





Mapeamento de dados como arcabouço para a Arquitetura de Dados da Avaliação da Pós-Graduação Brasileira no contexto da CAPES

Data mapping as a framework for the Data Architecture of the Brazilian Postgraduate Evaluation in the context of CAPES



Emanuelle Torino¹

 <http://lattes.cnpq.br/5042296869081637>
 <https://orcid.org/0000-0002-3791-9884>



Caio Saraiva Coneglian²

 <http://lattes.cnpq.br/3954065076810604>
 <https://orcid.org/0000-0002-6126-9113>



Mariana Baptista Brandt³

 <http://lattes.cnpq.br/3761037263199030>
 <https://orcid.org/0000-0001-8119-7527>

Talita Moreira de Oliveira⁴

 <http://lattes.cnpq.br/2374823276591654>
 <https://orcid.org/0000-0001-5740-4207>

Silvana Aparecida Borsetti Gregorio Vidotti⁵

 <http://lattes.cnpq.br/7390573927636069>
 <https://orcid.org/0000-0002-4216-0374>

¹ Doutora em Ciência da Informação, Universidade Estadual Paulista Júlio de Mesquita Filho (Unesp), Marília, São Paulo, Brasil. Bibliotecária da Universidade Tecnológica Federal do Paraná (UTFPR), Curitiba, Paraná, Brasil. Coordenadora da Coordenação de Serviços Bibliográficos (COBIB), Instituto Brasileiro de Informação em Ciência e Tecnologia (Ibict), Brasília, Distrito Federal, Brasil. emanuelle@utfpr.edu.br.

² Doutor em Ciência da Informação, Universidade Estadual Paulista Júlio de Mesquita Filho (Unesp), Marília, São Paulo, Brasil. Professor e coordenador dos cursos de graduação de Análise e Desenvolvimento de Sistemas e Ciência da Computação da Universidade de Marília (Unimar), Marília, São Paulo, Brasil. Professor do Programa de Pós-graduação em Ciência da Informação, Universidade Estadual Paulista Júlio de Mesquita Filho (Unesp), Marília, São Paulo, Brasil. caio.coneglian@gmail.com.

³ Doutora em Ciência da Informação, Universidade Estadual Paulista Júlio de Mesquita Filho (Unesp), Marília, São Paulo, Brasil. Analista Legislativo - Documentação e informação legislativa na Câmara dos Deputados, Brasília, DF, Brasil. mariana.brandt@unesp.br.

⁴ Doutora em Educação em Ciências, Universidade Federal do Rio Grande do Sul (UFRGS), Porto Alegre, Rio Grande do Sul, Brasil. Analista em Ciência e Tecnologia na Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), Brasília, Distrito Federal, Brasil. talita.oliveira@cpes.gov.br.

⁵ Doutora em Educação, Universidade Estadual Paulista Júlio de Mesquita Filho (Unesp), Marília, São Paulo, Brasil. Professora do Departamento e do Programa de Pós-graduação em Ciência da Informação, Universidade Estadual Paulista Júlio de Mesquita Filho (Unesp), Marília, São Paulo, Brasil. Coordenadora de Tecnologias Aplicadas (COTEA), Instituto Brasileiro de Informação em Ciência e Tecnologia (Ibict), Brasília, Distrito Federal, Brasil. silvana.vidotti@unesp.br.

Resumo

O Ecossistema de Avaliação da Pós-Graduação Brasileira busca aprimorar a disponibilização de dados para processos como a avaliação dos Programas de Pós-Graduação *stricto sensu*, o acompanhamento institucional e a transparência para a sociedade. No entanto, a diversidade estrutural dos sistemas de informação utilizados tem dificultado a padronização e a interoperabilidade dos dados, impactando sua integração e a tomada de decisão. Este estudo objetiva apresentar o mapeamento dos dados da avaliação da pós-graduação *stricto sensu* brasileira no contexto da CAPES, fornecendo subsídios para o desenvolvimento de uma Arquitetura de Dados como alicerce à integração de dados. Trata-se de uma pesquisa de natureza empírica utilizando-se de fonte documental, que investiga um fenômeno com escassez de literatura disponível em âmbito nacional e internacional, evidenciando sua relevância. O estudo fundamenta-se em análise exploratória da estrutura de dados alicerçada em referenciais teóricos e práticos das áreas de Ciência da Informação, Sistemas de Informação e Ciência da Computação. A metodologia adotada consistiu em uma análise detalhada da estrutura de dados utilizada na avaliação da pós-graduação, com ênfase na integração e interoperabilidade dos dados. Como resultado, obteve-se o mapeamento da estrutura de dados da avaliação da pós-graduação *stricto sensu* brasileira no contexto da CAPES, que servirá como base para o desenvolvimento de uma Arquitetura de Dados para esse processo.

Palavras-chave: Arquitetura de Dados; mapeamento de dados; integração de dados; avaliação da pós-graduação.

Abstract

The Brazilian Graduate Evaluation Ecosystem aims to improve data availability for processes such as the evaluation of *stricto sensu* Graduate Programs, institutional monitoring, and transparency for society. However, the structural diversity of the information systems used makes it difficult for data standardization and interoperability, impacting their integration and decision-making processes. This study's objective is to present the mapping of evaluation data of Brazilian *stricto sensu* graduate programs in the context of CAPES, providing support for the development of a Data Architecture to serve as the foundation for data integration. It is an empirical research using documentary sources, that investigates a phenomenon with limited available literature both nationally and internationally, highlighting its relevance. The study is based on an exploratory analysis of the data structure, grounded in theoretical and practical frameworks from the fields of Information Science, Information Systems, and Computer Science. The adopted methodology consisted of a detailed analysis of the data structure used in graduate evaluation, with an emphasis on data integration and interoperability. As a result, the study achieved the mapping of the data structure for the evaluation of Brazilian *stricto sensu* graduate programs in the CAPES context, which will serve as the basis for the development of a Data Architecture for this process.

Keywords: Data Architecture; data mapping; data integration; postgraduate evaluation.

1 INTRODUÇÃO

O ecossistema de avaliação da pós-graduação brasileira tem se desenvolvido buscando favorecer as formas de disponibilização dos dados para diferentes processos, dentre os quais destacam-se: a avaliação dos programas de pós-graduação *stricto sensu*, o acompanhamento do desenvolvimento destes programas pelas próprias instituições e a abertura de dados para a sociedade.

Diferentes sistemas de informação foram utilizados ao longo dos anos, com vistas a favorecer o processamento dos dados pelas instituições envolvidas na avaliação

da pós-graduação *stricto sensu* brasileira, seja por instituições de ensino, agências de fomento e pela própria Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES). Desta forma, a heterogeneidade estrutural desses sistemas apresenta desafios à padronização e integração de dados, o que dificulta a interoperabilidade entre as plataformas, além de tornar mais complexo o processo de tomada de decisão baseado em dados.

O estudo de Nóbrega, Torino, Vidotti, Halicki e Oliveira (2023, p. 2) observou que “No ecossistema brasileiro de Educação, Ciência, Tecnologia e Inovação (ECTI), a coleta de dados armazenados em diferentes sistemas de informação é complexa em função das estruturas e formas de armazenamento dos dados.” Além disso, o estudo possibilitou o desenvolvimento, pela CAPES, de Ferramentas de Apoio à Integração de Sistemas (FAIS) como solução para a integração de dados da avaliação da pós-graduação *stricto sensu* brasileira.

No entanto, tais ferramentas dependem da compreensão das estruturas de dados entre as instituições que oferecem programas de pós-graduação *stricto sensu* e a própria CAPES, que atuam neste processo, respectivamente como fornecedora e consumidora de dados.

Desta forma, este trabalho apresenta uma parte de um projeto de pesquisa que busca desenvolver a Arquitetura de Dados da Avaliação da Pós-Graduação Brasileira no contexto da CAPES. A proposição de tal Arquitetura apoiará a compreensão das estruturas de dados para otimizar os processos de integração.

Para tanto, é necessário compreender a estrutura atual dos dados e avaliá-la para que a Arquitetura de Dados a ser desenvolvida possa refletir a necessidade da CAPES no que tange à avaliação da pós-graduação *stricto sensu* brasileira.

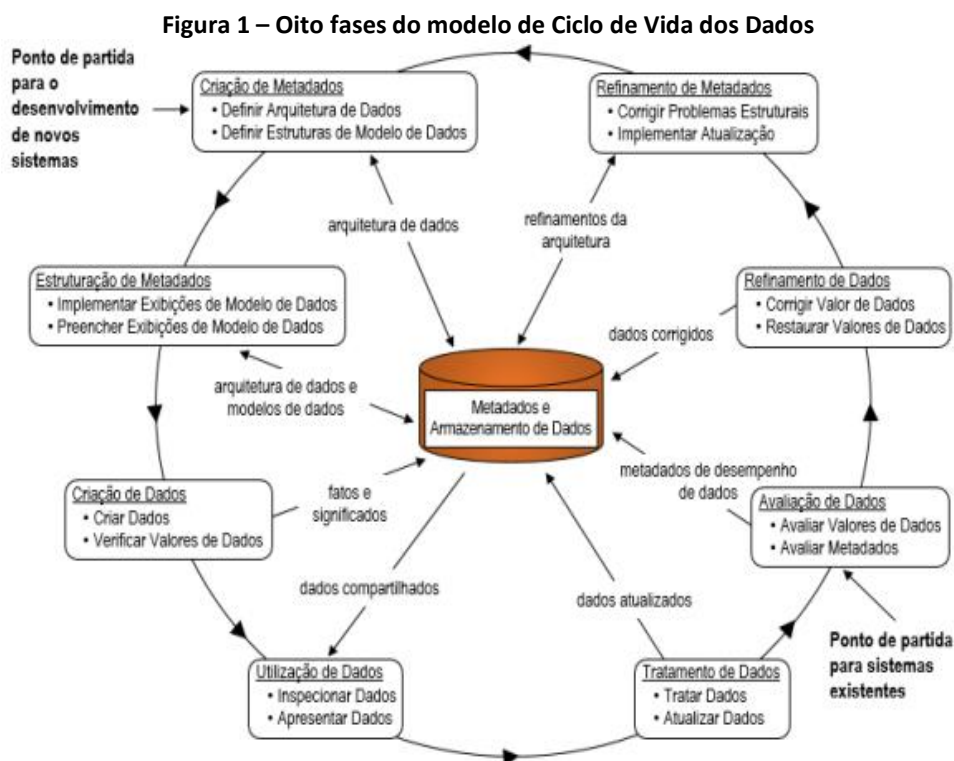
Tem-se assim, como objetivo deste estudo apresentar o mapeamento dos dados da avaliação da pós-graduação *stricto sensu* brasileira no contexto da CAPES, visando embasar a Arquitetura de Dados a ser desenvolvida como alicerce à integração de dados.

1.1 Arquitetura de Dados

Neste estudo, o arcabouço conceitual utilizado provém da Arquitetura de Dados no contexto da Ciência da Informação, conceituada por Torino (2023, p. 192) como:

A Arquitetura de Dados atua no projeto da infraestrutura de dados e do ciclo de vida dos dados, de forma holística e coesa, considerando que os dados necessitam de padrões de metadados, estruturas sintática e semântica e atendam a princípios éticos, legais e técnicos para que sejam consumidos e fornecidos para diferentes finalidades utilizando técnicas e tecnologias emergentes.

Destaca-se que a Arquitetura de Dados pode ser desenvolvida quando do planejamento de um sistema de informação ou considerando a estrutura de dados existente em um sistema já implantado. No âmbito da Arquitetura de Dados, destaca-se o ciclo de vida dos dados que, de acordo com o modelo proposto por Yoon, Aiken e Guimarães (2005), é constituído por oito fases (Figura 1) que podem ser iniciadas pelo "Ponto de partida para o desenvolvimento de novos sistemas" ou pelo "Ponto de partida para sistemas existentes".



Fonte: Yoon, Aiken e Guimarães (2005, p. 147) traduzido por Torino (2023, p. 198).

Neste estudo, considerando a existência de uma infraestrutura de sistemas de informação e uma estrutura de dados relativos ao processo de avaliação da pós-graduação *stricto sensu* brasileira no contexto da CAPES, adotou-se o "Ponto de partida para sistemas existentes". Neste cenário, o processo foi iniciado pela avaliação dos dados e, considerando a Arquitetura de Dados, a avaliação se aplicou à estrutura de dados composta por metadados, sobretudo aos atributos.

Os atributos são relevantes, uma vez que a Arquitetura de Dados projeta os dados para que sejam facilmente consumidos por sistemas de informação, desde que atendam aos requisitos estabelecidos (Torino, 2023). Assim, dados normalizados aumentam a possibilidade de compartilhamento, em especial quando orientados pelo uso e não pela aplicação (Tupper, 2011).

Essa abordagem reforça que a necessidade de integração de sistemas tem sido uma problemática que culminou na alteração do desenvolvimento de sistemas, cuja abordagem de sistemas isolados passa a dar lugar ao suporte multifuncional, que requer tratamento de dados para assegurar sua qualidade, integração e interoperabilidade (Yoon; Aiken; Guimarães, 2005). Destaca-se que, para que as integrações sejam satisfatórias, é necessário que a compreensão dos dados seja clara tanto para humanos quanto para aplicações computacionais.

Neste sentido, Torino (2023) afirma que a Arquitetura de Dados pode contribuir para a melhoria da sintaxe e da semântica dos dados, necessárias para a compreensão do sentido dos dados, para otimizar a comunicação dos dados em fluxos de integração, interoperabilidade, processamento, uso e reuso.

Desta forma, considerando que o processo de avaliação da pós-graduação *stricto sensu* no contexto da CAPES já possui uma estrutura de dados e sistemas existentes, tornou-se necessário mapear a estrutura de dados disponível, visando avaliá-la para que os refinamentos e correções necessários possam ser considerados na proposição de uma Arquitetura de Dados.

2 METODOLOGIA E MAPEAMENTO DOS DADOS

O presente estudo caracteriza-se como uma pesquisa empírica, que investiga a estrutura de dados da avaliação da pós-graduação *stricto sensu* brasileira, com foco na CAPES.

A pesquisa adota uma abordagem exploratória, pautada na análise da estrutura de dados utilizada na avaliação da pós-graduação, alicerçada em referenciais teóricos e práticos das áreas de Ciência da Informação, Sistemas de Informação e Ciência da Computação, bem como de fontes documentais. A análise foi conduzida a partir da coleta, tratamento e consolidação dos dados da CAPES, com o objetivo de identificar padrões, redundâncias e oportunidades para aprimoramento da integração e interoperabilidade dos dados.

Para realizar o processo do mapeamento dos dados de avaliação da pós-graduação, identificou-se, no contexto da CAPES, documentos e bancos de dados disponíveis que são capazes de apoiar a compreensão dos atributos e dos metadados que são parte do cenário tratado no âmbito desta pesquisa.

Inicialmente, para a realização do mapeamento dos dados de avaliação da pós-graduação brasileira, buscou-se compreender o estado atual e os avanços que a CAPES realizou nos últimos anos no que tange à integração de dados com as Instituições de Ensino Superior.

Neste cenário, em 2023, a CAPES lançou o Programa de Governança Colaborativa de Informações da Pós-Graduação (GoPG) para promover a interoperabilidade entre sistemas acadêmicos, científicos, tecnológicos e de inovação, além de fornecer ferramentas de gestão da informação e estabelecer padrões e serviços com controle e proteção de dados (Brasil, 2024). Integrada ao GoPG está a Rede de Integração da Comunidade Acadêmico-Científica | Pós-Graduação (RICA|PG), que tem como missão a construção de conexões para compartilhamento e reuso de dados da pós-graduação, sendo responsável pela definição de padrões para a comunicação de dados no processo de avaliação da pós-graduação *stricto sensu* no Brasil.

Considerando a missão da RICA|PG e a necessidade de desenvolvimento de uma Arquitetura de Dados para a avaliação da pós-graduação *stricto sensu* no contexto da

CAPES, iniciou-se o mapeamento dessa estrutura de dados, visando embasar a Arquitetura de Dados a ser desenvolvida como alicerce à integração de dados.

A partir da compreensão deste cenário, iniciou-se o processo de mapeamento, cuja coleta dos dados ocorreu entre outubro de 2023 e janeiro de 2024, utilizando-se fontes documentais oficiais disponibilizadas pela CAPES. O mapeamento considerou as seguintes fontes de dados: Inventário de Dados da Plataforma Sucupira, Proposta de Entidades derivadas do "Relatório da Rede RICA|PG", Dados Abertos da CAPES, Dicionário de Dados da Plataforma Sucupira, descritas a seguir.

Inventário de Dados da Plataforma Sucupira: elaborado com base na interface de uso da Plataforma Sucupira, fornece uma visão detalhada e abrangente dos conjuntos de dados, metadados e variáveis utilizados na Plataforma Sucupira. Esse inventário contém informações específicas sobre os dados coletados e armazenados pela CAPES. Desta fonte foram extraídos 509 atributos e incluídos na análise 438 atributos, que fornecem uma base para compreender a estrutura dos dados e as informações que são gerenciadas pela Plataforma Sucupira.

Proposta de Entidades derivadas do "Relatório da Rede RICA|PG":, representam uma fonte importante, pois refletem o trabalho desenvolvido pelas instituições que compõem a RICA|PG com o intuito de definir entidades e variáveis relevantes para a avaliação da pós-graduação *stricto sensu* brasileira, sob a ótica das Instituições que possuem programas de pós-graduação. Desta fonte, foram extraídos 246 atributos e incluídos na análise 153 atributos, que representam as diferentes dimensões e aspectos definidos pelas instituições partícipes da RICA|PG.

Dados Abertos da CAPES: fornecem uma perspectiva dos dados da avaliação da pós-graduação *stricto sensu* brasileira que são de interesse da comunidade, uma vez que são exibidos publicamente no Portal Dados Abertos CAPES. Desta fonte, foram coletados 1.443 atributos, dos quais foram incluídos na análise 330 atributos únicos.

Dicionário de Dados da Plataforma Sucupira: consiste nos esquemas e atributos presentes no banco de dados da Plataforma Sucupira, de forma que representa uma fonte fundamental que proporciona uma visão detalhada e estruturada dos atributos e variáveis presentes no referido banco. Desta fonte foram extraídos 8.411 atributos que

oferecem uma compreensão abrangente da estrutura e do conteúdo do banco de dados da Plataforma Sucupira. Para a análise foram incluídos 854 atributos únicos, visando contribuir para uma visualização detalhada e precisa.

No total, foram mapeados mais de 10 mil atributos das quatro fontes citadas, que, após processos de normalização e agrupamento, resultaram em 1.775 atributos únicos representativos do processo de avaliação da pós-graduação *stricto sensu*, necessários para compreender a estrutura de dados utilizada para a avaliação da pós-graduação brasileira no contexto da CAPES.

Destaca-se que a totalidade dos atributos disponíveis em cada fonte de dados foi extraída, no entanto, foram realizadas análises prévias para identificar aqueles que são utilizados no processo investigado neste trabalho. Além disso, houve um tratamento para que os atributos que estão presentes em mais de uma entidade fossem agrupados, reduzindo o quantitativo, por serem considerados o mesmo atributo, por isso, designados atributos únicos.

Tendo em vista as especificidades de cada fonte de dados, realizou-se um processo de extração para que as estruturas pudessem ser compreendidas e registradas de forma padronizada. Para tanto, foi desenvolvido um "Instrumento de Análise de Dados da Avaliação da Pós-Graduação Brasileira", utilizado para identificar os atributos únicos relacionados ao processo em análise. Tal instrumento foi construído utilizando como ferramenta o Google Planilhas. Aponta-se ainda que as terminologias adotadas em cada uma das fontes de dados analisadas foram mantidas, visando elucidar questões semânticas que estão relacionadas ao processo de avaliação da pós-graduação pela CAPES.

3 RESULTADOS

O "Instrumento de Análise de Dados da Avaliação da Pós-Graduação Brasileira" foi desenvolvido para registrar os elementos presentes em cada uma das fontes de dados analisadas. Para tanto, foram criadas categorias correspondentes às entidades do Modelo Conceitual da Plataforma Sucupira, considerado elemento central para a compreensão da estrutura de dados da avaliação da pós-graduação *stricto sensu* brasileira, composto por

Programa, Área de concentração, Linha de pesquisa, Projeto, Produção intelectual, Tese, Docente, Discente, Participante externo, Disciplina e Curso. A partir delas, as entidades, conjuntos de informações, metadados, variáveis e outros elementos relevantes, provenientes das quatro fontes de dados foram categorizadas conforme o tema das categorias e registradas em um único documento em um processo de compatibilização, considerando a correspondência dos atributos entre as fontes.

A Figura 2 apresenta um recorte do instrumento, que demonstra como as 4 fontes foram alinhadas e, que durante o processo, realizou-se o processo da identificação dos elementos correspondentes, para possibilitar uma visão mais clara de quais são os atributos disponíveis, além de como estes se relacionam no contexto das diversas fontes de dados.

Figura 2 – Recorte do instrumento de mapeamento de estruturas de dados

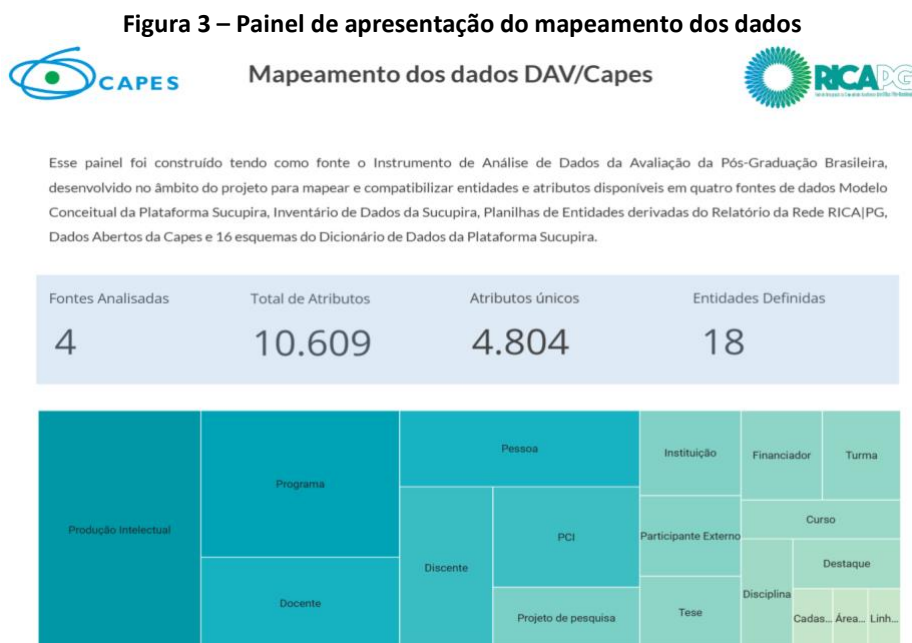
Entidade	Inventário	RICA-PG	Dados Abertos	Dicionário de Dados
9. Programa	Área básica		NM_AREA_BASICA	NM_AREA_BASICA
10. Produção Intelectual	ÁREA REPRESENTADA		DS_AREA_REPRESENTADA	
11. Área de concentração	Á(s) Áreas de concentração Ob...	NM_AREA_CONCENTRACAO	NM_AREA_CONCENTRACAO	NM_AREA_CONCENTRACAO
12. Disciplina	Á(s) Áreas de concentração Ob...	NM_AREA_CONCENTRACAO	NM_AREA_CONCENTRACAO	NM_AREA_CONCENTRACAO
13. Linha de pesquisa	Vínculo com Área de Concentra...		NM_AREA_CONCENTRACAO	NM_AREA_CONCENTRACAO

Fonte: Elaboração própria.

A inclusão das fontes, levou ao crescimento vertical do Instrumento, ao mesmo tempo, à medida que novos atributos e variáveis dessas fontes foram sendo acrescidos houve um crescimento horizontal, possibilitando uma visão macro do *status* dos dados. Tal processo foi inicialmente testado e validado para que fosse aplicado ao conjunto de atributos que compuseram a análise.

A partir dos dados estruturados no Instrumento de Análise de Dados da Avaliação da Pós-Graduação Brasileira, foi desenvolvido um ambiente de análise de dados utilizando a ferramenta de *Business Intelligence (BI) Looker Studio*, da *Google*, com o objetivo de apresentar os resultados alcançados. O ambiente foi designado como "Mapeamento dos Dados DAV/CAPES", que objetiva atuar como fonte de visualização dos dados das fontes (Inventário de Dados, RICA|PG, Dados Abertos, Dicionário de Dados) em suas estruturas originais e do mapeamento de dados das mesmas estruturas, realizado por meio do Instrumento de Análise de Dados da Avaliação da Pós-Graduação *stricto sensu* brasileira.

A Figura 3 apresenta o painel inicial do ambiente de BI desenvolvido, em que é possível identificar as diversas fontes analisadas, além de elementos como o quantitativo de atributos, e a sua divisão por meio das entidades disponíveis no contexto da avaliação.



Fonte: Elaboração própria.

Destaca-se que o ambiente de BI integra e apresenta cinco estruturas de dados, sendo as quatro fontes primárias e o Instrumento de Análise de Dados que as compatibiliza, totalizando 4.804 atributos únicos, exibidos com diferentes configurações em 20 páginas de painéis de visualização, para o apoio na análise quantitativa e qualitativa dos dados. Tais painéis foram utilizados para realizar análises da estrutura de dados mapeada, bem como possui potencial uso estratégico para visualizações e outros mapeamentos relacionados à Avaliação da pós-graduação brasileira.

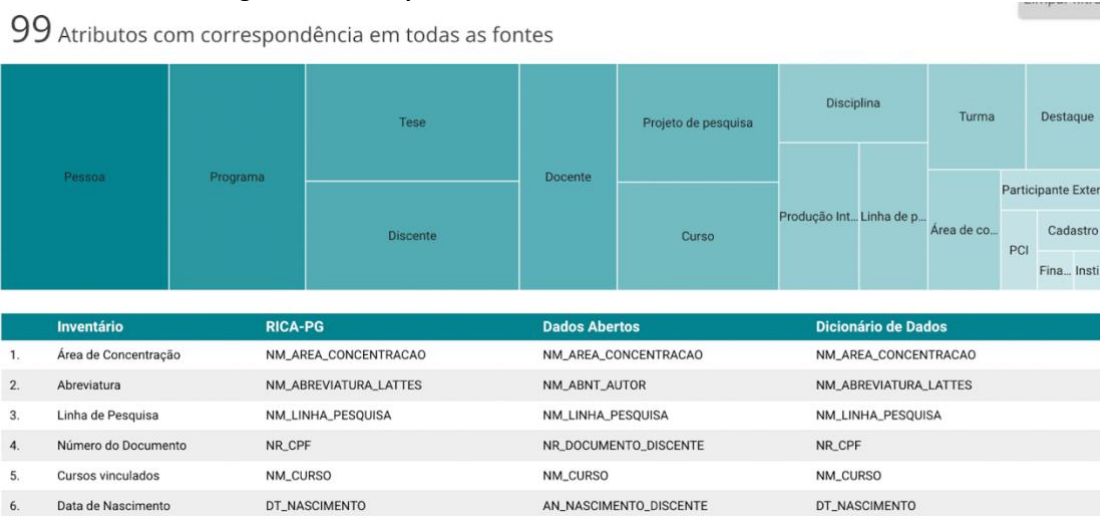
A partir do processo de mapeamento e da construção do instrumento e do ambiente de BI, verificou-se que, dos 4.804 atributos únicos, apenas 1.615 foram compatibilizados, ou seja, possuíam correspondência entre uma ou mais fontes. Dentre os atributos compatibilizados, a maioria refere-se à entidade Produção intelectual, seguida da entidade Programa de Pós-graduação.

Observou-se ainda que somente 99 atributos possuem correspondência em todas as estruturas analisadas, ou seja, dentre as estruturas das fontes de dados atuais da Capes, apenas 2% encontram correspondência completa. Esse percentual ressalta a

necessidade de integração entre as estruturas de dados. A Figura 4 apresenta esses atributos nas entidades mapeadas e alguns dos registros de atributos correspondentes nas 4 fontes analisadas.

Destaca-se que a compreensão e mapeamento dos atributos referem-se às estruturas de dados utilizadas para que os dados sejam armazenados nos sistemas de informação e tenham potencial de uso e reuso automatizados por processos de integração e interoperabilidade.

Figura 4 – Correspondência de atributos entre todas as fontes



Fonte: Elaboração própria.

Tal mapeamento é significativo, pois permitiu compreender a estrutura de dados utilizada para a representação de diferentes entidades e classes presentes na avaliação da pós-graduação *stricto sensu* brasileira no contexto da CAPES.

A partir dos dados mapeados, o estudo terá como continuidade o desenvolvimento da proposta de Arquitetura de Dados da Avaliação da Pós-Graduação Brasileira no contexto da CAPES, cuja aplicação poderá ser realizada nos sistemas de informação disponíveis, sobretudo na Plataforma Sucupira, além de servir de padrão para a comunicação de dados entre as Instituições e a CAPES, possibilitando que a integração de dados culmine em processos automatizados de alimentação de sistemas como a própria Plataforma Sucupira.

4 CONSIDERAÇÕES FINAIS

O presente estudo objetivou apresentar o mapeamento das estruturas de dados da avaliação da pós-graduação *stricto sensu* brasileira no contexto da CAPES, visando embasar a Arquitetura de Dados a ser desenvolvida como alicerce à integração de dados.

Esse processo possibilitará desenvolver uma Arquitetura de Dados da Avaliação da Pós-Graduação *stricto sensu* Brasileira no contexto da CAPES, por meio da qual as Instituições que ofertam Programas de Pós-Graduação poderão interagir, quer seja no processo de fornecimento ou de consumo dos dados de forma automatizada. Com isso, o uso e reuso dos dados poderá ser realizado com menor ruído de comunicação, gerando precisão, veracidade e velocidade ao processo.

A integração de dados de diferentes instituições e áreas do conhecimento pode se constituir em uma poderosa ferramenta de gestão, que possibilitará gerar análises, insights, apoiar a tomada de decisão, de forma eficiente e transparente, além de desenvolver pesquisas. Ademais, aponta-se o aspecto da transparência nos processos que envolvem a relação entre os Programas de Pós-Graduação e a CAPES sobretudo no que tange à avaliação.

Destaca-se ainda, que este trabalho contribui para o aprimoramento da integração entre os sistemas da CAPES e das instituições de Ensino Superior, e poderá apoiar a disseminação de boas práticas em integração, interoperabilidade e representação de dados. Este trabalho visa ainda otimizar os processos de avaliação da pós-graduação, além de contribuir para uma gestão mais eficiente e transparente dos dados da pós-graduação. Isto inclui os dados da produção científica, tendo as próprias instituições como fonte de proveniência dos dados, assegurando transparência, confiabilidade e eficácia ao processo de avaliação.

Portanto, o desenvolvimento do estudo contribui com o aprimoramento da gestão de dados da avaliação da pós-graduação *stricto sensu* brasileira pela CAPES, com potencial para impactar positivamente o processo de avaliação da pós-graduação *stricto sensu* brasileira.

REFERÊNCIAS

BRASIL. Ministério da Educação. **Programa de governança colaborativa de informações da pós-graduação (GoPG)**. [S. l.]: CAPES, [2024]. Disponível em:

<https://www.gov.br/capes/pt-br/aceso-a-informacao/acoes-e-programas/avaliacao/programa-de-governanca-colaborativa-de-informacoes-da-pos-graduacao-gopp>. Acesso em: 23 abr. 2024.

NÓBREGA, H. G. M.; TORINO, E.; VIDOTTI, S. A. B. G.; HALICKI, J. P. C.; OLIVEIRA, T. M. Ferramentas de apoio à integração de dados entre sistemas de informação no contexto da avaliação da pós-graduação stricto sensu pela CAPES. **Tendências da Pesquisa Brasileira em Ciência da Informação**, [S. l.], v. 16, n. 1, p. 1-21, dez. 2023. Disponível em: <https://revistas.ancib.org/index.php/tpbci/article/view/656>. Acesso em: 15 fev. 2024.

TORINO, E. **Arquitetura de dados no contexto da ciência da informação**. 2022. Tese (Doutorado em Ciência da Informação) – Universidade Estadual Paulista, Unesp, Marília, 2023. Disponível em: <https://repositorio.unesp.br/items/b2192b88-8362-488f-9b85-2c173eb66e48>. Acesso em: 15 fev. 2024.

TUPPER, C. D. **Data architecture: from zen to reality**. Amsterdam: Morgan Kaufmann, 2011.

YOON, V. Y.; AIKEN, P.; GUIMARAES, T. Applying a metadata framework to improve data quality. *In*: KHOSROW-POUR, M. (ed.). **Encyclopedia of information science and technology**. Hershey: Idea Group Reference, 2005. v. 1, p. 146-151.



[4.0 Internacional](https://creativecommons.org/licenses/by/4.0/)

Esta licença permite compartilhamento, remixe, adaptação e criação a partir do trabalho, mesmo para fins comerciais, desde que sejam atribuídos créditos ao(s) autor(es). Conteúdos elaborados por terceiros, citados e referenciados nesta obra não são cobertos pela licença.

Como citar este trabalho:

TORINO, Emanuelle; CONEGLIAN, Caio Saraiva; BRANDT, Mariana Baptista; OLIVEIRA, Talita Moreira de; VIDOTTI, Silvana Aparecida Borsetti Gregorio. Mapeamento de dados como arcabouço para a Arquitetura de Dados da Avaliação da Pós-Graduação Brasileira no contexto da CAPES. *In*: WORKSHOP DE INFORMAÇÃO DADOS E TECNOLOGIA, 8., 2025, Marília, SP. **Anais [...]**. Marília, SP: Universidade de Marília, 2025. DOI: <https://doi.org/10.22477/viii.widat.266>.