito mais detalhada de muitas funções.
- Desagregação das entradas e saídas em níveis mais detalhados.
- Modelagem de muitas das novas funções do sistema que surgiram como resultado da Internet (repositórios de acesso aberto, *harvesters*)⁷⁷. (BJÖRK, 2005, p. 175, tradução própria).

Uma das principais motivações para a elaboração de modelos na Ciência da Informação é a ruptura de estruturas monolíticas para possibilitar a integração dos dados. Com esse fim, utiliza-se o uso do modelo ER neste trabalho.

⁷⁷ Tradução própria do original em inglês: "Hierarchical structure of the model. More modelling constructs, i.e. controls and mechanisms. Much more detailed modelling of many of the functions. Disaggregation of inputs and outputs on more detailed levels. Modelling of many of the new system functions that have emerged as a result of the Internet (OA repositories, harvesters).

O modelo ER permite a representação de entidades em bancos de dados. Porém, também é necessário analisar os aspectos semânticos do banco de dados, já que “[...] dados e código em um banco de dados carecem de significado até que alguém lhes atribua um significado e alguém possa interpretá-los.”⁷⁸ (LIU, 2000, p. 119, tradução própria). Portanto, é requerido o uso de normas sociais e culturais para a interpretação das informações. Esse aspecto não é um aspecto técnico senão de organização e adaptação do modelo a um contexto determinado; neste caso, ao contexto da publicação científica.

Esse tipo de modelo vem sendo aplicado na área da Ciência da Informação com diferentes fins. Exemplos desta aplicação são a modelagem e descrição do recurso imagético apresentado na modelagem DILAM (Digital Images for Libraries Archives and Museums) (SIMIONATO, 2015); ou para a representação da estruturas existentes em coletas de dados de redes sociais, bem como para a representação de informações específicas sobre a privacidade de dados e análise de aspectos de privacidade no compartilhamento de dados (RODRIGUES, 2017).

A ideia do uso da modelagem de dados ER para construir um marco de trabalho para publicações ampliadas vem da necessidade de ter um modelo que auxilie na criação de sistemas. Um modelo conceitual ER pode auxiliar na criação de uma arquitetura para incluir as publicações ampliadas em ambientes digitais tais como os repositórios. Ao entender uma publicação ampliada como um sistema de conhecimento que pode adquirir características dinâmicas é importante ter em consideração as diferentes entidades envolvidas nelas e os atributos e relacionamentos presentes.

A **entidade** é cada coisa, objeto ou conceito real distinguível. Pode ter características concretas, com existência física (p.ex. uma pessoa); ou abstratas, com existência conceitual (p.ex. um nome). O **atributo** é cada propriedade ou característica identificativa da entidade. Cada atributo possui um valor para a identificação da entidade (p.ex. título, autor, data de publicação ou formato), sendo o atributo identificativo aquele que determina de forma unívoca à entidade (p.ex. um PID). Os atributos classificam-se em *simples* (não podem ser derivados, possuem valores atômicos), *compostos* (formados por mais de um

⁷⁸ Tradução própria do original em inglês: “Data and code in a database are meaningless until someone assigns a meaning to them and someone is able to interpret them”.

atributo simples), *derivados* (obtidos a partir de outros atributos da base, p.ex. o valor médio e a idade), *de valor único* (só pode existir um valor para esse atributo, p.ex. número de identificação pessoal), *multivaluados* (podem existir vários valores para o atributo, p.ex. instituições às que um pesquisador está vinculado e endereços de e-mail), *descritivos* (os atributos de um relacionamento). O **relacionamento** é a associação entre entidades

Entre entidades existe uma correspondência de cardinalidade, ou seja, o número de entidades B com a que pode estar relacionada a entidade A. A cardinalidade pode ser:

- um para um (1:1): Cada entidade envolvida relaciona-se apenas com um registro da outra;
- um para vários (1:N ou 1..*): A entidade A pode estar relacionada com vários registros de outra. Porém, a entidade B só pode estar ligada com um registro da entidade A;
- vários para um (N:1 ou *..1): Representa a situação oposta a “um para vários”;
- vários para vários (N:N ou *..*): Tanto a entidade A quanto a B podem referenciar múltiplos registros da outra.

Para a representação das entidades extraídas no Capítulo 4 em combinação com a estrutura proposta no Capítulo 5, que inclui a proposta de artigo modular, é composto um Diagrama Entidade – Relacionamento. Para a realização do diagrama, além da identificação das entidades, consideramos necessário realizar uma análise de semântica para entender qual é o relacionamento que existe entre as entidades e como se estabelecem as dependências ontológicas.

6.1 ANÁLISE SEMÂNTICA

Como apontado no Capítulo 4, a semântica analisa a relação entre signos e aquilo ao que se referem, bem como outorga significado aos dados para que possam ser interpretados.

Segundo Liu (2000) existem dois tipos de significados aplicáveis a predicados, proposições e sentenças: intenção (sentido) e extensão (referência).

A intenção de um predicado, ou seja, o seu sentido, é identificada com a propriedade que expressa; a sua extensão é a classe generalizada ou referências no mundo dos *affairs*⁷⁹ que possuem a propriedade. A intenção de uma proposição é determinada pelo significado dos predicados usados; a extensão é a sua correspondência com o mundo dos *affairs*⁸⁰. (LIU, 2000, p. 119, tradução própria).

Ao trasladar o conceito de semântica aos bancos de dados, introduzimos o modelo ontológico como esquema conceitual. As entidades e seus relacionamentos são organizados segundo esse modelo ontológico.

O modelo ontológico representa conceitos gerais e fundacionais do conhecimento, bem como terminologia dentro de um domínio específico. Ao categorizar os conceitos e a terminologia seguindo relacionamentos ontológicos, fornece-se um contexto semântico que permite entender cada entidade dentro do modelo (LIU, 2000).

Por sua vez, a análise semântica é utilizada para produzir modelos semânticos. Liu (2000) separa a análise semântica em fases:

- Definição do problema.
- Definição do problema de estudo (em muitos casos integrada na fase “definição do problema”).
- Identificação das unidades semânticas (entidades).
- Classificação das unidades semânticas.
- Identificação das dependências ontológicas.
- Finalização do modelo semântico.

A seguir, detalham-se as fases aplicadas ao nosso caso de estudo. Estas fases resumem também análises desenvolvidos em outros capítulos.

⁷⁹ A palavra *affair* deve ser entendida neste contexto como uma situação no mundo real ou imaginado.

⁸⁰ Tradução própria do original em inglês: “The intention of a predicate, i.e. its sense, is identified with the property it expresses; its extension is the generalised class or referents in the world of affairs that possess the property. The intension of a proposition is determined by the meaning of the predicates used; the extension is its correspondence to the world of affairs”.

6.1.1 DEFINIÇÃO DO PROBLEMA E PROBLEMA DE ESTUDO

Definir com clareza um problema genérico costuma ser uma tarefa árdua e resulta em um problema pouco preciso devido às considerações epistemológicas. Como aponta Liu (2000) as análises semânticas devem começar com esse planejamento defeituoso do problema. Resulta mais fatível definir o problema de estudo. Ao definir o problema de estudo, “o procedimento inteiro será iterado, aumentando gradualmente o escopo do problema examinado”⁸¹ (LIU, 2000, p. 166, tradução própria).

No presente trabalho, a problemática da integração de diferentes ativos/assets oriundos da pesquisa nos ciclos de comunicação científica no contexto da ciência aberta e da gestão de dados de pesquisa é a definição do problema. Como problema de estudo determina-se um tipo de mecanismo para a integração: a publicação ampliada.

6.1.2 IDENTIFICAÇÃO DAS UNIDADES SEMÂNTICAS

Esse passo toma a definição do problema como entrada ou ponto de partida (input), a saída ou resultado (output) é uma série de ativos/assets oriundos da pesquisa. O processo completo de identificação realizou-se nos Capítulos 4 e 5, com a análise Semiótica dos ambientes de publicação científica e a proposta de entidades da publicação ampliada, respectivamente.

Os resultados são descritos nos quadros 10 e 11, onde se mostra a identificação das seguintes entidades (entidades): autor, áudio, base do conhecimento, código, conjunto de dados brutos, arquivo multimídia, código, conjunto de dados brutos, arquivo multimídia, figura, imagem, planilha de dados, quadro, *software*, tabela, texto (sub-entidades: palavras chave, resumo, objetivos, agradecimentos, métodos, medições, resultados, discussões, conclusões, referências), vídeo e *workflow*.

É importante destacar que a informação que deva ser modelada sobre cada uma destas entidades é um aspecto variável. Uma das principais dependências é a disciplina para a qual se aplique a modelagem. Assim, por

⁸¹ Tradução própria do original em inglês: “The whole procedure will be iterated, gradually increasing the scope of the problem being examined”.

exemplo, os atributos da entidade “Imagem” serão mais exaustivos em um contexto de museologia ou arte, onde o recuso imagético pode ser central. Por sua vez, a entidade “Código” terá uma presença maior nas ciências experimentais e serão necessárias mais informações, por exemplo, sobre a versão da linguagem utilizada ou as funções inseridas e/ou definidas.

Para o propósito deste trabalho, utilizaremos as características mínimas do objeto digital como explicado no Capítulo 4. Consideramos, porém, que deve existir uma flexibilidade nas infraestruturas de informação científica para admitir metadados específicos, como apontado nos requerimentos do marco de trabalho para publicações ampliadas de Bardi e Manghi (2015) (capítulo 5).

6.1.3 CLASSIFICAÇÃO DAS UNIDADES SEMÂNTICAS

Para continuar com a análise das unidades semânticas, Lui (2000) as classifica em várias categorias.

- **Agente – *affordance*** (*Agent – affordance*). Consideramos agentes a quem realiza ações por si mesmo e pode se responsabilizar por elas. “Qualquer ação racional do agente é constrangida pelo repertório do comportamento e dirigida pelo seu conhecimento do mundo. Esse repertório de comportamentos pode ser visto como *affordances*.”⁸² (LIU, 2000, p. 61, tradução própria). Nos sistemas sociais, estas *affordances* “[...] são ações ou resultados das ações destes agentes”⁸³ (LIU, 2000, p. 167, tradução própria). Uma *affordance*, segundo a definição de Gibson (1979), é “[...] algo que se refere tanto ao ambiente como ao animal de uma forma que nenhum termo existente faz. Implica a complementariedade do animal e do ambiente”⁸⁴ (GIBSON, 1979, p.127, tradução própria). Na abordagem da TI (FARAJ E AZAD, 2012; LIU, 2000), estabelece-se uma relação (ou analogia) entre *affordance* e objetos cotidianos com suas correspondentes características e funções, outorgadas pelo criador do objeto.

⁸² Tradução própria do original em inglês: “Any rational action of the agent is constrained by the repertoire of behaviour and directed by his knowledge of the world. This repertoire of behaviour can be seen as *affordances*”.

⁸³ Tradução própria do original em inglês: “[...] are actions or results of actions of those agents”.

⁸⁴ Tradução própria do original em inglês: “[...] something that refers to both the environment and the animal in a way that no existing term does. It implies the complementarity of the animal and the environment.”

- **Universal – particular** (*Universal – particular*). Na descrição e análise da informação descrevemos tipos de coisas (universal). Às vezes, uma instancia particular resulta única ou necessária e possui uma descrição própria. Neste trabalho, as sub-entidades do “Texto” são particulares; porém, devem ser tratadas especialmente no modelo dadas a suas características próprias e sua relação direta com outras entidades universais.
- **Genérica – específica** (*Generic – specific*). Existem entidades genéricas com propriedades abrangentes e específicas que detalham mais alguma propriedade. “As específicas herdam propriedades das genéricas”⁸⁵ (LIU, 2000, p. 167, tradução própria). De novo, no caso das sub-entidades do “Texto” especificam aspectos do texto.
- **Inteira – parte** (*Whole – part*). Alguns fenômenos existem somente como parte de um todo. Por exemplo, um grupo de trabalho é parte de um departamento de pesquisa que, ao mesmo tempo é parte de uma faculdade, e esta faculdade de uma universidade.
- **Portador(a) do papel – nome do papel** (*role-carrier – role-name*). Se os agentes desempenham um papel determinado, adquirem certos nomes devido ao papel. Por exemplo, um(a) pesquisador(a) que cria um ativo/asset torna-se autor(a) ou criador(a) do mesmo. Ao mesmo tempo, em uma publicação textual esse(a) autor(a) pode ser o autor(a) de correspondência.

⁸⁵ Tradução própria do original em inglês: “The specifics the inherit properties of the generics”.

Seguindo estas categorias, as nossas entidades classificam-se da seguinte maneira.

Quadro 18 Classificação das entidades.

Autor	Universal	Agente Portador do papel	Genérica
Áudio	Universal	<i>Affordance</i>	Genérica
Base do conhecimento	Universal	<i>Affordance</i>	Genérica
Código	Universal	<i>Affordance</i>	Genérica
Conjunto de dados brutos	Universal	<i>Affordance</i>	Genérica
Arquivo multimídia	Universal	<i>Affordance</i>	Genérica
Figura	Universal	<i>Affordance</i>	Genérica
Imagem	Universal	<i>Affordance</i>	Genérica
Planilha de dados	Universal	<i>Affordance</i>	Genérica
Quadro	Universal	<i>Affordance</i>	Genérica
<i>Software</i>	Universal	<i>Affordance</i>	Genérica
Tabela	Universal	<i>Affordance</i>	Genérica
Texto	Universal	<i>Affordance</i>	Genérica
Texto – palavras chave	Particular	<i>Affordance</i>	Específica Parte de “Texto”
Texto – resumo	Particular	<i>Affordance</i>	Específica Parte de “Texto”
Texto – objetivos	Particular	<i>Affordance</i>	Específica Parte de “Texto”
Texto – agradecimentos	Particular	<i>Affordance</i>	Específica Parte de “Texto”
Texto – métodos	Particular	<i>Affordance</i>	Específica Parte de “Texto”
Texto – medições	Particular	<i>Affordance</i>	Específica Parte de “Texto”
Texto - resultados	Particular	<i>Affordance</i>	Específica Parte de “Texto”
Texto – discussões	Particular	<i>Affordance</i>	Específica Parte de “Texto”
Texto – conclusões	Particular	<i>Affordance</i>	Específica Parte de “Texto”
Texto - referências	Particular	<i>Affordance</i>	Específica Parte de “Texto”
Vídeo	Universal	<i>Affordance</i>	Genérica
<i>Workflow</i>	Universal	<i>Affordance</i>	Genérica

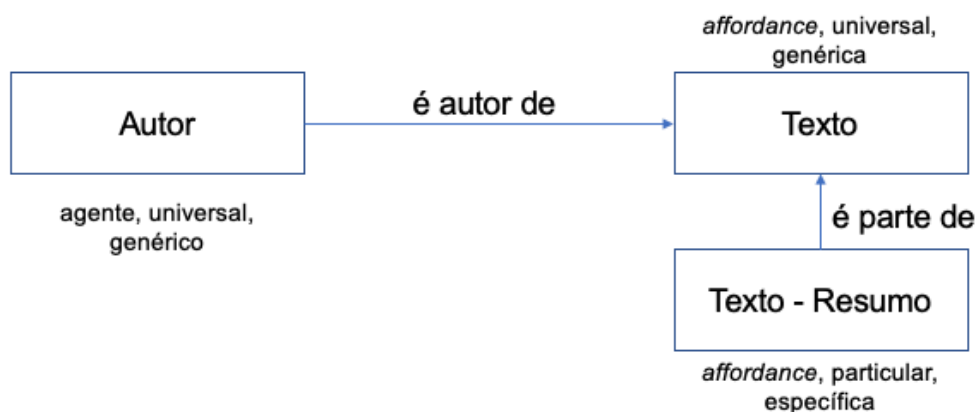
Fonte: Elaboração própria.

Esta categorização permite definir as propriedades e o tipo de dados vinculados a cada unidade, bem como a identificar claramente as dependências ontológicas. Estas dependências são estabelecidas entre unidades tanto particulares quanto universais.

6.1.4 IDENTIFICAÇÃO DAS DEPENDÊNCIAS ONTOLÓGICAS

Para desenvolver uma análise semântica, é essencial estabelecer dependências ontológicas entre os diferentes fenômenos (as entidades do nosso modelo). Esta dependência ontológica acontece quando a existência do fenômeno B depende da existência do fenômeno A. O fenômeno A é denominado “dependente” enquanto o fenômeno B conhece-se como “antecedente” (LIU, 2000).

Figura 16 Dependência entre as entidades “Autor”, “Texto” e “Texto – resumo”.



Fonte: Elaboração própria.

Na figura anterior, observa-se o relacionamento entre as entidades universais “Autor” e “Texto”, mas também no nível particular entre “Texto” e “Texto – resumo”. A entidade “texto – resumo”, bem como outras entidades textuais do modelo, consideram-se particulares, já que, se bem compartilham aspecto estruturais com o texto, relacionam-se de forma diferente com outras entidades. Além disso, especificam características ou parte concretas dele, por exemplo, no caso da entidade “Texto-resumo”, ela recolhe os principais pontos expostos no texto.

Os relacionamentos ontológicos entre as entidades modelam-se como muitos fragmentos independentes. O procedimento consiste na contextualização dos fragmentos tendo em consideração as limitações ontológicas existentes.

Estas limitações são definidas no modelo ER escolhido para esta modelagem como a cardinalidade entre as entidades.

Ainda, entendemos o relacionamento entre as entidades como um evento. Os eventos, como apresentados no ponto 5.3, são “processos relevantes para a proveniência dos recursos, tais como criação, curadoria, acesso, reivindicação, atualização, citação, revisão” (DAPPER *et al.*, 2017, p. 6). A identificação do relacionamento entre entidades permite compreender a origem da ação e o caminho seguido e garantir a reprodutibilidade do resultado. É também uma forma de rastrear a proveniência da entidade. A proveniência define-se como “[...] as fontes de informação, tais como entidades e processos, envolvidas na produção ou entrega de um artefato”⁸⁶ (W3C PROVENANCE INCUBATOR GROUP, 2005, tradução própria). Desta forma, é possível estabelecer relacionamentos precisos entre os agentes (autores) e cada uma das versões das *affordances* (ativos/assets).

Ainda, diversos bancos de dados de ativos/assets de pesquisa, tais como DataCite, Crossref ou Cobalmetrics⁸⁷, arquivam também os eventos entre ativos/assets e não exclusivamente a informação relativa cada ativo/asset. Não obstante, a maioria dos eventos arquivados são citações e não se incluem outros como criação, curadoria, acesso ou atualização. Neste modelo, ao considerar eventos além da citação consegue-se retratar a dimensão executável exposta no Capítulo 4, bem como o dinamismo da publicação ampliada.

6.1.5 FINALIZAÇÃO DO MODELO SEMÂNTICO

O passo final consiste em finalizar o modelo juntando os fragmentos entidade-relacionamento em um desenho completo. Para realizar esse último passo é necessário utilizar uma série de normas para completar a sintaxe do modelo. No nosso caso, as tecnologias da *web* semântica servem de base, que permitem identificar e normalizar a estrutura dos recursos presentes nos ambientes digitais, bem como descrever sua semântica computacionalmente (RAMALHO; OUCHI, 2012; WOUTERSEN-WINDHOUWER, 2009; MORATO *et al.*, 2008).

⁸⁶ Tradução própria do original em inglês: “[...]the sources of information, such as entities and processes, involved in producing or delivering an artifact”.

⁸⁷ Cobalmetrics: <https://cobalmetrics.com/>

Ramalho e Ouchi (2012) consideram as seguintes tecnologias como tecnologias da *web* semântica:

- *Extended Markup Language* (XML) – definição de marcações personalizadas;
- *Resource Description Framework* (RDF) – descrição de relacionamentos e declarações sobre tipos de objetos;
- *Resource Description Framework – Schema* (RDF-S) – representação de vocabulários;
- *Web Ontology Language* (OWL) – desenvolvimento de ontologias;
- *Rule Interchange Format* (RIF) – intercâmbio e interoperabilidade de regras lógicas;
- *SPARQL Protocol and RDF Query Language* (SPARQL) – consultas sobre estruturas RDF.

No contexto das publicações ampliadas, o uso preferente de vocabulários padronizados e controlados garante a interoperabilidade semântica dos ativos/assets e constituem normas dentro da comunidade científica. Junto com os padrões genéricos como Dublin Core Metadata Initiative⁸⁸, DataCite Schema⁸⁹, Core Scientific Metadata Model⁹⁰ ou Data Catalog Vocabulary⁹¹, existem padrões (também denominados *common standards*) específicos⁹² para diferentes áreas do conhecimento. Estes padrões devem ser tidos em consideração na construção de infraestrutura específica para uma área, bem como na adaptação de uma infraestrutura genérica para fins específicos.

Como indicam Bornea *et al.* (2013), a utilização do RDF permite o compartilhamento das informações contidas nos bancos de dados como *Linked Data* (ou dados ligados - LD). Os LD constituem um mecanismo semântico de organização do conhecimento necessário para a publicação de dados na *web* e contribuem para expandir a reutilização dos ativos/assets da publicação ampliada (MUCHERONI; SILVA; PALETTA, 2015). Hooland e Verborgh (2014) indicam que os LD não são uma tecnologia bem definida, senão um conjunto de

⁸⁸ Dublin Core Metadata Initiative: <https://www.dublincore.org/specifications/dublin-core/dces/2012-06-14/>

⁸⁹ DataCite Metadata Schema: <https://schema.datacite.org/meta/kernel-4.3/>

⁹⁰ Core Scientific Metadata Model: <http://icatproject-contrib.github.io/CSMD/csmd-4.0.html>

⁹¹ Data Catalog Vocabulary: <https://www.w3.org/TR/vocab-dcat/>

⁹² Disciplinary metadata: <http://www.dcc.ac.uk/drupal/resources/metadata-standards>

boas práticas para a publicação de dados estruturados na *web*. A utilização destas práticas no contexto da publicação ampliada contribui à identificação dos ativos/assets e à interoperabilidade dos metadados mediante padrões.

6.2. PROPOSTA DE MODELAGEM CONCEITUAL PARA PUBLICAÇÕES CIENTÍFICA AMPLIADAS

A modelagem conceitual proposta é o resultado da análise dos ambientes realizada no Capítulo 4 para extrair as entidades da publicação e a análise semântica examinada no ponto anterior deste capítulo.

Convém notar que o conjunto de entidades modeladas é geral e não versa representar uma publicação ampliada em uma área específica do conhecimento senão abranger o conjunto da publicação científica. Esse tipo de modelagem pode servir como base para repositórios institucional ou megaperiódicos (*megajournals*)⁹³ que abranjam vários domínios. Outros campos descritivos podem ser adicionados às entidades representadas ou novas entidades podem ser agregadas desde que respeitem as características mínimas do objeto digital.

Para a descrição das entidades e seus relacionamentos são utilizados os seguintes padrões.

Quadro 19 Padrões para a descrição das entidades e relacionamentos.

Abreviatura	Nome	URL
ali	Access and License Indicators	https://www.niso.org/schemas/ali/1.0
csmd	CSMD: the Core Scientific Metadata Model	http://icatproject-contrib.github.io/CSMD/csmd-4.0.html
cdesc	Ontology for content description	http://everest.expertsystemlab.com/vocabulary/cdesc/index-en.html
coar	Controlled Vocabulary for Resource Type Genres	http://vocabularies.coar-repositories.org/pubby/resource_type.html
datacite	DataCite Metadata Schema	https://schema.datacite.org/meta/kernel-4.3/
dcterms	DCMI Metadata Terms	http://purl.org/dc/terms/
dcat	Data Catalog Vocabulary (DCAT)	https://www.w3.org/TR/vocab-dcat/
dk	The Data Knowledge Vocabulary	http://www.data-knowledge.org/dk/1.1

⁹³ Os megaperiódicos são plataformas de publicação em Acesso Aberto e revisadas por pares que aceitam uma grande quantidade maior artigos, já que carecem de números e possuem publicação em fluxo contínuo. Exemplos de megaperiódicos são aqueles editados por PLOS ou F1000Research.

foaf	Friend of a Friend (FOAF)	http://xmlns.com/foaf/spec/
oco	Open Citation Ontology	https://w3id.org/oc/ontology/
owl	Ontology language web	https://www.w3.org/TR/owl-ref
schema	Schema.org	https://schema.org/
rdfs	RDF Schema	https://www.w3.org/TR/rdf-schema/

Fonte: Elaboração própria.

Estes padrões fornecem as propriedades necessárias para a descrição das entidades e possuem consenso de uso na comunidade científica. ALI permite a descrição precisa das licenças utilizada, o que esclarece as condições de reutilização do recurso. CSMD representa diferentes atividades científicas. COAR define os tipos de recursos com o objetivo de delimitar as classificações em repositórios. Datacite é um dos padrões mais utilizados no registro de DOIs em repositórios de dados e abrange muitas propriedades das entidades publicadas. DCTerms é amplamente aceito na área de biblioteconomia e desenvolvimento ontológico e serve de base nas descrições. DCAT facilita a interoperabilidade entre catálogos de dados publicados na *web*. FOAF permite a descrição de propriedades pessoais e de relacionamento entre agentes. OCC modela os relacionamentos entre autor e citação e entre as citações entre si. RDFS auxilia na modelagem de arquivos RDF, compostos de três componentes: recurso (ativo/asset), propriedade e valor (valor literal ou outro ativo/asset).

Além disso, na descrição da entidade “Autor”, inclui-se o tipo de contribuição realizada. Para isso, utiliza-se a taxonomia CRediT⁹⁴ (Contributor Roles Taxonomy). Esta taxonomia representa quatorze roles de contribuição: (1) conceituização, (2) curadoria de dados, (3) análise formal, (4) aquisição de financiamento, (5) pesquisa, (6) metodologia, (7) administração do projeto, (8) recursos, (9) *software*, (10) supervisão, (11) validação, (12) visualização, (13) escrita – esboço original, (14) escrita – revisão e edição.

A seguir, descrevem-se as entidades e os relacionamentos com as correspondentes cardinalidades. Quando possível na descrição das entidades, utilizam-se como valores de exemplo, valores pertencentes a esta tese.

⁹⁴ CRediT: <https://www.casrai.org/credit.html>

6.2.1 ENTIDADES

Propriedades de “Autor”

Quadro 20 Propriedades da entidade “Autor”.

Nome	Descrição	Tipo de dado	Cardinalidade	Padrão	Valor exemplo
author_id	Identificador do autor	Cadeia de caracteres	Exatamente um	foaf:account	https://orcid.org/0000-0001-7460-7794
author_id_type	Tipo de identificador	Termo de vocabulário controlado	Exatamente um	datacite:nameIdentifierScheme	HTTP-ORCID
name	Nome do autor	Cadeia de caracteres	Exatamente um	foaf:name	Paloma Marín Arraiza
mail	Endereço de e-mail	Cadeia de caracteres	Exatamente um	foaf:mbox	paloma.arraiza@unesp.br
contribution	Tipo de contribuição no trabalho. Termo selecionado de CRediT.	Termo de vocabulário controlado	Um ou mais	rdfs:label	Conceptualization, Investigation, Formal analysis, Methodology, Validation, Writing – original draft, Writing – review and editing
institution_id	Identificador da instituição	Cadeia de caracteres	Exatamente um	datacite:affiliationIdentifier	https://ror.org/00987cb86
Institution_id_type	Tipo de identificador	Termo de vocabulário controlado	Exatamente um	datacite:affiliationIdentifierScheme	HTTP-ROR

Fonte: Elaboração própria.

Para os fins deste modelo, o identificador de autor escolhido foi o ORCID iD. Porém, no caso de autores históricos (p.ex. Marie Curie) pode ser utilizado o identificador de Wikidata (p.ex. <https://www.wikidata.org/wiki/Q7186>). Também podem ser utilizados identificadores do controle de autoridades de bibliotecas (p.ex. *Gemeinsame Normdatei* – GND, da Biblioteca Nacional Alemã).

Propriedades de “Código” e “Software”.

As entidades “Código” e “Software” são modeladas de forma conjunta, já que a segunda (*Software*) é a versão compilada da primeira (Código). Em um projeto pode ser desenvolvido um código ou *software* específico para a coleta ou processamento de dados (p.ex. um código na linguagem Python) ou pode ser utilizado outro desenvolvido por terceiros, tanto proprietário quanto livre (p.ex. OpenWriter para o processamento de textos ou 3DMax para o desenvolvimento de modelo 3D e maquetes eletrônicas).

Quadro 21 Propriedades das entidades “Código” e “Software”.

Nome	Descrição	Tipo de dado	Cardinalidade	Padrão	Valor exemplo
id	Identificador do ativo/asset	Cadeia de caracteres	Exatamente um	dcterms:identifier	https://www.wikidata.org/wiki/Q996596
id_type	Tipo de identificador	Termo de vocabulário controlado	Exatamente um	rdfs:label	HTTP-WIKIDATA
title	Título	Cadeia de caracteres	Exatamente um	dcterms:title	StarUML
creator	Identificação do criador	Cadeia de caracteres	Um ou mais	dcterms:creator	MKLab
creator_id	Identificador do autor	Cadeia de caracteres	Exatamente um	foaf:account	https://www.wikidata.org/wiki/Q58614718
issued	Data de publicação	Data	Zero ou um	dcterms: date	2019-10-10
version	Versão utilizada	Cadeia de caracteres	Exatamente um	owl:versionInfo	Version: 3.1.0
type	Tipo de ativo/asset. Seguindo o vocabulário COAR.	Termo de vocabulário controlado	Exatamente um	dcterms:type	<i>Software</i>
license	Licença do ativo asset	Cadeia de caracteres	Exatamente um	dcterms:license	GNU General Public License ⁹⁵
license_ref	Referência à descrição da licença	URI	Exatamente um	rdfs:label	https://www.gnu.org/licenses/gpl-3.0.en.html

⁹⁵ O software StarUML possui na atualidade uma licença proprietária. Antigamente foi um software livre com licença GNU. A licença GNU tem sido utilizada aqui como exemplo. Embora o software possua uma licença proprietária, é possível utilizá-lo sem registro (versão de avaliação) de forma indefinida, desde que seja para fins não comerciais.

language	Idioma dos comentários, documentação ou interfase a partir da norma ISO 6391-1 (código do país com dois letras)	Termo de vocabulário controlado	Zero ou um	dcterms:language	en
description	Descrição do ativo/asset	Cadeia de caracteres	Zero ou um	dcterms:description	Ferramenta para desenho de diagramas UML desenvolvida pelo MKLab

Fonte: Elaboração própria.

Propriedades de “Conjunto de dados brutos”, “Planilha de dados” e “Tabela”

Estas três entidades são modeladas da mesma forma. A diferença entre “Conjunto de dados brutos”, “Planilha de dados” e “Tabela” reside no grau de processamento e no formato de apresentação final dos dados. Porém, os relacionamentos destas entidades com outras do modelo são distintos. O “Conjunto de dados brutos” abrange tanto aqueles dados positivos quanto negativos, quer dizer aqueles que não levam a resultados concluintes. Podemos ver esta entidade como uma via de publicação de dados negativos, outorgando mais transparência aos resultados da pesquisa correspondente.

Quadro 22 Propriedades da entidade “Conjunto de dados brutos”.

Nome	Descrição	Tipo de dado	Cardinalidade	Padrão	Valor exemplo
id	Identificador do ativo/asset	Cadeia de caracteres	Exatamente um	dcterms:identifier	https://doi.org/10.1234/5678
id_type	Tipo de identificador	Termo de vocabulário controlado	Exatamente um	rdfs:label	HTTP-DOI
title	Título	Cadeia de caracteres	Exatamente um	dcterms:title	Medidas da temperatura e humidade na cidade de Marília (SP)
creator	Identificação do criador	Cadeia de caracteres	Um ou mais	dcterms:creator	Maria da Silva
creator_id	Identificador do autor	Cadeia de caracteres	Exatamente um	foaf:account	https://orcid.org/0000-1111-2222-3333
issued	Data de publicação	Data	Zero ou um	dcterms: date	2019-10-10

publisher	Editora ou entidade que publica o ativo/asset	Cadeia de caracteres	Exatamente um	dcterms: publisher	Universidade Estadual Paulista
keywords	Palavras chave	Cadeia de caracteres	Zero ou mais	dcat:keyword	temperatura, humidade, Marília
type	Tipo de ativo/asset. Seguindo o vocabulário COAR.	Termo de vocabulário controlado	Exatamente um	dcterms:type	Text
license	Licença do ativo asset	Cadeia de caracteres	Exatamente um	dcterms:license	Creative Commons Attribution 4.0 International
license_ref	Referência à descrição da licença	URI	Exatamente um	rdfs:label	https://creativecommons.org/licenses/by/4.0/
language	Idioma do conjunto de dados a partir da norma ISO 6391-1 (código do país com duas letras)	Termo de vocabulário controlado	Zero ou um	dcterms:language	pt
description	Descrição do ativo	Cadeia de caracteres	Zero ou um	dcterms:description	Medições da temperatura e humidade na cidade de Marília (SP), no mês de agosto de 2019 em intervalos de uma hora.
format	Formato de ativo/asset. Uso do vocabulário IANA Media Types ⁹⁶	Termo de vocabulário controlado	Zero ou um	dcterms: format	text/csv
data_access	Modo de acesso ao ativo/asset (aberto, restrito, fechado)	Termo de vocabulário controlado	Exatamente um	ali:free_to_read ⁹⁷	aberto
metadata	Descrição dos metadados e dos padrões de metadados utilizados	Estrutura de dados aninhada	Zero ou mais		Ver Quadro 29
personal_data	Existência de dados pessoais. Valores	Termo de vocabulário controlado	Exatamente um		Não

⁹⁶ IANA Media Types: <https://www.iana.org/assignments/media-types/media-types.xhtml>

⁹⁷ ali:free_to_read permite delimitar as datas de acesso ao ativo/asset, especificar períodos de embargo ou estabelecer o acesso aberto ao conteúdo: https://groups.niso.org/apps/group_public/download.php/14226/rp-22-2015_ALI.pdf

	possíveis: sim / não / desconhecido				
sensitive_data	Existência de dados sensíveis. Valores possíveis: sim / não / desconhecido	Termo de vocabulário controlado	Exatamente um		Não
data_processing	Informação sobre o grau de processamento dos dados. Valores possíveis: bruto / intermediário / final	Termo de vocabulário controlado	Exatamente um		Bruto

Fonte: Elaboração própria.

Propriedades de “Arquivo multimídia”

O artigo multimídia se diferencia de outras entidades pelo grau de interatividade, definido no atributo “interaction_grade” a partir do esquema “schema.org”.

Quadro 23 Propriedades da entidade “Arquivo multimídia”.

Nome	Descrição	Tipo de dado	Cardinalidade	Padrão	Valor exemplo
id	Identificador do ativo/asset	Cadeia de caracteres	Exatamente um	dcterms:identifier	https://doi.org/10.1234/5678
id_type	Tipo de identificador	Termo de vocabulário controlado	Exatamente um	rdfs:label	HTTP-DOI
title	Título	Cadeia de caracteres	Exatamente um	dcterms:title	Visualização das trajetórias parabólicas
creator	Identificação do criador	Cadeia de caracteres	Um ou mais	dcterms:creator	Maria da Silva
creator_id	Identificador do autor	Cadeia de caracteres	Exatamente um	foaf:account	https://orcid.org/0000-1111-2222-3333
issued	Data de publicação	Data	Zero ou um	dcterms: date	2019-10-10

publisher	Editora ou entidade que publica o ativo/asset	Cadeia de caracteres	de	Exatamente um	dcterms: publisher	Universidade Estadual Paulista
keywords	Palavras chave	Cadeia de caracteres	de	Zero ou mais	dcat:keyword	Trajetórias parabólicas, movimento parabólico, física
type	Tipo de ativo/asset. Seguindo o vocabulário COAR.	Termo de vocabulário controlado	de	Exatamente um	dcterms:type	Interactive resource
license	Licença do ativo asset	Cadeia de caracteres	de	Exatamente um	dcterms:license	Creative Commons Attribution 4.0 International
license_ref	Referência à descrição da licença	URI		Exatamente um	rdfs:label	https://creativecommons.org/licenses/by/4.0/
language	Idioma dos metadados a partir da norma ISO 6391-1 (código do país com dois letras)	Termo de vocabulário controlado	de	Zero ou um	dcterms:language	en
description	Descrição do ativo	Cadeia de caracteres	de	Zero ou um	dcterms:description	Applet para o cálculo e visualização das trajetórias parabólicas na aula de física do ensino médio.
format	Formato de ativo/asset. Uso do vocabulário IANA Media Types ⁹⁸	Termo de vocabulário controlado	de	Zero ou um	dcterms: format	application/javascript
data_access	Modo de acesso ao ativo/asset (aberto, restrito, fechado)	Termo de vocabulário controlado		Exatamente um	ali:free_to_read ⁹⁹	aberto
Interaction_grade	Modo predominante de interatividade presente no recurso	Termo de vocabulário controlado	de	Exatamente um	schema:InteractioGrade	misto

⁹⁸ IANA Media Types: <https://www.iana.org/assignments/media-types/media-types.xhtml>

⁹⁹ ali:free_to_read permite delimitar as datas de acesso ao ativo/asset, especificar períodos de embargo ou estabelecer o acesso aberto ao conteúdo: https://groups.niso.org/apps/group_public/download.php/14226/rp-22-2015_ALI.pdf

	(ativo, expositivo, misto)				
--	----------------------------	--	--	--	--

Fonte: Elaboração própria.

Propriedades de “Áudio”, “Figura”, “Imagem”, “Vídeo” e “Workflow”.

Estas cinco entidades podem ser modeladas seguindo as mesmas propriedades. Se bem, na propriedade “metadata” devem ser inseridos mais metadados descritivos. Por exemplo, descrições da geolocalização, anotações, tamanho do arquivo, resolução, entre outros.

Quadro 24 Propriedades das entidades “Áudio”, “Figura”, “Imagem”, “Vídeo” e “Workflow”.

Nome	Descrição	Tipo de dado	Cardinalidade	Padrão	Valor exemplo
id	Identificador do ativo/asset	Cadeia de caracteres	Exatamente um	dcterms:identifier	https://doi.org/10.1234/5678
id_type	Tipo de identificador	Termo de vocabulário controlado	Exatamente um	rdfs:label	HTTP-DOI
title	Título	Cadeia de caracteres	Exatamente um	dcterms:title	Exemplo de projeto de publicação no limite superior
creator	Identificação do criador	Cadeia de caracteres	Um ou mais	dcterms:creator	Paloma Marín Arraiza
creator_id	Identificador do autor	Cadeia de caracteres	Exatamente um	foaf:account	https://orcid.org/0000-0001-7460-7794
issued	Data de publicação	Data	Zero ou um	dcterms: date	2019-10-10
publisher	Editores ou entidade que publica o ativo/asset	Cadeia de caracteres	Exatamente um	dcterms: publisher	Universidade Estadual Paulista
keywords	Palavras chave	Cadeia de caracteres	Zero ou mais	dcat:keyword	ambiente de publicação, vectors journal
type	Tipo de ativo/asset. Seguindo o vocabulário COAR.	Termo de vocabulário controlado	Exatamente um	dcterms:type	Video
license	Licença do ativo asset	Cadeia de caracteres	Exatamente um	dcterms:license	Creative Commons Attribution 4.0 International
license_ref	Referência à descrição da licença	URI	Exatamente um	rdfs:label	https://creativecommons.org/licenses/by/4.0/

language	Idioma do ativo/asset a partir da norma ISO 6391-1 (código do país com dois letras)	Termo de vocabulário controlado	Zero ou um	dcterms:language	pt
description	Descrição do ativo	Cadeia de caracteres	Zero ou um	dcterms:description	Resultado da análise do projeto de publicação “Vectors journal”. Mostra as características de interação da plataforma.
format	Formato de ativo/asset. Uso do vocabulário IANA Media Types ¹⁰⁰	Termo de vocabulário controlado	Zero ou um	dcterms:format	video/mp4
data_access	Modo de acesso ao ativo/asset (aberto, restrito, fechado)	Termo de vocabulário controlado	Exatamente um	ali:free_to_read ¹⁰¹	aberto
metadata	Descrição dos metadados e dos padrões de metadados utilizados	Estrutura de dados aninhada	Zero ou mais		Ver Quadro 29

Fonte: Elaboração própria.

Propriedades de “Quadro”

Quadro 25 Propriedades da entidade “Quadro”.

Nome	Descrição	Tipo de dado	Cardinalidade	Padrão	Valor exemplo
id	Identificador do ativo/asset	Cadeia de caracteres	Exatamente um	dcterms:identifier	https://doi.org/10.1234/5678
id_type	Tipo de identificador	Termo de vocabulário controlado	Exatamente um	rdfs:label	HTTP-DOI

¹⁰⁰ IANA Media Types: <https://www.iana.org/assignments/media-types/media-types.xhtml>

¹⁰¹ ali:free_to_read permite delimitar as datas de acesso ao ativo/asset, especificar períodos de embargo ou estabelecer o acesso aberto ao conteúdo: https://groups.niso.org/apps/group_public/download.php/14226/rp-22-2015_ALI.pdf

title	Título	Cadeia de caracteres	Exatamente um	dcterms:title	Características de cada nível de interação com a publicação e exemplos.
creator	Identificação do criador	Cadeia de caracteres	Um ou mais	dcterms:creator	Paloma Marín Arraiza
creator_id	Identificador do autor	Cadeia de caracteres	Exatamente um	foaf:account	https://orcid.org/0000-0001-7460-7794
issued	Data de publicação	Data	Zero ou um	dcterms: date	2019-10-10
publisher	Editora ou entidade que publica o ativo/asset	Cadeia de caracteres	Exatamente um	dcterms: publisher	Universidade Estadual Paulista
type	Tipo de ativo/asset. Seguindo o vocabulário COAR.	Termo de vocabulário controlado	Exatamente um	dcterms:type	Text
license	Licença do ativo asset	Cadeia de caracteres	Exatamente um	dcterms:license	Creative Commons Attribution 4.0 International
license_ref	Referência à descrição da licença	URI	Exatamente um	rdfs:label	https://creativecommons.org/licenses/by/4.0/
language	Idioma do ativo/asset a partir da norma ISO 6391-1 (código do país com dois letras)	Termo de vocabulário controlado	Zero ou um	dcterms:language	pt
description	Descrição do ativo	Cadeia de caracteres	Zero ou um	dcterms:description	Quadro com a classificação de ambientes de publicação científica a partir das categorias fenomenológicas de Charles Peirce.
format	Formato de ativo/asset. Uso do vocabulário IANA Media Types ¹⁰²	Termo de vocabulário controlado	Zero ou um	dcterms: format	text/rtf
data_access	Modo de acesso ao ativo/asset (aberto, restrito, fechado)	Termo de vocabulário controlado	Exatamente um	ali:free_to_read ¹⁰³	aberto

Fonte: Elaboração própria.

¹⁰² IANA Media Types: <https://www.iana.org/assignments/media-types/media-types.xhtml>

¹⁰³ ali:free_to_read permite delimitar as datas de acesso ao ativo/asset, especificar períodos de embargo ou estabelecer o acesso aberto ao conteúdo: https://groups.niso.org/apps/group_public/download.php/14226/rp-22-2015_ALI.pdf

Propriedades de “Texto”

Os atributos descritos para a entidade “texto” são aplicáveis a todas as sub-entidades dele, tais como “texto – medições” ou “texto – resumo”. Na publicação científica atual é comum achar documentos cujo resumo (abstract) está disponível em aberto; porém o resto do texto possui um caráter fechado. Nesta proposta, o acesso aos dados poderia ser definido de forma distinta em cada uma das entidades textuais se for necessário.

Quadro 26 Propriedades da entidade “Texto”.

Nome	Descrição	Tipo de dado	Cardinalidade	Padrão	Valor exemplo
id	Identificador do ativo/asset	Cadeia de caracteres	Exatamente um	dcterms:identifier	https://doi.org/10.1234/5678
id_type	Tipo de identificador	Termo de vocabulário controlado	Exatamente um	rdfs:label	HTTP-DOI
title	Título	Cadeia de caracteres	Exatamente um	dcterms:title	Multimodalidade na publicação científica ampliada: considerações semióticas e modelo de representação
creator	Identificação do criador	Cadeia de caracteres	Um ou mais	dcterms:creator	Paloma Marín Arraiza
creator_id	Identificador do autor	Cadeia de caracteres	Exatamente um	foaf:account	https://orcid.org/0000-0001-7460-7794
issued	Data de publicação	Data	Zero ou um	dcterms: date	2019-10-10
publisher	Editora ou entidade que publica o ativo/asset	Cadeia de caracteres	Exatamente um	dcterms: publisher	Universidade Estadual Paulista
type	Tipo de ativo/asset. Seguindo o vocabulário COAR.	Termo de vocabulário controlado	Exatamente um	dcterms:type	Doctoral thesis
license	Licença do ativo asset	Cadeia de caracteres	Exatamente um	dcterms:license	Creative Commons Attribution 4.0 International
license_ref	Referência à descrição da licença	URI	Exatamente um	rdfs:label	https://creativecommons.org/licenses/by/4.0/

language	Idioma do ativo/asset a partir da norma ISO 6391-1 (código do país com dois letras)	Termo de vocabulário controlado	Zero ou um	dcterms:language	pt
description	Descrição do ativo	Cadeia de caracteres	Zero ou um	dcterms:description	Tese de doutorado do PPGCI da UNESP.
format	Formato de ativo/asset. Uso do vocabulário IANA Media Types ¹⁰⁴	Termo de vocabulário controlado	Zero ou um	dcterms:format	application/pdf
data_access	Modo de acesso ao ativo/asset (aberto, restrito, fechado)	Termo de vocabulário controlado	Exatamente um	ali:free_to_read ¹⁰⁵	aberto

Fonte: Elaboração própria.

Propriedades de “Texto - referências”

A entidade “texto – referencias” é descrita individualmente, já que estabelece vínculos diretos com entidades externas. Para o estabelecimento destes vínculos, utilizam-se os eventos de citação. Estes eventos de citação, seguindo ontologias como Open Citations Ontology, armazenam-se em bancos abertos de dados sobre citações; por exemplo, o banco da agência de registro de DOIs, Crossref. A adequada modelagem e disponibilização destas citações facilita as tarefas de avaliação da ciência mediante métodos cientométricos.

¹⁰⁴ IANA Media Types: <https://www.iana.org/assignments/media-types/media-types.xhtml>

¹⁰⁵ ali:free_to_read permite delimitar as datas de acesso ao ativo/asset, especificar períodos de embargo ou estabelecer o acesso aberto ao conteúdo: https://groups.niso.org/apps/group_public/download.php/14226/rp-22-2015_ALI.pdf

Quadro 27 Propriedades da entidade “Texto - referências”.

Nome	Descrição	Tipo de dado	Cardinalidade	Padrão	Valor exemplo
audio_id	Identificador do ativo/asset	Cadeia de caracteres	Exatamente um	dcterms:identifier	https://doi.org/10.1234/5678
audio_id_type	Tipo de identificador	Termo de vocabulário controlado	Exatamente um	rdfs:label	HTTP-DOI
title	Título	Cadeia de caracteres	Exatamente um	dcterms:title	Referência de: Multimodalidade na publicação científica ampliada: considerações semióticas e modelo de representação
creator	Identificação do criador	Cadeia de caracteres	Um ou mais	dcterms:creator	Paloma Marín Arraiza
creator_id	Identificador do autor	Cadeia de caracteres	Exatamente um	foaf:account	https://orcid.org/0000-0001-7460-7794
publisher	Editora ou entidade que publica o ativo/asset	Cadeia de caracteres	Exatamente um	dcterms:publisher	Universidade Estadual Paulista
type	Tipo de ativo/asset. Seguindo o vocabulário COAR.	Termo de vocabulário controlado	Exatamente um	dcterms:type	Bibliography
license	Licença do ativo asset	Cadeia de caracteres	Exatamente um	dcterms:license	Creative Commons Attribution 4.0 International
license_ref	Referência à descrição da licença	URI	Exatamente um	rdfs:label	https://creativecommons.org/licenses/by/4.0/
language	Idioma do ativo/asset a partir da norma ISO 6391-1	Termo de vocabulário controlado	Zero ou um	dcterms:language	pt

	(código do país com dois letras)				
description	Descrição do ativo	Cadeia de caracteres	Zero ou um	dcterms:description	Referência realizada na tese que conecta com entidades externas.
format	Formato de ativo/asset. Uso do vocabulário IANA Media Types ¹⁰⁶	Termo de vocabulário controlado	Zero ou um	dcterms:format	text/uri-list
cited_document	Entidade citada	Cadeia de caracteres	Exatamente um	oco:cites	
cited_document_id	Identificador do ativo/asset citado	Cadeia de caracteres	Exatamente um	dcterms:identifier	https://doi.org/10.2218/ijdc.v4i1.72
citation_date	Data da citação	Data	Zero ou um	oco:hasCitationCreationDate	2019 – 09 - 03

Fonte: Elaboração própria.

As restantes sub-entidades do texto possuem a mesma descrição básica, embora variem no conteúdo. Porém, os relacionamentos descritos no ponto 6.2.2. variam segundo a entidade de texto e retratam a estrutura das publicações ampliadas proposta no Quadro 11 (Capítulo 5).

Além das entidades próprias do modelo, consideramos importante incluir uma entidade adicional aplicável a todas as anteriores: a entidade “Metadados”. Como expõem os princípios FAIR, é igualmente necessário identificar os metadados. Ainda, como comentado nos Capítulos 3 e 5, no caso de dados fechados ou de acesso restrito, é possível publicar os metadados tanto descritivos como específicos da área do conhecimento.

¹⁰⁶ IANA Media Types: <https://www.iana.org/assignments/media-types/media-types.xhtml>

Propriedades de “Metadados”

Quadro 28 Propriedades da entidade “Metadados”.

Nome	Descrição	Tipo de dado	Cardinalidade	Padrão	Valor exemplo
description	Descrição do tipo de metadados publicados: descritivos, técnicos, próprios da área, próprios do instrumento de medição	Cadeia de caracteres	Zero ou um	dcterms:description	Metadados descritivos do recurso
language	Idioma dos metadados a partir da norma ISO 6391-1 (código do país com dois letras)	Termo de vocabulário controlado	Zero ou um	dcterms:language	pt
metadata_id	Identificador	Cadeia de caracteres	Exatamente um	dcterms:identifier	http://doi.org/10.111/1234
metadata_id_type	Tipo de identificador	Termo de vocabulário controlado	Exatamente um	rdfs:label	HTTP-DOI

Fonte: Elaboração própria.

6.2.2 RELACIONAMENTOS

Quadro 29 Descrição dos relacionamentos

Entidades relacionadas	Relacionamento	Cardinalidade	Descrição
Autor – ativo/asset	is author of (cdesc)	N:N	A entidade descrita possui a autoria da entidade relacionada. Cada autor pode ter vários artigos e cada artigo vários autores
Texto – Texto (sub-entidade)	has part (dcterms)	1:1	A entidade relacionada está incluída física ou logicamente na entidade descrita. Cada texto central pode ter várias sub-entidades. Porém, apenas uma de cada tipo (objetivos, métodos, medições, resultados, discussões, conclusões, referências) e todas elas estão referidas a um único texto central
Texto (palavras chave) – base do conhecimento	denotes (oco)	1:1	A entidade descrita relaciona um ponteiro de referência à entrada bibliográfica que é indicada pelo ponteiro de referência. Cada palavra chave aponta a uma base do conhecimento. Esta base arquivava diferentes fontes relacionadas com a palavra.
Workflow - Texto (métodos)	documents (datacite)	1:N	A entidade relacionada serve de documentação da entidade descrita (p.ex. passos na coleta e processamento dos dados). A metodologia pode estar dividida em vários fluxos de trabalho.
Texto (métodos) – Código/ <i>Software</i>	requires (datacite)	1:N	A entidade descrita requer o uso da entidade relacionada. No trabalho pode ser utilizado mais de um código ou <i>software</i>
Texto (medições) – Código / <i>Software</i>	(re)uses (cdesc)	1:N	A entidade descrita faz uso da entidade relacionada para algum fim (p.ex. análise de dados). Nas medições podem ser utilizados (ou reutilizados) vários <i>softwares</i> ou código.
Texto (resultados) – Dados brutos	is supplemented by (datacite)	1:N	A entidade relacionada suplementa a informação da entidade descrita. Os resultados podem ser suplementados com vários conjuntos de dados brutos.
Texto (resultados) – Planilhas de dados	is supplemented by (datacite)	1:N	Os resultados podem ser suplementados com planilhas de dados processados subjacentes da pesquisa.
Texto (resultados) – Tabela	has part (dcterms)	1:N	Cada apartado de resultados pode possuir várias tabelas
Texto (resultados) – Figura	has part (dcterms)	1:N	Cada apartado de resultados pode possuir várias figuras
Texto (resultados) – Áudio	has part (dcterms)	1:N	Cada apartado de resultados pode possuir vários áudios

Texto (resultados) – Arquivo multimídia	has part (dcterms)	1:N	Cada apartado de resultados pode possuir vários arquivos multimídia
Texto (resultados) – Vídeo	has part (dcterms)	1:N	Cada apartado de resultados pode possuir vários vídeos
Texto (resultados) – Quadro	has part (dcterms)	1:N	Cada apartado de resultados pode possuir vários quadros
Texto (discussões) - Figura	has part (dcterms)	1:N	Cada apartado de discussões pode possuir várias figuras
Texto (discussões) - Imagem	has part (dcterms)	1:N	Cada apartado de discussões pode possuir várias imagens
Texto (discussões) - Quadro	has part (dcterms)	1:N	Cada apartado de discussões pode possuir vários quadros
Texto (referência) – Entidades externas	cites (oco)	N:N	A entidade descrita cita a entidade relacionada, seja direta e explicitamente, indiretamente ou implicitamente. Cada entidade pode citar várias entidades e receber citações de várias entidades que também citam outras entidades.
Dados brutos – Código/ <i>Software</i>	reuses (cdesc)	N:N	Os conjuntos de dados coletados podem ser processados com vários <i>softwares</i> diferentes
Código/ <i>Software</i> – Planilha de dados	Processes (dk)	N:N	A entidade descrita processa a informação da entidade relacionada. Vários <i>softwares</i> /códigos podem ser usados para processar várias planilhas de dados.
Planilha de dados - Dados brutos	hadPrimarySource (oco)	N:N	Especifica a fonte primária da entidade descrita. Várias planilhas de dados podem derivar de vários conjuntos de dados brutos.
Planilha de dados - Tabela	source (dcterm)	N:N	A entidade descrita pode ser derivada da entidade relacionada relacionado. Várias podem derivar de várias planilhas de dados
Tabela - Figura	isVisualisedIn	N:N	A entidade descrita é visualizada na entidade relacionada. As tabelas podem ser visualizadas em diferentes figuras após análise

Fonte: Elaboração própria.

Cada um destes relacionamentos deve ser identificado de forma persistente. Entendemos assim os relacionamentos como “eventos” dentro da nossa publicação.

A identificação persistente do relacionamento entre entidades permite compreender a origem da ação e o caminho seguido e garantir a reprodutibilidade do resultado. Podemos estabelecer esta afirmação como análoga às recomendações 7, 8 e 9 do grupo de trabalho “Data Citation” da Research Data Alliance (RDA):

R7 – Marcação temporal da consulta: Atribuir uma marcação temporal à consulta baseada na última atualização de toda a base de dados (ou à última atualização da seleção de dados afetados pelo ou o tempo de execução da consulta). Isto permite recuperar os dados tal como existiam no momento em que um usuário emitiu uma consulta.

R8 – PID para a consulta: Atribuir um novo PID à consulta se a consulta for nova ou se o conjunto de resultados retornado de uma consulta idêntica anterior for diferente devido a alterações nos dados. Caso contrário, retornar o PID existente.

R9 – Arquivamento da consulta: Armazenar a consulta e os metadados (por exemplo, PID, consulta original e normalizada, soma de verificação da consulta e do conjunto de resultados, data e hora, PID do superconjunto, descrição do conjunto de dados e outros) no arquivo de consultas.¹⁰⁷ (RAUBER *et al.*, 2015, tradução própria).

Estas três recomendações ajudam a garantir a reprodutibilidade desde a consulta, pois permitem reestabelecer as condições do momento que a consulta foi feita. Dessa forma, garante-se que sempre será acessado o conjunto de dados na forma que possuía no momento da marcação temporal.

Alguns repositórios como Zenodo, já trabalham com marcações temporais para o controle de versões e a garantia da integridade do recurso. O repositório Zenodo¹⁰⁸ trabalha com o sistema *fingerprint* “MD5¹⁰⁹ checksum”. A comparação entre *checksums* (somadas de verificação) ajuda a garantir que cada cópia do arquivo seja genuína e livre de erros. Outras iniciativas que trabalham nesta

¹⁰⁷ Tradução própria do original em inglês: **R7 – Query Timestamping:** Assign a timestamp to the query based on the last update to the entire database (or the last update to the selection of data affected by the query or the query execution time). This allows retrieving the data as it existed at the time a user issued a query. **R8 – Query PID:** Assign a new PID to the query if either the query is new or if the result set returned from an earlier identical query is different due to changes in the data. Otherwise, return the existing PID. **R9 – Store Query:** Store query and metadata (e.g. PID, original and normalized query, query & result set checksum, timestamp, superset PID, data set description, and other) in the query store).

¹⁰⁸ Zenodo: <https://zenodo.org/>

¹⁰⁹ MD5: <https://pt.wikipedia.org/wiki/MD5>

corrente são “OriginStamp¹¹⁰” e “Blockchain for Peer Review¹¹¹”. Não obstante, aprofundar nos aspectos da marcação temporal excede o escopo desta tese.

Analogamente, neste modelo, a atribuição de um PID ao relacionamento entre entidades permite identificar a proveniência e acessar o grafo desde o vínculo.

A identificação dos relacionamentos permite também transladar o conceito de dependência ontológica ao modelo. Para tal fim, pode ser utilizado qualquer sistema PID (p.ex. ARK, DOI, Handle ou PURL) e deve ser incluída a entidade de entrada (input) e de saída (output). Como descrito no ponto 6.2.1, cada uma destas entidades possui informação sobre a data na que foram criadas e, no caso de entidades com várias versões, também se inclui informação sobre a versão. Ainda, ao identificar o relacionamento, faz-se constância de quais entidades estão envolvidas e quando se estabeleceu tal relacionamento.

Da mesma forma que as entidades se armazenam em bancos de dados, os relacionamentos armazenam-se como eventos. Neste caso, os eventos constituem em si uma entidade encontrável e recuperável.

6.2.3 DIAGRAMA DO MODELO

A anterior descrição das entidades e relacionamentos estão representadas no seguinte diagrama (ver Figura 17). Esse diagrama mostra a ampliação da publicação a partir da ideia de publicação modular proposta por Kircz (1998; 2002).

No diagrama é possível visualizar graficamente os relacionamentos entre entidades e a cardinalidade desses relacionamentos. Por sua vez, as não todos os atributos descritos para as entidades aparecem no diagrama. Para a visualização, escolheram-se apenas os atributos mínimos para garantir a integridade dos ativos/assets representados em casa uma das entidades. Estes atributos são o identificador persistente, designado no modelo como “PK” (*Principal Key* ou identificador principal); o criador e a instituição editora, dois atributos essenciais para marcar a proveniência do ativo/asset; a data de

¹¹⁰ OriginStamp: <https://originstamp.org>

¹¹¹ Blockchain for Peer Review: <https://www.blockchainpeerreview.org/>

publicação, atributo que também permite estabelecer uma marca temporal essencial para o controle das possíveis versões do ativo/asset; o título e o tipo de ativo/asset, atributo que define a natureza do ativo/asset e que, em combinação com outro atributos como o formato, baseará a estratégia de preservação ao longo prazo e curadoria.

Para a visualização consideramos essenciais os relacionamentos, já que através deles apresenta-se uma parte da nova abordagem ao conceito de publicação ampliada: a atribuição de um PID para cada relacionamento como forma de translação da estrutura ontológica ao modelo.

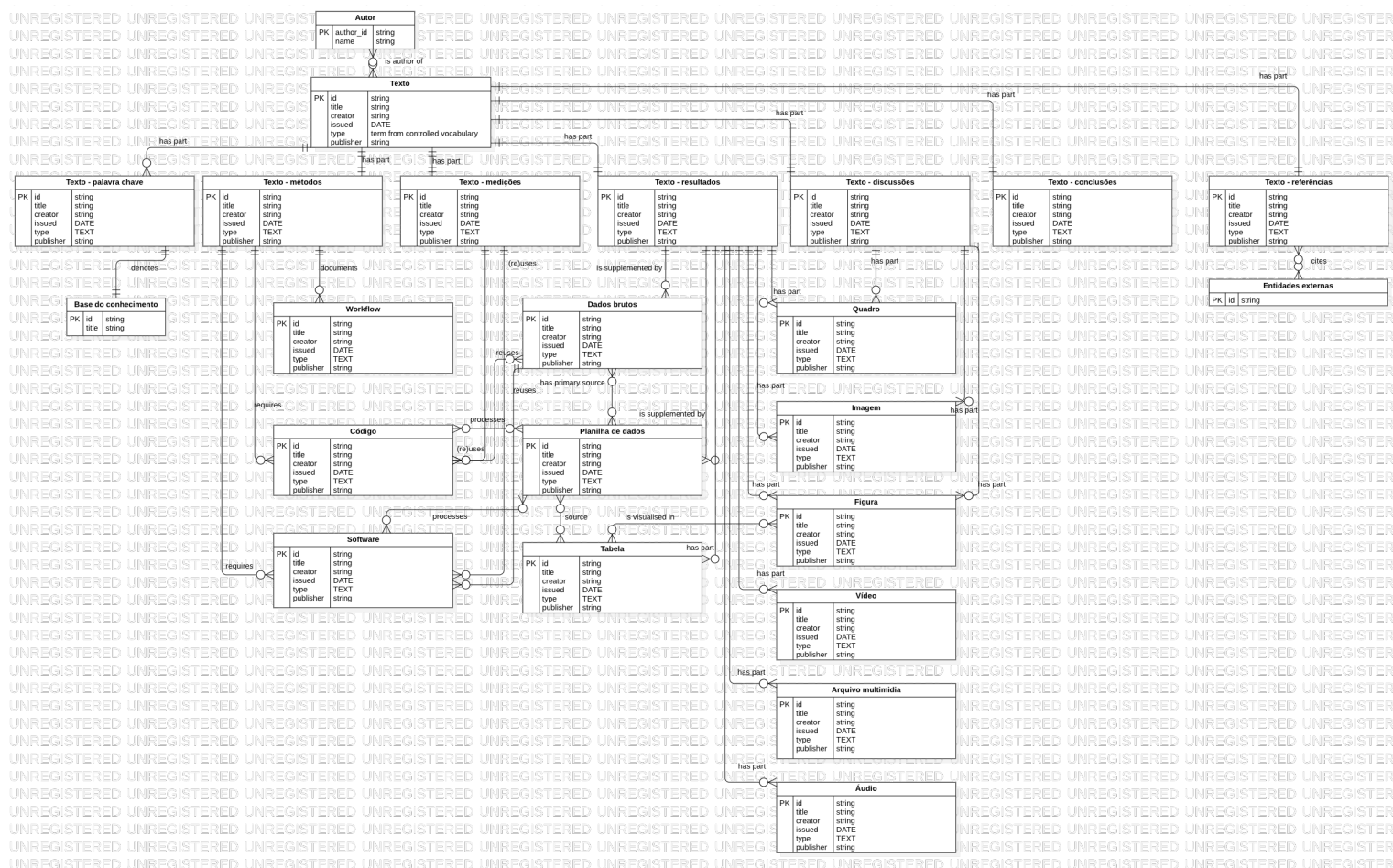
Para utilizar o modelo como base para desenhar uma infraestrutura, devem ser considerados todos os atributos descritos no ponto 6.2.1.

No diagrama aparece também uma “entidade externa” que se relaciona com a publicação ampliada a partir das referências citadas no texto e nos outros ativos/assets. O relacionamento entre a entidade “texto – referencias” e as “entidades externas” é um evento de citação, como definido na Open Citation Ontology.

Aparentemente, esse diagrama mantém uma estrutura com foco na parte narrativa, o texto. Porém, entendemos que esta parte narrativa pode ser desenvolvida em outros formatos também, por exemplo, no formato audiovisual o que geraria um artigo em vídeo (*vídeo article*). Por sua vez, não todas as entidades representadas têm que formar parte de uma modelagem de uma publicação ampliada, senão que podem ser selecionadas aquelas que resultem de interesse. Por exemplo, se utilizarmos apenas as entidades “texto - medições” e “texto - métodos” junto com as entidades “*workflow*”, “*software*”, “dados brutos” e “dados processados”, bem como os relacionamentos entre elas, estaríamos representando a modelagem de um artigo de dados.

Ainda, representar a modelagem dos dados de uma publicação ampliada seguindo o modelo ER facilita a translação depois a um sistema de banco de dados que possa arquivar esse tipo de publicações e se ajusta aos requerimentos para infraestruturas do marco de trabalho proposto por Bardi e Manghi (2015). Além disso, a partir do modelo ER é possível estruturar os dados em triplas seguindo o marco RDF, o que permitiria a disponibilização desses dados em um entorno de *Linked Data*.

Figura 17 Modelo ER para publicações ampliadas.



Fonte: Elaboração própria. Acesso arquivo do projeto StarUML e ao diagrama do modelo: <https://doi.org/10.6084/m9.figshare.11663802>

6.3 CONSIDERAÇÕES SOBRE O MODELO

Como exposto no início do capítulo este modelo pretende ser genérico e adaptável segundo o domínio e o tipo de entidade que precise ser modelada. Ainda, oferece a descrição mínima para garantir a integridade de cada entidade (como apresentado no Capítulo 4) e torna-la um objeto FAIR (apresentado no Capítulo 5). Nem todas as entidades descritas precisam ser incluídas em todas as representações. Por exemplo, se o texto não possui uma parte específica de medições, não será necessário incluir esta entidade individualmente.

Segundo a área do conhecimento a que pertence a publicação modelada, encontram-se uns ativos/assets e outros, como exposto no Capítulo 5 ao introduzir os dados de pesquisa segundo áreas. Por exemplo, na área de climatologia, haverá um trabalho maior com conjuntos de dados brutos observacionais resultado de medições em centrais meteorológicas. Consequentemente, estes conjuntos de dados e as planilhas após processamento seriam entidades do modelo. Por sua vez, a área de artes ou museologia, trabalhará mais com o recurso imagético e precisará uma descrição mais detalhada dele¹¹². Portanto, será necessário modelar esse recurso imagético; porém, é possível que não exista a entidade “*Software*”.

Ainda, a granularidade na representação e identificação de ativos/assets —ou seja, se estes devem ser considerados ou não citáveis e preserváveis de forma independente— é uma decisão que deve ser tomada como parte da planificação do projeto de pesquisa. Da mesma forma, devem ser definidos os direitos de acesso de cada um dos ativos/assets.

Este modelo serve como base conceitual para ambientes de publicação genéricos, como um repositório institucional ou um megaperiódico. Porém, também pode basear o desenvolvimento de ambientes mais específicos, tendo em consideração que estes ambientes mais específicos precisarão de mais campos de metadados ajustados à área de conhecimento.

Além disso, recuperando a definição de publicação ampliada como espaço semiótico, é possível tanto gerar núcleos dentro desta modelagem

¹¹² Para mais informação, pode ser consultada a tese de doutorado de Ana Carolina Simionato do PPGCI da UNESP.

quanto expandir esta modelagem para incluir estruturas externas ao espaço semiótico.

As entidades “Texto – métodos”, “Texto – medições”, “*Workflow*” e “Dados brutos” e seus correspondentes relacionamentos constituem uma publicação ampliada por si só: um artigo de dados. Também, as entidades “Conjunto de dados brutos” e “Planilha de dados” fazem parte da modelagem para planos de gestão de dados legíveis por máquinas (maDMP)¹¹³ proposta pela Research Data Alliance. Juntar as duas modelagens permite traspassar as fronteiras definidas no nosso modelo, que faz referência apenas a fase de publicação pós-projeto, e incluir informações de fases anteriores, como é a fase pré-projeto caracterizada pelo planejamento da pesquisa e o uso de planos de gestão de dados.

Finalmente, o modelo pode se expandir se considerarmos mais entidades físicas, como instrumentos de medição utilizados na pesquisa ou instalações onde a pesquisa foi desenvolvida e que contam com os instrumentos necessários. É possível modelar estas entidades partindo do padrão CSMD (The core Scientific Metadata Model) que apresenta entidades como “Instalação” (csmd:Facility) ou “Instrumento” (csmd:Instrument). Por exemplo, nas pesquisas desenvolvidas com Eye-Tracker no grupo de pesquisa Novas Tecnologias da Informação da UNESP, o aparelho do Eye-Tracker é o instrumento e o Laboratório 2 do prédio de informática do Campus de Filosofia e Ciências da UNESP, a instalação.

¹¹³ maDMP Common Standard RDA: <https://github.com/RDA-DMP-Common/RDA-DMP-Common-Standard>

7 CONSIDERAÇÕES FINAIS

Em virtude das novas necessidades de vinculação e integração de ativos/assets de pesquisa, esta pesquisa traz uma abordagem para o contexto da comunicação científica e a publicação de resultados.

A arquitetura hipertextual que o ambiente *web* proporciona permite a vinculação de recursos de forma general. Porém, para o caso dos ativos/assets de pesquisa observa-se uma carência de boas orientações de como descrever estes ativos/assets e vinculá-los entre si para garantir sua encontrabilidade e acesso. Ao mesmo tempo, observa-se que, no âmbito da comunicação científica, a publicação textual —o artigo—ainda é considerado o veículo principal de comunicação de resultados, o que impede, às vezes, a valoração de outros ativos/assets. Dada também a necessidade de aportar uma base sólida para o processo de pesquisa, garantir a sua transparência deste e reprodutibilidade, é necessário considerar o valor de todos os ativos/assets oriundos do processo.

A transparência também pode ser considerada desde um ponto de vista social, já que a adequada disponibilização dos ativos/assets é uma forma de devolver para a sociedade o conhecimento científico-acadêmico desenvolvido nas instituições públicas de pesquisa. Não obstante, consideramos que somente disponibilizar os ativos/assets não é suficiente para que exista um retorno do conhecimento para a sociedade. As atividades de disseminação e divulgação da ciência e da atividade acadêmica são essenciais também para esse fim. Combinando estas duas ações é possível conseguir um maior engajamento da sociedade nas atividades científicas.

Por sua vez, cada ativo/assets pode ser de maior ou menor interesse dependendo da comunidade discursiva a qual pertença o leitor ou pessoa que interaja com o conteúdo. Assim, os dados brutos numéricos podem ser de maior interesse para um cientista de dados ou alguém que trabalhe com aspectos da reprodutibilidade de resultados; enquanto um resumo em vídeo pode chamar mais a atenção de um jornalista científico que trabalhe em atividades de divulgação do conhecimento.

Consequentemente, foram estabelecidas duas hipóteses para esta tese:

1. a publicação ampliada pode ser conceitualizada como um espaço semiótico composto por ativos/assets heterogêneos e que abrangem as interações dentro do contexto da comunicação científica.
2. a completa descrição dos ativos/assets dentro de um modelo de publicação ampliada serve como mecanismo de publicação dos diferentes ativos/assets científico-acadêmico.

Estabelece-se assim a publicação ampliada como objeto de pesquisa desta tese, onde é contextualizado com um mecanismo para a publicação e vinculação dos ativos/assets de pesquisa. Porém, reconhecem-se os desafios que esse tipo de publicação apresenta desde uma perspectiva epistemológica, teórica e técnica.

Por isso, escolhe-se o método quadripolar como procedimento metodológico. Como exposto no Capítulo 2, o método quadripolar possui uma plasticidade que se adapta à complexidade das pesquisas da Ciência da Informação.

Epistemologicamente, a mudança de paradigma na comunicação e publicação científica, motivado pelo movimento de ciência aberta e pelas considerações pós-custodiais de disponibilização da informação, requerem de novas abordagens para entender os sistemas e ciclos de publicação. Assim, define-se uma nova abordagem à publicação científica que entende a publicação ampliada como um espaço semiótico e não somente como um documento público. Não obstante, continua-se utilizando a palavra “publicação” como metáfora para indicar o ato de fazer pública uma informação e pela associação prática que temos à ideia de publicação.

Teoricamente, analisam-se conceitos semióticos e estruturas de dados para estabelecer nos marcos nos que a publicação ampliada se integra. Tecnicamente, aborda-se a modelagem conceitual dos ativos/assets de uma publicação ampliada considerando a integridade de cada um deles e as dependências semânticas entre eles.

Estas perspectivas foram abordadas no método quadripolar, que se adapta à complexidade das pesquisas em Ciência da Informação e divide as fases de estudo em quadro polos conectados entre si. Os quatro polos, o epistemológico, o teórico, o técnico e o morfológico, como exposto no Capítulo

2, abrangem a totalidade das tarefas desenvolvidas na pesquisa. Os polos observam-se ao longo do texto.

O polo epistemológico, principalmente presente no Capítulo 3, aborda a mudança de paradigma no fazer científico como principal catalizador na necessidade de publicação de diversos ativos/assets. Neste capítulo, destaca-se também como a mudança de paradigma não é exclusiva do movimento de ciência aberta, senão também da importância atual da gestão de dados de pesquisa como maneira de garantir a disponibilidade, reuso e reprodutibilidade dos ativos/assets de pesquisa. Ainda, esta mudança entende uma ciência desenvolvida nos ambientes web e com um foco no tratamento de dados de diferente natureza e com distintos marcos legais.

O polo teórico está presente nos Capítulos 4 e 5, onde apresentam-se os conceitos semióticos e de publicação ampliada que resultarão na definição Semiótica de publicação ampliada e na proposta de estrutura para a mesma.

O polo técnico observa-se no Capítulo 4 com a análise Semiótica dos ambientes web de publicação e no Capítulo 6 com a análise semântica das entidades e os relacionamentos e a descrição dos atributos das entidades.

O polo morfológico abrange o resultado final completo, ou seja, esta tese com os seus correspondentes ativos/assets, bem como a modelagem e definição propostas para publicações ampliadas. Neste polo, entendendo o processo de doutorado de forma global, incluem-se também aquelas produções científicas desenvolvidas ao longo do doutorado e que, direta ou indiretamente, contribuem ao resultado final.

Desta forma, atinge-se o objetivo geral desta pesquisa fixado na proposta, considerando os aspectos semióticos, de um modelo para publicações ampliadas que sirva mecanismo para a integração dos diferentes ativos/assets científicos na comunicação científica.

Para a consecução do objetivo principal, estabeleceram-se quatro objetivos específicos.

O primeiro deles, contextualizar as mudanças que a ciência aberta leva ao contexto da comunicação científica, foi atendido no Capítulo 3. Neste capítulo aprofunda-se no movimento de ciência aberta, sua taxonomia e escolas de pensamento. Ainda, expõem-se algumas situações que impossibilitariam a publicação dos ativos/assets de pesquisa ou, pelo menos, de uma parte deles.

Isto acontece especialmente quando tratarmos com os denominados dados sensíveis ou confidenciais. Também se expõem quais são as implicações que o movimento de ciência aberta traz para a publicação ampliada, especialmente na descrição de ativos/assets diferentes do texto.

O segundo deles, definir a publicação ampliada desde a ótica da Semiótica Peirceana considerando conceitos como a linguagem híbrida e incluído conceitos como a semiosfera da Semiótica da linguagem foi atendido no Capítulo 4. Porém, as implicações semióticas estendem-se também ao longo do Capítulo 6. O Capítulo 4 apresenta a Semiótica Peirceana como teoria dos signos, analisa a classificação dos signos e as categorias fenomenológicas e como estas permeiam no estudo das matrizes da linguagem de Santaella e na análise de interatividade dos ambientes *web* de publicação. Ainda, apresentam-se as arquiteturas hipertextuais que permitem a ruptura com a leitura linear, favorecendo as estruturas reticulares nas quais se baseia uma publicação ampliada. Neste capítulo, discutem-se também as características mínimas de um objeto digital para garantir sua integridade e as dimensões do objeto digital segundo o seu nível de interação e processamento intelectual. É neste ponto onde se esclarece o uso de ativo/asset ao longo desta tese para denominar os produtos oriundos de uma pesquisa. O ativo/asset é o objeto da dimensão pragmática e é um objeto útil que expõe características apropriadas para os sentidos humanos.

Neste capítulo, introduz-se também o conceito de semiosfera, a topografia da semiosfera e o conceito de fronteira. Estes conceitos serviram para entender a publicação ampliada como um espaço semiótico e basearam a definição proposta.

A publicação ampliada é definida desde uma abordagem Semiótica como um espaço semiótico de topografia tanto heterogênea quanto homogênea. Nele, a heterogeneidade é representada pelos ativos/assets de distinta natureza e pertencentes a qualquer matriz da linguagem, cuja identidade é garantida por meio de informação descritiva, identificação persistente, descrição da proveniência e da utilização. A homogeneidade resulta ao entender os ativos/assets como integrantes do mesmo processo de pesquisa. As fronteiras da publicação ampliada são permeáveis, já que cada ativo/asset está em contato e intercâmbio com outros ativos/assets pertencentes à outras publicações

ampliadas, bem como com outros elementos da web. Esse intercâmbio ativa os processos de semiose e transforma em signos informações externas ao espaço semiótico inicial. O dinamismo da publicação ampliada envolve estas interações, enriquecimentos e consequentes evoluções, resultantes da comunicação dentro da comunidade científica e com outras comunidades discursivas. Estas evoluções conectam-se com o interpretante dinâmico peirceano (PEIRCE, 1958, CP 8.315, 1909), o signo adicional que resulta da interpretação de uma mente (humana ou computacional) do signo.

Esta definição garante o cumprimento do segundo objetivo da tese. Além disso, garante a base dos aspectos semióticos considerados na hipótese e que serão aplicados no desenvolvimento do modelo conceitual.

Os objetivos terceiro e quarto, identificar as tecnologias da *web* semântica que estão sendo aplicadas ou podem ser aplicadas nos modelos de dados para publicação ampliada e analisar o ciclo de pesquisa científica para identificar as entidades necessárias na modelagem conceitual para publicações ampliadas, foram atendidos nos capítulos 4, 5 e 6.

Em primeiro lugar, no Capítulo 4, analisam-se cinco ambientes de publicação segundo o seu grau de interatividade. Esta análise abrange as categorias fenomenológicas de Peirce e a ideia de espectro de níveis de interação (limite baixo, meio do espectro e limite superior) de Breure, Voorbij e Hoogerwerf (2011). Como resultado, além da classificação dos ambientes, obtém-se uma lista de entidades que compõem as publicações e que servem para a estrutura proposta para a publicação.

Em segundo lugar, no Capítulo 5, apresenta-se o referencial teórico sobre a publicação ampliada e sua evolução desde as primeiras propostas sobre publicação eletrônica em 1995 até os projetos atuais de integração mediante identificadores persistentes. Como parte deste referencial aparecem os modelos de dados para publicações ampliadas e os marcos de trabalho para sua integração em sistemas da informação. Neste capítulo se detalha o conceito de artigo modular que auxiliará na especificação das entidades textuais de uma publicação ampliada. Ainda, esse capítulo identifica o valor adicional do uso de PIDs para diferentes ações do processo científico, define-se o conceito de PID confiável e se apresentam as características de um sistema PID. Também, comentam-se o papel dos PIDs dentro dos princípios FAIR e na definição de

objeto FAIR. O objeto FAIR está composto por camadas, sendo a primeira a camada dos dados, a segunda a camada dos PIDs, a terceira a camada dos padrões e código e a última a camada dos metadados. Entender os ativos/assets como objetos FAIR e mais uma forma de garantir a sua adequada descrição nos ambientes web. Essas considerações são levadas também ao modelo proposto para publicações ampliadas.

Em terceiro lugar, no Capítulo 6, descrevem-se os passos da análise semântica —Definição do problema e do problema de estudo; identificação das unidades semânticas (entidades); classificação das unidades semânticas; identificação das dependências ontológicas; e finalização do modelo— e aplicam-se no contexto desta pesquisa. Como parte da identificação das entidades, analisam-se padrões comuns aceitados na comunidade científica para a descrição de atividades. Para a descrição das entidades e relacionamentos, foram considerados os seguintes: Access and License Indicators, CSMD: the Core Scientific Metadata Model, Controlled Vocabulary for Resource Type Genres, DataCite Metadata Schema, DCMI Metadata Terms, Data Catalog Vocabulary (DCAT), Friend of a Friend (FOAF), Open Citation Ontology, Schema.org e RDF Schema. Seu uso especifica-se nos pontos 6.2.1 e 6.2.2.

A modelagem conceitual apresentada concebe o objetivo geral deste trabalho e abrange três dos objetivos específicos já elucidados. Esta proposta auxilia na publicação de ativos/assets no contexto de uma publicação ampliada identificando as características mínimas para a integridade de cada ativo/asset. Além disso, apresenta uma questão maior: a identificação persistente dos relacionamentos. Esta identificação é considerada um mecanismo mais para trasladar o conceito ontológico à representação da publicação ampliada, mas também, e inclusive mais importante, para auxiliar na reprodutibilidade. Os ativos/assets são entidades dinâmicas submetidas a versões, a identificação do relacionamento permite reproduzir a estrutura (ou condições) da publicação ampliada independentemente do dinamismo do ativo/asset.

Desta forma, a definição Semiótica de publicação ampliada e o modelo conceitual desenvolvido a partir da análise semântica permitem a confirmação da hipótese de partida desta tese. Consideramos, portanto, que com base neste modelo e as ampliações do mesmo, podem ser desenvolvidos ou melhorados

ambientes para integração dos diferentes ativos/assets científicos na comunicação científica. Entendemos também que continuar desenvolvendo esta proposta aportará ao entendimento das estruturas de publicação e comunicação científica como ecologias informacionais complexas nas que exista “uma conjunção sinérgica de ambientes analógicos, digitais e híbridos, tecnologias analógicas e digitais, utilizadas de maneira holística, em multicanais pelos sujeitos em determinado contexto cultural” (OLIVEIRA; LIMA, 2016, p. 53). Assim os ambientes analógicos podem ser os espaços físicos onde são desenvolvidas as pesquisas, bem como os materiais utilizados para as mesmas. Os espaços digitais são os ambientes web onde são disponibilizados, compartilhados e publicados os ativos/assets, que reúnem uma descrição e são caracterizados como objetos FAIR. Finalmente, os espaços híbridos serão aquelas pontes entre os aspectos digitais e físicos das pesquisas. Por exemplo, a descrição digital de objetos físicos.

Esta tese iniciou-se incidindo na necessidade de abordar as publicações ampliadas desde uma perspectiva epistemológica, teórica e técnica. Seguindo estas perspectivas identificam-se também futuras linhas de trabalho ou aspectos que, ao nosso ver, deveriam ser aprofundados.

Epistemologicamente, a mudança de paradigma entre custodial e pós-custodial deve ser considerada para se entender como será o desenvolvimento de futuras práticas de curadoria digital que abranjam a publicação ampliada. Como exposto por Dallas (2015), estamos frente a uma “fronteira selvagem” onde atores, objetos, processos e ambientes mudam. Torna-se necessário pensar que tipo de infraestruturas sociotécnicas é requerido para que as práticas de curadoria sejam pervasivas. A análise desta “fronteira selvagem” parte da perspectiva pragmática, que reflete nesta tese ao considerar o ativo/asset. Também, esta fronteira contextualiza-se de forma análoga à fronteira da semiosfera.

Outra consideração é a análise e evolução epistemológica do conceito de unidade mínima de informação, bem como as características para a integridade dessa unidade mínima. Nesta tese, a entidade tem adquirido o valor de unidade mínima. Porém, existem abordagens que consideram qualquer afirmação identificada univocamente e atribuível a um autor como unidade mínima. Esse

conceito é o conhecido como “nano publicação¹¹⁴” e pode ser estabelecido como a base de processo de mineração de textos.

Teoricamente, é necessário continuar definindo marcos de trabalho e estabelecendo definições e modelos que permitam trabalhar em um ambiente muito mais orientado por dados, tanto em nível científico-acadêmico quanto em social e de negócios. Desta forma, podem ser desenhadas novas práticas para o tratamento de ativos/assets que não sejam uma adaptação do sistema atual de publicação textual. De forma comum utiliza-se o termo “publicação de dados” considerando os dados como um tipo de artigo acadêmico e, conseqüentemente, pensamos na aplicação dos mesmos modelos de avaliação (CALLAGHAN, 2019). Nesse sentido, convém desenvolver dimensões e métricas para avaliar a qualidade dos dados, especialmente do ponto de vista da sua descrição e da integridade dos metadados.

Além disso, convém analisar a transição do conceito de “publicação ou compartilhamento de dados” ao conceito de “visita de dados” (*data visiting*). A partir deste conceito, deve estabelecer-se uma Internet de máquinas sociais, onde haja mais interação humana e computacional mediante uma melhor definição, representação, curadoria e armazenamento dos dados, baseados em padrões mínimos.

Tecnicamente, deve ser considerado o investimento em infraestruturas de pesquisa (incluindo a ciberinfraestruturas), de preferência abertas¹¹⁵ e na implementação dos objetos FAIR. As implementações técnicas abrangem diferentes aspectos, desde o desenvolvimento e uso de infraestruturas confiáveis e certificadas até a elaboração de planos de sustentabilidade.

¹¹⁴ Nanopub: <http://nanopub.org/wordpress/>

¹¹⁵ Invest in Open: <https://investinopen.org/>

REFERÊNCIAS¹¹⁶

AALBERSBERG, Ij. J.; HEEMAN, F.; KOERS, H.; ZUDILOVA-SEINSTRA, E. Elsevier's Article of the Future enhancing the user experience and integrating data through applications. **Insights: the UKSG journal**, [s. l.], v. 25, n. 1, p. 33–43, 2012. Disponível em: <http://doi.org/10.1629/2048-7754.25.1.33>. Acesso em: 9 mar. 2019. **OA**.

ABNT. **NBR 6022: informação e documentação - artigo em publicação periódica científica impressa - apresentação.**, ABNT, 2003. **OAA**: <https://posticsenasp.ufsc.br/files/2014/04/abntnbr6022.pdf>

ABRAMS, S. A foundational framework for digital curation: The Sept domain model. In: 2015, **Anais...** : UC Office of the President, 2015. Disponível em: <https://escholarship.org/uc/item/75v3z67n>. **OA**.

ABRAMS, S.; CRUSE, P.; KUNZE, J. Preservation Is Not a Place. **International Journal of Digital Curation**, [s. l.], v. 4, n. 1, p. 8–21, 2009. Disponível em: <https://doi.org/10.2218/ijdc.v4i1.72>. Acesso em: 10 mar. 2019. **OA**.

ALBAGLI, S.; CLINIO, A.; RAYCHTOCK, S. Ciência Aberta: correntes interpretativas e tipos de ação | Open Science: interpretive trends and types of action. **Liinc em Revista**, [s. l.], v. 10, n. 2, 2014. Disponível em: <https://doi.org/10.18617/liinc.v10i2.749>. Acesso em: 10 mar. 2019. **OA**.

ALI-KHAN, S. E.; JEAN, A.; MACDONALD, E.; GOLD, E. R. Defining Success in Open Science. **MNI Open Research**, [s. l.], v. 2, p. 2, 2018. Disponível em: <https://doi.org/10.12688/mniopenres.12780.2>. Acesso em: 2 jan. 2019. **OA**.

ANGLADA, L.; ABADAL, E. ¿Qué es la ciencia abierta? **Anuario ThinkEPI**, [s. l.], v. 12, p. 292, 2018. Disponível em: <https://doi.org/10.3145/thinkepi.2018.43>. Acesso em: 10 mar. 2019. **OA**.

ARAYA, E. R. M. **Comunicação científica: agregação, compartilhamento e reúso de elementos informacionais**. 2014. Universidade Estadual Paulista, Marília, 2014. Disponível em: <<https://repositorio.unesp.br/handle/11449/121981>>. **OA**.

ATTWOOD, T. K.; KELL, D. B.; MCDERMOTT, P.; MARSH, J.; PETTIFER, S. R.; THORNE, D. Calling International Rescue: knowledge lost in literature and data landslide! **Biochemical Journal**, [s. l.], v. 424, n. 3, p. 317–333, 2009. Disponível em: <https://doi.org/10.1042/BJ20091474>. Acesso em: 9 mar. 2019. **OA**.

BAIRON, S. **O que é hipermídia**. São Paulo: Brasiliense, 2011.

¹¹⁶ As referências digitais (ou digitalizadas) marcadas com “**OA**” (*open access*) estão disponíveis em acesso aberto direto. As marcadas com “**OAA**” (*open access alternative*) não estão disponíveis em acesso aberto no site do editor; porém existe uma versão aberta no enlace disponibilizado. As marcadas com “**CA**” (*closed access*) não estão disponíveis em acesso aberto e não possuem nenhuma versão aberta.

BARDI, A.; MANGHI, P. Enhanced Publications: Data Models and Information Systems. **LIBER Quarterly**, [s. l.], v. 23, n. 4, p. 240, 2014. Disponível em: <https://doi.org/10.18352/lq.8445>. Acesso em: 9 mar. 2019. **OA**.

BARDI, A.; MANGHI, P. A Framework Supporting the Shift from Traditional Digital Publications to Enhanced Publications. **D-Lib Magazine**, [s. l.], v. 21, n. 1/2, 2015. Disponível em: <https://doi.org/10.1045/january2015-bardi>. Acesso em: 9 mar. 2019. **OA**.

BARTLING, S.; FRIESIKE, S. Towards Another Scientific Revolution. In: BARTLING, S.; FRIESIKE, S. (Eds.). **Opening Science**. Cham: Springer International Publishing, 2014. p. 3–15 Disponível em: https://doi.org/10.1007/978-3-319-00026-8_1. Acesso em: 9 mar. 2019. **OA**.

BECHHOFFER, S.; BECHHOFFER, S.; DE ROURE, D.; GAMBLE, M.; GOBLE, C.; BUCHAN, I. Research Objects: Towards Exchange and Reuse of Digital Knowledge. **Nature Precedings**, [s. l.], 2010. Disponível em: <http://doi.org/10.1038/npre.2010.4626.1>. Acesso em: 9 mar. 2019. **OA**.

BECKER, C. Metaphors We Work By: Reframing Digital Objects, Significant Properties, and the Design of Digital Preservation Systems. **Archivaria**, [s. l.], n. 85, p. 6–37, 2018. Disponível em: <https://archivaria.ca/index.php/archivaria/article/view/13628>. Acesso em: 9 mar. 2019. **OAA**: <http://hdl.handle.net/1807/87826>.

BJÖRK, B.-C. A lifecycle model of the scientific communication process. **Learned Publishing**, [s. l.], v. 18, n. 3, p. 165–176, 2005. Disponível em: <http://doi.org/10.1087/0953151054636129>. Acesso em: 29 set. 2019. **CA**.

BORGMAN, C. L. Data, disciplines, and scholarly publishing. **Learned Publishing**, [s. l.], v. 21, n. 1, p. 29–38, 2008. Disponível em: <http://doi.org/10.1087/095315108X254476>. Acesso em: 9 mar. 2019. **CA**.

BORNEA, M. A.; DOLBY, J.; KEMENTSIETSIDIS, A.; SRINIVAS, K.; DANTRESSANGLE, P.; UDREA, O.; BHATTACHARJEE, B. Building an efficient RDF store over a relational database. In: PROCEEDINGS OF THE 2013 INTERNATIONAL CONFERENCE ON MANAGEMENT OF DATA - SIGMOD '13 2013, New York, New York, USA. **Anais...** In: THE 2013 INTERNATIONAL CONFERENCE. New York, New York, USA: ACM Press, 2013. Disponível em: <https://doi.org/10.1145/2463676.2463718>
Acesso em: 30 set. 2019. **OAA**: <https://cs.uwaterloo.ca/~gweddell/cs848/papers/Bornea.pdf>.

BORNMAN, L.; MUTZ, R. Growth rates of modern science: A bibliometric analysis based on the number of publications and cited references. **arXiv:1402.4578 [physics, stat]**, [s. l.], 2014. Disponível em: <http://arxiv.org/abs/1402.4578>. Acesso em: 10 mar. 2019. **OA**.

BRAMMER, G. R.; CROSBY, R. W.; MATTHEWS, S. J.; WILLIAMS, T. L. Paper

Mâché: Creating Dynamic Reproducible Science. **Procedia Computer Science**, [s. l.], v. 4, p. 658–667, 2011. Disponível em: <https://doi.org/10.1016/j.procs.2011.04.069>. Acesso em: 9 mar. 2019. **OA**.

BREURE, L. Transforming a research paper into a rich internet publication. **Information Services & Use**, [s. l.], v. 34, n. 3–4, p. 335–344, 2014. Disponível em: <https://doi.org/10.3233/ISU-140757>. Acesso em: 9 mar. 2019. **OA**.

BREURE, L.; VOORBIJ, H.; HOOGERWERF, M. Rich Internet Publications: “Show what you tell”. **Journal of Digital Information**, [s. l.], v. 12, n. 1, 2011. Disponível em: <https://journals.tdl.org/jodi/index.php/jodi/article/view/1606>. **OA**.

BUFREM, L. S. Configurações da pesquisa em Ciência da Informação. [s. l.], v. 14, n. 6, p. 13, 2013. Disponível em: <http://www.brapci.inf.br/index.php/article/download/50777>. **OA**.

CALLAGHAN, S. Research Data Publication: Moving Beyond the Metaphor. **Data Science Journal**, n.18, v.1, p.39-46, 2019. Disponível em: <http://doi.org/10.5334/dsj-2019-039>. Acesso em: 10 ago. 2019. **OA**.

CALLAGHAN, S.; DONEGAN, S.; PEPLER, S.; THORLEY, M.; CUNNINGHAM, N.; KIRSCH, P.; AULT, L.; BELL, P.; BOWIE, R.; LEADBETTER, A.; LOWRY, R.; MONCOIFFÉ, G.; HARRISON, K.; SMITH-HADDON, B.; WEATHERBY, A.; WRIGHT, D. Making Data a First Class Scientific Output: Data Citation and Publication by NERC’s Environmental Data Centres. **International Journal of Digital Curation**, [s. l.], v. 7, n. 1, p. 107–113, 2012. Disponível em: <https://doi.org/10.2218/ijdc.v7i1.218>. Acesso em: 16 mar. 2019. **OA**.

CARBONELL-CARRERA, C.; SAORÍN, J.-L.; MEIER, C.; MELIÁN-DÍAZ, D.; DE-LA-TORRE-CANTERO, J. Tecnologías para la incorporación de objetos 3D en libros de papel y libros digitales. **El Profesional de la Información**, [s. l.], v. 25, n. 4, p. 661, 2016. Disponível em: <https://doi.org/10.3145/epi.2016.jul.16>. Acesso em: 9 mar. 2019. **OA**.

CASATI, F.; GIUNCHIGLIA, F.; MARCHESE, M. Liquid Publications: Scientific Publications meet the web. [s. l.], 2007. Disponível em: <http://eprints.biblio.unitn.it/1313/>. **OA**.

CAULFIELD, T.; HARMON, S. H.; JOLY, Y. Open science versus commercialization: a modern research conflict? **Genome Medicine**, [s. l.], v. 4, n. 2, p. 17, 2012. Disponível em: <http://doi.org/10.1186/gm316>. Acesso em: 10 mar. 2019. **OA**.

CHEN, P. The Entity-Relationship Model – Toward a Unified View of Data. **ACM Transactions on Data Systems**, [s. l.], v. 1, n. 1, p. 9–36, 1976. Disponível em: <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.526.369&rep=rep1&type=pdf>. **OA**.

CHEN, P. Entity-Relationship Modeling: Historical Events, Future Trends, and Lessons Learned. In: BROY, M.; DENERT, E. (Eds.). **Software Pioneers**. Berlin,

Heidelberg: Springer Berlin Heidelberg, 2002. p. 296–310. Disponível em: https://link.springer.com/chapter/10.1007/978-3-642-59412-0_17. **OAA:** http://www.csc.lsu.edu/~chen/pdf/Chen_Pioneers.pdf.

CHEN, X.; DALLMEIER-TIESSEN, S.; DASLER, R.; FEGER, S.; FOKIANOS, P.; GONZALEZ, J. B.; HIRVONSALO, H.; KOUSIDIS, D.; LAVASA, A.; MELE, S.; RODRIGUEZ, D. R.; ŠIMKO, T.; SMITH, T.; TRISOVIC, A.; TRZCINSKA, A.; TSANAKTSIDIS, I.; ZIMMERMANN, M.; CRANMER, K.; HEINRICH, L.; WATTS, G.; HILDRETH, M.; LLORET IGLESIAS, L.; LASSILA-PERINI, K.; NEUBERT, S. Open is not enough. **Nature Physics**, [s. l.], v. 15, n. 2, p. 113–119, 2019. Disponível em: <https://doi.org/10.1038/s41567-018-0342-2>. **OA**.

CHESBROUGH, H. **From Open Science to Open Innovation**, Science/Business Publishing, 2015. Disponível em: <https://www.fosteropenscience.eu/sites/default/files/pdf/1798.pdf>. **OA**.

COMISSÃO EUROPEIA (ED.). **Open innovation, open science, open to the world: a vision for Europe**. Luxembourg: Publications Office of the European Union, 2016. Disponível em: <https://ec.europa.eu/digital-single-market/en/news/open-innovation-open-science-open-world-vision-europe>. **OA**.

CORDEIRO, D.; BRAGHETTO, K. R.; GOLDMAN, A.; KON, F. Da ciência à eciência: paradigmas da descoberta do conhecimento. **Revista USP**, [s. l.], v. 0, n. 97, p. 71, 2013. Disponível em: <https://doi.org/10.11606/issn.2316-9036.v0i97p71-81>. Acesso em: 10 mar. 2019. **OA**.

CORTÁZAR, J. **Rayuela**. Buenos Aires: Editorial Sudamericana, 1963.

DALLAS, C. Digital curation beyond the “wild frontier”: a pragmatic approach. *Arch Sci*, v.16, n.4, p.421-457, 2015. Disponível em: <https://doi.org/10.1007/s10502-015-9252-6>. Acesso em: 10 ago. 2019. **OA**.

DALLMEIER-TIESSEN, S.; LAVASA, A.; HERTERICH, P.; RUEDA, L.; KOTARSKI, R.; NEWBOLD, E. A comparative analysis of disciplinary data management workflows. In: IEEE/ACM JOINT CONFERENCE ON DIGITAL LIBRARIES 2014, London, United Kingdom. **Anais...** In: 2014 IEEE/ACM JOINT CONFERENCE ON DIGITAL LIBRARIES (JCDL). London, United Kingdom: IEEE, 2014. Disponível em: <http://ieeexplore.ieee.org/document/6970180/>. Acesso em: 16 mar. 2019. **CA**.

DAPPER, A.; FARQUHAR, A.; KOTARSKI, R.; HEWLETT, K. Connecting the Persistent Identifier Ecosystem: Building the Technical and Human Infrastructure for Open Research. **Data Science Journal**, [s. l.], v. 16, 2017. Disponível em: <http://doi.org/10.5334/dsj-2017-028>. Acesso em: 9 mar. 2019. **OA**.

DE BRUYNE, P.; HERMAN, J.; DE SCHOUTHEETE, M. **Dinâmica da Pesquisa em Ciências Sociais. Os pólos da prática metodológica**. Rio de Janeiro: F.Alves., 1982.

DE ROURE, D. The future of scholarly communications: Based on a paper

presented at the 37th UKSG Conference, Harrogate, April 2014. **Insights: the UKSG journal**, [s. l.], v. 27, n. 3, p. 233–238, 2014. Disponível em: <http://doi.org/10.1629/2048-7754.171>. Acesso em: 10 mar. 2019. **OA**.

DE ROURE, D.; GOBLE, C.; ALEKSEJEVS, S.; BECHHOFFER, S.; BHAGAT, J.; CRUICKSHANK, D.; FISHER, P.; HULL, D.; MICHAELIDES, D.; NEWMAN, D.; PROCTER, R.; LIN, Y.; POSCHEN, M. Towards open science: the myExperiment approach. **Concurrency and Computation: Practice and Experience**, [s. l.], v. 22, n. 17, p. 2335–2353, 2010. Disponível em: <http://doi.org/10.1002/cpe.1601>. Acesso em: 10 mar. 2019. **OA**.

DE ROURE, D.; GOBLE, C.; STEVENS, R. The design and realisation of the Virtual Research Environment for social sharing of workflows. **Future Generation Computer Systems**, [s. l.], v. 25, n. 5, p. 561–567, 2009. Disponível em: <https://doi.org/10.1016/j.future.2008.06.010>. Acesso em: 11 mar. 2019. **OAA**: <https://eprints.soton.ac.uk/265709/1/fgcs.pdf>

DE SALVADOR AGRA, S. Semiosis en el entorno hipertextual. **AdVersus: Revista de Semiótica**, [s. l.], n. 30, p. 66–89, 2016. Disponível em: <http://www.adversus.org/indice/nro-30/articulos/XIII3003.pdf>. **OA**.

DE WAARD, A. From Proteins to Fairytales: Directions in Semantic Publishing. **IEEE Intelligent Systems**, [s. l.], v. 25, n. 2, p. 83–88, 2010. Disponível em: <https://doi.org/10.1109/MIS.2010.49>. Acesso em: 17 mar. 2019. **CA**.

DE WAARD, A.; BUCKINGHAM SHUM, S.; CARUSI, A.; PARK, J.; SAMWALD, M.; SÁNDOR, Á. Hypotheses, evidence and relationships: The Hyper approach for representing scientific knowledge claims. In: PROCEEDINGS 8TH INTERNATIONAL SEMANTIC WEB CONFERENCE 2009, Washington DC. **Anais...** In: 8TH INTERNATIONAL SEMANTIC WEB CONFERENCE. Washington DC.: Springer Verlag, 2009. **OAA**: <http://oro.open.ac.uk/18563/>

DELFANTI, A. **Biohackers: the politics of open science**. London: Pluto Press, 2013. Disponível em: <http://delfanti.org/biohackers/>. **OA**.

DESTRO BISOL, G.; ANAGNOSTOU, P.; CAPOCASA, M.; BENCIVELLI, S.; CERRONI, A.; CONTRERAS, J.; ENKE, N.; FANTINI, B.; GRECO, P.; HEENEY, C.; LUZI, D.; MANGHI, P.; MASCALZONI, D.; MOLLOY, J.; PARENTI, F.; WICHERTS, J.; BOULTON, G. Perspectives on Open Science and scientific data sharing: an interdisciplinary workshop. **Journal of anthropological sciences = Rivista di antropologia : JASS**, [s. l.], v. 92, p. 179–200, 2014. Disponível em: <http://www.isita-org.com/jass/Contents/2014vol92/Destro/25020017.pdf>. **OA**.

FARACE, D.; STOCK, C.; FRANTZEN, J.; SESINK, L.; RABINA, D. L.; GREYNET-GREY LITERATURE NETWORK SERVICE Linking full-text grey literature to underlying research and post-publication data: An Enhanced Publications Project 2011-2012. . [s.l.] : Data Archiving and Networked Services (DANS), 2013. Disponível em: <https://easy.dans.knaw.nl/ui/datasets/id/easy-dataset:53456>. Acesso em: 9 mar. 2019. **OA**.

FARAJ, S; AZAD, B. The Materiality of Technology: an Affordance Perspective. Em: Materiality and Organizing: Social Interaction in a Technological World, 2012.

OAA: https://www.researchgate.net/publication/236591952_The_Materiality_of_Technology_An_Affordance_Perspective

FECHER, B.; FRIESIKE, S. Open Science: One Term, Five Schools of Thought. In: BARTLING, S.; FRIESIKE, S. (Eds.). **Opening Science**. Cham: Springer International Publishing, 2014. p. 17–47. Disponível em: https://doi.org/10.1007/978-3-319-00026-8_2. OA.

FEINSTEIN, A. D.; SCHLIEDER, J. E.; LIVINGSTON, J. H.; CIARDI, D. R.; HOWARD, A. W.; ARNOLD, L.; BARENTSEN, G.; BRISTOW, M.; CHRISTIANSEN, J. L.; CROSSFIELD, I. J. M.; DRESSING, C. D.; GONZALES, E. J.; KOSIAREK, M.; J. LINTOTT, C.; MILLER, G.; MORALES, F. Y.; PETIGURA, E. A.; THACKERAY, B.; AULT, J.; BAETEN, E.; JONKEREN, A. F.; LANGLEY, J.; MOSHINALY, H.; PEARSON, K.; TANNER, C.; TREASURE, J. K2-288Bb: A Small Temperate Planet in a Low-mass Binary System Discovered by Citizen Scientists. **The Astronomical Journal**, [s. l.], v. 157, n. 2, p. 40, 2019. Disponível em: <https://doi.org/10.3847/1538-3881/aafa70>. Acesso em: 10 mar. 2019. OA.

FENNER, M. Altmetrics and Other Novel Measures for Scientific Impact. In: BARTLING, S.; FRIESIKE, S. (Eds.). **Opening Science**. Cham: Springer International Publishing, 2014. p. 179–189. Disponível em: https://doi.org/10.1007/978-3-319-00026-8_12. OA.

FOSTER OPEN SCIENCE. **Open Science definition**, 2018. Disponível em: <https://www.fosteropenscience.eu/foster-taxonomy/open-science-definition>

FRIESIKE, S.; WIDENMAYER, B.; GASSMANN, O.; SCHILDHAUER, T. Opening science: towards an agenda of open science in academia and industry. **The Journal of Technology Transfer**, [s. l.], v. 40, n. 4, p. 581–601, 2015. Disponível em: <http://doi.org/10.1007/s10961-014-9375-6>. Acesso em: 10 mar. 2019. OA.

GIBSON, J.J. The Ecological Approach to Visual Perception. Boston: Houghton Mifflin Harcourt (HMH), 1979. CA.

HEUVEL, H. Van den; HORIK, R. van; SCAGLIOLA, S. I.; SANDERS, E. P.; WITKAMP, P. The VeteranTapes: Research Corpus, Fragment Processing Tool, and Enhanced Publications for the e-Humanities. In: PROCEEDINGS OF LREC 2009, **Anais...** . In: 7TH INTERNATIONAL CONFERENCE ON LANGUAGE RESOURCES AND EVALUATION (LREC). [s.l.: s.n.]. Disponível em: <http://hdl.handle.net/2066/85921>. OA.

HEY, T.; HEY, J. e-Science and its implications for the library community. **Library Hi Tech**, [s. l.], v. 24, n. 4, p. 515–528, 2006. Disponível em: <https://www.emeraldinsight.com/doi/10.1108/07378830610715383>. Acesso em: 10 mar. 2019. OAA: http://eprints.rclis.org/9202/1/heyhey_final_web.pdf

HEY, T.; TANSLEY, S.; TOLLE, K. **The fourth paradigm: data-intensive scientific discovery**. Redmond , Washington: Microsoft Research, 2009. Disponível em: <https://www.immagic.com/eLibrary/ARCHIVES/EBOOKS/M091000H.pdf>.

Acesso em: 10 mar. 2019. **OA**.

HEY, T.; TREFETHEN, A. Cyberinfrastructure for e-Science. **Science**, [s. l.], v. 308, n. 5723, p. 817–821, 2005. Disponível em: <http://doi.org/10.1126/science.1110410>. Acesso em: 9 mar. 2019. **OA**.

HIGMAN, R.; BANGERT, D.; JONES, S. Three camps, one destination: the intersections of research data management, FAIR and Open. **Insights the UKSG journal**, [s. l.], v. 32, p. 18, 2019. Disponível em: <http://doi.org/10.1629/uksq.468/>. Acesso em: 9 jun. 2019. **OA**.

HODSON, S.; JONES, S.; COLLINS, S.; GENOVA, F.; HARROWER, N.; LAAKSONEN, L. *et al.* Turning FAIR data into reality: interim report from the European Commission Expert Group on FAIR data. 2018. Disponível em: <https://doi.org/10.5281/record.1285272>. Acesso: 10 jun. 2019. **OA**.

HOOLAND, S. Van; VERBORGH, R. **Linked data for libraries, archives and museums: how to clean, link and publish your metadata**. London: Facet Publishing, 2014.

HUNTER, J. Scientific Publication Packages – A Selective Approach to the Communication and Archival of Scientific Output. **International Journal of Digital Curation**, [s. l.], v. 1, p. 33–52, 2006. Disponível em: <https://doi.org/10.2218/ijdc.v1i1.4>. Acesso em: 9 mar. 2019. **OA**.

INTERNATIONAL DOI FOUNDATION. **DOI Handbook**. [s.l.: s.n.]. Disponível em: <https://www.doi.org/hb.html>

JANKOWSKI, N. W.; SCHARNHORST, A.; TATUM, C.; TATUM, Z. Enhancing Scholarly Publications: Developing Hybrid Monographs in the Humanities and Social Sciences. **SSRN Electronic Journal**, [s. l.], 2012. Disponível em: <http://dx.doi.org/10.2139/ssrn.1982380>. Acesso em: 9 mar. 2019. **OA**.

JOHNSTON, L. R.; CARLSON, J.; HUDSON-VITALE, C.; IMKER, H.; KOZLOWSKI, W.; OLENDORF, R.; STEWART, C. How Important is Data Curation? Gaps and Opportunities for Academic Libraries. **Journal of Librarianship and Scholarly Communication**, [s. l.], v. 6, n. 1, p. 2198, 2018. Disponível em: <http://doi.org/10.7710/2162-3309.2198>. Acesso em: 9 mar. 2019. **OA**.

KIRCZ, J. G. Modularity: the next form of scientific information presentation? **Journal of Documentation**, [s. l.], v. 54, n. 2, p. 210–235, 1998. Disponível em: <http://doi.org/10.1108/EUM0000000007185>. Acesso em: 9 mar. 2019. **CA**.

KIRCZ, J. G. New practices for electronic publishing 2: New forms of the scientific

paper. **Learned Publishing**, [s. l.], v. 15, n. 1, p. 27–32, 2002. Disponível em: <http://doi.org/10.1087/095315102753303652>. Acesso em: 9 mar. 2019. **CA**.

KLEIN, M.; VAN DE SOMPEL, H. Discovering Scholarly Orphans Using ORCID. **arXiv:1703.09343** [cs], [s. l.], 2017. Disponível em: <http://arxiv.org/abs/1703.09343>. Acesso em: 9 mar. 2019. **OA**.

KRAFT, A.; DREYER, B.; LÖWE, P.; ZIEDORN, F. 14 Years of PID Services at the German National Library of Science and Technology (TIB): Connected Frameworks, Research Data and Lessons Learned from a National Research Library Perspective. **Data Science Journal**, [s. l.], v. 16, 2017. Disponível em: <http://doi.org/10.5334/dsj-2017-036/>. Acesso em: 9 mar. 2019. **OA**.

KRAMER, B.; BOSMAN, J. 101 Innovations in Scholarly Communication - the Changing Research Workflow. . [s.l.] : Figshare, 2015. Disponível em: <https://figshare.com/articles/101_Innovations_in_Scholarly_Communication_the_Changing_Research_Workflow/1286826/1>. Acesso em: 28 abr. 2019. **OA**.

KUHN, T. S. **The structure of scientific revolutions**. Fourth edition ed. Chicago London: The University of Chicago Press, 1962.

KUNZE, J. The ARK Identifier Scheme. 2013. Disponível em: <https://tools.ietf.org/html/draft-kunze-ark-18>. Acesso em: 10 jun. 2019. **OA**.

LANCASTER, F. W. The Evolution of Electronic Publishing. **Library Trends**, [s. l.], v. 43, n. 4, p. 518–527, 1995. Disponível em: <http://hdl.handle.net/2142/7981> **OA**.

LEMKE, J. L. Travels in hypermodality. **Visual Communication**, [s. l.], v. 1, n. 3, p. 299–325, 2002. Disponível em: <http://doi.org/10.1177/147035720200100303>. Acesso em: 16 mar. 2019. **CA**.

LERU. **Open Science and its role in universities: A roadmap for cultural change**. [s.l.: s.n.]. Disponível em: <https://www.leru.org/files/LERU-AP24-Open-Science-full-paper.pdf>. **OA**.

LEVY, P. **As tecnologias da inteligência: o futuro do pensamento na era da informática**. Tradução Carlos Irineu Da Costa. Rio de Janeiro: Editora 34, 1993. **OAA**: <http://www.mozo.pt/tesp/livros/LEVY-Pierre-1998-Tecnologias-da-Inteligencia.pdf>

LIU, K. **Semiotics in information systems engineering**. Cambridge; New York: Cambridge University Press, 2000.

LÓPEZ-BORRULL, A. “Plan S”: La velocitat de l'accés obert depèn del punt de referència? **COMeIN Revista dels Estudis de Ciències de la Informació i de la Comunicació**, [s. l.], n. 84, 2019. Disponível em: <http://comein.uoc.edu/divulgacio/comein/ca/numero84/articles/plan-S-velocitat-acces-obert-punt-referencia.html>. **OA**.

LÓPEZ-PELLICER, F.; BARRERA, J.; GONZÁLEZ, J.; ZARAZAGA-SORIA, F. J.; LÓPEZ, E.; ABAD, P.; RODRIGUEZ, A. F. El desafío de los identificadores persistentes y accionables. In: 2016, **Anais...** In: VII JORNADAS IBÉRICAS DE INFRAESTRUCTURAS DE DATOS ESPACIALES. [s.l.: s.n.] Disponível em: http://www.jiide.org/Jiide-theme/resources/docs/pdf/articulos/09_art_IAAA_IdentificadoresPersistentesAccionables.pdf. OA.

LOTMAN, J. On the semiosphere. **Sign System Studies**, [s. l.], v. 33, n. 1, p. 205–226, 2005. Disponível em: <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.693.9961&rep=rep1&type=pdf>. OA.

LOTMAN, Y. M. **La semiosfera. La Semiótica de la cultura**. Tradução Desiderio Navarro. Valencia: Cátedra, 1996.

MARCONDES, C. H. From Scientific Communications to public Knowledge: The scientific Article Web Published as a Knowledge Base. In: 2005, Leuven (Belgium). **Anais...** In: INTERNATIONAL CONFERENCE ON ELECTRONIC PUBLISHING. Leuven (Belgium) Disponível em: http://eprints.rclis.org/7389/1/ELPUB_2005-Marcondes.pdf. OA.

MARCONDES, C. H. Um modelo semântico de publicações eletrônicas | A semantic model for electronic publishing. **Liinc em Revista**, [s. l.], v. 7, n. 1, 2011. Disponível em: <https://doi.org/10.18617/liinc.v7i1.404>. Acesso em: 10 mar. 2019. OA.

MARÍN ARRAIZA, P.; VIDOTTI, S. El vídeo como ampliación de publicaciones científicas: Descripción y modelos de datos. **Informação & Tecnologia**, [s. l.], v. 5, n. 1, p. 52–64, 2018. Disponível em: <https://doi.org/10.22478/ufpb.2358-3908.2018v5n1.38381>. OA.

MASUZZO, P.; MARTENS, L. Do you speak open science? Resources and tips to learn the language. [s. l.], PeerJ Preprints 5:e2689v1, 2017. Disponível em: <https://doi.org/10.7287/peerj.preprints.2689v1>. Acesso em: 10 mar. 2019. OA.

MEADOWS, A. J.; LEMOS, A. A. B. de L. **A comunicação científica**. Brasília: Briquet de Lemos/livros, 1999.

MCMURRY, J.A.; JUTY, N.; BLOMBERG, N.; BURDETT, T.; CONLIN, T.; CONTE, N.; *et al.* Identifiers for the 21st century: How to design, provision, and reuse persistent identifiers to maximize utility and impact of life science data. *PLoS Biology*, v.15, n.6, e2001414, 2017. Disponível em: <http://doi.org/10.1371/journal.pbio.2001414> . Acesso em: 10 jun. 2019. OA.

MÉNDEZ RODRÍGUEZ, E. **Open Science? Darling, we need to talk**. Berlín.

MERRELL, F. Charles Peirce y sus signos. **Signos en Rotación**, [s. l.], v. Año III, n. 181, 2001. Disponível em: <http://www.unav.es/gep/Articulos/SRotacion3.html>. Acesso: 10 jun. 2019. OA.

MONTEIRO, S. D. A forma eletrônica do hipertexto. **Ciência da Informação**, [s. l.], v. 29, n. 1, p. 25–39, 2000. Disponível em: <http://revista.ibict.br/ciinf/article/view/896>. Acesso em: 16 mar. 2019. **OA**.

MONTEIRO, S. D. Semiótica peirciana e a questão da informação e do conhecimento. **Encontros Bibli: revista eletrônica de biblioteconomia e Ciência da Informação**, [s. l.], v. 11, n. 2, 2007. Disponível em: <http://doi.org/10.5007/1518-2924.2006v11nesp3p43>. Acesso em: 16 mar. 2019. **OA**.

MONTEIRO, S. D.; FIDENCIO, M. V. As dobras semióticas do ciberespaço: da web visível à invisível. **Transinformação**, [s. l.], v. 25, n. 1, p. 35–46, 2013. Disponível em: <http://dx.doi.org/10.1590/S0103-37862013000100004>. Acesso em: 16 mar. 2019. **OA**.

MORATO, J.; SÁNCHEZ-CUADRADO, S.; FRAGA, A.; MORENO-PELAYO, V. Hacia una web semántica social. **El Profesional de la Información**, [s. l.], v. 17, n. 1, p. 78–85, 2008. Disponível em: <https://doi.org/10.3145/epi.2008.ene.09>. Acesso em: 6 out. 2019. **OA**.

MORRIS, C. W. **Fundamentos de la teoría de los signos**. Barcelona: Paidós, 1985.

MUCHERONI, M. L.; SILVA, F. J. M. Da; PALETTA, C. F. Entre a publicação ampliada e a multimodalidade. In: ANAIS DO XVII ENANCIB 2015, **Anais...**. In: XVII ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIAS DA INFORMAÇÃO. [s.l.: s.n.] Disponível em: <http://www.brapci.inf.br/index.php/article/view/0000017566/ce274ff0aecf67a2b5df818c5a1bff3d>. **OA**.

NARANJO, S. Una lectura Semiótica del hipertexto y la comunicación digital. **Intertextos**, [s. l.], n. 4, 2010.

NELSON, T. H. **Literary machines: edition 87.1**. (S.l.: Published by the author), 1987.

NÖTH, W. The topography of Yuri Lotman's semiosphere. **International Journal of Cultural Studies**, [s. l.], v. 18, n. 1, p. 11–26, 2015. Disponível em: <http://doi.org/10.1177/1367877914528114>. Acesso em: 31 ago. 2019. **CA**.

ODIN Consortium. D4.1 Conceptual model of interoperability .Figshare; 2013. Disponível em: <https://doi.org/10.6084/m9.figshare.824314.v1>. Acesso em: 10 jun. 2019. **OA**.

O'HALLORAN, K.; MARIANI, R.; TISSE, C.; PODLASOV, A.; SMITH, B.; NAGARAJAN, A.; FASCIANI, S.; XIAOLING, Q.; ROUSSEL, R. **Multimodal Semiosis, Multimodal Semiotics: Digital Technologies and Techniques for Studying Multimodal Communication**, 2010. Disponível em: <https://semioticon.com/semiotix/2010/03/multimodal-semiosis-multimodal->

[semiotics-digital-technologies-and-techniques-for-studying-multimodal-communication/](#). OA.

OECD. **Making Open Science a Reality**. 2015. Disponível em: <https://doi.org/10.1787/5jrs2f963zs1-en> . Acesso em: 9 mar. 2019. OA.

OLIVEIRA, A. C. S. De; SILVA, E. M. Da. Ciência aberta: dimensões para um novo fazer científico. **Informação & Informação**, [s. l.], v. 21, n. 2, p. 5, 2016. Disponível em: <http://dx.doi.org/10.5433/1981-8920.2016v21n2p5>. Acesso em: 10 mar. 2019. OA.

OLIVEIRA, H. P. C. De; LIMA, I. F. De. Navegando na Arquitetura da Informação Pervasiva: O Artigo Científico como Ecologia Informacional Complexa. [s. l.], v. 1, n. 1, p. 47–61, 2016. Disponível em: <https://revistas.ufrj.br/index.php/rca/article/view/3542/2784>. OA.

PAMPEL, H.; VIERKANT, P.; SCHOLZE, F.; BERTELMANN, R.; KINDLING, M.; KLUMP, J.; GOEBELBECKER, H.-J.; GUNDLACH, J.; SCHIRMBACHER, P.; DIEROLF, U. Making Research Data Repositories Visible: The re3data.org Registry. **PLoS ONE**, [s. l.], v. 8, n. 11, p. e78080, 2013. Disponível em: <https://doi.org/10.1371/journal.pone.0078080>. Acesso em: 9 mar. 2019. OA.

PARK, H.; WOLFRAM, D. An examination of research data sharing and re-use: implications for data citation practice. **Scientometrics**, [s. l.], v. 111, n. 1, p. 443–461, 2017. Disponível em: <http://doi.org/10.1007/s11192-017-2240-2>. Acesso em: 17 mar. 2019. OA.

PEIRCE, C. S. **Principles of philosophy: two volumes in one**. 5. [printing] ed. Cambridge, Mass.: Belknap Press of Harvard Univ. Press, 1932.

PEIRCE, C. S. **The Collected Papers of Charles Sanders Peirce**. Cambridge, MA, USA: Harvard University Press, 1958. v. I–VIII Disponível em: <https://bit.ly/2UjsKFP>

PIWOWAR, H. Value all research products. **Nature**, [s. l.], v. 493, p. 159, 2013. Disponível em: <https://doi.org/10.1038/493159a>. OA.

POMERANTZ, J.; PEEK, R. Fifty shades of open. **First Monday**, [s. l.], v. 21, n. 5, 2016. Disponível em: <https://doi.org/10.5210/fm.v21i5.6360>. Acesso em: 10 mar. 2019. OA.

PONTIKA, N.; KNOTH, P.; CANCELLIERI, M.; PEARCE, S. Fostering open science to research using a taxonomy and an eLearning portal. In: PROCEEDINGS OF THE 15TH INTERNATIONAL CONFERENCE ON KNOWLEDGE TECHNOLOGIES AND DATA-DRIVEN BUSINESS - I-KNOW '15 2015, Graz, Austria. **Anais...** In: THE 15TH INTERNATIONAL CONFERENCE. Graz, Austria: ACM Press, 2015. Disponível em: <https://doi.org/10.1145/2809563.2809571>. Acesso em: 10 mar. 2019. OAA: <http://oro.open.ac.uk/44719/>

PRIEM, J.; GROTH, P.; TARABORELLI, D. The Altmetrics Collection. **PLoS ONE**, [s. l.], v. 7, n. 11, p. e48753, 2012. Disponível em: <https://doi.org/10.1371/journal.pone.0048753>. Acesso em: 10 mar. 2019. **OA**.

RAMALHO, R. A. S.; OUCHI, M. T. Tecnologias Semânticas: Novas Perspectivas para a Representação de Recursos Informacionais; Tecnologías Semánticas: Nuevas Perspectivas para la Representación de los Recursos de Información. **Informação & Informação**, [s. l.], v. 16, n. 3, p. 60–75, 2012. Disponível em: <http://dx.doi.org/10.5433/1981-8920.2011v16n3p60>. Acesso em: 29 abr. 2019. **OA**.

RAUBER, A.; ASMI, A.; VAN UYTVANCK, D.; PRÖLL, S. **Data Citation of Evolving Data: Recommendation of the Working Group on Data Citation: Re.** [s.l.: s.n.]. Disponível em: https://rd-alliance.org/system/files/RDA-DC-Recommendations_151020.pdf. **OA**.

RENTIER, B. Open science: a revolution in sight? **Interlending & Document Supply**, [s. l.], v. 44, n. 4, p. 155–160, 2016. Disponível em: <https://doi.org/10.1108/ILDS-06-2016-0020>. Acesso em: 10 mar. 2019. **OAA**: https://orbi.uliege.be/bitstream/2268/198865/1/Open%20Science%20a%20revolution%20in%20sight_.pdf

ROBINSON, S. **Conceptual modeling for discrete-event simulation**. Boca Raton [u.a.: Taylor & Francis, 2011. Disponível em: <https://doi.org/10.1201/9781439810385>. Acesso em: 29 set. 2019. **OAA**: <http://www.flexsimbrasil.com.br/downloads/ConceptualModelingDES.pdf>

RODRIGUES, F. de A.; SANT'ANA, R. Publicação Ampliada no Contexto de Teses e Dissertações. **Informação & Tecnologia**, [s. l.], v. 3, n. 1, p. 4–26, 2016. Disponível em: <http://www.periodicos.ufpb.br/ojs/index.php/itec/article/view/38248/20145>. Acesso em: 10 mar. 2019. **OA**.

RODRIGUES, Fernando de Assis. **COLETA DE DADOS EM REDES SOCIAIS: Privacidade de dados pessoais no acesso via Application Programming Interface**. 678 f. Tese (Doutorado) - Faculdade de Filosofia e Ciências, Universidade Estadual Paulista, Marília, 2017. Disponível em: <http://hdl.handle.net/11449/149768>. Acesso em: 10 ago. 2019. **OA**.

ROSS, J. S.; KRUMHOLZ, H. M. Ushering in a New Era of Open Science Through Data Sharing: The Wall Must Come Down. **JAMA**, [s. l.], v. 309, n. 13, p. 1355, 2013. Disponível em: <http://doi.org/10.1001/jama.2013.1299>. Acesso em: 10 mar. 2019. **OA**.

ROYAL SOCIETY (GREAT BRITAIN); POLICY STUDIES UNIT. **Science as an open enterprise**. 2012. Disponível em: https://royalsociety.org/~media/Royal_Society_Content/policy/projects/sape/2012-06-20-SAOE.pdf. Acesso em: 9 mar. 2019. **OA**.

SALES, L. F. **Integração semântica de publicações científicas e dados de**

pesquisa: proposta de modelo de publicação ampliada para a área de Ciências Nucleares. 2014. Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2014. Disponível em: <http://hdl.handle.net/ien/853>. Acesso em 9 mar. 2019. **OA**.

SALES, L. F.; SAYÃO, L. F. Enhanced publications: a new model of scientific publications in the nuclear area. **IENT - Progress Report 2013-2014**, [s. l.], v. 2, p. 1, 2015. Disponível em: <http://revistas.ien.gov.br/index.php/ienprogressreport/article/view/131>. Acesso em: 9 mar. 2019. **OA**.

SALES, L. F.; SAYÃO, L. F. Uma proposta de taxonomia para dados de pesquisa. **Conhecimento em Ação**, [s. l.], v. 4, n. 1, p. 31–48, 2019. Disponível em: <https://revistas.ufrj.br/index.php/rca/article/view/26337/14573>. Acesso em: 10 sep. 2019. **OA**.

SANTAELLA, M. L. S. **O que é Semiótica**. São Paulo: Brasiliense, 1983.

SANTAELLA, M. L. S. **Matrizes da linguagem e pensamento: sonora, visual, verbal. Aplicações na hipermídia**. São Paulo: Ed. Iluminuras, 2001.

SIMIONATO, A. C. Modelagem conceitual DILAM: princípios descritivos de arquivos, bibliotecas e museus para o recurso imagético digital. 2015. 200 f. Tese (doutorado) - Universidade Estadual Paulista Júlio de Mesquita Filho, Faculdade de Filosofia e Ciências, 2015. Disponível em: <http://hdl.handle.net/11449/123318>. Acesso em: 10 ago. 2019. **OA**.

SCHLEUBINGER, M.; REX, J. Forschungsdaten veröffentlichen? [s. l.], 2019. Disponível em: <https://doi.org/10.5281/record.3368293>. Acesso em: 24 ago. 2019. **OA**.

SCHMIDT, B.; BERTINO, A.; BEUCKE, D.; BRINKEN, H.; JAHN, N.; MATTHIAS, L.; MIMKES, J.; MÜLLER, K.; ORTH, A.; BARGHEER, M. Open Science Support as a Portfolio of Services and Projects: From Awareness to Engagement. **Publications**, [s. l.], v. 6, n. 2, p. 27, 2018. Disponível em: <https://doi.org/10.3390/publications6020027>. Acesso em: 10 mar. 2019. **OA**.

SHANNON, C. E.; WEAVER, W. **The mathematical theory of communication**. Urbana: Univ. of Illinois Press, 1949.

SILVA, A. M. Da. **A informação da compreensão do fenómeno e construção do objeto científico**. Oporto (Portugal): Edições Afrontamento, 2006.

SILVA, A. M. Da; RIBERO, F. **Das “Ciências” documentais á ciências da informação ensaio epistemológico para um novo modelo curricular**. Oporto (Portugal: Edicoes Afrontamento, 2008.

SPICER, S. Exploring Video Abstracts in Science Journals: An Overview and Case Study. **Journal of Librarianship and Scholarly Communication**, [s. l.], v. 2, n. 2, 2014. Disponível em: <http://doi.org/10.7710/2162-3309.1110>. Acesso

em: 16 mar. 2019. **OA**.

STAMPER, R. **Information in business and administrative systems**. New York: Wiley, 1973.

STEIN, L. D. Towards a cyberinfrastructure for the biological sciences: progress, visions and challenges. **Nature Reviews Genetics**, [s. l.], v. 9, n. 9, p. 678–688, 2008. Disponível em: <https://doi.org/10.1038/nrg2414>. Acesso em: 9 mar. 2019. **CA**.

TENOPIR, C.; ALLARD, S.; DOUGLASS, K.; AYDINOGLU, A. U.; WU, L.; READ, E.; MANOFF, M.; FRAME, M. Data Sharing by Scientists: Practices and Perceptions. **PLoS ONE**, [s. l.], v. 6, n. 6, p. e21101, 2011. Disponível em: <https://doi.org/10.1371/journal.pone.0021101>. Acesso em: 10 mar. 2019. **OA**.

VAN DE SOMPEL, H.; KLEIN, M.; JONES, S. M. Persistent URIs Must Be Used To Be Persistent. **arXiv:1602.09102 [cs]**, [s. l.], 2016. Disponível em: <http://arxiv.org/abs/1602.09102>. Acesso em: 17 mar. 2019. **OA**.

VAN DE SOMPEL, H.; PAYETTE, S.; ERICKSON, J.; LAGOZE, C.; WARNER, S. Rethinking Scholarly Communication: Building the System that Scholars Deserve. **D-Lib Magazine**, [s. l.], v. 10, n. 9, 2004. Disponível em: <http://www.dlib.org/dlib/september04/vandesompel/09vandesompel.html>. Acesso em: 17 mar. 2019. **OA**.

VAN DE SOMPEL, H.; SANDERSON, R.; SHANKAR, H.; KLEIN, M. Persistent Identifiers for Scholarly Assets and the Web: The Need for an Unambiguous Mapping. **International Journal of Digital Curation**, [s. l.], v. 9, n. 1, p. 331–342, 2014. Disponível em: <https://doi.org/10.2218/ijdc.v9i1.320>. Acesso em: 9 mar. 2019. **OA**.

VERHAAR, P. **Report on Object Models and Funtionalities: DRIVER**, Digital Repository Infrastructure Vision for European Research II. 2008. Disponível em: <http://hdl.handle.net/1887/16018>. **OA**.

VICENTE-SAEZ, R.; MARTINEZ-FUENTES, C. Open Science now: A systematic literature review for an integrated definition. **Journal of Business Research**, [s. l.], v. 88, p. 428–436, 2018. Disponível em: <https://doi.org/10.1016/j.jbusres.2017.12.043>. Acesso em: 10 mar. 2019. **OA**.

W3C PROVENANCE INCUBATOR GROUP. What is provenance? In: **W3C Wiki**, 2005. Disponível em: https://www.w3.org/2005/Incubator/prov/wiki/W3C_Provenance_Incubator_Group_Wiki. Acesso em: 19 ago. 2019. **OA**.

WAGENSBERG, J. **Ideas para la imaginación impura: 53 reflexiones en su propia sustancia**. Barcelona: Tusquets, 1998. **CA**.

WATSON, M. When will 'open science' become simply 'science'? **Genome Biology**, [s. l.], v. 16, n. 1, 2015. Disponível em: <https://doi.org/10.1186/s13059-015-0669-2>.

Acesso em: 10 mar. 2019. **OA**.

WHYTE, A.; TEDDS, J. Making the Case for Research Data Management. **DCC Briefing Papers**. Edinburgh: **Digital Curation Centre**., [s. l.], 2011. Disponível em: <http://www.dcc.ac.uk/resources/briefing-papers>. Acesso em: 10 mar. 2019. **OA**.

WILKINSON, M. D. *et al.* The FAIR Guiding Principles for scientific data management and stewardship. **Scientific Data**, [s. l.], v. 3, p. 160018, 2016. Disponível em: <https://doi.org/10.1038/sdata.2016.18>. Acesso em: 10 mar. 2019. **OA**.

WOUTERSEN-WINDHOUWER, S. (ED.). **Enhanced publications: linking publications and research data in digital repositories**. Amsterdam: Amsterdam Univ. Press, 2009. Disponível em: <http://arno.uva.nl/document/150723>. Acesso em: 10 mar. 2019. **OA**.