

Universidade Federal do Triângulo Mineiro
Programa de Mestrado Profissional em Inovações e Tecnologias

Ulisses Fagundes de Sousa

**Plataforma de indicadores acadêmicos utilizando integração de dados por
microserviços**

UBERABA - MG
2026

Ulisses Fagundes de Sousa

**Plataforma de indicadores acadêmicos utilizando integração de dados por
microserviços**

Dissertação apresentada ao Programa de Mestrado Profissional em Inovações e Tecnologias da Universidade Federal do Triângulo Mineiro, na área de concentração: Propriedade Intelectual, como requisito parcial para obtenção do título de Mestre em Inovações e Tecnologias.

Orientador: Prof. Dr. Julio Cesar de Souza Inácio Gonçalves

Coorientadora: Profa. Dra. Ana Paula Fernandes

UBERABA - MG

2026

**Catálogo na fonte: Biblioteca da Universidade Federal do
Triângulo Mineiro**

S698p Sousa, Ulisses Fagundes de
Plataforma de indicadores acadêmicos utilizando integração de dados
por microsserviços / Ulisses Fagundes de Sousa. -- 2026.
127 f. : il., graf., tab.

Dissertação (Mestrado Profissional em Inovações e Tecnologias) --
Universidade Federal do Triângulo Mineiro, Uberaba, MG, 2026
Orientador: Prof. Dr. Júlio Cesar de Souza Inácio Gonçalves
Coorientadora: Profa. Dra. Ana Paula Fernandes

1. Educação. 2. Publicações científicas. 3. Professores universitários.
4. Integração de dados (Computação). 5. Bibliometria. I. Gonçalves, Júlio
Cesar de Souza Inácio. II. Universidade Federal do Triângulo Mineiro.
III. Título.

CDU 37

ULISSES FAGUNDES DE SOUSA**PLATAFORMA DE INDICADORES ACADÊMICOS UTILIZANDO INTEGRAÇÃO DE DADOS POR
MICROSSERVIÇOS**

Dissertação apresentada ao Programa de Pós-graduação Profissional em Inovações e Tecnologias da Universidade Federal do Triângulo Mineiro como requisito parcial para obtenção do título de mestre.

Uberaba, 26 de fevereiro de 2026

Banca Examinadora:

Dr. Julio Cesar de Souza Inácio Gonçalves – Orientador
Universidade Federal do Triângulo Mineiro

Dr. Esdras Viggiano de Souza
Universidade Federal do Triângulo Mineiro

Dra. Fernanda Barbosa Borges Jardim
Instituto Federal do Triângulo Mineiro



Documento assinado eletronicamente por **JULIO CESAR DE SOUZA INACIO GONCALVES, Professor do Magistério Superior**, em 26/02/2026, às 16:34, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#) e no art. 34 da [Portaria Reitoria/UFTM nº 215, de 16 de julho de 2024](#).



Documento assinado eletronicamente por **ESDRAS VIGGIANO DE SOUZA, Professor do Magistério Superior**, em 26/02/2026, às 16:37, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#) e no art. 34 da [Portaria Reitoria/UFTM nº 215, de 16 de julho de 2024](#).



Documento assinado eletronicamente por **FERNANDA BARBOSA BORGES JARDIM, Usuário Externo**, em 05/03/2026, às 11:15, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#) e no art. 34 da [Portaria Reitoria/UFTM nº 215, de 16 de julho de 2024](#).



A autenticidade deste documento pode ser conferida no site http://sei.ufmt.edu.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0, informando o código verificador **1698489** e o código CRC **FFE9ABA7**.

RESUMO

A produção científica institucional é registrada em fontes heterogêneas e dispersas, o que dificulta a consolidação, a análise e a verificabilidade de indicadores. Este trabalho apresenta o desenvolvimento e a aplicação da Plataforma Capivara, uma plataforma computacional voltada à curadoria, integração e exploração interativa da produção científica institucional. A metodologia proposta organiza um fluxo de ingestão de dados institucionais e curriculares, extração e normalização de currículos da Plataforma Lattes, consolidação e deduplicação rastreável das produções, enriquecimento bibliométrico hierárquico com fontes externas e disponibilização de indicadores com navegação do indicador ao registro e exportação consistente. Com extração em 24/01/2026, o estudo analisou 882 currículos Lattes de docentes estatutários da Universidade Federal do Triângulo Mineiro (UFTM), incluindo docentes aposentados, totalizando 169.657 declarações de produções no intervalo 1971–2026. O enriquecimento de identificadores persistentes elevou a cobertura de DOI de 7,17% para 8,58% no escopo global e de 51,83% para 62,11% no núcleo bibliográfico (artigos, livros e capítulos), com apoio de bases externas como a Crossref. A deduplicação apresentou taxa global aproximada de 11,4% no recorte de vínculo institucional, evidenciando redundância declarativa e permitindo análises comparativas por múltiplas facetas (temporal, organizacional, tipológica, colaborativa e bibliométrica). Conclui-se que a plataforma reduz a distância entre indicadores agregados e registros subjacentes, oferecendo base verificável para diagnóstico exploratório e suporte à gestão da informação científica institucional.

Palavras-chave: produção científica; plataforma Lattes; integração de dados; indicadores institucionais; enriquecimento bibliométrico; rastreabilidade.

ABSTRACT

Institutional scientific output is recorded in heterogeneous and dispersed sources, which makes the consolidation, analysis, and verifiability of indicators more difficult. This study presents the development and application of the Capivara Platform, a computational platform designed for the curation, integration, and interactive exploration of institutional scientific output. The proposed methodology organizes a workflow for the ingestion of institutional and curricular data, extraction and normalization of records from the Lattes Platform, traceable consolidation and deduplication of outputs, hierarchical bibliometric enrichment using external sources, and the provision of indicators with drill-down navigation from indicator to record and consistent export. Based on data extracted on January 24, 2026, the study analyzed 882 Lattes CVs of statutory faculty members at the Federal University of Triângulo Mineiro (UFTM), including retired faculty, totaling 169,657 production records in the 1971–2026 period. Persistent identifier enrichment increased DOI coverage from 7.17% to 8.58% in the overall scope and from 51.83% to 62.11% in the bibliographic core (articles, books, and book chapters), supported by external sources such as Crossref. Deduplication reached an approximate overall rate of 11.4% within the institutional affiliation scope, evidencing declarative redundancy and enabling comparative analyses across multiple facets (temporal, organizational, typological, collaborative, and bibliometric). It is concluded that the platform reduces the distance between aggregated indicators and their underlying records, providing a verifiable basis for exploratory diagnosis and support for the management of institutional scientific information.

Keywords: scientific output; Lattes platform; data integration; institutional indicators; bibliometric enrichment; traceability.

LISTA DE FIGURAS

Figura 1	Série temporal de artigos publicados no ICS (2016–2025).	64
Figura 2	Docentes produtores e artigos por docente no ICS (2016–2025). . . .	65
Figura 3	Série anual de artigos por departamento no ICS.	66
Figura 4	Série temporal do agregado “eventos acadêmicos” (2016–2025).	68
Figura 5	Série temporal de artigos publicados por instituto (2022–2024).	72
Figura 6	Série temporal de livros+capítulos por instituto (2022–2024).	73
Figura 7	Composição por categoria de interesse por instituto (2022–2024). . . .	77
Figura 8	Produção técnica por instituto: composição por tipo (2022–2024). . .	77
Figura 9	Tipos raros de produção técnica concentrados: exemplos (2022–2024). .	78
Figura 10	Orientações por instituto e natureza (2022–2024).	79
Figura 11	Quantidade de trabalhos no Top 1% por classe de colaboração (2021–2025). .	81
Figura 12	Total de citações recebidas por ano e por fonte (OpenAlex, Open- Citations e Scholar, 2021–2025).	82
Figura 13	Qualis × total de citações (OpenAlex) por ano e estrato (2019–2025; Qualis 2021–2024).	83
Figura 14	Header com os controles de fonte/modo noturno e links de navegação. . .	87
Figura 15	Footer com links institucionais.	88
Figura 16	Painel de cards de indicadores agregados na Home.	88
Figura 17	Um indicador na Home mostrando: título/subtítulo, barra de ações e filtros visíveis.	89
Figura 18	Popover do Download com opções (JSON/CSV/XLSX/imagem; ex- press/custom).	90
Figura 19	Modal do "i"(definição + fontes + observações).	91
Figura 20	Indicador expandido (tela cheia).	91
Figura 21	Tela de Produções com filtros e alternância declarada/deduplicada. . .	93
Figura 22	Tela de Produções: resultado paginado.	93
Figura 23	Tela Instituto/ID: cabeçalho, cartões de indicadores e início da ta- bela de departamentos.	94
Figura 24	Tela Instituto/ID: drawer de seleção de indicadores e continuação da tabela.	94
Figura 25	Cabeçalho do professor + abas.	95
Figura 26	Aba Impacto com citações por fonte e evolução temporal.	95
Figura 27	Sequência de drill-through: gráfico com tooltip e Tela de Produções resultante.	96

LISTA DE TABELAS

Tabela 1	Cobertura de DOI antes e após enriquecimento	44
Tabela 2	Evolução temporal da produção deduplicada no vínculo (1971–2025) .	47
Tabela 3	Principais tipos de produção após deduplicação — dez categorias mais frequentemente no período de vínculo	48
Tabela 4	Distribuição por categorias de interesse (agrupamento semântico de tipos) — deduplicada no vínculo	49
Tabela 5	Volume absoluto e relativo por instituto (participações em produções deduplicadas)	50
Tabela 6	Comparação de rankings por instituto (total <i>versus</i> ALC)	50
Tabela 7	Participações em produções deduplicadas por programa de pós-graduação (período de vínculo)	52
Tabela 8	Completeness de DOI na produção institucional (vínculo, deduplicada)	53
Tabela 9	Grupos consolidados com múltiplos declarantes (coautoria interna) .	54
Tabela 10	Produção classificada como “não identificado” (lacunas organizaci- onais)	55
Tabela 11	Síntese do papel das bases externas no enriquecimento bibliométrico	58
Tabela 12	Cobertura de integração no núcleo ALC (vínculo) por base externa .	60
Tabela 13	Convergência de citações entre fontes bibliométricas (agregado) . . .	61
Tabela 14	Distribuição global de tipos bibliográficos na UFTM (2022–2024). .	71
Tabela 15	Proporção de livros+capítulos no ALC por instituto (2022–2024). .	71
Tabela 16	Decomposição do ALC no IELACHS por departamento (2022–2024; apenas depts com $ALC \geq 10$).	74
Tabela 17	Engajamento discente em artigos publicados por instituto (2016–2024).	75
Tabela 18	Engajamento discente em livros+capítulos por instituto (2016–2024).	76
Tabela 19	Engajamento discente por tipo bibliográfico (2016–2024).	76
Tabela 20	Coautoria (OpenAlex) e frequência de alto impacto normalizado (Top 10% e Top 1%). <i>Percentuais calculados sobre o subconjunto com valor normalizado disponível.</i>	80
Tabela 21	Mapa de facetas analíticas operacionalizadas no artefato (recortes e onde são aplicados).	92
Tabela 22	Cronograma executado por marcos e entregáveis	108
Tabela 23	Mapeamento de tipos do Lattes para categoria de interesse	122

LISTA DE ABREVIACÕES, SIGLAS E TERMOS

Termo	Significado
ALC	Artigos publicados, Livros e Capítulos (núcleo bibliográfico)
API	Application Programming Interface (interface de programação de aplicação)
AWS	Amazon Web Services
CAFe	Comunidade Acadêmica Federada
CAPES	Coordenação de Aperfeiçoamento de Pessoal de Nível Superior
CEFORES	Centro de Educação Profissional
CEP	Comitê de Ética em Pesquisa
CNPq	Conselho Nacional de Desenvolvimento Científico e Tecnológico
CRIS	Current Research Information Systems (sistema de informação de pesquisa)
Crossref	Agência de registro de DOI e infraestrutura de metadados acadêmicos
CSV	Comma-Separated Values (formato de valores separados por vírgula)
DBIO	Departamento de Biomedicina
DCC	Departamento de Clínica Cirúrgica
DCES	Departamento de Ciências do Esporte
DCM	Departamento de Clínica Médica
DE	Departamento de Educação
DEAH	Departamento de Enfermagem na Assistência Hospitalar
DEESC	Departamento de Enfermagem em Educação e Saúde Comunitária
DEGEO	Departamento de Geografia
DEL	Departamento de Estudos Literários
DESCO	Departamento de Saúde Coletiva
DFAP	Departamento de Fisioterapia Aplicada
DFICS	Departamento de Filosofia e Ciências Sociais
DGO	Departamento de Ginecologia e Obstetrícia
DHIS	Departamento de História
DLE	Departamento de Línguas Estrangeiras
DLLP	Departamento de Linguística e Língua Portuguesa
DNUTRI	Departamento de Nutrição
Docker	Plataforma de containerização de aplicações
DOI	Digital Object Identifier (identificador digital de objeto)
DPED	Departamento de Pediatria

DPSICO	Departamento de Psicologia
DRCA	Divisão de Registro e Controle Acadêmico
DSR	Design Science Research (pesquisa em ciência do design)
DSS	Departamento de Serviço Social
DTO	Departamento de Terapia Ocupacional
Elasticsearch	Mecanismo distribuído de busca e análise de dados
ELT	Extract, Load, Transform (extrair, carregar, transformar)
ETL	Extract, Transform, Load (extrair, transformar, carregar)
FAIR	Findable, Accessible, Interoperable, Reusable (localizável, acessível, interoperável, reutilizável)
GCP	Google Cloud Platform
Google Scholar	Motor de busca acadêmica do Google
HTTP	HyperText Transfer Protocol (protocolo de transferência de hipertexto)
IC	Iniciação Científica
ICAEBI	Instituto de Ciências Agrárias, Exatas e Biológicas de Iturama
ICBN	Instituto de Ciências Biológicas e Naturais
ICENE	Instituto de Ciências Exatas, Naturais e Educação
ICS	Instituto de Ciências da Saúde
ICTE	Instituto de Ciências Tecnológicas e Exatas
IELACHS	Instituto de Educação, Letras, Artes, Ciências Humanas e Sociais
IES	Instituições de Ensino Superior
ISBN	International Standard Book Number (número padrão internacional de livro)
ISSN	International Standard Serial Number (número padrão internacional de periódico)
JSON	JavaScript Object Notation (notação de objetos JavaScript)
JSONB	Formato binário para armazenamento de JSON no PostgreSQL
KPI	Key Performance Indicator (indicador-chave de desempenho)
Lattes	Plataforma de currículos acadêmicos mantida pelo CNPq
Next.js	Framework React para aplicações web
Node.js	Ambiente de execução JavaScript no servidor
OpenAlex	Base aberta de metadados acadêmicos e bibliométricos
OpenCitations	Infraestrutura aberta de dados de citações
ORCID	Open Researcher and Contributor ID (identificador aberto de pesquisadores)
PDF	Portable Document Format (formato de documento portátil)
PDI	Plano de Desenvolvimento Institucional
PID	Persistent Identifier (identificador persistente de recurso digital)

PLN	Processamento de Linguagem Natural
PMPIT	Programa de Mestrado Profissional em Inovações e Tecnologias
PostgreSQL	Sistema de gerenciamento de banco de dados relacional
PPGAS	Programa de Pós-Graduação em Atenção à Saúde
PPGCF	Programa de Pós-Graduação em Ciências Fisiológicas
PPGCS	Programa de Pós-Graduação em Ciências da Saúde
PPGCTA	Programa de Pós-Graduação em Ciência e Tecnologia Ambiental
PPGCTM	Programa de Pós-Graduação em Ciência e Tecnologia de Materiais
PPGE	Programa de Pós-Graduação em Educação
PPGECM	Programa de Pós-Graduação em Educação em Ciências e Matemática
PPGEF	Programa de Pós-Graduação em Educação Física
PPGFISIO	Programa de Pós-Graduação em Fisioterapia
PPGMQ-MG	Programa de Pós-Graduação Multicêntrico em Química de Minas Gerais
PPGMTI	Programa de Pós-Graduação em Medicina Tropical e Infectologia
PPGP	Programa de Pós-Graduação em Psicologia
Profiap	Mestrado Profissional em Administração Pública em Rede Nacional
Profletras	Mestrado Profissional em Letras em Rede Nacional
Profmat	Mestrado Profissional em Matemática em Rede Nacional
Profqui	Mestrado Profissional em Química em Rede Nacional
PROGEPE	Pró-Reitoria de Gestão de Pessoas
PROPPG	Pró-Reitoria de Pesquisa e Pós-Graduação
Qualis	Sistema de classificação de periódicos da CAPES
React	Biblioteca JavaScript para construção de interfaces de usuário baseadas em componentes
Redis	Sistema de cache e armazenamento de dados em memória
REST	Representational State Transfer (transferência de estado representacional)
Scopus	Base bibliométrica e de citações mantida pela Elsevier
SISCAD	Sistema institucional de gestão acadêmica
SQL	Structured Query Language (linguagem de consulta estruturada)
Sucupira	Plataforma da CAPES para coleta, análise e avaliação da pós-graduação
TCC	Trabalho de Conclusão de Curso
TypeScript	Linguagem de programação com tipagem estática, baseada em JavaScript
UFTM	Universidade Federal do Triângulo Mineiro
URL	Uniform Resource Locator (localizador uniforme de recursos)

XML

eXtensible Markup Language (linguagem de marcação extensível)

SUMÁRIO

1	Introdução	19
1.1	Justificativa	20
1.1.1	Público-alvo e volume.	21
1.2	Contribuições esperadas e impacto	21
1.3	Objetivos	22
1.3.1	Objetivo geral	22
1.3.2	Objetivos específicos	22
1.4	Metodologia e procedimentos técnicos	23
1.5	Estrutura do trabalho	23
2	Referencial Teórico	25
2.1	Curadoria e Gestão da Informação Científica em Instituições de Ensino Superior	25
2.1.1	Informação científica como ativo institucional estratégico . .	25
2.1.2	Fragmentação de dados e desafios de integração	25
2.1.3	Sistemas de Informação de Pesquisa Corrente (CRIS) como resposta institucional	25
2.1.4	A Plataforma Lattes no contexto brasileiro	26
2.1.5	Desafios técnicos e limitações do Lattes	26
2.2	Qualidade de Dados e Rastreabilidade em Sistemas de Informação Científica	28
2.2.1	Processos de garantia de qualidade em repositórios científicos	28
2.2.2	Rastreabilidade e proveniência de fluxos de trabalho analíticos	28
2.2.3	Prestígio institucional e confiabilidade de dados	28
2.2.4	Múltiplas dimensões da qualidade de dados	29
2.3	Bibliometria e Indicadores Científicos: Usos e Limites na Avaliação Institucional	30
2.3.1	Fundamentos: sensibilidade às bases e limites conceituais .	30
2.3.2	Contexto brasileiro: cobertura, Sucupira e efeitos institucionais	31
2.4	Interoperabilidade e Infraestrutura Aberta para Avaliação Científica	32
2.4.1	Princípios FAIR aplicados à informação de pesquisa	32
2.4.2	Integração de bases heterogêneas: abordagens técnicas . . .	32

2.4.3	Escolha de bases abertas: cobertura e completude de metadados	32
2.4.4	Grafos de conhecimento, citações abertas e avaliação reprodutível	32
3	Metodologia	34
3.1	Natureza da pesquisa e abordagem metodológica	34
3.2	Problema de pesquisa e solução proposta	34
3.3	Visão geral do método	34
3.4	Fontes de dados e critérios de uso	35
3.5	Conjunto de registros, elegibilidade e recorte analítico	36
3.5.1	Universo institucional de referência	36
3.5.1.1	<i>Casos sem vinculação departamental e estratégia de complementação.</i>	37
3.5.2	Elegibilidade dos currículos Lattes	38
3.5.2.1	<i>Critérios operacionais (resumo).</i>	38
3.5.3	Recorte analítico principal e implicações	38
3.5.3.1	<i>Definição operacional do início e término do vínculo.</i>	39
3.6	Tratamento, integração e consolidação das produções	39
3.7	Enriquecimento bibliométrico	40
3.8	Consolidação analítica e geração de indicadores	40
3.9	Construção dos indicadores	40
3.10	Estratégias de exploração e visualização	41
3.11	Avaliação do artefato	41
3.12	Aspectos éticos	42
4	Resultados e Discussão	43
4.1	Síntese descritiva do corpus consolidado	43
4.1.1	Universo e cobertura (síntese)	43
4.1.2	A base de dados Lattes analisada	43
4.2	Consolidação e deduplicação das produções	44
4.2.1	Unidade de análise: declaração versus produção consolidada	44
4.2.2	Identificadores persistentes e cobertura de DOI	44
4.2.3	Divergência de ano entre declarações vinculadas por DOI .	44
4.2.4	Coautoria interna: grupos com múltiplos declarantes	45
4.2.5	Síntese	45
4.2.5.1	<i>Rastreabilidade (cadeia de confiança).</i>	45
4.3	Perfil institucional da produção no período de vínculo	46
4.3.1	Auditoria de consistência do recorte (controle de universo)	46

4.3.1.1	<i>Discussão: efeito global da deduplicação.</i>	46
4.3.2	Distribuição temporal e marcos interpretativos	46
4.3.3	Composição tipológica da produção	48
4.3.4	Distribuição organizacional	49
4.3.4.1	<i>Comparação de ranking (total versus ALC)</i>	49
4.3.4.2	<i>Coautoria interinstitucional e lacunas de vincu-</i> <i>lação</i>	50
4.3.4.3	<i>Departamentos e programas de pós-graduação</i>	51
4.3.4.4	<i>Síntese do perfil institucional</i>	51
4.4	Qualidade e cobertura dos dados	53
4.4.1	Compleitude de metadados e capacidade de integração	53
4.4.1.1	<i>Discussão: enriquecimento de DOI e viés de co-</i> <i>bertura.</i>	54
4.4.2	Cobertura de coautoria e redundância declarativa	54
4.4.3	Lacunas de identificação institucional e cobertura organiza-	
	cional	54
4.4.4	Atualidade do Lattes e efeitos de consolidação recente	55
4.4.5	Síntese de qualidade e cobertura	56
4.5	Enriquecimento bibliométrico e integração com bases externas	57
4.5.1	Escopo do enriquecimento e recortes de análise	57
4.5.2	Fontes externas e critérios de priorização	57
4.5.3	Identificadores de integração: DOI declarado <i>versus</i> DOI	
	final	58
4.5.3.1	<i>Crossref como mecanismo de fallback para recu-</i> <i>peração de DOI.</i>	59
4.5.4	Cobertura de integração por base externa	59
4.5.5	Convergência de citações entre fontes	60
4.5.6	Ganho informacional: temas e atributos de colaboração	
	(OpenAlex)	62
4.5.7	Classificação nacional: Qualis (periódicos) e escopo de apli-	
	cação	62
4.5.8	Atualização, latência e completude do enriquecimento	62
4.5.9	Limitações e vieses de cobertura	62
4.6	Análises comparativas e padrões emergentes	64
4.6.1	Padrões temporais e organizacionais	64
4.6.1.1	<i>Caso âncora: variação recente de artigos publi-</i> <i>cados no ICS (2016–2025).</i>	64
4.6.1.2	<i>Decomposição da produtividade: taxa de publica-</i> <i>dores e intensidade.</i>	65

4.6.1.3	<i>Síntese do achado (ICS).</i>	66
4.6.1.4	<i>Contraponto institucional: crescimento no ICTE.</i>	67
4.6.1.5	<i>Panorama interinstitucional da variação 2022–2024.</i>	67
4.6.1.6	<i>Nota metodológica: janela trienal.</i>	68
4.6.1.7	<i>Caso 3 (padrão transversal): ruptura em tipos associados a eventos acadêmicos (2016–2024).</i>	68
4.6.1.8	<i>Discussão: impacto da pandemia, persistência parcial e heterogeneidade.</i>	69
4.6.2	<i>Padrões por tipologia e colaboração</i>	70
4.6.2.1	<i>Escopo e princípio de verificabilidade.</i>	70
4.6.2.2	<i>Janela temporal.</i>	70
4.6.2.3	<i>Fundamentação: perfis tipológicos e avaliação.</i>	70
4.6.2.4	<i>Padrão global do triênio e núcleo bibliográfico.</i>	70
4.6.2.5	<i>Caso 1 (excentricidade tipológica): livros e capítulos no IELACHS.</i>	70
4.6.2.6	<i>Contraste temporal (triênio) e interpretação do desvio.</i>	71
4.6.2.7	<i>Decomposição organizacional (drill-down): quais departamentos do IELACHS sustentam o perfil.</i>	73
4.6.2.8	<i>Janela específica para engajamento discente.</i>	73
4.6.2.9	<i>Caso 2 (engajamento discente): padrões de formação e produção.</i>	74
4.6.2.10	<i>Decomposição por tipo bibliográfico: hierarquia de engajamento.</i>	74
4.6.2.11	<i>Evolução temporal do engajamento discente (global).</i>	74
4.6.2.12	<i>Nota de interpretação e robustez.</i>	75
4.6.2.13	<i>Ampliação: excentricidade para além do núcleo bibliográfico.</i>	75
4.6.2.14	<i>Produção técnica: variedade tipológica e concentração de tipos raros.</i>	77
4.6.2.15	<i>Orientações concluídas: evidência tipológica da dimensão formativa.</i>	78
4.6.2.16	<i>Síntese (tipologia e colaboração).</i>	79
4.6.3	<i>Padrões com enriquecimento bibliométrico</i>	80
4.6.3.1	<i>Escopo e recorte dos dados bibliométricos.</i>	80
4.6.3.2	<i>Coautoria internacional e frequência de alto impacto (OpenAlex).</i>	80

4.6.3.3	<i>Discussão do padrão (colaboração e impacto).</i>	81
4.6.3.4	<i>Condições de validade e vieses do recorte.</i>	81
4.6.3.5	<i>Citações recebidas por ano e por fonte (série temporal).</i>	82
4.6.3.6	<i>Qualis e intensidade de citações em artigos recentes.</i>	82
4.6.4	<i>Discussão dos padrões observados</i>	84
4.6.4.1	<i>Complementaridade de lentes analíticas.</i>	84
4.6.4.2	<i>Heterogeneidade institucional como achado central.</i>	84
4.6.4.3	<i>Padrões transversais e choques externos.</i>	84
4.6.4.4	<i>Engajamento discente: evidência de articulação entre produção e formação.</i>	85
4.6.4.5	<i>Colaboração internacional e impacto: evidência de associação robusta.</i>	85
4.6.4.6	<i>Limitações interpretativas e múltiplas leituras possíveis.</i>	85
4.7	<i>Exploração interativa e verificabilidade dos resultados</i>	87
4.7.1	<i>Dimensões de análise disponíveis no artefato</i>	87
4.7.1.1	<i>Arquitetura de navegação por telas.</i>	87
4.7.1.2	<i>Núcleo analítico reutilizável: seção de indicadores.</i>	88
4.7.1.3	<i>Ações e controles padrão dos indicadores.</i>	89
4.7.1.4	<i>Filtros e recortes: mapa de facetas operacionalizadas.</i>	92
4.7.1.5	<i>Tela de Produções como mecanismo de auditoria.</i>	92
4.7.1.6	<i>Telas institucionais e composição por blocos.</i>	93
4.7.1.7	<i>Tela do professor e exploração por abas.</i>	94
4.7.2	<i>Rastreabilidade e validação por consulta</i>	95
4.7.2.1	<i>Princípio: do agregado ao registro, mantendo o recorte.</i>	95
4.7.2.2	<i>Observação sobre filtros passíveis de drill-through.</i>	96
4.7.2.3	<i>Fluxo padrão de verificação (5 passos).</i>	96
4.7.2.4	<i>Exemplo concreto de rastreabilidade (vinculado ao 4.6).</i>	97
4.7.2.5	<i>Modal de “Info” como contrato metodológico.</i>	97
4.7.3	<i>Exemplos de uso e implicações institucionais</i>	97
4.7.3.1	<i>Cenário 1: gestão institucional e dinâmica recente de publicações.</i>	97
4.7.3.2	<i>Cenário 2: coordenação acadêmica e engajamento discente.</i>	97

4.7.3.3	<i>Cenário 3: internacionalização e evidência bibliométrica de alto impacto.</i>	98
4.7.3.4	<i>Síntese: verificabilidade como propriedade de engenharia.</i>	98
4.8	Síntese e limitações	99
4.8.1	Síntese dos principais achados	99
4.8.2	Limitações e ameaças à validade	99
4.8.2.1	<i>(L1) Cobertura e viés de DOI-chave.</i>	99
4.8.2.2	<i>(L2) Cobertura de afiliações e classificação de colaboração.</i>	100
4.8.2.3	<i>(L3) Janela temporal e maturação de citações.</i>	100
4.8.2.4	<i>(L4) Deduplicação e dependência de metadados.</i>	100
4.8.2.5	<i>(L5) Atribuição organizacional e lacunas históricas.</i>	100
4.8.2.6	<i>(L6) Associação não implica causalidade.</i>	100
4.8.3	Implicações para uso e direções de aprimoramento	101
5	Conclusão	102
5.1	Contribuições principais	102
5.1.1	<i>(C1) Metodológica: consolidação rastreável e deduplicação auditável.</i>	102
5.1.2	<i>(C2) Analítica: padrões observáveis e verificáveis em múltiplas dimensões.</i>	102
5.1.3	<i>(C3) Técnica: interoperabilidade e arquitetura modular.</i>	103
5.1.4	<i>(C4) Institucional: diagnóstico exploratório e suporte à decisão informada.</i>	103
5.2	Limitações e contexto de aplicabilidade	103
5.3	Alinhamento com objetivos e pergunta de pesquisa	103
5.4	Considerações finais	104
6	Trabalhos Futuros	105
6.1	Ampliação do escopo de dados e indicadores	105
6.1.1	<i>(TF1) Expansão para domínios além da produção científica registrada no Lattes.</i>	105
6.2	Aprofundamento analítico e exploração interativa	105
6.2.1	<i>(TF2) Rotulagem temática e taxonomias institucionais.</i>	105
6.2.2	<i>(TF3) Módulos avançados de análise com rastreabilidade.</i>	105
6.2.3	<i>(TF4) Painéis longitudinais e análises de sensibilidade.</i>	105
6.3	Transferibilidade	106
6.3.1	<i>(TF5) Adaptação multi-institucional e comparabilidade.</i>	106

6.3.2 Síntese.	106
7 Orçamento	107
8 Cronograma	108
8.1 Cronograma executado (marcos)	108
REFERÊNCIAS	109
APÊNDICE A Construção e Catálogo de Indicadores	113
APÊNDICE B Exploração Analítica e Visualização	116
APÊNDICE C Regras de Integração e Deduplicação das Produções	119
APÊNDICE D Taxonomias e regras de agregação	121
APÊNDICE E Arquitetura da Plataforma Capivara	124
APÊNDICE F Estratégias de Enriquecimento Bibliométrico	127

1 INTRODUÇÃO

O crescimento do volume e da complexidade da informação científica (Fortunato *et al.*, 2018) ampliou a necessidade de infraestruturas institucionais capazes de gerenciar informação científica como ativo estratégico (Conway, 2008). Instituições de Ensino Superior (IES) enfrentam hoje o desafio não apenas de preservar e organizar registros acadêmicos dispersos, mas de transformá-los em evidência estruturada e verificável, capaz de subsidiar processos decisórios institucionais. No entanto, estudos mostram que decisões em contextos universitários são frequentemente tomadas com evidência parcial ou fragmentada (Söderlind; Geschwind, 2019; Thiedig; Wegner, 2024).

Registros científicos permanecem dispersos em sistemas funcionais internos e bases bibliográficas externas (Maltese, 2018; Mena-Chalco *et al.*, 2014), e sistemas de informação de pesquisa (CRIS – *Current Research Information Systems*) surgem como resposta a essa fragmentação (Castro, 2018), demandando processos de integração e governança de dados que vão além da mera agregação técnica. Esse desafio envolve curadoria institucional, qualidade e rastreabilidade dos dados, limites do uso de indicadores e interoperabilidade entre sistemas e bases. A confiabilidade de indicadores não depende apenas da integração de fontes, mas da rastreabilidade das transformações aplicadas aos dados e da documentação das escolhas metodológicas (Mitchell *et al.*, 2022). A questão central não é apenas “ter dados”, mas “saber de onde vieram” e “como foram processados”.

No campo da bibliometria, indicadores científicos, embora consolidados como ferramentas de gestão universitária, apresentam limitações estruturais relacionadas à cobertura das bases, vieses de indexação e efeitos comportamentais de sistemas de avaliação (Bar-Ilan, 2008; Szomszor *et al.*, 2021). A construção responsável de indicadores exige reconhecimento explícito dessas limitações e transparência metodológica (Szomszor *et al.*, 2021). Gestores universitários reconhecem que métricas auxiliam na aquisição de recursos e conferem legitimidade, mas interpretação contextual dos resultados de acordo com as particularidades do ambiente institucional (Söderlind; Geschwind, 2019).

No contexto da interoperabilidade, a adoção dos princípios FAIR (*Findable, Accessible, Interoperable, Reusable* — encontráveis, acessíveis, interoperáveis e reutilizáveis) (Wilkinson *et al.*, 2016), originalmente propostos para dados de pesquisa, estende-se à informação de pesquisa (Hauschke *et al.*, 2021), fundamentando escolhas metodológicas que priorizam transparência e reprodutibilidade.

No contexto brasileiro, essas dimensões adquirem contornos específicos. A Plataforma Lattes constitui infraestrutura nacional de registro científico (Mena-Chalco *et al.*, 2014), e o sistema de avaliação da CAPES estabelece parâmetros para programas de pós-graduação (Capes, 2025). Contudo, a limitada cobertura de bases bibliográficas internacionais para a produção científica brasileira demanda estratégias de integração entre fontes locais e globais. Estudos indicam que a Web of Science indexa cerca de 46% dos

programas de pós-graduação do país e que esse percentual varia significativamente entre áreas, podendo ser muito baixa em alguns casos (Melo; Trinca; Maricato, 2021). Essas limitações motivam a busca por estratégias alternativas de integração, como o uso de bases abertas. Nesse sentido, abordagens que articulam a Plataforma Lattes com bases abertas internacionais, como OpenAlex (Culbert *et al.*, 2025), emergem como alternativa metodológica para ampliar a cobertura e a visibilidade institucional. Essas abordagens alinham-se a princípios de infraestrutura aberta e reprodutibilidade (Mazoni *et al.*, 2025).

A Universidade Federal do Triângulo Mineiro (UFTM) reflete esse cenário nacional. O Plano de Desenvolvimento Institucional (PDI) 2025–2029 estabelece metas de fortalecimento da pesquisa, incluindo ampliação de indexadores, incentivo à publicação qualificada e aumento da visibilidade científica institucional (Universidade Federal do Triângulo Mineiro, 2025). No entanto, registros acadêmicos e científicos encontram-se dispersos entre sistemas internos e bases externas (Lattes, Sucupira, OpenAlex, Scopus), dificultando análises em nível agregado (institutos, departamentos, programas). Levantamentos que hoje dependem de consultas pontuais e planilhas demandam esforço contínuo de consolidação, sem garantir rastreabilidade ou documentação metodológica adequada. Essa dinâmica gera retrabalho recorrente, comprometendo a eficiência administrativa e o uso de recursos institucionais — tanto de pessoal quanto orçamentários. Este trabalho propõe uma aplicação de integração e análise — a plataforma Capivara — que consolida essas fontes e disponibiliza indicadores institucionais com rastreabilidade e documentação metodológica.

Diante desse cenário, esta pesquisa busca responder às seguintes questões:

- a) Como integrar registros institucionais e bibliográficos dispersos em múltiplas fontes (Lattes, Sucupira, bases internacionais e sistemas internos) para apoiar a gestão da pesquisa e da pós-graduação, assegurando rastreabilidade das transformações aplicadas?
- b) De que forma princípios de curadoria, qualidade de dados e interoperabilidade podem fundamentar a construção de indicadores estratégicos úteis à análise institucional, preservando níveis adequados de agregação?
- c) Quais soluções tecnológicas são adequadas para viabilizar essa integração de forma responsiva e reprodutível, reconhecendo as limitações inerentes aos indicadores bibliométricos?

1.1 JUSTIFICATIVA

Esta pesquisa se justifica pela convergência de desafios conceituais e operacionais enfrentados por IES brasileiras. Do ponto de vista conceitual, a literatura evidencia que a fragmentação de dados institucionais compromete a capacidade analítica (Maltese, 2018),

e que indicadores bibliométricos, embora úteis, são sensíveis à escolha da base (Bar-Ilan, 2008). Estudos sobre uso de evidência em decisões universitárias (Thiedig; Wegner, 2024) mostram que ter dados não equivale a usá-los de forma informada, pois esse uso depende de fatores organizacionais, da compreensão do escopo e das limitações dos próprios dados e indicadores e de sua adequação ao contexto decisório. Gestores universitários reconhecem que métricas conferem sustentação às decisões institucionais, mas requerem a interpretação dos resultados de acordo com as particularidades do ambiente institucional (Söderlind; Geschwind, 2019).

No plano operacional, a proposta responde à necessidade de organizar informações oriundas de sistemas administrativos e acadêmicos (PROGEPE, DRCA) e de bases externas (Lattes, Sucupira, OpenAlex, Scopus), com vistas à análise em nível agregado. Uma aplicação orientada a painéis e consultas, fundamentada em princípios de rastreabilidade e transparência metodológica (Hauschke *et al.*, 2021; Mitchell *et al.*, 2022), pode tornar esse processo mais sistemático e reproduzível.

A escolha de infraestruturas abertas — privilegiando bases como OpenAlex (Culbert *et al.*, 2025) e adotando princípios FAIR (Wilkinson *et al.*, 2016) — alinha-se aos requisitos de transparência da Lei de Acesso à Informação (Lei n. 12.527/2011) (Brasil, 2011) e aos debates internacionais sobre uso responsável de métricas. A integração de múltiplas fontes aplica metodologias recentes de articulação entre fontes locais e globais (Mazoni *et al.*, 2025), adaptando-as ao contexto institucional da UFTM. Sob essa perspectiva, a curadoria institucional é compreendida como prática de transformação de dados dispersos em informação estruturada e verificável, conforme proposto por Conway (2008).

1.1.1 Público-alvo e volume.

O público-alvo inclui gestores institucionais (pró-reitorias, direções de institutos, chefias de departamento e coordenações de programas), comissões de acompanhamento, profissionais de apoio técnico e analítico (como bibliotecas e setores de planejamento), e pesquisadores. O sistema também atende agências de fomento, órgãos reguladores, colaboradores externos e o público em geral, no contexto de transparência e visibilidade da produção científica institucional. O universo de referência envolve aproximadamente 882 docentes estatutários e 36.988 discentes com registros nas bases institucionais da UFTM, abrangendo tanto vínculos ativos quanto registros históricos, conforme dados da PROGEPE e da DRCA. No recorte analítico de produção científica, a base de dados apresenta produções datadas a partir de 1971, conforme registros retrospectivos informados nos currículos Lattes, considerando a extração realizada em 24/01/2026.

1.2 CONTRIBUIÇÕES ESPERADAS E IMPACTO

- a) Contribuir para a organização analítica de registros institucionais e bibliográficos, favorecendo consultas e análises em nível agregado com garantia de rastreabilidade

e documentação metodológica;

- b) Viabilizar mecanismos de acesso aos indicadores institucionais de pesquisa, incluindo painéis interativos e acesso programático, com atenção à reprodutibilidade dos procedimentos;
- c) Apoiar processos decisórios por meio de indicadores sobre produção científica, co-autorias, internacionalização e atualização curricular, posicionando-os como ferramentas de suporte — não substitutos — ao julgamento acadêmico;
- d) Alinhar a disponibilização de resultados aos princípios de transparência em nível agregado, preservando informações pessoais e adotando infraestruturas abertas compatíveis com princípios FAIR;
- e) Aplicar metodologias recentes de integração entre fontes locais (Lattes, Sucupira, sistemas institucionais) e bases abertas internacionais (OpenAlex) ao contexto institucional brasileiro, contribuindo para a visibilidade e a governança da informação científica institucional.

1.3 OBJETIVOS

1.3.1 Objetivo geral

Desenvolver o artefato Plataforma Capivara, uma aplicação para integração e análise de dados institucionais e bibliográficos, voltada à visualização de indicadores estratégicos, ao suporte à tomada de decisão e à transparência institucional.

1.3.2 Objetivos específicos

1. Definir e implementar regras de integração, consolidação e agregação de dados institucionais e bibliográficos;
2. Projetar um ambiente analítico que permita a geração e exploração controlada de indicadores institucionais;
3. Avaliar o funcionamento da aplicação por meio da caracterização da base consolidada e da análise de indicadores agregados;
4. Documentar e validar tecnicamente o artefato, considerando extensibilidade e reprodutibilidade;
5. Analisar padrões observacionais derivados dos indicadores produzidos pela plataforma desenvolvida, incluindo trajetórias organizacionais, colaboração e impacto bibliométrico, com ênfase em verificabilidade e rastreabilidade dos resultados.

1.4 METODOLOGIA E PROCEDIMENTOS TÉCNICOS

Trata-se de pesquisa aplicada, de caráter documental e tecnológico, centrada na análise e integração de registros já existentes em sistemas internos e bases públicas ou de acesso institucional. Não há recrutamento ou contato direto com participantes, e não são tratados dados pessoais sensíveis.

A pesquisa foi desenvolvida em cinco etapas principais:

1. **Estruturação de dados institucionais:** organização de registros de institutos, departamentos, programas e cursos a partir de fontes institucionais (PROGEPE, DRCA, relatórios públicos);
2. **Vinculação de docentes e discentes:** associação de vínculos funcionais e acadêmicos aos respectivos registros institucionais;
3. **Consulta a bases bibliográficas:** extração de informações de produção acadêmica a partir de Plataforma Lattes, Scupira, OpenAlex, Scopus, Periódicos CAPES e outras bases públicas ou de acesso autorizado;
4. **Padronização e consolidação:** normalização dos registros em ambiente analítico, com tratamento de duplicidades e documentação das transformações aplicadas (Mitchell *et al.*, 2022);
5. **Geração de indicadores agregados:** elaboração e disponibilização de indicadores em níveis institucionais (institutos, departamentos, programas) e individualizado, em painéis interativos e por meio de mecanismos de acesso programático.

Os resultados esperados incluem a disponibilização de painéis e indicadores analíticos voltados a dimensões como volume e tipologia de produções, coautorias, internacionalização e atualização curricular, apresentados como instrumentos de apoio à análise institucional — e não como substitutos do julgamento acadêmico (Söderlind; Geschwind, 2019). A validação da aplicação ocorre de forma distribuída ao longo do trabalho: a qualidade e a cobertura dos dados consolidados são examinadas na Subseção 4.2, enquanto a consistência das regras analíticas é verificada por meio da exploração interativa documentada na Subseção 4.7.

1.5 ESTRUTURA DO TRABALHO

Este trabalho está organizado da seguinte forma:

O **Capítulo 2 – Referencial Teórico** (Seção 2) está estruturado em quatro dimensões conceituais: **(2.1)** Curadoria institucional e CRIS (Subseção 2.1), que contextualiza práticas de organização e governança de informação científica institucional; **(2.2)**

Qualidade e rastreabilidade de dados sobre produção científica (Subseção 2.2), que examina critérios de confiabilidade, documentação metodológica e princípios FAIR; **(2.3)** Bibliometria e uso responsável de indicadores (Subseção 2.3), que discute limitações estruturais de métricas científicas e os requisitos para seu uso informado; e **(2.4)** Interoperabilidade e infraestrutura aberta (Subseção 2.4), que fundamenta escolhas técnicas orientadas à integração entre fontes locais e bases bibliográficas globais.

O **Capítulo 3 – Metodologia** (Seção 3) descreve os procedimentos técnicos adotados, incluindo critérios de elegibilidade dos dados, estratégias de consolidação e deduplicação rastreável, métodos de enriquecimento bibliométrico hierárquico e diretrizes para construção de indicadores agregados, explicitando escolhas metodológicas e seus impactos sobre a interpretação dos resultados.

O **Capítulo 4 – Resultados e Discussão** apresenta a aplicação desenvolvida e os achados empíricos organizados em oito seções: **(4.1)** Caracterização geral da base consolidada (Subseção 4.1); **(4.2)** Consolidação e deduplicação dos registros do Lattes (Subseção 4.2); **(4.3)** Perfil institucional da produção (Subseção 4.3); **(4.4)** Qualidade e cobertura dos dados (Subseção 4.4); **(4.5)** Enriquecimento bibliométrico (Subseção 4.5); **(4.6)** Análises comparativas e padrões emergentes (Subsubseção 4.6.1, Subsubseção 4.6.2, Subsubseção 4.6.3 e Subsubseção 4.6.4); **(4.7)** Exploração interativa e verificabilidade (Subseção 4.7); e **(4.8)** Síntese e limitações (Subseção 4.8).

O **Capítulo 5 – Conclusão** (Seção 5) retoma os objetivos da pesquisa, sintetiza as principais contribuições (Subseção 5.1), reconhece as limitações do estudo (Subseção 5.2), discute o alinhamento com os objetivos institucionais (Subseção 5.3) e apresenta considerações finais sobre o conceito de *engenharia de evidência* aplicado à gestão da informação científica institucional (Subseção 5.4).

O **Capítulo 6 – Trabalhos Futuros** (Seção 6) apresenta direções de continuidade organizadas em três eixos estratégicos: expansão de escopo (Subseção 6.1), aprofundamento analítico (Subseção 6.2) e transferibilidade (Subseção 6.3), delineando possibilidades de extensão metodológica, técnica e institucional da abordagem proposta.

2 REFERENCIAL TEÓRICO

2.1 CURADORIA E GESTÃO DA INFORMAÇÃO CIENTÍFICA EM INSTITUIÇÕES DE ENSINO SUPERIOR

Instituições de ensino superior acumulam e gerenciam volumes substanciais de informação científica provenientes de suas atividades de pesquisa. A forma como esses registros são organizados, preservados e disponibilizados impacta diretamente a gestão institucional e a formulação de políticas acadêmicas. Nesse sentido, Conway (2008) argumenta que a informação científica deve ser tratada como ativo institucional estratégico, cuja gestão deliberada torna-se central na universidade digital.

A digitalização da produção acadêmica e a multiplicidade de fontes deslocaram o tema da curadoria para desafios técnicos e organizacionais associados à qualidade, governança e sustentabilidade dos processos (Kindling; Strecker, 2022).

2.1.1 Informação científica como ativo institucional estratégico

Conway (2008) propõe compreender a informação científica produzida nas universidades como ativo estratégico, exigindo práticas sistemáticas de gestão: reunir, cuidar, disponibilizar e preservar recursos digitais em ambientes universitários. Essa perspectiva conecta comunicação científica e gestão institucional, reforçando que a utilidade dos dados depende de organização, acesso contínuo e capacidade de integração a processos de avaliação e decisão.

2.1.2 Fragmentação de dados e desafios de integração

Universidades enfrentam dificuldades para consolidar informação científica devido à fragmentação e diversidade de dados. Maltese (2018) descreve esse cenário em termos de silos informacionais, duplicações e baixa correlacionabilidade por diferenças de formatos, metadados e convenções.

No contexto brasileiro, a fragmentação envolve plataformas nacionais (Lattes, Scupira), repositórios institucionais, sistemas internos (RH e gestão acadêmica) e bases externas, com duplicações e inconsistências entre declarações. Como consequência, análises comparativas e indicadores demandam esforços relevantes de extração, limpeza e reconciliação (Kindling; Strecker, 2022). Além disso, por serem dinâmicos, esses registros exigem atualização contínua e governança para manter consistência em múltiplas fontes de verdade (Szomszor *et al.*, 2021).

2.1.3 Sistemas de Informação de Pesquisa Corrente (CRIS) como resposta institucional

Sistemas CRIS (*Current Research Information Systems*) são uma resposta institucional recorrente à fragmentação, atuando como plataformas integradoras entre entidades

do ciclo de vida da pesquisa (pesquisadores, projetos, publicações, financiamentos), com suporte a *reporting*, análises e cumprimento de políticas.

Nesse sentido, Castro (2018) ilustram esse papel ao discutir o uso de um CRIS (Pure, plataforma comercial de CRIS adotada por universidades) para apoiar políticas de Acesso Aberto e RDM, incluindo articulação organizacional e maior coerência entre processos. Contudo, CRIS não esgotam as necessidades de avaliação baseada em múltiplas fontes heterogêneas: persistem desafios de integração com plataformas nacionais/internacionais, produção de indicadores comparáveis e governança de dados em ecossistemas diversos. No presente trabalho, os CRIS são adotados como referência conceitual para a integração de registros institucionais e bibliográficos, orientando a definição das entidades institucionais e a disponibilização de indicadores analíticos. Nesse sentido, a informação de pesquisa é organizada a partir de entidades e de suas relações (como pessoas, vínculos, unidades e produções), sendo a integração dos dados tratada como etapa distinta da geração de indicadores, permitindo maior controle, rastreabilidade e interpretação contextual dos resultados.

2.1.4 A Plataforma Lattes no contexto brasileiro

A Plataforma Lattes, mantida pelo CNPq desde 1999, funciona como currículo nacional padronizado e ocupa posição central na avaliação acadêmica no Brasil, com impacto em processos CAPES e CNPq. Além de infraestrutura administrativa, é fonte relevante para análises bibliométricas em larga escala. Mena-Chalco *et al.* (2014) demonstram empiricamente seu uso para caracterização da ciência brasileira e redes de coautoria.

Em princípio, instituições podem extrair do Lattes dados consolidados sobre produção e colaboração, inclusive via mecanismos institucionais de extração mediante solicitação formal, conforme regulamentado (Conselho Nacional de Desenvolvimento Científico e Tecnológico, 2025).

2.1.5 Desafios técnicos e limitações do Lattes

Apesar de sua relevância, o Lattes apresenta desafios para análises automatizadas. Estudos apontam heterogeneidade e inconsistências decorrentes do preenchimento manual, além de dificuldades de extração e normalização por sua natureza semiestruturada (Digiampietri *et al.*, 2015). A redundância de registros é recorrente: uma mesma produção pode ser declarada de formas distintas por coautores, dificultando consolidação e deduplicação (Digiampietri *et al.*, 2015). Evidências recentes indicam que, em cenários de integração entre bases locais e globais, a ambiguidade de nomes permanece um desafio persistente, com impacto na integridade de entidades em sistemas integrados (Mazoni *et al.*, 2025).

Há ainda lacunas críticas para interoperabilidade: ausência parcial de identificadores persistentes (como DOI), além de inconsistências de preenchimento, escassez de

metadados bibliométricos e inconsistência de filiações e coautores (Mena-Chalco *et al.*, 2014). Mesmo quando identificadores estão disponíveis, sua cobertura pode ser incompleta e enviesada, o que leva ao uso de identificadores como “pistas” para pareamento parcial e validação (Mazoni *et al.*, 2025). Por isso, estratégias de integração com bases externas (OpenAlex, Scopus, Web of Science) e classificações como Qualis tornam-se relevantes para complementar e validar metadados, embora exijam desambiguação, pareamento aproximado e processos explícitos de validação e rastreabilidade.

Em síntese, a curadoria institucional envolve consolidar dados heterogêneos em informação estruturada e verificável, com governança, atualização e rastreabilidade das decisões metodológicas. No Brasil, o Lattes é uma base valiosa, mas limitada, e seu uso institucional robusto depende de integração e enriquecimento. Para ser sustentável, esse esforço requer responsabilidades definidas e rotinas institucionais de atualização e verificação de consistência, de modo que a curadoria se consolide como prática institucional e não apenas como iniciativa técnica pontual. A próxima seção discute qualidade, rastreabilidade e proveniência como fundamentos para indicadores defensáveis.

2.2 QUALIDADE DE DADOS E RASTREABILIDADE EM SISTEMAS DE INFORMAÇÃO CIENTÍFICA

A construção de indicadores científicos confiáveis requer, além de integração entre fontes, processos sistemáticos de validação, documentação de proveniência e garantia de qualidade ao longo do ciclo de vida dos dados. A literatura destaca que qualidade não é atributo binário, mas resultado de processos contínuos e escolhas institucionais explícitas (Kindling; Strecker, 2022; Liaw *et al.*, 2021).

2.2.1 Processos de garantia de qualidade em repositórios científicos

Kindling e Strecker (2022) analisam práticas de garantia da qualidade (QA, *quality assurance*) em repositórios e mostram que certificações formais não garantem qualidade: estratégias de QA variam, dependem de recursos e envolvem escolhas e limitações. Um ponto central é que a ausência de metadados sobre validação e critérios de inclusão dificulta que usuários avaliem confiabilidade. Para sistemas institucionais, a implicação é direta: consolidar dados não basta; é necessário documentar origem, transformações, critérios e limitações, para sustentar o uso responsável dos indicadores.

2.2.2 Rastreabilidade e proveniência de fluxos de trabalho analíticos

Rastreabilidade e proveniência permitem responder à pergunta: “de onde veio esse número?” Mitchell *et al.* (2022) discutem gerenciamento de proveniência em fluxos de trabalho, defendendo a ligação explícita entre dados originais, transformações intermediárias e outputs, registrada por metadados estruturados e identificadores persistentes. Em sistemas institucionais, isso implica documentar filtragem, deduplicação, enriquecimento e agregação, além de permitir auditoria e verificação por terceiros (Mitchell *et al.*, 2022).

No contexto deste trabalho, esses requisitos são operacionalizados na Plataforma Capivara por quatro mecanismos: (i) preservação de metadados de origem, declarantes e divergências durante a consolidação; (ii) registro de versões, processamentos e regras analíticas na base relacional; (iii) explicitação de recortes e definições de indicadores em camadas documentadas; e (iv) verificabilidade operacional por exploração interativa e exportação consistente dos dados. A implementação desses mecanismos é detalhada na Metodologia (Capítulo 3) e sua aplicação é demonstrada nos Resultados (Capítulo 4).

2.2.3 Prestígio institucional e confiabilidade de dados

A confiabilidade atribuída a dados e publicações é influenciada por concepções de prestígio. Brembs (2018) argumenta que ranking/impacto de periódicos não se correlaciona necessariamente com maior reprodutibilidade. Para avaliação institucional, isso reforça que confiabilidade deve ser sustentada por processos de validação e curadoria, e não apenas pela reputação das fontes ou veículos.

2.2.4 Múltiplas dimensões da qualidade de dados

A qualidade de dados é multidimensional, envolvendo aspectos intrínsecos, técnicos e contextuais. Liaw *et al.* (2021) sintetizam esse entendimento e destacam que critérios variam conforme o uso pretendido. Para sistemas institucionais, isso implica explicitar escolhas (completude versus acurácia, curadoria manual versus validação automática, padronização versus flexibilidade) e documentar limitações de forma compreensível aos usuários.

Na prática, essa multidimensionalidade aparece na variabilidade de atualidade e consistência dos registros entre fontes e mecanismos de ingestão. Bases bibliográficas diferem em políticas de atualização, correção e exclusão, tornando seus registros inerentemente voláteis e, por vezes, inconsistentes (Culbert *et al.*, 2025). Mesmo quando um DOI está disponível, a qualidade dos metadados recuperados a partir dele pode variar entre provedores e mecanismos de importação, produzindo registros mais completos ou mais degradados conforme o estágio de curadoria da fonte. Diante dessa variabilidade, o registro de proveniência e de transformações ao longo do fluxo de processamento é um requisito para indicadores passíveis de verificação, pois permite rastrear a origem e a atualidade das informações incorporadas.

Em síntese, qualidade e confiabilidade de indicadores dependem de processos: validação, transparência metodológica, proveniência e documentação de limitações. A próxima seção discute usos e limites da bibliometria, enfatizando como escolhas de base e critérios de agregação afetam validade e interpretação.

2.3 BIBLIOMETRIA E INDICADORES CIENTÍFICOS: USOS E LIMITES NA AVALIAÇÃO INSTITUCIONAL

Indicadores bibliométricos são amplamente usados para caracterizar produção e impacto, mas sua validade depende das bases utilizadas, critérios de agregação e do contexto disciplinar e nacional. A literatura aponta oportunidades (comparações em larga escala, identificação de padrões) e riscos (vieses, interpretações simplistas e uso acrítico), exigindo transparência e contextualização.

2.3.1 Fundamentos: sensibilidade às bases e limites conceituais

Indicadores baseados em publicações e citações são sensíveis à base de dados. O índice h , que combina produtividade e impacto ao considerar o número de publicações com pelo menos o mesmo número de citações, apresenta variações substanciais entre bases. Bar-Ilan (2008) mostram variações substanciais do índice h entre Web of Science, Scopus e Google Scholar, devido a diferenças de cobertura e critérios de indexação. Isso implica que análises bibliométricas devem explicitar fontes e limitações, sob risco de comparações inadequadas.

Szomszor *et al.* (2021) defendem interpretação responsável: indicadores não falam por si, devem ser comunicados com clareza, acompanhados de limitações e combinados com evidências qualitativas. Complementarmente, Belter (2015) alertam que métricas são frequentemente aplicadas sem domínio metodológico, levando a interpretações equivocadas.

Uma crítica clássica é que citação não equivale a qualidade: Seglen (1991) argumenta que citações medem utilidade/atenção e sofrem vieses (por exemplo, disciplina, reputação, autocitação), o que exige leitura crítica para evitar reduzir a avaliação científica a contagens numéricas. Também podem ocorrer incentivos editoriais que influenciam padrões de citação, inclusive com encorajamento de autocitações em nível de periódico. Em ambientes de “publicar ou perecer” (alta pressão por produtividade e desempenho em métricas), práticas de citação podem ser influenciadas por incentivos e pressões do ecossistema editorial (Taşkın, 2025).

A comparabilidade entre indicadores também é limitada por diferenças disciplinares nos regimes de produção e autoria. No Brasil, padrões de coautoria variam entre grandes áreas e influenciam diretamente métricas de produtividade e colaboração (Mena-Chalco *et al.*, 2014). Por isso, processos avaliativos tendem a combinar evidências quantitativas e qualitativas, e a interpretação de indicadores deve considerar parâmetros disciplinares e objetivos institucionais, evitando comparações diretas entre áreas com regimes distintos de produção.

2.3.2 Contexto brasileiro: cobertura, Sucupira e efeitos institucionais

No contexto brasileiro, divergências entre bases também se refletem em análises institucionais. Um estudo conduzido por pesquisadores da UFTM identificou diferenças no índice h conforme a fonte considerada (Lattes, Web of Science, Scopus, Google Scholar) (Gouvêa *et al.*, 2022), evidenciando que resultados bibliométricos variam de acordo com as bases e critérios adotados, o que reforça a importância da explicitação metodológica em processos formais.

Bases internacionais apresentam sub-representação de periódicos usados por programas brasileiros: Melo, Trinca e Maricato (2021) evidenciam cobertura incompleta e desigual por área, introduzindo vieses em indicadores derivados exclusivamente dessas fontes.

No cenário nacional, a Plataforma Sucupira é a principal base do sistema de avaliação da pós-graduação e depende de dados fornecidos pelos programas; embora tenda a ser mais completa para a produção nacional, pode apresentar imperfeições associadas ao preenchimento descentralizado (Melo; Trinca; Maricato, 2021). Nesse sistema, as notas atribuídas aos programas variam de 1 a 7, e notas 1 e 2 implicam descredenciamento (Melo; Trinca; Maricato, 2021). Nesse cenário, inconsistências de preenchimento não afetam apenas a qualidade do dado, mas também podem produzir efeitos avaliativos adversos para os programas.

Mugnaini e Quoniam (2004) discutem a incorporação progressiva de indicadores bibliométricos na avaliação e no *reporting* científico no Brasil, destacando que essas métricas podem apoiar decisões institucionais, mas também introduzem tensões quando passam a orientar práticas e prioridades acadêmicas de forma excessivamente quantitativa.

O Qualis é um instrumento de classificação de periódicos adotado pela CAPES na avaliação da pós-graduação brasileira. Pires *et al.* (2020) observam concentração em periódicos com pontuação elevada, indicando que sistemas de avaliação podem moldar comportamentos e gerar efeitos colaterais. Isso reforça a necessidade de avaliações multidimensionais e de indicadores rastreáveis, com premissas explicitadas.

Em síntese, bibliometria é útil, mas limitada por vieses de base, interpretação e contexto. Esses limites deslocam parte do problema para a infraestrutura que torna indicadores rastreáveis e verificáveis — tema da próxima seção sobre interoperabilidade e infraestrutura aberta.

2.4 INTEROPERABILIDADE E INFRAESTRUTURA ABERTA PARA AVALIAÇÃO CIENTÍFICA

A construção responsável de indicadores institucionais depende de infraestruturas que permitam integrar fontes heterogêneas, documentar transformações e viabilizar verificação metodológica. Nesse contexto, interoperabilidade é requisito central para transparência e confiabilidade.

2.4.1 Princípios FAIR aplicados à informação de pesquisa

Hauschke *et al.* (2021) discutem FAIR aplicado à **informação de pesquisa** (metadados de pesquisadores, projetos, publicações e financiamentos), destacando identificadores persistentes (como DOI e ORCID¹), vocabulários controlados e metadados estruturados como fundamentos de interoperabilidade. Esses elementos favorecem integração e rastreabilidade, ao permitir que processos de curadoria e transformação sejam documentados e reproduzidos.

2.4.2 Integração de bases heterogêneas: abordagens técnicas

A integração entre fontes como Lattes, Sucupira e bases internacionais exige lidar com inconsistências e ausência parcial de identificadores. Mazoni *et al.* (2025) exemplificam abordagem híbrida: estratégias determinísticas quando disponíveis e estratégias probabilísticas de pareamento, baseadas em similaridade textual quando identificadores persistentes não estão disponíveis. Essa combinação viabiliza enriquecimento com citações, acesso aberto e coautoria, mas requer validação e documentação explícita dos limites e implicações metodológicas.

2.4.3 Escolha de bases abertas: cobertura e completude de metadados

Bases proprietárias impõem custos e restrições que afetam reprodutibilidade. Bases abertas como OpenAlex oferecem acesso irrestrito, e Culbert *et al.* (2025) discutem competitividade em cobertura e completude de metadados, com ganhos de verificabilidade, na medida em que dados e processos são mais transparentes. Contudo, a abertura não elimina problemas de qualidade: duplicações e erros persistem, exigindo tratamento e validação dos dados pela própria instituição.

2.4.4 Grafos de conhecimento, citações abertas e avaliação reprodutível

A interoperabilidade facilita estruturar metadados como grafos de conhecimento, explicitando relações entre entidades acadêmicas. Manghi (2024) associam metadados

¹ORCID (Open Researcher and Contributor ID) é um identificador persistente e único de pesquisadores, usado para desambiguação de autoria e interoperabilidade entre sistemas, facilitando a vinculação entre perfis e produções em diferentes bases.

abertos e grafos à avaliação reprodutível, desde que escolhas metodológicas (fontes integradas, deduplicação, atribuição institucional, desambiguação) sejam documentadas.

Dados de citação abertos ampliam transparência e reduzem dependência de bases proprietárias. Taşkın (2025) discutem citações abertas (*open citations*) como condição para análises independentes, ainda que sujeitas a vieses estruturais (idioma, regionalidade e cobertura).

Além disso, mesmo quando metadados são acessíveis, mecanismos de ranqueamento e recomendação podem permanecer opacos. Filtros baseados em métricas (por exemplo, limiares de índice h) podem introduzir efeitos de retroalimentação, concentrando visibilidade em atores já estabelecidos (Taşkın, 2025). Esse cenário reforça a necessidade de transparência metodológica e explicitação de premissas na produção e no uso de indicadores (Szomszor *et al.*, 2021).

Interoperabilidade, FAIR e infraestrutura aberta criam condições para indicadores mais transparentes e reprodutíveis, mas não substituem governança e curadoria locais. No contexto brasileiro, integrar Lattes, Qualis e bases abertas, como OpenAlex, é estratégia viável para superar limitações de fontes isoladas, fundamentando o desenvolvimento do sistema Capivara, cujo desenho e implementação são detalhados no próximo capítulo.

3 METODOLOGIA

3.1 NATUREZA DA PESQUISA E ABORDAGEM METODOLÓGICA

Este trabalho caracteriza-se como uma pesquisa aplicada de natureza tecnológica, cujo objetivo é a concepção e implementação de um artefato computacional denominado *Capivara*, destinado à consolidação e análise de indicadores de produção científica institucional.

A abordagem metodológica adotada fundamenta-se na *Design Science Research* (DSR), conforme proposta por Hevner et al. (2004), por se tratar de um estudo orientado à solução de um problema prático por meio da construção e avaliação de um sistema de informação. Nesse contexto, o conhecimento científico é produzido a partir do desenvolvimento do artefato, de suas decisões metodológicas e de sua capacidade de responder às necessidades institucionais identificadas.

3.2 PROBLEMA DE PESQUISA E SOLUÇÃO PROPOSTA

O problema de pesquisa deste trabalho consiste em **como integrar e tratar registros institucionais e bibliográficos dispersos em múltiplas fontes, de modo a viabilizar um sistema de indicadores institucionais rastreável e passível de verificação**. As debilidades informacionais observadas — fragmentação, inconsistência e dispersão de dados de produção acadêmica declarados individualmente por docentes em plataformas curriculares, em especial o currículo Lattes — constituem o ponto de partida do problema e dificultam análises institucionais consolidadas, comparações entre unidades acadêmicas, identificação de padrões de colaboração e avaliação de impacto científico.

Como solução, propõe-se o desenvolvimento da plataforma *Capivara*, um sistema de indicadores de pesquisa capaz de integrar, tratar, consolidar e enriquecer dados curriculares, transformando registros declarativos individuais em uma base analítica institucional estruturada e explorável, **com registro de proveniência e rastreabilidade das transformações aplicadas**.

3.3 VISÃO GERAL DO MÉTODO

A metodologia organiza-se como um fluxo de transformação de dados, no qual as informações percorrem etapas sucessivas até se tornarem aptas à análise institucional. De forma sintética, o fluxo metodológico compreende:

1. Preparação das entidades institucionais, incluindo docentes, discentes e estrutura organizacional;
2. Extração e normalização dos currículos Lattes, considerados fonte primária de validade institucional da produção acadêmica;

3. Integração e deduplicação das produções, consolidando coautorias e estabelecendo identidade institucional das produções;
4. Enriquecimento bibliométrico, por meio da integração com bases externas, sem substituição dos dados declarados;
5. Consolidação analítica, com preparação dos dados para exploração e geração de indicadores.

Esse fluxo é conceitualmente sequencial, mas admite reprocessamentos e atualizações incrementais, garantindo consistência e rastreabilidade dos dados. As decisões de validade, deduplicação e enriquecimento são registradas com origem e data, permitindo auditoria.

Essa governança é sustentada por uma base relacional central em PostgreSQL, adotada como base mestra dos dados consolidados. A partir dela, são executados os processos de integração, enriquecimento, deduplicação e indexação que sustentam a camada analítica e a exploração interativa da Plataforma Capivara.

3.4 FONTES DE DADOS E CRITÉRIOS DE USO

As fontes de dados utilizadas foram organizadas de forma estratificada, conforme seu papel metodológico:

- a) **Fontes institucionais:** dados oficiais de docentes, discentes e estrutura organizacional, utilizados para definição de identidade e vínculos institucionais;
- b) **Fonte curricular declarativa:** a Plataforma Lattes, adotada como fonte primária declaratória da produção acadêmica dos docentes;
- c) **Fontes externas de enriquecimento:** bases como OpenAlex, Crossref, OpenCitations, Scopus e Qualis, e serviços de identificação como ORCID, utilizadas para enriquecimento contextual e obtenção de metadados e métricas quando disponíveis.

As extrações dessas fontes foram realizadas em um período de referência documentado nos metadados de processamento, de modo a permitir reexecução, comparação e auditoria dos indicadores.

Para fins de consistência do recorte, as fontes de referência — bases institucionais (UFTMNet e SISCAD) e a Plataforma Lattes — foram extraídas em 24/01/2026, data utilizada para caracterizar o universo de docentes e discentes e delimitar o período analisado. As fontes externas de enriquecimento (OpenAlex, Crossref, OpenCitations, Scopus, Qualis e ORCID) foram consultadas em janelas de coleta, uma vez que as coletas podem ocorrer em dias distintos; as datas e versões consultadas são registradas nos metadados de processamento para controle de reexecução e comparação. No presente

estudo, as consultas às bases abertas (OpenAlex, Crossref e OpenCitations) ocorreram em jan. 2026 (janela mensal); a coleta via Scopus compreendeu uma janela trimestral, em função de limitações operacionais de acesso à API (restrições de cota e disponibilidade de consulta); e a classificação Qualis fundamentou-se no quadriênio 2021–2024, conforme o ciclo disponível e vigente no período de desenvolvimento do estudo.

Um princípio central da metodologia é o da não substituição: dados provenientes de fontes externas não invalidam nem substituem informações declaradas no currículo Lattes, atuando apenas como camadas complementares de informação. Em particular, as fontes externas não são usadas para acrescentar novas produções ao conjunto analisado; elas apenas enriquecem registros já declarados (por exemplo, por DOI). Métricas agregadas fornecidas por bases bibliométricas (por exemplo, total de citações e índice h) são tratadas como atributos externos do pesquisador e podem refletir cobertura diferente da base declaratória, sem implicar incorporação de novas produções. Assim, a prioridade metodológica é consistência institucional e rastreabilidade, e não exaustividade bibliométrica.

3.5 CONJUNTO DE REGISTROS, ELEGIBILIDADE E RECORTE ANALÍTICO

Esta subseção define o **conjunto de registros** analisado e explicita os **recortes metodológicos** adotados para garantir consistência e verificabilidade dos indicadores produzidos. Em particular, estabelece: (i) o universo institucional de referência; (ii) o processo de elegibilidade dos currículos Lattes; e (iii) o recorte analítico principal utilizado na interpretação dos resultados.

3.5.1 Universo institucional de referência

Para fins operacionais e de verificabilidade, o universo de docentes considerado neste estudo corresponde aos professores cadastrados no sistema institucional da UFTM denominado UFTMNet (Sistema Integrado da UFTM), que constitui a base administrativa oficial utilizada para gestão funcional e acadêmica. Esse recorte não representa a totalidade histórica de docentes que já atuaram na instituição, mas o conjunto formalmente registrado no sistema na **data de referência** (conforme período de referência das fontes descrito na Seção 3.4).

Para contextualização do escopo, na **data de referência** (24/01/2026), a UFTM possuía **1.287 docentes cadastrados** (fonte: PROGEPE), abrangendo diferentes regimes e perfis de vínculo. Este estudo delimita-se ao subconjunto de **docentes estatutários**, incluindo **ativos e aposentados**.

No universo estatutário, **107 docentes** não apresentavam vinculação departamental identificável na base funcional. Esses casos concentram-se predominantemente em registros mais antigos, refletindo práticas administrativas anteriores à consolidação do modelo departamental atualmente adotado. Para evitar perda de cobertura e manter a

rastreabilidade, esses registros foram preservados e tratados explicitamente como categoria *não identificado* nas análises organizacionais, conforme estratégia de complementação descrita a seguir.

Para contextualizar indicadores de **participação discente** (por exemplo, coautoria e envolvimento de alunos em publicações e demais registros), considera-se, na **data de referência** (24/01/2026), o universo de **36.988 discentes cadastrados** no SISCAD (sistema institucional de gestão acadêmica), abrangendo tanto vínculos ativos quanto registros históricos de discentes que já passaram pela instituição (fonte: DRCA).

3.5.1.1 Casos sem vinculação departamental e estratégia de complementação.

A identificação departamental (lotação/unidade acadêmica) tem como **fontes institucionais** os sistemas mantidos pela PROGEPE, com destaque para o UFTMNet e o SISCAD, que possuem finalidades e escopos temporais distintos. Enquanto o UFTMNet reflete prioritariamente a situação funcional mais recente, o SISCAD preserva vínculos acadêmicos históricos, incluindo registros de docentes já desligados.

No momento da extração e consolidação dos dados (24/01/2026), uma fração dos docentes estatutários não apresentava departamento identificado no UFTMNet, situação mais frequente em registros antigos. Para reduzir perda de cobertura analítica, adotou-se uma estratégia de complementação em duas etapas: (i) verificação a partir do próprio UFTMNet; e (ii) quando necessário, consulta ao SISCAD como fonte auxiliar para identificação de vínculos departamentais pretéritos.

Em uma etapa intermediária do processo, foram identificados **201 docentes** estatutários sem departamento no UFTMNet; nessa mesma etapa, o cruzamento com o SISCAD permitiu identificar **132 vínculos departamentais** adicionais. Considerando que ambos os sistemas estão sujeitos a atualizações administrativas ao longo do tempo, esses quantitativos são interpretados como **marcos do processo de consolidação**, e não como valores absolutos históricos.

No conjunto analítico final, os casos remanescentes foram mantidos como categoria *não identificado*, garantindo rastreabilidade e evitando inferências não verificáveis.

Além da ausência de identificação departamental em parte dos registros antigos, há uma limitação adicional relacionada à mobilidade organizacional de docentes ao longo do tempo e às mudanças na própria estrutura departamental da instituição. Neste estudo, a vinculação departamental é tratada de forma estática, com base na lotação identificada nas bases institucionais utilizadas na etapa de consolidação, privilegiando uma leitura alinhada à configuração organizacional vigente na UFTM. Adicionalmente, como os registros administrativos disponíveis priorizam a situação funcional mais recente e não preservam, de forma suficientemente estruturada, todo o histórico de movimentações internas, mudanças de denominação, desmembramentos, fusões ou extinções de departamentos, não foi possível rastrear de modo sistemático essas transformações ao longo do tempo. Assim,

nas análises organizacionais, a produção de cada docente é associada à lotação identificada na base consolidada, e não à reconstrução exaustiva de seu histórico departamental.

3.5.2 Elegibilidade dos currículos Lattes

A base inicial compreendeu **4.290 registros** fornecidos pela PROPPG, incluindo diferentes perfis profissionais vinculados à instituição. Desse universo, foram **processados 1.206 currículos** referentes a **docentes em geral**, e então selecionados **882 currículos Lattes** correspondentes a **docentes estatutários**. Esse conjunto compreende **963 registros funcionais** pertencentes a **948 pessoas**, refletindo situações em que uma mesma pessoa tem múltiplos vínculos ao longo do tempo (por exemplo, recontrações ou mudanças de regime). Este conjunto constitui o universo docente analisado.

3.5.2.1 Critérios operacionais (resumo).

A elegibilidade considera: (i) presença de identificadores mínimos para vinculação entre registros de pessoa no universo institucional; (ii) pertencimento ao subconjunto estatutário no período analisado; e (iii) consistência mínima dos metadados necessários à consolidação e deduplicação. Os detalhes de vinculação e validação são descritos nas subseções de integração e consolidação.

3.5.3 Recorte analítico principal e implicações

O recorte analítico principal considera as produções declaradas **durante o período de vínculo institucional**, por refletirem diretamente a contribuição acadêmica no contexto da UFTM. Assim, análises institucionais e comparativas são construídas prioritariamente sobre o subconjunto de produções associado ao vínculo à UFTM. Operacionalmente, a classificação “durante o vínculo” é obtida a partir do cruzamento entre ano do registro e janelas de vínculo funcional derivadas do UFTMNet (detalhadas na etapa de consolidação).

Produções declaradas **fora do período de vínculo** são preservadas na base para fins de contextualização (trajetória acadêmica), mas não compõem o núcleo das análises institucionais. Essa decisão metodológica reduz vieses de atribuição institucional, ao evitar que produções anteriores ou posteriores ao vínculo sejam interpretadas como resultado da atuação na UFTM. Em consonância com o princípio de não substituição (Seção 3.4), dados externos são utilizados apenas como enriquecimento contextual, sem invalidar declarações curriculares.

A base contempla os **sete institutos** da UFTM, incluindo o **CEFORES** (*Centro de Educação Profissional*), e, no eixo temporal, os **registros de produção** cobrem o intervalo de **1971–2026**.

3.5.3.1 Definição operacional do início e término do vínculo.

Para operacionalizar a janela de vínculo utilizada no recorte ‘durante o vínculo’, o UFTMNet disponibiliza múltiplas datas funcionais com semânticas distintas. Neste estudo, o **início do vínculo institucional** é estimado pela **data de exercício**, descrita na fonte como a data em que o servidor ‘entrou em atividade’ no cargo. Essa escolha foi adotada por ser a referência mais adequada disponível para delimitar o início do vínculo analisado.

Quando a data de exercício não está disponível, utiliza-se a **data de ingresso** como aproximação (*fallback*), pois o sistema também registra essa data para o vínculo funcional. Segundo a descrição do sistema, a data de ingresso pode representar a entrada do servidor no sistema, e não necessariamente o início efetivo do vínculo considerado na análise. Por essa razão, ela não é adotada como referência principal quando a data de exercício está presente.

O **término do vínculo** é definido como a primeira ocorrência entre **data de exclusão** e **data de inatividade**, quando existentes. Na ausência de ambas, o vínculo é tratado como **sem data de término**, permitindo a inclusão de docentes com vínculo em aberto no período analisado.

Em síntese, o estudo analisa **882 currículos Lattes de docentes estatutários** da UFTM, com recorte principal nas produções declaradas **durante o vínculo institucional**. Esse enquadramento metodológico orienta a interpretação dos resultados apresentados no Capítulo 4.

3.6 TRATAMENTO, INTEGRAÇÃO E CONSOLIDAÇÃO DAS PRODUÇÕES

As produções acadêmicas extraídas dos currículos Lattes são inicialmente tratadas de forma individual, sendo posteriormente integradas para identificação de coautorias institucionais e eliminação de duplicidades decorrentes de múltiplas declarações da mesma produção.

A consolidação é realizada por meio de critérios hierárquicos, priorizando identificadores persistentes (como DOI) e, quando aplicável, combinações de atributos normalizados (título, ano e tipo de produção). Esse processo permite manter simultaneamente:

- a) uma visão declarativa individual, útil para análises por docente; e
- b) uma visão institucional consolidada, adequada à geração de indicadores agregados.

Como a consolidação se apoia em metadados declarados, permanecem casos-limite em que produtos relacionados a uma mesma pesquisa podem coexistir como registros distintos na base consolidada. Isso pode ocorrer, por exemplo, quando um trabalho apresentado em evento é posteriormente expandido e publicado em outro formato. Nesses

casos, a inclusão do tipo de produção na chave de consolidação busca preservar manifestações editoriais distintas, evitando a fusão indevida de registros que representam etapas diferentes da trajetória de comunicação científica (Bar-Ilan, 2008; Manghi, 2024). Por essa razão, os critérios adotados visam sobretudo eliminar duplicidades evidentes, sem pretensão de resolver exaustivamente todos os possíveis desdobramentos editoriais de um mesmo conteúdo intelectual.

3.7 ENRIQUECIMENTO BIBLIOMÉTRICO

O enriquecimento bibliométrico é realizado de forma incremental, utilizando identificadores persistentes sempre que disponíveis. O OpenAlex atua como principal *hub* de integração, permitindo acesso a metadados, perfis de autores e informações de citação, complementados por consultas ao Crossref, ORCID, OpenCitations e, quando aplicável, ao Scopus.

As métricas obtidas são armazenadas como camadas adicionais de informação, preservando a distinção entre dados institucionais declarados e dados externos enriquecidos.

3.8 CONSOLIDAÇÃO ANALÍTICA E GERAÇÃO DE INDICADORES

Após o processamento e enriquecimento, os dados são consolidados em uma camada analítica otimizada para consultas agregadas e exploração multidimensional. Essa etapa viabiliza a construção dos indicadores de pesquisa disponibilizados pela plataforma *Capivara*, permitindo análises por período, unidade acadêmica, tipo de produção, impacto bibliométrico e padrões de colaboração.

A metodologia adotada garante reprodutibilidade, rastreabilidade e alinhamento institucional, assegurando que os indicadores gerados reflitam tanto a produção declarada quanto seu contexto científico ampliado.

3.9 CONSTRUÇÃO DOS INDICADORES

Os indicadores construídos pela plataforma organizam-se em tipologias analíticas não excludentes, permitindo múltiplas leituras do mesmo conjunto de dados conforme o contexto institucional. De forma conceitual, distinguem-se indicadores de natureza **institucional** (volume, distribuição e evolução da produção), **bibliométrica** (impacto, citações e qualidade percebida) e **colaborativa** (coautoria, internacionalização e participação discente).

As métricas adotadas baseiam-se em operações fundamentais de contagem, agregação e distribuição, organizadas por eixos analíticos recorrentes, como tempo, unidades organizacionais e tipos de produção. Dimensões temáticas adicionais (idioma, país, áreas de conhecimento, classificação Qualis) são incorporadas quando coerentes com o indicador analisado.

Nas análises organizacionais por programas de pós-graduação, a vinculação é estabelecida com base na associação do docente ao respectivo programa no período analisado. Trata-se de um critério de participação organizacional, que não implica, por si só, que cada produção tenha sido formalmente desenvolvida no âmbito do programa.

Um princípio central da metodologia é a manutenção explícita da **dualidade analítica** entre a visão declarativa (preservando as declarações individuais dos currículos) e a visão deduplicada (institucional), evitando sobrecontagens em análises agregadas. Essa decisão permite atender simultaneamente a demandas de auditoria curricular e de avaliação institucional.

O detalhamento das tipologias, métricas, critérios de agregação e regras de cálculo dos indicadores é apresentado no Apêndice A.

3.10 ESTRATÉGIAS DE EXPLORAÇÃO E VISUALIZAÇÃO

A exploração dos indicadores é organizada a partir de dimensões analíticas explícitas e filtros controlados, permitindo a navegação entre diferentes níveis de agregação sem alteração das regras metodológicas de cálculo. As dimensões centrais incluem tempo, unidades organizacionais, tipo de produção e recortes temáticos específicos de cada indicador.

O sistema suporta operações de ir do agregado para os registros (*drill-through*), troca de eixos analíticos e seleção interativa de categorias, respeitando os limites metodológicos definidos para cada indicador. Essas operações possibilitam aprofundamento analítico progressivo, do nível institucional agregado a subconjuntos específicos.

As funcionalidades de exportação preservam o recorte analítico aplicado (filtros, dimensões e perspectiva deduplicada ou declarativa), assegurando reprodutibilidade e transparência dos resultados. O detalhamento das estratégias de exploração, operações suportadas e mecanismos de exportação é apresentado no Apêndice B.

3.11 AVALIAÇÃO DO ARTEFATO

A avaliação da plataforma *Capivara* foi conduzida pelo pesquisador, combinando **validação técnica** e **avaliação por cenários institucionais representativos**, com ênfase em **verificabilidade por rastreabilidade** (do dado de origem ao indicador) e ***drill-through*** (do agregado aos registros subjacentes).

A **validação técnica** verificou (i) a consistência dos dados consolidados em relação às fontes, (ii) a idempotência dos processos de integração e reprocessamento, e (iii) a rastreabilidade completa das transformações aplicadas ao longo do pipeline (extração, normalização, deduplicação e enriquecimentos). Essa validação foi operacionalizada por checagens de integridade e reexecuções controladas, confirmando estabilidade dos resultados e possibilidade de auditoria dos passos de transformação.

A **avaliação por cenários representativos** consistiu na execução, pelo pesquisador, de recortes analíticos inspirados em demandas institucionais típicas (por exemplo: caracterização de produção por unidade, evolução temporal, perfis tipológicos e recortes por programa/departamento), verificando a coerência dos indicadores gerados por meio da navegação do agregado até os registros subjacentes na Tela de Produções, com filtros pré-aplicados (*drill-through*), e conferência de contagens e subconjuntos exportados. Assim, a validade de uso foi tratada como **capacidade do artefato de sustentar exploração analítica reprodutível e verificável**, mesmo sem participação direta de gestores.

Os resultados dessa avaliação são apresentados no capítulo seguinte, incluindo a caracterização da base consolidada, análises de qualidade e cobertura, e exemplos de exploração analítica e verificabilidade.

3.12 ASPECTOS ÉTICOS

Este projeto foi submetido ao Comitê de Ética em Pesquisa (CEP) e aprovado sob o CAAE 94775125.7.0000.5154, conforme Parecer nº 8.090.187, de 05 de janeiro de 2026 (Uberaba–MG). A pesquisa utiliza dados institucionais e registros bibliográficos, sem recrutamento, intervenção ou contato direto com participantes, e sem coleta ou tratamento de dados pessoais sensíveis, observando-se os princípios éticos e as diretrizes aplicáveis.

4 RESULTADOS E DISCUSSÃO

4.1 SÍNTESE DESCRITIVA DO CORPUS CONSOLIDADO

Esta seção apresenta uma síntese descritiva da base de dados analítica resultante do processo metodológico definido na Seção 3.5, incluindo composição do universo analisado e estatísticas de cobertura necessárias para interpretar as análises subsequentes.

4.1.1 Universo e cobertura (síntese)

O universo analisado corresponde ao subconjunto de **docentes estatutários** cadastrados no UFTMNet (Sistema Integrado da UFTM), incluindo **ativos e aposentados**. No período de referência definido na Metodologia, a UFTM possuía **1.287 docentes cadastrados** (fonte: PROGEPE), e considera-se ainda um universo de **36.988 discentes** (fonte: DRCA) para contextualização de indicadores de participação discente.

Registros sem vinculação departamental identificável (**107**) foram preservados e aparecem como categoria *não identificado* nas análises organizacionais.

4.1.2 A base de dados Lattes analisada

Considerando o período de referência estabelecido na Metodologia, foram selecionados **882 currículos Lattes** de docentes estatutários, que constituem o universo docente analisado.

A base abrange os **sete institutos** da UFTM (incluindo o **CEFORES**), e as produções cadastradas no Lattes durante o vínculo institucional cobrem o período representado pelos registros disponíveis, compreendendo o intervalo **1971–2026**.

4.2 CONSOLIDAÇÃO E DEDUPLICAÇÃO DAS PRODUÇÕES

Esta seção apresenta evidências empíricas sobre a consolidação das produções extraídas do Lattes, motivando as escolhas de deduplicação adotadas no recorte institucional.

4.2.1 Unidade de análise: declaração versus produção consolidada

No Lattes, uma produção é inicialmente observada como uma **declaração individual** associada a um currículo. Assim, produções coautoradas por diferentes docentes podem aparecer como múltiplas declarações redundantes no nível institucional. Para evitar inflação em indicadores agregados, as declarações são agrupadas em **grupos de produção**, permitindo manter duas visões complementares: (i) a visão *declarativa* (por indivíduo) e (ii) a visão *institucional consolidada* (por grupo). Os critérios de agrupamento e de definição da versão principal são descritos detalhadamente no Apêndice C.

4.2.2 Identificadores persistentes e cobertura de DOI

A consolidação prioriza identificadores persistentes quando disponíveis. No recorte de vínculo institucional, a presença de DOI nas declarações é de **7,17%**, percentual que se eleva para **8,58%** após enriquecimento via Crossref.

Na Tabela 1, detalha-se o efeito do enriquecimento de DOI em dois escopos analíticos distintos. No conjunto total de produções (que inclui, por exemplo, orientações, patentes, artigos, traduções e outros tipos declarados no Lattes), o enriquecimento resulta em ganho modesto de cobertura. Em contraste, no recorte das produções ALC (artigos, livros e capítulos), o impacto do enriquecimento é substancialmente maior, refletindo a maior padronização bibliográfica desse tipo de produção.

Tabela 1 – Cobertura de DOI antes e após enriquecimento

Escopo	<i>n</i> total	<i>n</i> DOI bruto	% bruto	<i>n</i> DOI enriq.	% enriq.
Global	169.657	12.160	7,17%	14.556	8,58%
ALC	16.276	8.436	51,83%	10.109	62,11%

Fonte: Elaboração própria.

Nota: DOI = *Digital Object Identifier*; ALC = artigos, livros e capítulos (núcleo bibliográfico).

No núcleo bibliográfico, o DOI apresenta alta incidência e ganho expressivo após enriquecimento, confirmando seu papel como critério robusto de vinculação entre declarações consolidadas. O contraste com o universo total do Lattes reflete a heterogeneidade tipológica da base, não a limitação do identificador no recorte analítico de interesse.

4.2.3 Divergência de ano entre declarações vinculadas por DOI

Mesmo quando o DOI coincide, declarações podem divergir quanto ao campo *ano*. No núcleo bibliográfico, aproximadamente **2,58%** ($n = 360$ de 13.949) das declarações

com DOI apresentam divergência de ano, indicando que inconsistências temporais existem, mas afetam uma fração restrita dos registros e não comprometem o uso do identificador como critério de vinculação.

4.2.4 Coautoria interna: grupos com múltiplos declarantes

No recorte de vínculo institucional, **16,93%** ($n = 2.767$ de 16.344) dos grupos do núcleo bibliográfico apresentam dois ou mais declarantes, confirmando a presença significativa de redundância declarativa associada à coautoria interna e justificando a consolidação das produções.

4.2.5 Síntese

Os achados apresentados nas subseções anteriores podem ser sintetizados em três pontos principais, relativos à cobertura de identificadores, à consistência temporal e à estrutura de coautoria interna:

- a) O DOI apresenta baixa cobertura no universo total, mas alta incidência no núcleo bibliográfico e ganho adicional após enriquecimento;
- b) Divergências de ano entre declarações vinculadas por DOI existem, porém são limitadas;
- c) A presença de múltiplos declarantes é significativamente maior no núcleo, reforçando a necessidade de consolidar coautorias internas.

Esses achados fundamentam a interpretação do efeito global da deduplicação na base consolidada e preparam a análise dos perfis temporal, tipológico e organizacional da produção institucional.

4.2.5.1 Rastreabilidade (cadeia de confiança).

À luz desses achados, quando disponível, o DOI funciona como âncora persistente que conecta declarações redundantes ao mesmo objeto bibliográfico, tornando a consolidação e a deduplicação auditáveis: do indicador agregado no painel é possível retornar ao grupo e às declarações de origem. Essa rastreabilidade é coerente com a ênfase em identificadores persistentes e metadados/proveniência para suportar localização, reúso e verificabilidade (Wilkinson *et al.*, 2016; Mitchell *et al.*, 2022).

4.3 PERFIL INSTITUCIONAL DA PRODUÇÃO NO PERÍODO DE VÍNCULO

Esta seção descreve o perfil institucional da produção **no período de vínculo** a partir de três perspectivas complementares: (i) distribuição temporal, (ii) composição tipológica e (iii) distribuição organizacional. Os resultados utilizam a base consolidada apresentada na Seção 4.1 e a lógica de consolidação discutida na Seção 4.2.

4.3.1 Auditoria de consistência do recorte (controle de universo)

Antes de detalhar distribuições e comparações, realizou-se uma verificação de consistência do universo utilizado nesta seção, de modo a garantir que os totais reportados nas tabelas subsequentes correspondam ao mesmo recorte analítico (*período de vínculo* e base consolidada).

No recorte adotado, há **169.657 declarações** classificadas como ocorridas no período de vínculo e **150.495 produções deduplicadas** (um registro representativo por grupo). Diferenças residuais podem surgir quando o critério de vínculo é aplicado no nível do grupo (i.e., grupos com ao menos uma declaração no vínculo), pois a classificação de vínculo é calculada por declaração, enquanto a deduplicação seleciona uma versão representativa por grupo.

4.3.1.1 Discussão: efeito global da deduplicação.

Do ponto de vista interpretativo, a diferença quantitativa observada entre **169.657** declarações brutas e a base consolidada de **150.495** produções deduplicadas evidencia empiricamente a fragmentação e redundância típicas de ecossistemas informacionais distribuídos, nos quais dados se dispersam em silos e passam a coexistir em versões potencialmente conflitantes (Maltese, 2018). Nesse contexto, a taxa global de deduplicação (**11,29%**) no período de vínculo institucional não deve ser interpretada como perda de informação, mas como a consolidação de uma base institucional de análise, na qual métodos de resolução de entidades são necessários para vincular registros duplicados e evitar a inflação artificial de indicadores (Digiampietri *et al.*, 2015).

4.3.2 Distribuição temporal e marcos interpretativos

A série temporal da produção deduplicada no vínculo abrange **1971–2025**. A escolha de **2006** como marco divisório reflete o contexto institucional: a Universidade Federal do Triângulo Mineiro (UFTM) foi criada em **2005** (Lei nº 11.152, de 29 de julho de 2005), a partir da federalização das antigas Faculdades Integradas do Triângulo Mineiro (FITM). Assim, 2006 representa o primeiro ano completo após a federalização, permitindo distinguir um período anterior de menor densidade (1971–2005, com média de **359,7 produções/ano**) de um período subsequente de consolidação como universidade federal.

Na Tabela 2, apresenta-se a evolução desde 1971, com o período pré-UFTM (1971–2005) agregado e os anos 2006–2025 detalhados individualmente. Observa-se crescimento contínuo de 2006 (2.308 produções) até o pico em **2019** (9.744 produções), seguido de redução acentuada em **2020** (6.569 produções, **-32,59%** em relação ao ano anterior). A queda observada em 2020 ocorre no mesmo período do início da pandemia de COVID-19, devendo ser interpretada como um contexto externo potencialmente associado a alterações na dinâmica de registro da produção acadêmica. Após 2020, verifica-se estabilização entre 2021–2024. O ano de 2025 apresenta volume menor, compatível com efeito de consolidação e atualização tardia de currículos, devendo ser interpretado com cautela por ainda poder receber atualizações. A série anual é apresentada até 2025 por 2026 ser um ano parcial no período de extração.

Quanto ao núcleo bibliográfico (ALC), sua participação relativa tende a crescer no período recente: enquanto os recortes 1971–2005 e 2006–2019 apresentam aproximadamente **10,53%** e **9,06%** de ALC, o período 2020–2025 atinge **14,64%**. No recorte anual (Tabela 2), o percentual de ALC varia entre **9,00%** e **18,79%**, reforçando o crescimento da produção bibliográfica e a necessidade de cautela ao interpretar anos mais recentes.

Tabela 2 – Evolução temporal da produção deduplicada no vínculo (1971–2025)

Ano/Período	Produções dedup.	Produções ALC	Docentes no ano
1971–2005	12.588	1.325	54–129
2006	2.308	164	176
2007	2.646	167	178
2008	3.116	209	192
2009	4.574	291	269
2010	6.678	484	375
2011	7.087	540	398
2012	6.350	635	408
2013	8.097	714	515
2014	8.657	685	536
2015	8.142	839	567
2016	8.157	734	562
2017	9.270	1.048	618
2018	9.480	1.020	623
2019	9.744	1.012	630
2020	6.569	1.115	630
2021	8.413	1.185	628
2022	7.521	1.107	632
2023	8.247	1.037	631
2024	7.789	994	629
2025	4.961	932	645

Fonte: Elaboração própria.

Nota: ALC = artigos, livros e capítulos (núcleo bibliográfico). Período 1971–2005: valores agregados (total de produções, docentes em range min–max). A partir de 2006: valores anuais. “Docentes no ano” corresponde ao número de docentes estatutários com vínculo cobrindo o ano.

4.3.3 Composição tipológica da produção

A produção institucional no vínculo apresenta composição tipológica diversificada, refletindo a pluralidade de atividades acadêmicas registradas no Lattes. Essa diversidade implica que leituras baseadas apenas em volume total tendem a capturar atividades acadêmicas amplas, enquanto subconjuntos específicos são mais adequados a análises bibliométricas e comparativas. Na Tabela 3, são apresentados os tipos mais frequentes após deduplicação.

Tabela 3 – Principais tipos de produção após deduplicação — dez categorias mais frequentes no período de vínculo

Tipo	Qtd. dedup.
Trabalho em Eventos	20.037
Outras Orientações Concluídas	19.155
Apresentação de Trabalho	16.854
Artigo Publicado	12.802
Participação em Banca de Graduação	11.196
Participação em Congresso	7.137
Outras Participações em Eventos/Congressos	6.148
Participação em Banca de Mestrado	6.064
Participação em Banca de Exame de Qualificação	5.378
Organização de Evento	4.932

Fonte: Elaboração própria.

Em nível agregado, uma leitura mais interpretável é obtida ao agrupar os tipos do Lattes em *categorias de interesse* (Tabela 4). Observa-se predominância de registros associados a atividades relacionadas a eventos, seguidas por bancas e orientações, que em conjunto respondem por **mais de dois terços** do volume total deduplicado no vínculo.

Nesse contexto, observa-se que a produção bibliográfica estrita (artigos, capítulos e livros) representa parcela minoritária do volume total, o que reflete a natureza heterogênea das atividades acadêmicas registradas no Lattes. Ainda assim, esse subconjunto apresenta maior padronização descritiva e uso de identificadores persistentes, sendo adotado como referência analítica nas análises de integração e enriquecimento bibliométrico desenvolvidas nas seções subsequentes.

Por categoria nativa do Lattes, observam-se como mais frequentes as atividades complementares (**39,00%**), a produção bibliográfica (**24,91%**), a produção técnica (**20,49%**) e as orientações concluídas (**15,16%**).

Tabela 4 – Distribuição por categorias de interesse (agrupamento semântico de tipos) — deduplicada no vínculo

Categoria de interesse	Qtd. dedup.	%
Participações em eventos	43.697	29,04
Bancas	34.462	22,90
Orientações	21.443	14,25
Trabalhos em eventos	20.037	13,31
Artigos	13.073	8,69
Produção técnica	9.051	6,01
Outras	3.940	2,62
Capítulos	2.793	1,86
Textos bibliográficos	906	0,60
Livros	681	0,45
Total	150.495	100,00

Fonte: Elaboração própria.

Nota: Categorias obtidas por mapeamento dos tipos do Lattes para agrupamentos analíticos (categoria_interesse, Apêndice D).

4.3.4 Distribuição organizacional

A distribuição organizacional foi calculada em modo de *participação* (Apêndice D), considerando-se que uma unidade é associada a uma produção quando ao menos um docente com vínculo institucional válido participa do respectivo registro. Uma produção deduplicada pode ser contabilizada em mais de uma unidade (instituto, departamento ou PPG) quando houver coautoria interna entre unidades distintas, sem duplicar a contagem dentro da mesma unidade. No período de vínculo, observou-se cobertura de **7 institutos**, **43 departamentos** e **17 programas de pós-graduação** com ao menos uma participação em produções deduplicadas.

É importante ressaltar uma implicação interpretativa desse procedimento: os totais reportados por unidade organizacional representam participações em produções deduplicadas e, portanto, não devem ser somados diretamente para se obter o total institucional. Em casos de coautoria entre unidades distintas, uma mesma produção pode ser contabilizada em mais de uma unidade. Assim, o total institucional corresponde ao número de produções deduplicadas distintas, e não à soma dos totais unitários.

Na Tabela 5, apresenta-se o volume absoluto e relativo por instituto, separando-se o total de participações, o recorte de trabalhos classificados como produção (Apêndice D) e o núcleo ALC. O **ICS** concentra **43,01%** das participações totais, seguido por **IELACHS** (**18,65%**) e pelos institutos **ICENE**, **ICTE** e **ICBN**.

4.3.4.1 Comparação de ranking (total *versus* ALC)

Na Tabela 6, compara-se o ranking de institutos por volume total de participações e por participações em ALC. A diferença de ranking sugere que escala total e densidade

Tabela 5 – Volume absoluto e relativo por instituto (participações em produções deduplicadas)

Sigla	Total	%	Produção	%	ALC	%
ICS	63.255	43,01	38.661	43,08	6.591	40,93
IELACHS	27.429	18,65	16.508	18,39	2.848	17,69
ICENE	17.880	12,16	11.116	12,39	1.788	11,10
ICTE	15.190	10,33	8.681	9,67	1.842	11,44
ICBN	15.137	10,29	9.815	10,94	2.192	13,61
CEFORES	4.485	3,05	2.518	2,81	306	1,90
ICAEBI	3.680	2,50	2.444	2,72	535	3,32

Fonte: Elaboração própria.

Nota: As colunas referem-se a participações: uma mesma produção deduplicada pode contribuir para mais de um instituto quando houver coautoria interinstitucional. ALC = artigos, livros e capítulos (núcleo bibliográfico). **Siglas:** ICS = Instituto de Ciências da Saúde; IELACHS = Instituto de Educação, Letras, Artes, Ciências Humanas e Sociais; ICENE = Instituto de Ciências Exatas, Naturais e Educação; ICTE = Instituto de Ciências Tecnológicas e Exatas; ICBN = Instituto de Ciências Biológicas e Naturais; CEFORES = Centro de Educação Profissional; ICAEBI = Instituto de Ciências Agrárias, Exatas e Biológicas de Iturama.

bibliográfica não são equivalentes: o **ICBN** apresenta maior peso relativo no núcleo ALC (3º em ALC, 5º no total), enquanto o **ICENE** reduz sua posição relativa no ALC (5º em ALC, 3º no total).

Tabela 6 – Comparação de rankings por instituto (total *versus* ALC)

Sigla	Total	Rank	Produção	Rank	ALC	Rank
ICS	63.255	1	38.661	1	6.591	1
IELACHS	27.429	2	16.508	2	2.848	2
ICENE	17.880	3	11.116	3	1.788	5
ICTE	15.190	4	8.681	5	1.842	4
ICBN	15.137	5	9.815	4	2.192	3
CEFORES	4.485	6	2.518	6	306	7
ICAEBI	3.680	7	2.444	7	535	6

Fonte: Elaboração própria.

Nota: ALC = artigos, livros e capítulos (núcleo bibliográfico). **Siglas:** ICS = Instituto de Ciências da Saúde; IELACHS = Instituto de Educação, Letras, Artes, Ciências Humanas e Sociais; ICENE = Instituto de Ciências Exatas, Naturais e Educação; ICTE = Instituto de Ciências Tecnológicas e Exatas; ICBN = Instituto de Ciências Biológicas e Naturais; CEFORES = Centro de Educação Profissional; ICAEBI = Instituto de Ciências Agrárias, Exatas e Biológicas de Iturama.

4.3.4.2 Coautoria interinstitucional e lacunas de vinculação

Como aproximação de coautoria interna entre unidades, identificou-se que **4,64%** das produções com instituto identificado possuem participação de **2 ou mais institutos** (6.495 de 140.029 produções com instituto mapeado), evidenciando que a maior parte da produção deduplicada é intraunidade no nível de instituto.

Apesar da elevada cobertura, **6,95%** das produções deduplicadas no vínculo (**10.466** de 150.495) não possuem instituto identificado pelo procedimento de participação, constituindo uma lacuna organizacional relevante para interpretação das comparações entre unidades. Essa lacuna é retomada na discussão de qualidade e cobertura (Seção 4.4).

4.3.4.3 Departamentos e programas de pós-graduação

No nível departamental, a cobertura por instituto varia de **12** departamentos (ICS) a **1** (CEFORES), com **43 departamentos** no total apresentando ao menos uma participação em produções deduplicadas.

No nível de pós-graduação, observam-se **17 PPGs** com participação no período (Tabela 7). A vinculação considera a associação do docente ao programa no período analisado, conforme descrito na Seção 3.9 (Capítulo 3), não implicando necessariamente que a produção tenha sido desenvolvida no âmbito formal do respectivo PPG.

Os programas com maior volume de participações são **PPGAS** (13.709), **PPGCF** (12.979), **PPGE** (12.241) e **PPGCS** (12.142). A proporção de participações em ALC varia entre **3,33%** (Profmato) e **17,58%** (PPGMQ-MG), evidenciando perfis produtivos distintos entre programas.

4.3.4.4 Síntese do perfil institucional

Em conjunto, os resultados indicam: (i) crescimento temporal com marcos interpretativos (2006 como início do período pós-federalização, pico em 2019 e queda em 2020); (ii) composição tipológica plural, na qual atividades não bibliográficas dominam o volume total; e (iii) heterogeneidade organizacional relevante, com diferenças de escala entre unidades e variações no peso relativo do subconjunto bibliográfico. Essas características orientam a análise de qualidade e cobertura do recorte (Seção 4.4) e o aprofundamento bibliométrico em subconjuntos mais padronizados (Seção 4.5).

Tabela 7 – Participações em produções deduplicadas por programa de pós-graduação (período de vínculo)

Programa	Total	Produção	ALC	%ALC
PPGAS	13.709	8.789	1.629	11,88
PPGCF	12.979	7.967	1.936	14,92
PPGE	12.241	7.219	1.416	11,57
PPGCS	12.142	7.980	1.936	15,94
PPGEF	9.411	6.111	1.484	15,77
PPGP	8.977	5.720	1.190	13,26
PMPIT	8.443	4.916	1.124	13,31
PPGFISIO	8.169	4.924	811	9,93
PPGMTI	8.056	4.756	1.238	15,37
PPGCTA	6.231	3.733	864	13,87
PPGCTM	5.888	3.566	698	11,85
Profiap	5.204	3.497	589	11,32
Profletras	4.868	2.860	465	9,55
PPGECM	4.401	2.684	514	11,68
Profqui	3.997	2.290	277	6,93
PPGMQ-MG	3.970	2.532	698	17,58
Profmat	1.953	995	65	3,33

Fonte: Elaboração própria.

Nota: Contagens em modo participação: uma mesma produção deduplicada pode contribuir para mais de um programa quando houver coautoria interna entre docentes vinculados a PPGs distintos. O recorte por programa considera o vínculo institucional no ano da produção. ALC = artigos, livros e capítulos (núcleo bibliográfico). **Siglas:** PPGAS = Atenção à Saúde; PPGCF = Ciências Fisiológicas; PPGE = Educação; PPGCS = Ciências da Saúde; PPGEF = Educação Física; PPGP = Psicologia; PMPIT = Mestrado Profissional em Inovações e Tecnologias; PPGFISIO = Fisioterapia; PPGMTI = Medicina Tropical e Infectologia; PPGCTA = Ciência e Tecnologia Ambiental; PPGCTM = Ciência e Tecnologia de Materiais; Profiap = Administração Pública em Rede Nacional; Profletras = Letras em Rede Nacional; PPGECM = Educação em Ciências e Matemática; Profqui = Química em Rede Nacional; PPGMQ-MG = Multicêntrico em Química de Minas Gerais; Profmat = Matemática em Rede Nacional.

4.4 QUALIDADE E COBERTURA DOS DADOS

Esta seção caracteriza limitações e forças do recorte consolidado, fornecendo contexto metodológico prévio às análises comparativas e ao enriquecimento bibliométrico. São examinadas três dimensões: (i) completude de identificadores persistentes (ênfase em DOI), (ii) redundância declarativa como indício operacional de coautoria interna, e (iii) cobertura organizacional e atualidade dos currículos.

4.4.1 Completude de metadados e capacidade de integração

A disponibilidade de DOI condiciona (i) a integração com bases externas e (ii) a própria confiabilidade de vínculos entre registros. Neste recorte, distingue-se *DOI declarado* (informado no Lattes, após normalização de formato) e *DOI final* (identificador utilizado para integração, que prioriza o DOI declarado e, quando ausente, incorpora o DOI recuperado por integração externa e igualmente normalizado).

Na Tabela 8, sintetiza-se a completude em ambos os recortes.

Tabela 8 – Completude de DOI na produção institucional (vínculo, deduplicada)

Recorte	DOI declarado (%)	DOI final (%)
Produção global (vínculo, dedup)	5,82	6,99
Núcleo ALC (vínculo, dedup)	51,83	62,11

Nota: DOI = *Digital Object Identifier*; ALC = artigos, livros e capítulos (núcleo bibliográfico). DOI declarado = informado no Lattes (normalizado); DOI final = DOI declarado quando presente ou, na sua ausência, DOI recuperado externamente.

Observa-se que, no recorte deduplicado do período de vínculo, a completude de DOI é baixa na produção global (5,82% declarado; 6,99% final), refletindo a heterogeneidade tipológica do Lattes (orientações, bancas, apresentações e outras atividades que tendem a não possuir DOI). Em contraste, no núcleo bibliográfico (ALC), a completude é substancialmente maior (51,83% declarado; 62,11% final), indicando ganho relevante do enriquecimento no subconjunto com maior potencial bibliométrico.

Do ponto de vista de interoperabilidade, o DOI funciona como identificador persistente e globalmente único, viabilizando recuperação automática e pareamentos rastreáveis entre bases, em conformidade com os princípios FAIR (Wilkinson *et al.*, 2016). Operacionalmente, ele também atua como *pista* para reduzir ambiguidade e restringir o espaço de busca em tarefas de desambiguação e integração entre bases heterogêneas (Mazoni *et al.*, 2025).

Por isso, as rotinas de integração externa concentram-se no ALC, onde a presença de DOI é substancialmente maior, enquanto a leitura institucional ampla permanece dependente do Lattes e de bases administrativas.

4.4.1.1 Discussão: enriquecimento de DOI e viés de cobertura.

O incremento na cobertura de DOI no núcleo bibliográfico (ALC), de **51,83%** para **62,11%** após o enriquecimento, é consistente com a premissa de que bases locais podem ser utilizadas para qualificar a integração com bases globais, ao recuperar identificadores ausentes e melhorar a consistência dos metadados (Mazoni *et al.*, 2025). Além disso, ao adotar o Crossref como mecanismo de *fallback*, torna-se possível reduzir a dependência exclusiva de identificadores previamente declarados, mitigando vieses de cobertura associados a estratégias de pareamento baseadas estritamente em DOI (Culbert *et al.*, 2025).

4.4.2 Cobertura de coautoria e redundância declarativa

A existência de múltiplas declarações para o mesmo grupo consolidado é esperada e funciona como um indício de participação compartilhada de docentes da UFTM em um mesmo registro do Lattes. No período de vínculo, **2,27%** dos grupos consolidados possuem **2 ou mais declarantes**. No núcleo ALC, esse percentual sobe para **20,30%**.

Esse contraste sugere que a redundância declarativa está mais associada à produção bibliográfica formal, na qual a coautoria é mais frequente e mais facilmente identificável, do que ao conjunto heterogêneo de registros do Lattes (Tabela 9).

Tabela 9 – Grupos consolidados com múltiplos declarantes (coautoria interna)

Recorte	Grupos com 2+ declarantes (%)
Produção global (vínculo, dedup)	2,27
Núcleo ALC (vínculo, dedup)	20,30

Nota: ALC = artigos, livros e capítulos (núcleo bibliográfico). Grupos = produções deduplicadas; declarantes = docentes que reportaram o item no Lattes.

Na literatura, é evidenciado que metadados curriculares/bibliográficos permitem análises de coautoria (Mena-Chalco *et al.*, 2014). Neste trabalho, a métrica é utilizada para caracterizar a redundância declarativa associada à produção bibliográfica formal e sinalizar o potencial de análises relacionais futuras, como indicadores de colaboração.

4.4.3 Lacunas de identificação institucional e cobertura organizacional

A associação de cada produção a instituto/departamento no ano correspondente depende de informação organizacional nas bases administrativas. No recorte deduplicado do período de vínculo, **6,95%** das produções (**10.466** de **150.495**) não puderam ser associadas a instituto, sendo classificadas na categoria “não identificado”.

Entre as produções “não identificadas”, **91,99%** possuem informante estatutário no ano da produção, indicando que a lacuna não decorre, em geral, de ausência de vínculo temporal. O diagnóstico aponta como causa predominante a ausência de informação

departamental no registro funcional no ano correspondente, responsável por **100%** dos casos identificados nessa etapa. Na Tabela 10, sintetiza-se essa lacuna organizacional.

Tabela 10 – Produção classificada como “não identificado” (lacunas organizacionais)

Categoria	Produções dedup	% do total
Não identificado (sem instituto)	10.466	6,95
Total geral (vínculo, dedup)	150.495	100,00

Nota: Percentual calculado sobre a base deduplicada no período de vínculo institucional.

A fração de produções classificadas como “não identificado” explicita uma zona residual de cobertura organizacional, típica de ambientes em que registros administrativos foram produzidos e mantidos ao longo de diferentes ciclos de sistemas e regras, o que dificulta a consistência histórica de metadados institucionais (Maltese, 2018). Em termos metodológicos, a opção por preservar essa categoria, em vez de imputar vínculos por heurísticas não verificáveis, favorece a transparência do indicador e delimita com precisão o subconjunto que ainda demanda curadoria administrativa. Esse princípio é coerente com abordagens orientadas à proveniência, nas quais a confiança nos resultados depende de explicitar origens, transformações e limitações do dado ao longo do processo (Mitchell *et al.*, 2022).

4.4.4 Atualidade do Lattes e efeitos de consolidação recente

A atualização do Lattes é descentralizada, o que produz defasagens heterogêneas. Entre docentes estatutários com vínculo ativo no momento da extração, **61,62%** atualizaram o currículo nos últimos 6 meses e **90,04%** nos últimos 12 meses; **5,17%** apresentam defasagem superior a 24 meses.

Entre docentes que não apresentavam vínculo ativo no momento da extração, a defasagem é maior: **49,78%** atualizaram em até 12 meses e **42,29%** estão há mais de 24 meses sem atualização. As estatísticas resumidas confirmam uma defasagem substancialmente maior nesse grupo. Docentes ativos apresentam média de **8,27** meses (mediana **3**, p90 **11**), enquanto docentes não ativos apresentam média de **48,98** meses (mediana **13**).

Quanto à base local utilizada na análise, a mediana de dias desde a coleta foi **49** (média **31**), caracterizando o recorte como recente do ponto de vista do repositório local, embora a atualização individual permaneça variável.

Os resultados evidenciam que a atualidade do Lattes é heterogênea e depende do engajamento do pesquisador, com perfis que atualizam com alta frequência e outros que permanecem longos períodos sem revisão. A literatura observa essa assimetria e destaca que mesmo currículos atualizados recentemente podem conter registros desatualizados, o que afeta a confiabilidade temporal de análises baseadas no registro declarativo (Di-giampietri *et al.*, 2015). Assim, variações de volume nos anos mais recentes devem ser

interpretadas com cautela, pois parte da produção pode ainda não ter sido registrada no currículo no momento da *extração*, especialmente quando o recorte envolve docentes com baixa frequência de atualização. Nesse sentido, a Plataforma Capivara atua também como instrumento de diagnóstico da qualidade e atualidade do registro, sem inferir inatividade real, mas evidenciando atrasos de documentação que podem orientar ações institucionais de estímulo à manutenção dos perfis.

4.4.5 Síntese de qualidade e cobertura

Em síntese, observam-se três aspectos centrais: (i) o DOI é raro na produção global, mas tem completude elevada no núcleo ALC, com ganho adicional via normalização/enriquecimento; (ii) múltiplos declarantes capturam coautoria interna, sobretudo no ALC; e (iii) lacunas organizacionais concentram-se em casos sem departamento recuperável no ano, enquanto a atualidade do Lattes é adequada entre docentes ativos e mais heterogênea entre não ativos. Ressalta-se, ainda, que a ausência de DOI pode ocorrer por limitações editoriais e de indexação, como em publicações locais e regionais ainda não integradas a sistemas internacionais de indexação, bem como em anais de eventos antigos ou extintos, cujo acervo pode estar disponível apenas em meio físico, sem versão digital. Esses aspectos de cobertura e atualidade delimitam as condições de interpretação dos resultados apresentados nas seções seguintes.

4.5 ENRIQUECIMENTO BIBLIOMÉTRICO E INTEGRAÇÃO COM BASES EXTERNAS

Esta seção apresenta os resultados do enriquecimento bibliométrico aplicado ao corpus consolidado no período de vínculo institucional, descrevendo (i) o *ganho incremental de identificadores* utilizáveis para integração, (ii) a *cobertura efetiva* obtida em bases externas e (iii) os *atributos bibliométricos e contextuais* que passaram a estar disponíveis para análises comparativas posteriores.

4.5.1 Escopo do enriquecimento e recortes de análise

O enriquecimento foi executado sobre as produções deduplicadas no período de vínculo institucional, com ênfase analítica no subconjunto **ALC** (artigos publicados, livros publicados ou organizados e capítulos de livro publicados), por representar um *recorte operacional* de produção potencialmente indexável em bases bibliográficas externas. Por construção, atividades de natureza estruturalmente pouco indexável (como bancas, orientações e outras atividades complementares) tendem a apresentar baixa cobertura fora do Lattes e, portanto, são tratadas principalmente como dimensão institucional e tipológica nas seções anteriores.

A cobertura em bases externas é reportada sempre **condicionada à existência de um identificador de integração** (principalmente o DOI), pois esse é o mecanismo que reduz ambiguidade e viabiliza pareamentos reprodutíveis entre bases.

4.5.2 Fontes externas e critérios de priorização

O sistema integra múltiplas bases externas, cada uma priorizada conforme o tipo de atributo bibliométrico ou contextual fornecido e sua viabilidade operacional no contexto institucional do estudo:

- a) **OpenAlex**: fonte primária de enriquecimento bibliométrico e temático. Utilizada para classificação por conceitos, domínios e áreas do conhecimento, identificação de padrões de colaboração (instituições e países), métricas de citação (totais, séries temporais e percentis normalizados) e como ponto de articulação para descoberta de identificadores externos (por exemplo, Scopus e ORCID), quando disponíveis.
- b) **OpenCitations**: fonte dedicada ao grafo aberto de citações, empregada para obtenção de contagens agregadas de citações por DOI. Atua como camada bibliométrica complementar e independente, favorecendo análises reprodutíveis desacopladas de bases proprietárias.
- c) **Scopus**: fonte bibliográfica de cobertura seletiva, conforme critérios editoriais. Quando acessível via API, é utilizada para recuperação de metadados do registro

indexado e métricas associadas, como contagem de citações e índice h, respeitando limitações de acesso e cobertura.

- d) **Crossref**: infraestrutura de registro de DOI, utilizada principalmente como *fall-back* para recuperação, validação e normalização de identificadores e metadados quando ausentes ou inconsistentes no Lattes. Não é tratada como fonte primária de indicadores bibliométricos.
- e) **Google Scholar**: utilizado de forma conservadora como sinal complementar de cobertura bibliográfica e métricas agregadas (por exemplo, citações e índice h), em razão da heterogeneidade dos registros, ausência de API pública e limitações de reprodutibilidade.

Esses critérios não implicam hierarquia absoluta de qualidade entre as fontes, mas refletem a priorização por tipo de atributo, consistência e escalabilidade no contexto do sistema desenvolvido. Na Tabela 11, apresenta-se uma síntese do papel atribuído a cada base no processo de enriquecimento.

Tabela 11 – Síntese do papel das bases externas no enriquecimento bibliométrico

Fonte	Atributos incorporados / papel principal
OpenAlex	Classificação temática, colaboração institucional e internacional, métricas de citação e articulação com identificadores externos.
OpenCitations	Contagem agregada de citações por DOI (grafo de citações aberto).
Scopus	Metadados e métricas bibliográficas sob cobertura seletiva.
Crossref	Recuperação, validação e normalização de DOI e metadados.
Google Scholar	Sinal complementar de cobertura e métricas agregadas.

Nota: DOI = *Digital Object Identifier*.

4.5.3 Identificadores de integração: DOI declarado *versus* DOI final

Um resultado central do processo de integração é a distinção entre o **DOI declarado** no currículo Lattes e o **DOI final** utilizado pelo sistema para pareamento. O DOI declarado corresponde ao identificador informado pelo docente no Lattes, após normalização de formato. O DOI final corresponde ao identificador efetivamente empregado para integração: quando o DOI declarado está presente, ele é utilizado; quando está ausente, o sistema pode recuperar um DOI por integração externa e, em seguida, normalizá-lo.

No recorte global do vínculo, a presença de DOI passa de **5,82%** (DOI declarado) para **6,99%** (DOI final), correspondendo a **1.765** registros adicionais com DOI utilizável para integração.

No núcleo **ALC**, o ganho é mais expressivo: a cobertura passa de **51,83%** (DOI declarado) para **62,11%** (DOI final), com **1.673** registros incorporados ao conjunto com

DOI utilizável. Esse ganho é relevante porque o DOI é o identificador persistente preferencial para integração, reduzindo ambiguidades e falhas de pareamento associadas a variações descritivas.

4.5.3.1 Crossref como mecanismo de *fallback* para recuperação de DOI.

Como estratégia de ampliação de cobertura, o Crossref foi utilizado como mecanismo de *fallback* para recuperar DOI quando o identificador não foi informado no Lattes. Em nível de **declaração** (registros reportados por diferentes docentes, incluindo duplicatas), havia **170.695** declarações no período de vínculo, das quais **157.812 (92,4%)** não continham DOI declarado. Nesse universo, o *fallback* via Crossref recuperou **2.562** DOIs, correspondendo a **1,62%** das declarações sem DOI.

Uma parcela dessas recuperações é **redundante** do ponto de vista do grupo deduplicado, pois reflete lacunas de preenchimento em cenários de coautoria: **708** recuperações (**27,6%**) ocorreram em grupos nos quais **outro declarante já havia informado DOI**. Ainda assim, observa-se **ganho líquido** relevante: **1.854** recuperações (**72,4%**) ocorreram em declarações pertencentes a grupos em que **nenhum declarante havia informado DOI**.

No nível de **grupo deduplicado** (corpus final), dos **150.495** grupos consolidados, **141.473 (94,0%)** não possuíam DOI declarado por nenhum declarante. Após o *fallback*, **1.505** grupos (**1,06%** desse universo) passaram a apresentar DOI final utilizável para integração. A diferença entre as **1.854** recuperações inéditas no nível de declaração e os **1.505** casos incorporados ao corpus deduplicado decorre do processo de seleção de versão principal: nem toda declaração com DOI recuperado é a que representa o grupo no documento deduplicado.

Em termos práticos, o *fallback* via Crossref contribui diretamente para o aumento observado na cobertura de DOI final no corpus deduplicado, viabilizando **1.505** grupos adicionais com DOI utilizável para integração.

4.5.4 Cobertura de integração por base externa

Como identificador persistente, o DOI é o elo mais estável para integração automatizada entre bases; por isso, as análises de cobertura nesta seção são condicionadas à presença de DOI final (Wilkinson *et al.*, 2016).

Na Tabela 12, sintetiza-se a cobertura observada no núcleo ALC **condicionada à existência de DOI final**, reportando a proporção de registros com correspondência válida em cada base externa. Para evitar que a ausência de identificador distorça os resultados, todos os percentuais utilizam como denominador apenas os registros com DOI final não vazio. Observa-se alta cobertura em **OpenAlex (95,77%)** e **OpenCitations (97,03%)**, cobertura intermediária em **Scopus (50,94%)** e presença mais limitada no **Scholar (23,71%)**.

Tabela 12 – Cobertura de integração no núcleo ALC (vínculo) por base externa

Fonte	Cobertura (% do ALC com DOI final)
OpenAlex	95,77%
Scopus	50,94%
OpenCitations	97,03%
Scholar	23,71%

Nota: DOI = *Digital Object Identifier*; ALC = artigos, livros e capítulos (núcleo bibliográfico). A cobertura é calculada sobre o núcleo ALC com DOI final e ano válido. O Crossref não é reportado nesta tabela por atuar predominantemente como *fallback* para recuperação de DOI quando o identificador está ausente, e não como fonte primária de enriquecimento de registros já integráveis.

A diferença de cobertura entre as fontes (por exemplo, OpenAlex com **95,77%** versus Scopus com **50,94%**) reflete distinções de escopo entre os sistemas de indexação. Bases abertas, como OpenAlex e OpenCitations, privilegiam a amplitude de indexação, enquanto bases seletivas, como o Scopus, operam com recortes mais restritivos do universo de publicações com DOI. Esse resultado não indica menor qualidade da produção, mas diferenças nos escopos de indexação e nas políticas de inclusão das bases. A literatura aponta que indicadores baseados em bases seletivas podem sub-representar áreas e contextos específicos, recomendando interpretação cautelosa (Melo; Trinca; Maricato, 2021); assim, a diferença observada aqui tende a refletir critérios de indexação e representatividade, e não um atributo intrínseco da produção analisada. Análises comparativas também indicam maior abrangência em bases abertas como o OpenAlex, em contraste com bases proprietárias de escopo seletivo (Culbert *et al.*, 2025).

O Scholar apresenta cobertura inferior às demais bases, em razão da ausência de API oficial e por depender de informações declaradas e recuperadas de forma não sistemática.

Observação metodológica: Os percentuais são calculados sobre o conjunto ALC com DOI final conforme armazenado na view analítica. Diferenças residuais de contagem podem ocorrer em função do tratamento de valores vazios ou inconsistentes, sendo adotado um critério conservador de exclusão de registros sem DOI final válido nas análises comparativas.

4.5.5 Convergência de citações entre fontes

Na Tabela 13, apresenta-se a sobreposição de citações entre as principais fontes bibliométricas consideradas, restringindo-se a registros com ao menos uma citação (≥ 1). O objetivo não é comparar métricas entre bases, mas avaliar o grau de convergência da informação bibliométrica associada aos DOIs do corpus. Como cada base observa subconjuntos distintos do grafo de referências e adota estratégias próprias de coleta/curadoria, divergências entre fontes são esperadas e não implicam, por si, falha técnica do pipeline (Culbert *et al.*, 2025).

Observa-se que **65,81%** das produções possuem citações em pelo menos uma fonte, e que **53,52%** apresentam convergência em duas ou mais fontes, indicando consistência inter-base. Em particular, **38,60%** das produções concentram-se em padrões de alta convergência (três ou quatro fontes), o que reforça a robustez da camada bibliométrica derivada. Casos de cobertura isolada por uma única fonte representam parcela minoritária (**12,09%**), enquanto **34,19%** dos registros não apresentam citações no período considerado, compatível com produções recentes ou ainda sem impacto mensurável.

A camada bibliométrica incorporada ao corpus armazena, quando disponíveis, as contagens de citações separadamente por fonte (OpenAlex, Scopus e OpenCitations). Para fins de visualização e indicadores que requerem uma métrica única por produção, o sistema adota uma regra determinística de consolidação, utilizando como *citações consolidadas* o maior valor observado entre as fontes disponíveis para o mesmo DOI chave. Essa escolha evita dupla contagem entre bases (que não são independentes) e preserva rastreabilidade, pois o valor consolidado pode ser decomposto nas contagens originais por fonte; conforme o objetivo analítico, podem ser utilizadas tanto as contagens por fonte quanto a contagem consolidada. Como consequência, as contagens por fonte são preservadas para inspeção e análises de sensibilidade quando necessário.

Tabela 13 – Convergência de citações entre fontes bibliométricas (agregado)

Padrão de cobertura	Total	%
<i>Alta convergência (3–4 fontes)</i>		
4 fontes (OC+OA+SC+GS)	1.282	12,20
3 fontes	2.775	26,40
<i>Convergência moderada (2 fontes)</i>		
Principais pares (OC+OA, OA+SC)	1.257	11,96
Outras combinações	335	3,19
<i>Fonte isolada (1 fonte)</i>		
OpenAlex	894	8,51
Scholar	271	2,58
Scopus	97	0,92
OpenCitations	3	0,03
Sem citações em nenhuma fonte	3.593	34,19
Com citações em ≥ 1 fonte	6.915	65,81

Notas: DOI = *Digital Object Identifier*; ALC = artigos, livros e capítulos (núcleo bibliográfico). Base de 10.508 produções ALC com DOI chave. Citações: registros com contagem ≥ 1 na respectiva fonte.

OC = OpenCitations; OA = OpenAlex; SC = Scopus; GS = Google Scholar. Alta convergência (38,60%) indica consenso multi-base e sinaliza robustez bibliométrica. Cobertura isolada (12,04%) reflete especialização ou alcance parcial das fontes.

4.5.6 Ganho informacional: temas e atributos de colaboração (OpenAlex)

Além de métricas, o enriquecimento amplia o corpus com atributos pouco padronizados no Lattes, como **classificações temáticas** e **atributos de afiliação institucional e colaboração**. Esses atributos não estão disponíveis de forma estruturada no Lattes e viabilizam análises temáticas e de colaboração em escala institucional. Entre as produções ALC com match em OpenAlex, **97,55%** possuem lista de **conceitos** (áreas/tópicos) não vazia, viabilizando análises temáticas sistemáticas. Ainda nesse subconjunto, **81,63%** apresentam informações de **country_code** associadas a instituições nas autorias, o que habilita análises de colaboração internacional baseadas em afiliação quando o dado está presente.

4.5.7 Classificação nacional: Qualis (periódicos) e escopo de aplicação

Além das fontes internacionais, o conjunto de dados foi enriquecido com a **classificação Qualis dos periódicos**, atribuída por meio do pareamento do veículo de publicação via **ISSN** com a tabela de estratos **correspondente ao quadriênio avaliativo 2017–2020**. Cabe destacar que, por se tratar do **Qualis de periódicos**, essa classificação se aplica **principalmente ao subconjunto de artigos**, não sendo pertinente, nos mesmos termos, a livros e capítulos presentes no ALC. No conjunto consolidado, **12.321 registros de artigos** receberam classificação Qualis com base no **quadriênio 2017–2020**.

A distribuição por estrato (A1–C e “Sem Qualis”) é utilizada como *camada adicional* de caracterização do perfil de publicação sob referencial nacional.

4.5.8 Atualização, latência e completude do enriquecimento

Os resultados de cobertura e de disponibilidade de métricas dependem das janelas de coleta e atualização de cada fonte utilizada. No âmbito deste estudo, as datas mais recentes observadas para coleta ou atualização dos dados foram: **OpenAlex** (31 de dezembro de 2025), **Scopus** (13 de janeiro de 2026), **Crossref** (29 de dezembro de 2025), **OpenCitations** (16 de janeiro de 2026) e **Scholar** (29 de dezembro de 2025). Essas datas indicam que o enriquecimento bibliométrico reflete um retrato recente e consistente entre fontes, ainda que não perfeitamente sincronizado.

4.5.9 Limitações e vieses de cobertura

Ressalta-se que a integração com bases externas apresenta vieses estruturais: (i) o DOI favorece a cobertura de artigos e parte da produção bibliográfica, enquanto livros e capítulos podem ter menor padronização e presença em bases externas; (ii) fontes diferem em escopo, granularidade e política de atualização, o que afeta cobertura observada; e (iii) a ausência de match em determinada base não implica ausência de relevância, mas

sem limitações de indexabilidade, identificadores ou acesso.

4.6 ANÁLISES COMPARATIVAS E PADRÕES EMERGENTES

Esta seção apresenta análises comparativas que articulam as dimensões temporal, organizacional, tipológica e bibliométrica do corpus consolidado, com o objetivo de identificar e interpretar padrões emergentes na produção científica institucional. A análise é organizada em três subseções complementares: **(4.6.1)** Padrões temporais e organizacionais, que examina trajetórias institucionais e variações por unidade; **(4.6.2)** Padrões por tipologia e colaboração, que caracteriza perfis de comunicação científica e engajamento discente; e **(4.6.3)** Padrões com enriquecimento bibliométrico, que explora relações entre colaboração internacional e impacto científico. A seção se encerra com **(4.6.4)** Discussão dos padrões observados, que sintetiza os achados e explicita suas limitações interpretativas.

4.6.1 Padrões temporais e organizacionais

4.6.1.1 Caso âncora: variação recente de artigos publicados no ICS (2016–2025).

Como exemplo verificável de padrão temporal e organizacional, analisa-se a série de **artigos publicados no Instituto de Ciências da Saúde (ICS)** ao longo de 2016–2025. O objetivo é ilustrar como o artefato permite identificar, decompor e interpretar variações institucionais específicas, sem pretensão de generalização automática para outros institutos. A leitura é feita em três níveis complementares: (i) evolução do volume anual de artigos, (ii) normalização pelo número de docentes ativos no instituto no ano, e (iii) desagregação por departamentos.

Na Figura 1, apresenta-se a série temporal de artigos publicados no ICS entre 2016 e 2025.

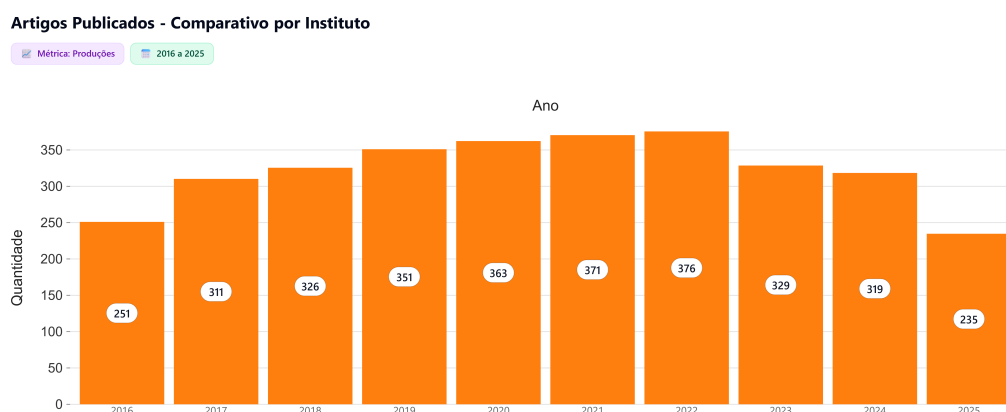


Figura 1 – Série temporal de artigos publicados no ICS (2016–2025).

Fonte: Elaboração própria.

Os resultados indicam crescimento até um patamar entre 2020–2022, seguido por redução em 2023–2025. Embora 2025 deva ser interpretado com cautela (anos recentes

são mais sensíveis a efeitos de atraso e janela de observação (Belter, 2015)), observa-se que a queda já se manifesta entre 2023 e 2024, sugerindo que o padrão não se explica apenas por defasagem do ano mais recente.

Para distinguir queda de volume absoluto de queda de produtividade média, normalizou-se a série pelo número de **docentes ativos** do ICS em cada ano, conforme ilustrado na Figura 2.²

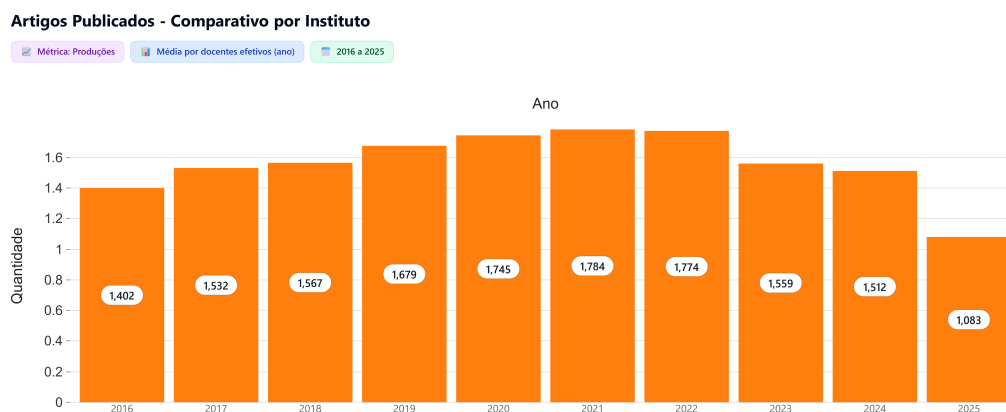


Figura 2 – Docentes produtores e artigos por docente no ICS (2016–2025).

Fonte: Elaboração própria.

Essa razão capta a produtividade institucional média, incluindo docentes que não publicaram no período.

Em particular, após o patamar observado entre 2020 e 2022 (**1,745** em 2020; 1,784 em 2021; **1,774** em 2022), verifica-se **declínio a partir de 2023** (1,559 em 2023; **1,512** em 2024, -14,8% em relação a 2022), indicando queda de produtividade média, e não apenas variação no tamanho do quadro.

4.6.1.2 Decomposição da produtividade: taxa de publicadores e intensidade.

Para interpretar o declínio da quantidade de artigos, separa-se a produtividade em dois componentes: (i) **participação**, isto é, a proporção de docentes que publicaram ao menos um artigo no ano; e (ii) **intensidade**, isto é, quantos artigos, em média, foram publicados por quem publicou.

Entre 2022 e 2024, a produtividade média (artigos por docente ativo) cai de **1,774** para **1,512**. Nesse mesmo intervalo, a proporção de docentes que publicaram ao menos um artigo reduz-se de **62,7%** para **58,3%** (-4,4 pontos percentuais), enquanto a intensidade entre publicadores cai de **2,827** para **2,593** (-8,3%). Em síntese, o padrão é compatível com a hipótese de que a queda recente no volume agregado do instituto resulta primariamente de **menor proporção de docentes publicando**, e não apenas de redução na produtividade dos que permanecem ativos na publicação.

²Considera docentes estatutários com vínculo ativo no instituto no ano.

Para verificar se a redução é concentrada ou difusa internamente, decompôs-se a série de artigos publicados por departamentos do ICS, conforme a Figura 3.

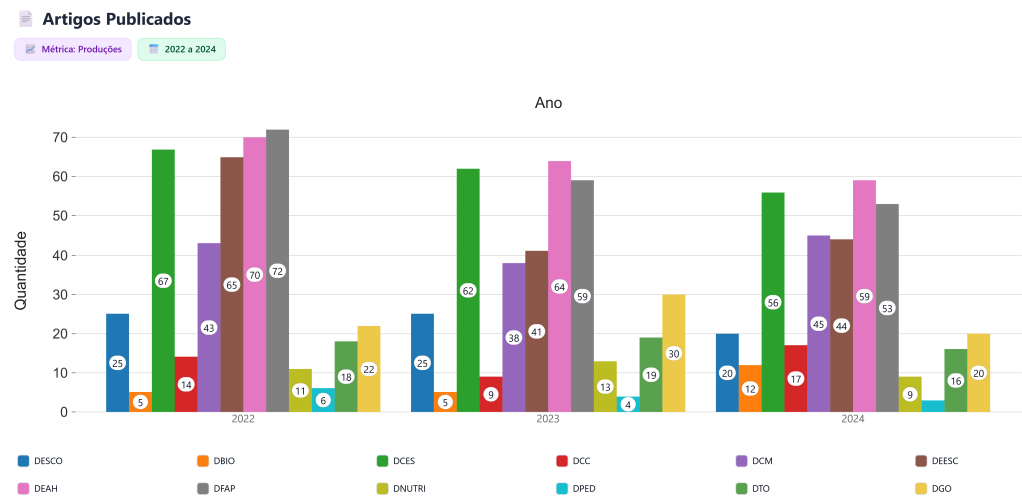


Figura 3 – Série anual de artigos por departamento no ICS.

Fonte: Elaboração própria.

Nota: DESCO = Departamento de Saúde Coletiva; DBIO = Departamento de Biomedicina; DCES = Departamento de Ciências do Esporte; DCC = Departamento de Clínica Cirúrgica; DCM = Departamento de Clínica Médica; DEESC = Departamento de Enfermagem em Educação e Saúde Comunitária; DEAH = Departamento de Enfermagem na Assistência Hospitalar; DFAP = Departamento de Fisioterapia Aplicada; DNUTRI = Departamento de Nutrição; DPED = Departamento de Pediatria; DTO = Departamento de Terapia Ocupacional; DGO = Departamento de Ginecologia e Obstetrícia.

A comparação entre 2022 e 2024 indica que a queda do instituto é **multifatorial**, com reduções expressivas em alguns departamentos e quedas moderadas em outros, acompanhadas de compensações parciais por crescimento em unidades específicas. A série anual por departamento reforça que a queda recente não é homogênea: coexistem departamentos com declínio acentuado após 2022 e departamentos com comportamento mais resiliente, sugerindo que o padrão agregado do instituto resulta de uma combinação de comportamentos departamentais assimétricos.

4.6.1.3 Síntese do achado (ICS).

O caso do ICS evidencia um padrão verificável em três camadas: (i) queda recente do volume anual de artigos publicados a partir de 2022 (de 376 em 2022 para 319 em 2024, -15,2%), (ii) redução da produtividade institucional (artigos/docente ativo) no período recente, passando de 1,774 (2022) para 1,512 (2024, -14,8%), com decomposição indicando primariamente redução na taxa de publicadores (62,7% → 58,3%, -4,4 p.p.), e (iii) heterogeneidade departamental na explicação do agregado, indicando que métricas no nível do instituto podem ocultar variação intrainstitucional.

4.6.1.4 Contraponto institucional: crescimento no ICTE.

Para contextualizar o padrão observado no ICS e verificar se a queda recente em artigos publicados constitui um fenômeno institucional generalizado, examina-se o comportamento do **Instituto de Ciências Tecnológicas e Exatas (ICTE)** no mesmo horizonte. O ICTE apresenta trajetória oposta ao ICS: enquanto o ICS registra queda de **15,2%** no volume (376 → 319 artigos) e de **14,8%** na métrica normalizada (1,774 → 1,512 artigos/docente ativo) entre 2022 e 2024, o ICTE exibe **crescimento de 27,8%** no volume (90 → 115 artigos) e **crescimento de 26,8%** na métrica normalizada (0,698 → 0,885 artigos/docente ativo).³

Em termos de série temporal, o ICTE apresenta volumes relativamente estáveis entre **2019–2021** (117 em 2019; 111 em 2020; 127 em 2021), atinge um **mínimo em 2022** (90 artigos; 0,698 art/doc ativo; taxa de publicadores 42,6%) e, em seguida, apresenta **recuperação em 2023** (116 artigos; 0,899 art/doc ativo; taxa 47,3%) e **manutenção do volume em 2024** (115 artigos; 0,885 art/doc ativo; taxa 36,9%), com o volume sustentado apesar da queda na taxa de publicadores por maior intensidade entre os que publicam.⁴ Esse padrão contrasta diretamente com o platô seguido de queda observado no ICS após 2022, indicando que a recuperação do ICTE não decorre apenas de efeito compositivo (variação no número de docentes), mas de incremento na produtividade entre os docentes que publicam. A decomposição revela que, no ICTE, a recuperação de 2022 para 2024 combina **crescimento na intensidade entre publicadores** (de 1,636 para 2,396 artigos/publicador, +46,5%) com **variação na taxa de publicadores** (42,6% → 47,3% → 36,9%), padrão compatível com um crescimento do volume sustentado principalmente por maior produtividade dos docentes que publicam, mesmo com redução proporcional de participantes em 2024.

4.6.1.5 Panorama interinstitucional da variação 2022–2024.

Para situar o contraste ICS–ICTE no conjunto institucional, apresenta-se o quadro de variação de artigos publicados no biênio 2022–2024 para todos os institutos da UFTM: **ICS**: 376→319 (-15,2%); **ICENE**: 120→105 (-12,5%); **ICBN**: 118→114 (-3,4%); **ICA-EBI**: 38→34 (-10,5%); **IELACHS**: 117→118 (+0,9%); **CEFORES**: 10→14 (+40,0%); **ICTE**: 90→115 (+27,8%). Observa-se que a maioria dos institutos apresenta redução ou estabilidade, com destaque para a estabilidade do IELACHS (+0,9%) e crescimento expressivo do ICTE (+27,8%) e CEFORES (+40,0%, embora com volumes reduzidos). Esse panorama reforça que o recuo observado no ICS não é uniforme entre institutos, destacando a necessidade de análises organizacionais desagregadas. A heterogeneidade de trajetórias sugere que fatores de dinâmica interna, composição de equipes e perfis de

³Variação calculada sobre artigos publicados por participação institucional no período 2022–2024.

⁴Série anual de artigos publicados por instituto, com denominador de docentes ativos estatutários por instituto/ano.

área modulam o comportamento de artigos publicados de forma distinta entre unidades, mesmo sob condições institucionais compartilhadas.

4.6.1.6 Nota metodológica: janela trienal.

Para reduzir ruído anual e tornar comparações interinstitucionais mais interpretáveis, adota-se, quando pertinente, a agregação em janelas trienais. A escolha da janela temporal afeta a densidade de relações observadas (janelas muito curtas subestimam interações; janelas longas as superestimam), o que justifica seu uso como escala de leitura para dinâmicas recentes (Mena-Chalco *et al.*, 2014).

4.6.1.7 Caso 3 (padrão transversal): ruptura em tipos associados a eventos acadêmicos (2016–2024).

Além de trajetórias específicas por instituto, observa-se um padrão temporal **transversal** no conjunto institucional: a ruptura em tipos de produção associados a **eventos acadêmicos**. Considerando o agregado de *trabalhos em eventos*, *apresentações de trabalho* e *organização de eventos*, o volume institucional atingiu pico em 2019 (**3.080** produções), sofreu queda de **53,1%** em 2020 (**1.445**) e apresentou recuperação gradual no triênio 2022–2024 (**1.549** → **1.710** → **1.712**), permanecendo **44,4%** abaixo do patamar pré-pandemia (2024 em relação a 2019). A proporção desses tipos no total anual também declinou de **50,03%** (2019) para **34,54%** (2020) e estabilizou em **36,57%** (2024), sugerindo mudanças persistentes na composição tipológica do corpus após 2020, sem inferir causalidade direta entre pandemia e perfil institucional permanente. Na Figura 4, ilustra-se essa trajetória no período 2016–2025 (com ressalva para o ano de 2025, cuja consolidação pode ser parcial).

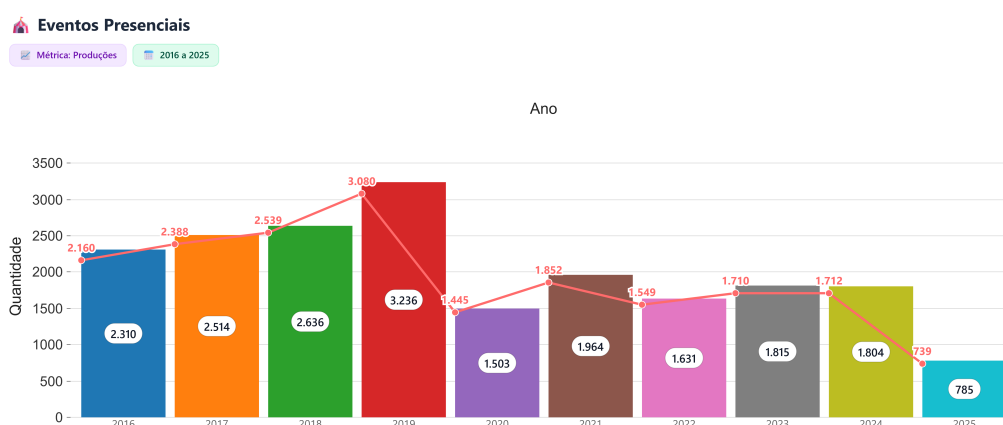


Figura 4 – Série temporal do agregado “eventos acadêmicos” (2016–2025).

Fonte: Elaboração própria.

Nota: o ano de 2025 é exibido no gráfico para referência visual, mas não é incluído na análise quantitativa em razão da consolidação ainda parcial dos currículos Lattes nesse período.

4.6.1.8 Discussão: impacto da pandemia, persistência parcial e heterogeneidade.

A ruptura observada em 2020 em tipos associados a eventos acadêmicos é compatível com as restrições e mudanças de rotina impostas pela pandemia de COVID-19, com recuperação parcial nos anos seguintes. Ainda assim, o contraste entre institutos (e.g., queda no ICS e recuperação no ICTE) reforça que métricas agregadas podem ocultar efeitos localizados por unidade, o que demanda leitura desagregada por tipologia e por recorte organizacional para evitar inferências indevidas.

4.6.2 Padrões por tipologia e colaboração

4.6.2.1 Escopo e princípio de verificabilidade.

Esta subseção explora padrões emergentes associados à **tipologia da produção** e a dimensões correlatas de colaboração/atividade formativa, com ênfase em **casos ilustrativos** (2–3) e evidências **rastreadas diretamente nas telas do artefato**. As análises utilizam a base consolidada e deduplicada, considerando participação institucional quando agregada por instituto/departamento.⁵

4.6.2.2 Janela temporal.

Adota-se como referência o triênio **2022–2024**, por reduzir ruído anual e mitigar efeitos de subatualização do último ano.

4.6.2.3 Fundamentação: perfis tipológicos e avaliação.

Nesse contexto, a literatura indica que perfis tipológicos variam entre áreas porque diferentes campos privilegiam veículos distintos de comunicação científica, com peso relativo maior de livros e capítulos em Ciências Humanas e Sociais, além de cobertura desigual em bases internacionais centradas em periódicos. Assim, a composição por tipos observada no triênio descreve perfis de comunicação e formação, úteis para contextualizar comparações entre unidades e indicadores, sem reduzir a avaliação a uma métrica única (Hicks *et al.*, 2015; Seglen, 1991; Melo; Trinca; Maricato, 2021). Sistemas avaliativos também podem influenciar escolhas de publicação, alterando a distribuição tipológica ao longo do tempo (Pires *et al.*, 2020).

4.6.2.4 Padrão global do triênio e núcleo bibliográfico.

Como linha de base, observa-se que a **produção bibliográfica** no triênio 2022–2024 é dominada por **artigos publicados** e **trabalhos em eventos**, que juntos respondem pela maior parte do volume deduplicado, enquanto **capítulos** e **livros** formam uma parcela menor porém substantiva. Esse panorama fornece o “padrão UFTM” necessário para interpretar desvios (excentricidades) como **perfil** e não como juízo de qualidade. No interior do triênio, nota-se variação moderada na participação de *trabalhos em eventos*, o que reforça a utilidade de uma janela agregada para caracterizar perfis tipológicos.

Na Tabela 14, apresenta-se a distribuição dos tipos bibliográficos no triênio.

4.6.2.5 Caso 1 (excentricidade tipológica): livros e capítulos no IELACHS.

Para operacionalizar "excentricidade tipológica" de forma verificável e estável, restringe-se inicialmente ao núcleo **ALC** (*artigos publicados, livros publicados/organizados e capí-*

⁵Por “participação institucional”, entende-se que uma mesma produção pode contribuir para mais de uma unidade em caso de coautorias interinstitucionais, o que é adequado para descrever perfis tipológicos associados à participação das unidades.

Tabela 14 – Distribuição global de tipos bibliográficos na UFTM (2022–2024).

Tipo	Qtde. (grupos)	% do total
Artigo publicado	2.358	43,7
Trabalho em eventos	2.098	38,9
Capítulo de livro	641	11,9
Livro publicado/organizado	139	2,6
Artigo aceito	74	1,4
Outra produção bibliográfica	43	0,8

tulos de livro), comparando a proporção de **livros+capítulos** no conjunto ALC por instituto (triênio 2022–2024). O **IELACHS** se destaca com **44,1%** de livros+capítulos no ALC (**269** de **610**), contrastando com institutos mais "padrão-artigo", como o **ICS** (**19,7%**; 252 de 1.276) e o **ICTE** (**16,4%**; 63 de 384). Esse desvio é interpretado como **perfil tipológico** (ênfase relativa em livro/capítulo), e não como indicador de volume absoluto, produtividade média por docente ou "melhor/pior". A Tabela 15 apresenta a comparação entre institutos por essa proporção. Como lente complementar fora do ALC, observa-se que o **ICTE** apresenta um perfil **orientação-cêntrico** no triênio (**49%** de orientações, comparado a **32%** no padrão global), compatível com perfis distintos de produção e formação: enquanto o IELACHS se diferencia por maior diversificação bibliográfica (livros/capítulos), o ICTE se destaca por maior ênfase relativa em atividades de **formação**.

Tabela 15 – Proporção de livros+capítulos no ALC por instituto (2022–2024).

Instituto	ALC total	Livros+Capítulos	% no ALC
IELACHS	610	269	44,1
ICAEBI	153	45	29,4
ICENE	457	117	25,6
CEFORES	51	12	23,5
ICS	1.276	252	19,7
ICTE	384	63	16,4
ICBN	381	53	13,9

Fonte: Elaboração própria.

Nota: ALC = artigos publicados, livros e capítulos (núcleo bibliográfico). **Siglas:** IELACHS = Instituto de Educação, Letras, Artes, Ciências Humanas e Sociais; ICAEBI = Instituto de Ciências Agrárias, Exatas e Biológicas de Iturama; ICENE = Instituto de Ciências Exatas, Naturais e Educação; CEFORES = Centro de Educação Profissional; ICS = Instituto de Ciências da Saúde; ICTE = Instituto de Ciências Tecnológicas e Exatas; ICBN = Instituto de Ciências Biológicas e Naturais.

4.6.2.6 Contraste temporal (triênio) e interpretação do desvio.

A série anual (2022–2024) mostra que a composição do IELACHS permanece consistentemente “puxada” por capítulos (e, em menor grau, livros), enquanto institutos como

o ICS mantém predominância de artigos. No IELACHS, o ALC em 2022 apresenta 117 artigos, 23 livros e 80 capítulos (total 220), o que corresponde a aproximadamente **46,8%** de livros+capítulos; em 2023, 107/25/58 (total 190; **43,7%**); e em 2024, 118/16/68 (total 202; **41,6%**). O padrão é estável o suficiente para ser tratado como **perfil tipológico recorrente** no triênio, e não como oscilação pontual. Na Figura 5, apresenta-se a série temporal de artigos publicados por instituto.

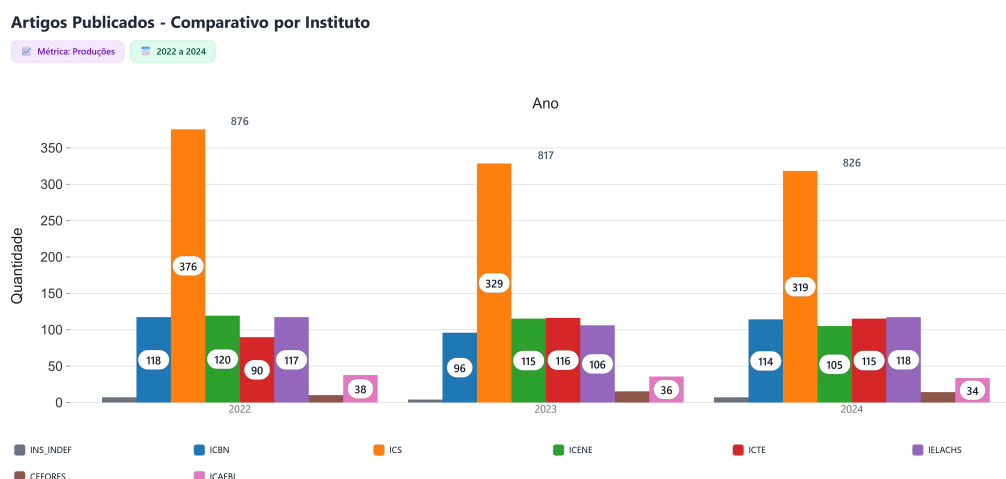


Figura 5 – Série temporal de artigos publicados por instituto (2022–2024).

Fonte: Elaboração própria.

Nota: ICS = Instituto de Ciências da Saúde; IELACHS = Instituto de Educação, Letras, Artes, Ciências Humanas e Sociais; ICENE = Instituto de Ciências Exatas, Naturais e Educação; ICTE = Instituto de Ciências Tecnológicas e Exatas; ICBN = Instituto de Ciências Biológicas e Naturais; CEFORES = Centro de Educação Profissional; ICAEBI = Instituto de Ciências Agrárias, Exatas e Biológicas de Iturama; INS_DEF = instituto indefinido (produção sem associação organizacional identificada).

Na Figura 6, apresenta-se a série temporal de livros+capítulos por instituto.

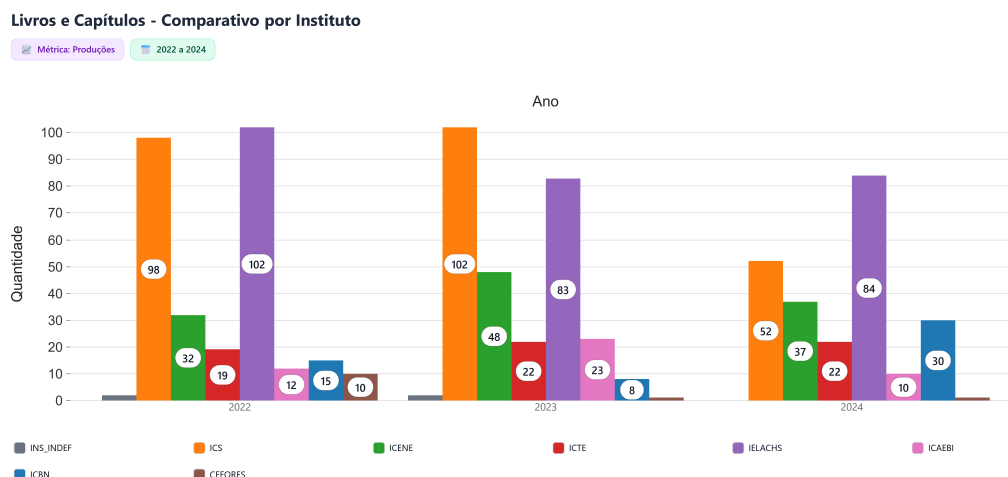


Figura 6 – Série temporal de livros+capítulos por instituto (2022–2024).

Fonte: Elaboração própria.

Nota: ICS = Instituto de Ciências da Saúde; IELACHS = Instituto de Educação, Letras, Artes, Ciências Humanas e Sociais; ICENE = Instituto de Ciências Exatas, Naturais e Educação; ICTE = Instituto de Ciências Tecnológicas e Exatas; ICBN = Instituto de Ciências Biológicas e Naturais; CEFORES = Centro de Educação Profissional; ICAEBI = Instituto de Ciências Agrárias, Exatas e Biológicas de Iturama; INS_DEF = instituto indefinido (produção sem associação organizacional identificada).

4.6.2.7 Decomposição organizacional (drill-down): quais departamentos do IELACHS sustentam o perfil.

Para verificar se a excentricidade é difusa ou concentrada internamente, decompõe-se o ALC do IELACHS por departamentos (2022–2024). Observa-se heterogeneidade: departamentos como **DLE** (83,3% de livros+capítulos no ALC; 10 de 12) e **DHIS** (65,1%; 28 de 43) apresentam predominância forte de livros/capítulos, enquanto unidades maiores (e.g., **DPSICO**, 42,1%; 85 de 202) sustentam o perfil com volume elevado. Esse resultado reforça que o “perfil do instituto” combina componentes **concentrados** (departamentos menores com alta proporção) e componentes **estruturais** (departamentos maiores com participação absoluta relevante). Na Tabela 16, detalha-se essa decomposição.

4.6.2.8 Janela específica para engajamento discente.

Enquanto a excentricidade tipológica é descrita no triênio 2022–2024 para reduzir ruído anual, o engajamento discente é analisado em janela ampliada (2016–2024), pois envolve trajetórias de formação (ingresso, conclusão e participação ao longo dos anos, incluindo egressos conforme regra operacional do indicador). A janela ampliada busca capturar um padrão institucional mais estável de participação discente, mantendo a verificabilidade nas telas do artefato.

Tabela 16 – Decomposição do ALC no IELACHS por departamento (2022–2024; apenas deptos com ALC ≥ 10).

Depto	ALC total	Artigos	Livros+Capítulos	% no ALC
DLE	12	2	10	83,3
DHIS	43	15	28	65,1
DEL	53	25	28	52,8
DE	77	40	37	48,1
DPSICO	202	117	85	42,1

Fonte: Elaboração própria.

Nota: ALC = artigos publicados, livros e capítulos (núcleo bibliográfico). **Siglas:** DLE = Departamento de Línguas Estrangeiras; DHIS = Departamento de História; DEL = Departamento de Estudos Literários; DE = Departamento de Educação; DPSICO = Departamento de Psicologia.

4.6.2.9 Caso 2 (engajamento discente): padrões de formação e produção.

A análise de engajamento discente (2016–2024) revela padrões consistentes e verificáveis. Em **artigos publicados**, o **ICS** apresenta maior participação discente (**62,56%** dos artigos com ao menos um discente na equipe), seguido por ICBN (58,27%) e CEFORES (56,93%), enquanto IELACHS registra engajamento moderado (43,76%) e ICAEBI apresenta engajamento baixo (17,84%). Em **livros e capítulos**, observa-se um padrão distinto: ICTE (53,01%) e ICBN (52,59%) apresentam engajamento elevado, ao passo que o **IELACHS**, apesar de concentrar maior proporção de livros+capítulos no ALC no triênio (44,1%, conforme Caso 1), exibe **menor engajamento discente** nesse conjunto (19,39%), sugerindo que livros e capítulos no instituto tendem a ser produtos menos associados à formação discente, em comparação a artigos e eventos. Nas Tabelas 17 e 18, apresenta-se o engajamento discente por instituto em artigos e em livros+capítulos, respectivamente.

4.6.2.10 Decomposição por tipo bibliográfico: hierarquia de engajamento.

A decomposição por tipo reforça uma hierarquia clara: **trabalhos em eventos** apresentam maior engajamento discente (**69,14%**), compatível com participação discente em congressos e iniciação científica; **artigos publicados** exibem engajamento intermediário (**49,07%**); **capítulos** apresentam engajamento menor (**32,56%**); e **livros** têm menor engajamento (**16,46%**), compatível com produtos de autoria docente consolidada. Na Tabela 19, apresenta-se o engajamento por tipo bibliográfico.

4.6.2.11 Evolução temporal do engajamento discente (global).

No plano temporal, observa-se crescimento do engajamento global de **45,48%** (2016) para **66,96%** (2024), com salto em 2019 (58,55%) e consolidação acima de 60% desde 2021. Esse padrão sugere maior presença discente nas produções no período recente, embora a interpretação deva considerar a definição operacional de discente (p.ex.,

Tabela 17 – Engajamento discente em artigos publicados por instituto (2016–2024).

Instituto	Total (grupos)	Com discente (grupos)	%
ICS	2.997	1.875	62,56
ICBN	901	525	58,27
CEFORES	137	78	56,93
IELACHS	1.033	452	43,76
ICTE	1.035	446	43,09
ICENE	923	370	40,09
ICAEBI	342	61	17,84

Fonte: Elaboração própria.

Nota: Contagens em modo participação: uma mesma produção deduplicada pode contribuir para mais de um instituto quando houver coautoria interinstitucional. “Com discente” indica presença de ao menos um discente ou egresso na equipe da produção. **Siglas:** ICS = Instituto de Ciências da Saúde; ICBN = Instituto de Ciências Biológicas e Naturais; CEFORES = Centro de Educação Profissional; IELACHS = Instituto de Educação, Letras, Artes, Ciências Humanas e Sociais; ICTE = Instituto de Ciências Tecnológicas e Exatas; ICENE = Instituto de Ciências Exatas, Naturais e Educação; ICAEBI = Instituto de Ciências Agrárias, Exatas e Biológicas de Iturama.

inclusão de egressos dentro de horizonte temporal) e a natureza de participação institucional (uma mesma produção pode contribuir para múltiplos institutos em coautorias). Embora a série não permita atribuição causal direta, o crescimento observado é compatível com movimentos institucionais de estímulo à participação discente na produção científica e à qualificação dos programas de pós-graduação. Esse eixo de valorização aparece formalizado no PDI 2025–2029 da UFTM, que estabelece, na Meta 1.19, a ampliação de publicações em coautoria entre docentes e discentes nos programas de pós-graduação stricto sensu (Universidade Federal do Triângulo Mineiro, 2025).

4.6.2.12 Nota de interpretação e robustez.

Os percentuais elevados não significam "proporção de discentes ativos no ano", e sim **presença de ao menos um discente/egresso identificado na equipe** conforme a regra do indicador e a capacidade de pareamento de autores. Como verificação de robustez (opcional), recomenda-se uma análise de sensibilidade com critério mais estrito, restringindo a contagem a discentes **ativos no ano da produção** (ano $\in$ [ingresso, conclusão] ou conclusão+ k), para avaliar a estabilidade das ordens relativas entre institutos sem alterar a estrutura do argumento.

4.6.2.13 Ampliação: excentricidade para além do núcleo bibliográfico.

A excentricidade tipológica do IELACHS não se restringe ao núcleo bibliográfico (ALC). Na decomposição por categoria de interesse, o instituto apresenta **16%** de produção técnica no triênio, proporção única entre os institutos (padrão global: **9%**). Esse comportamento é explicado, em grande medida, pela concentração de **programas de rádio/TV** (**93** no triênio, aproximadamente **45%** do total institucional nesse tipo),

Tabela 18 – Engajamento discente em livros+capítulos por instituto (2016–2024).

Instituto	Total (grupos)	Com discente (grupos)	%
ICTE	166	88	53,01
ICBN	116	61	52,59
CEFORES	37	17	45,95
ICS	689	223	32,37
ICAEBI	115	34	29,57
ICENE	306	90	29,41
IELACHS	753	146	19,39

Fonte: Elaboração própria.

Nota: Contagens em modo participação: uma mesma produção deduplicada pode contribuir para mais de um instituto quando houver coautoria interinstitucional. “Com discente” indica presença de ao menos um discente ou egresso na equipe da produção. **Siglas:** ICTE = Instituto de Ciências Tecnológicas e Exatas; ICBN = Instituto de Ciências Biológicas e Naturais; CEFORES = Centro de Educação Profissional; ICS = Instituto de Ciências da Saúde; ICAEBI = Instituto de Ciências Agrárias, Exatas e Biológicas de Iturama; ICENE = Instituto de Ciências Exatas, Naturais e Educação; IELACHS = Instituto de Educação, Letras, Artes, Ciências Humanas e Sociais.

Tabela 19 – Engajamento discente por tipo bibliográfico (2016–2024).

Tipo de produção	Total (grupos)	Com discente (grupos)	%
Trabalho em eventos	7.952	5.498	69,14
Artigo aceito para publicação	142	77	54,23
Artigo publicado	7.116	3.492	49,07
Capítulo de livro publicado	1.729	563	32,56
Texto em jornal ou revista	112	23	20,54
Livro publicado/organizado	407	67	16,46

compatível com um perfil de **divulgação e extensão** associado a áreas com produção midiática. Na Figura 7, ilustra-se a composição por categoria de interesse por instituto.

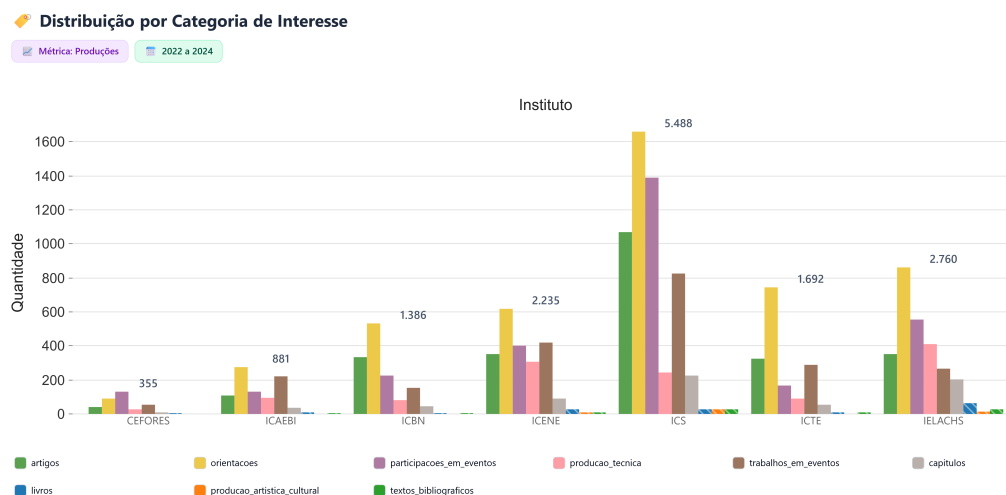


Figura 7 – Composição por categoria de interesse por instituto (2022–2024).

Fonte: Elaboração própria.

4.6.2.14 Produção técnica: variedade tipológica e concentração de tipos raros.

Ao ampliar o olhar para **produção técnica** (2022–2024), o padrão geral é dominado por tipos recorrentes (e.g., apresentação de trabalho e organização de evento), mas há sinais de especialização institucional em tipos menos frequentes. Dois achados são particularmente verificáveis: (i) o **ICTE** apresenta presença comparativamente maior de **software** e **patente** no triênio, é consistente com maior presença de artefatos técnicos; e (ii) o **ICS** concentra parte relevante de tipos menos frequentes no conjunto institucional, como **mídia social**, **website ou blog** e **desenvolvimento de material didático ou instrucional**, indicando heterogeneidade de repertórios mesmo em institutos com perfil bibliográfico centrado em artigos. Na Figura 8, apresenta-se a composição da produção técnica por instituto.

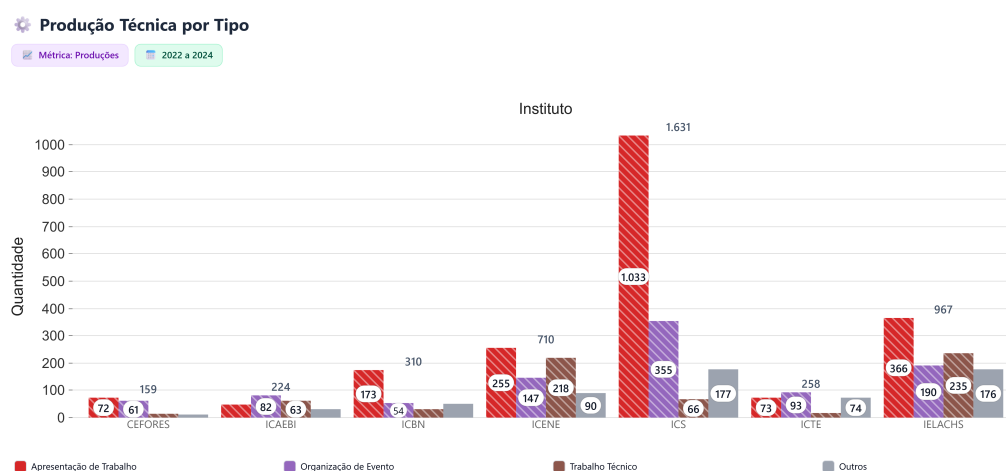


Figura 8 – Produção técnica por instituto: composição por tipo (2022–2024).

Fonte: Elaboração própria.

Na Figura 9, exemplificam-se tipos raros de produção técnica que apresentam concentração em institutos específicos.

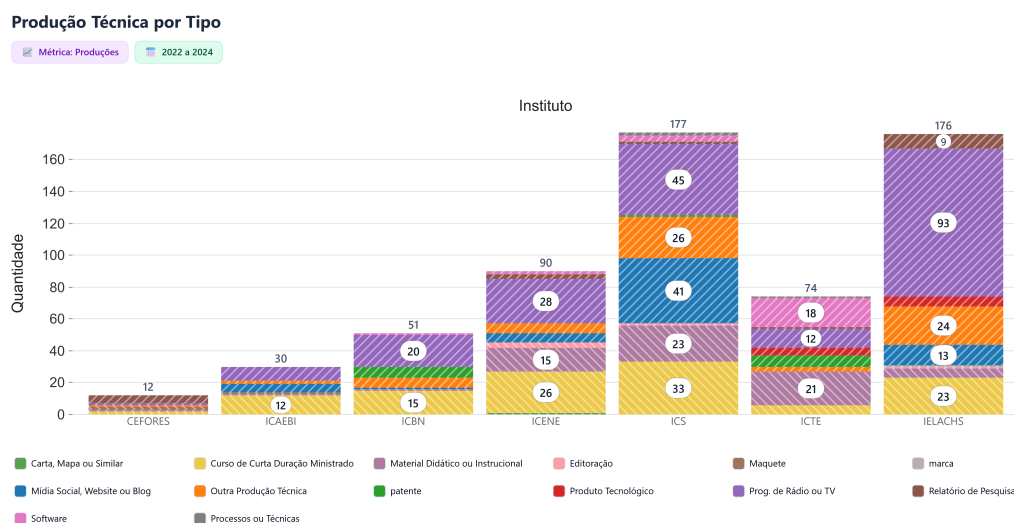


Figura 9 – Tipos raros de produção técnica concentrados: exemplos (2022–2024).

Fonte: Elaboração própria.

4.6.2.15 Orientações concluídas: evidência tipológica da dimensão formativa.

As **orientações concluídas** reforçam a dimensão formativa que o núcleo bibliográfico não captura. Como os tipos são mais heterogêneos nesse subconjunto, utiliza-se a classificação por **natureza** da orientação. No triênio 2022–2024, prevalecem **TCC de graduação** e **iniciação científica**, com variações por instituto: no **ICS**, por exemplo, destacam-se volumes elevados de TCC e IC, enquanto no **IELACHS** observa-se TCC, além de participação relevante de dissertações e IC. Esses padrões são consistentes com a expectativa de que a produção institucional combina resultados bibliográficos com atividades de formação, variando em intensidade e composição entre unidades. Na Figura 10, apresenta-se a distribuição de orientações por instituto e natureza.

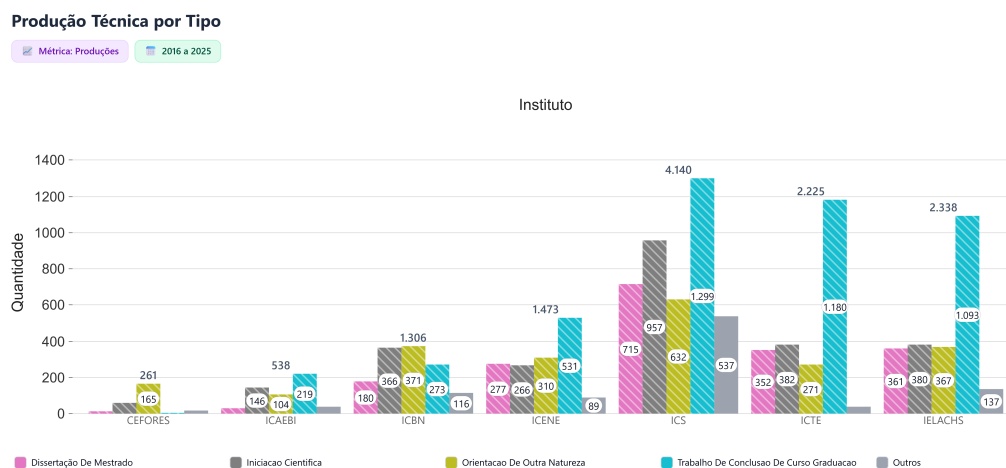


Figura 10 – Orientações por instituto e natureza (2022–2024).

Fonte: Elaboração própria.

4.6.2.16 Síntese (tipologia e colaboração).

No triênio 2022–2024, o perfil tipológico global é dominado por artigos e trabalhos em eventos, mas a comparação por unidade revela **excentricidades consistentes**, como a maior proporção de livros e capítulos no IELACHS (Caso 1). Essas excentricidades não indicam "qualidade", e sim **perfil de comunicação científica** e repertório de produtos, ampliado por produção técnica (especialmente no IELACHS, com concentração em rádio/TV). Ao incorporar orientações e engajamento discente (Caso 2), evidencia-se que institutos diferem também no **modo** como a produção se conecta à formação: há perfis mais orientações-cêntricos (ICTE) e perfis em que a participação discente varia fortemente conforme o tipo (eventos > artigos > capítulos > livros), reforçando a utilidade de múltiplas lentes para comparação de unidades e para interpretação cautelosa de diferenças tipológicas e formativas.

4.6.3 Padrões com enriquecimento bibliométrico

4.6.3.1 Escopo e recorte dos dados bibliométricos.

Nesta subseção, o foco recai sobre padrões observáveis quando a produção deduplicada é enriquecida com (i) citações (por fonte e no tempo) e (ii) variáveis bibliométricas auxiliares (impacto normalizado do OpenAlex e estrato Qualis do periódico). Para manter coerência com a estratégia de confiabilidade discutida na Seção 4.2 e com a necessidade de chave de ligação entre fontes, os resultados que dependem de OpenAlex são condicionados ao subconjunto com **DOI-chave**, e a colaboração internacional depende da **cobertura de afiliações** no OpenAlex (instituições e país associados às autorias).

4.6.3.2 Coautoria internacional e frequência de alto impacto (OpenAlex).

Para avaliar se a colaboração internacional se associa a maior impacto relativo, as produções do núcleo bibliográfico (artigos, livros e capítulos) com DOI-chave foram classificadas em três classes: *sem coautoria externa*, *coautoria nacional* e *coautoria internacional*. A classe é derivada a partir das afiliações institucionais no OpenAlex, excluindo a instituição de origem (UFTM) e marcando como internacional quando há instituição externa com código de país diferente de BR (Brasil).

Como métrica de "alto impacto", foi usado o percentil normalizado de citações do OpenAlex (normalizado por campo e ano de publicação) e, em especial, os indicadores de classificação nas faixas Top 10% e Top 1%, derivados automaticamente pelo OpenAlex a partir do percentil normalizado. Na Tabela 20, constam o total de produções por classe e a frequência de alto impacto.

Tabela 20 – Coautoria (OpenAlex) e frequência de alto impacto normalizado (Top 10% e Top 1%). *Percentuais calculados sobre o subconjunto com valor normalizado disponível.*

Classe de colaboração	N total	N com valor	Top 10%	% Top 10	Top 1%	% Top 1
Coautoria internacional	1.755	1.724	504	29,23%	98	5,68%
Coautoria nacional	4.147	4.107	697	16,97%	55	1,34%
Sem coautoria externa	3.779	3.678	308	8,37%	23	0,63%

Dois padrões são diretamente verificáveis na Tabela 20: (i) a **coautoria internacional** apresenta maior concentração de trabalhos em faixas de alto impacto normalizado; e (ii) há um **gradiente monotônico** entre as três classes. No recorte analisado, a proporção de itens no *Top 10%* é de 29,23% (coautoria internacional) contra 8,37% (sem coautoria externa), o que corresponde a uma **razão de proporções** de aproximadamente **3,5×**; para o *Top 1%*, a razão é de aproximadamente **9,0×** (5,68% vs. 0,63%). Os percentuais são calculados apenas sobre o subconjunto com valor normalizado disponível (*N com valor*); diferenças residuais entre *N total* e *N com valor* refletem registros com

DOI-chave que possuem match no OpenAlex, mas não apresentam o campo normalizado preenchido no snapshot.

4.6.3.3 Discussão do padrão (colaboração e impacto).

O gradiente observado (sem coautoria externa → coautoria nacional → coautoria internacional) é compatível com a literatura que descreve a ciência como um sistema em rede, no qual colaborações ampliam a recombinação de competências e a difusão de resultados, frequentemente associadas a maior impacto em métricas de larga escala (Fortunato *et al.*, 2018), o que é coerente com o aumento observado nas faixas *Top 10%* e *Top 1%*. No escopo deste trabalho, o resultado deve ser lido como **associação observacional**: a classificação depende das afiliações registradas no OpenAlex e o recorte está condicionado ao subconjunto com DOI-chave.

4.6.3.4 Condições de validade e vieses do recorte.

O efeito é estimado apenas para produções com **DOI-chave** e **afiliação recuperável** no OpenAlex, o que tende a privilegiar circuitos editoriais e áreas com maior inserção e rastreabilidade em bases globais; essa assimetria é consistente com as limitações de cobertura e representatividade de bases internacionais quando aplicadas a contextos nacionais (Melo; Trinca; Maricato, 2021). Por outro lado, a estratégia de integração local–global (Lattes/CAPES ↔ OpenAlex) viabiliza uma leitura mais auditável do fenômeno ao explicitar a chave de ligação (DOI) e o critério operacional de colaboração (país nas afiliações), permitindo inspecionar o recorte e reproduzir a classificação no artefato (Mazoni *et al.*, 2025). Na Figura 11, ilustra-se a evolução temporal da quantidade de trabalhos no Top 1% por classe de colaboração.

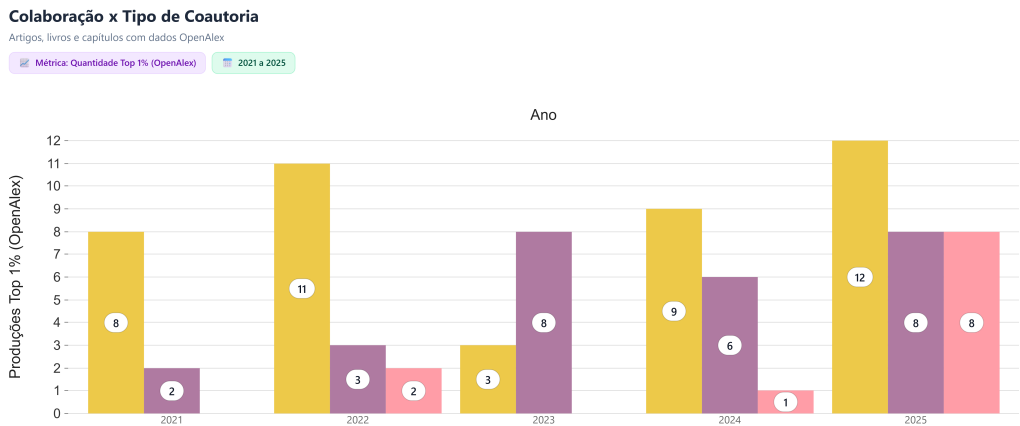


Figura 11 – Quantidade de trabalhos no Top 1% por classe de colaboração (2021–2025).

Fonte: Elaboração própria.

4.6.3.5 Citações recebidas por ano e por fonte (série temporal).

Como complemento, foi construída uma série temporal de **citações recebidas** agregadas por *ano de citação* e por fonte, combinando: (i) **OpenCitations** (citações por ano consolidadas), (ii) **Google Scholar** (citações por ano) e (iii) **OpenAlex** (contagens anuais de citações). Essa visão permite duas leituras úteis: (a) a maturação temporal das citações (crescimento acumulado ao longo do tempo) e (b) a diferença de cobertura entre fontes. O recorte exhibe os **últimos 5 anos (2021–2025)** para focar no período recente e facilitar a comparação entre fontes. Como as fontes têm coberturas e critérios distintos, as curvas não são diretamente comparáveis em nível absoluto; o objetivo é evidenciar tendência temporal e diferenças de cobertura, evitando interpretar contagens brutas como medida única de impacto (Szomszor *et al.*, 2021). Na Figura 12, apresenta-se a série de citações recebidas por ano e por fonte no período 2021–2025.

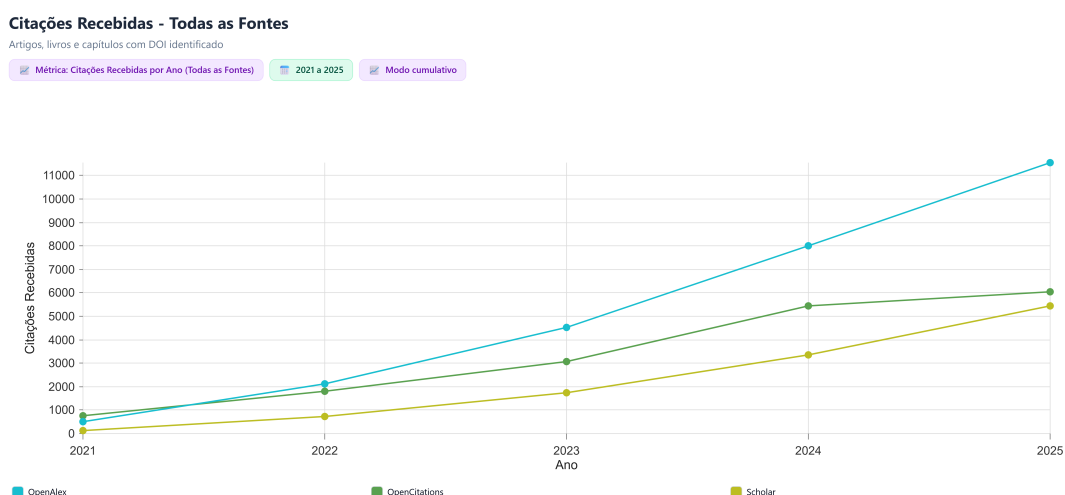


Figura 12 – Total de citações recebidas por ano e por fonte (OpenAlex, OpenCitations e Scholar, 2021–2025).

Fonte: Elaboração própria.

4.6.3.6 Qualis e intensidade de citações em artigos recentes.

Para verificar se o estrato Qualis se relaciona com maior intensidade de citações no recorte recente, foram considerados **artigos publicados** com **Qualis definido** e DOI-chave, utilizando citações do **OpenAlex** e agregando por *ano* e *estrato*. O recorte abrange **2019–2025** para capturar a maturação das citações, pois artigos mais antigos têm maior janela temporal para acumulação, permitindo melhor visualização do gradiente entre estratos. A classificação Qualis utilizada corresponde ao **quadriênio 2021–2024**, conforme definido na Seção 3.4. O padrão observado é consistente: estratos superiores (A1/A2) acumulam **volume total de citações** substancialmente maior do que estratos intermediários/baixos (B1/B2/C) no período em que há tempo suficiente para acumulação. Tomando **2019** como referência (ano com maior janela temporal e, portanto, melhor para

ilustrar as diferenças), artigos A1 acumulam **4.700 citações** no OpenAlex (123 artigos), enquanto B1 totaliza **899 citações** (137 artigos) e C totaliza **106 citações** (51 artigos) — evidenciando o gradiente $A1 \gg B1 \gg C$. Em **2023**, a mesma ordenação permanece, porém com totais menores (A1: 431 citações, 76 artigos), o que é compatível com a janela temporal mais curta para acumulação. Esse padrão também é compatível com o papel do Qualis/CAPES como sistema de classificação que influencia estratégias de publicação e direciona esforço para periódicos em estratos superiores (Pires *et al.*, 2020). Na Figura 13, apresenta-se a relação entre estrato Qualis e total de citações por ano.

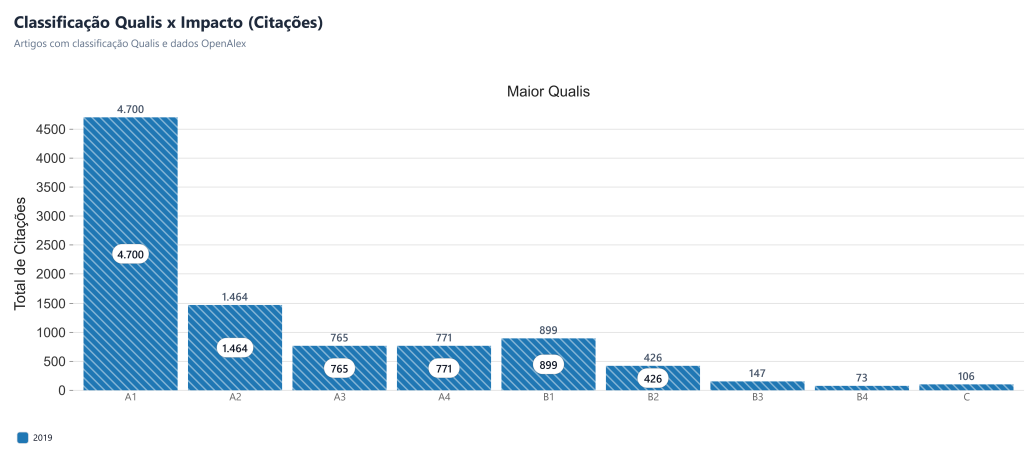


Figura 13 – Qualis × total de citações (OpenAlex) por ano e estrato (2019–2025; Qualis 2021–2024).

Fonte: Elaboração própria.

4.6.4 Discussão dos padrões observados

4.6.4.1 Complementaridade de lentes analíticas.

As três subseções anteriores (4.6.1–4.6.3) indicam que a caracterização da produção institucional exige **múltiplas lentes complementares**, pois cada uma evidencia aspectos diferentes da produção institucional. A lente **temporal-organizacional** mostra variações ao longo do tempo; a lente **tipológico-colaborativa** descreve perfis de comunicação científica e padrões de colaboração, além de evidenciar dimensões formativas via engajamento discente; e a lente **bibliométrica** permite observar associações entre padrões de coautoria, estratificação de periódicos e visibilidade científica no recorte bibliométrico-integrável (DOI-chave). Esse enquadramento é consistente com recomendações de uso responsável de indicadores, que enfatizam contexto e pluralidade de métricas em vez de avaliação por medida única (Hicks *et al.*, 2015).

4.6.4.2 Heterogeneidade institucional como achado central.

O achado transversal mais saliente é a **heterogeneidade institucional persistente** em múltiplas dimensões. As subseções 4.6.1–4.6.3 sugerem que institutos podem apresentar tendências temporais divergentes, perfis tipológicos distintos no triênio e padrões diferenciados de articulação entre produção e formação, além de diferenças na presença de colaboração internacional e em métricas bibliométricas no recorte com DOI.

A implicação é que os perfis institucionais não convergem para um padrão único e não são adequadamente descritos por uma única medida agregada. Assim, comparações úteis dependem de observar o **perfil** (tipologia, colaboração, formação, recorte bibliométrico) e o nível de desagregação adotado.

Esse ponto dialoga com críticas clássicas ao uso de indicadores como instrumentos universais e com a necessidade de interpretação contextual (Seglen, 1991; Hicks *et al.*, 2015).

Do ponto de vista organizacional, o resultado também é compatível com a caracterização de universidades como organizações *fracamente acopladas*, em que unidades acadêmicas operam com relativa autonomia e lógicas próprias, o que tende a preservar heterogeneidade entre institutos (Thiedig; Wegner, 2024).

4.6.4.3 Padrões transversais e choques externos.

Apesar da heterogeneidade entre unidades, as subseções 4.6.1–4.6.3 mostram a presença de padrões **transversais** no conjunto institucional. O caso mais claro é a ruptura em tipos associados a **eventos presenciais** durante a pandemia de COVID-19, seguida de recuperação parcial nos anos posteriores. Ao mesmo tempo, o comportamento não é homogêneo entre institutos, o que indica que o padrão transversal convive com dinâmicas institucionais próprias (Szomszor *et al.*, 2021).

4.6.4.4 Engajamento discente: evidência de articulação entre produção e formação.

Os resultados de 4.6.2 (Caso 2) sugerem que a articulação entre produção e formação **não é uniforme**: ela varia de modo sistemático conforme o **tipo de produção** e conforme o **perfil institucional**. Em termos interpretativos, a hierarquia observada entre tipos é compatível com processos típicos de formação, nos quais atividades como trabalhos em eventos e publicações seriadas tendem a envolver mais discentes em momentos de iniciação e desenvolvimento da pesquisa, ao passo que livros e parte dos capítulos podem refletir produtos mais autorais e consolidados. Esse padrão também é compatível com pressões normativas na pós-graduação, em que programas podem exigir ao menos uma publicação em periódicos bem avaliados no sistema vigente antes da defesa, o que tende a favorecer produtos de ciclo mais curto (artigos e trabalhos em eventos) (Pires *et al.*, 2020). Além disso, a comparação entre institutos indica que um perfil tipológico (p.ex., maior presença de livros/capítulos) não implica, por si só, maior engajamento discente nesse subconjunto, reforçando que o “perfil de produção” e o “perfil formativo” são dimensões **complementares** e devem ser analisadas em conjunto (Hicks *et al.*, 2015).

4.6.4.5 Colaboração internacional e impacto: evidência de associação robusta.

No recorte bibliométrico-integrável (produções com DOI-chave e afiliações recuperáveis no OpenAlex), os resultados de 4.6.3 mostram um **padrão ordenado** entre classes de colaboração: a coautoria internacional concentra maior presença relativa nas faixas de alto impacto normalizado (Top 10% e Top 1%) do que a coautoria nacional, que por sua vez supera as produções sem coautoria externa. No recorte analisado, isso torna a internacionalização uma dimensão descritiva relevante do subconjunto integrável, embora se trate de uma associação observada no recorte com DOI e afiliações recuperáveis. O padrão é compatível com evidências de que trabalhos em equipe tendem a combinar especialidades e ampliar visibilidade via redes de coautoria, o que se associa a maior impacto médio em diferentes domínios (Fortunato *et al.*, 2018).

4.6.4.6 Limitações interpretativas e múltiplas leituras possíveis.

Os padrões descritos nesta seção são **emergentes** na base consolidada e deduplicada e não devem ser lidos como relações determinísticas. Isto é, eles não implicam que um mesmo mecanismo explique todos os casos, nem autorizam inferências causais diretas a partir dos dados observados. Parte das variações pode refletir diferenças disciplinares, estratégias locais e dinâmicas de equipes, enquanto os resultados bibliométricos dependem de condições de integrabilidade (p.ex., DOI-chave e cobertura de afiliações), o que afeta a representatividade do recorte analisado. Assim, a contribuição do estudo é oferecer uma **descrição verificável** de perfis e evoluções, explicitando recortes, métricas e restrições, e criando base para leituras e hipóteses que podem ser aprofundadas por triangulação com evidências adicionais (Hicks *et al.*, 2015). Em síntese, os resultados indicam que a

produção institucional requer leitura multidimensional e que a verificabilidade da base consolidada é condição para que essas leituras possam ser reproduzidas e auditadas por diferentes usuários.

4.7 EXPLORAÇÃO INTERATIVA E VERIFICABILIDADE DOS RESULTADOS

A seção descreve os mecanismos de interação que permitem ao usuário reproduzir recortes, conferir contagens e navegar do agregado aos registros subjacentes (*drill-through*), incluindo exportação do subconjunto filtrado.

Drill-through é definido aqui como a transição do agregado (elemento do gráfico) para a listagem dos registros subjacentes na Tela de Produções, com filtros pré-aplicados.

4.7.1 Dimensões de análise disponíveis no artefato

4.7.1.1 Arquitetura de navegação por telas.

O sistema é organizado em um conjunto de telas interligadas: **Home**; telas institucionais (**Institutos**, **Instituto/ID**, **Departamento/ID**, **Programas**, **Programa/ID**); **Professor**; e **Produções**. Há ainda páginas estáticas (Sobre, Ajuda, Termos e Créditos). Um menu de cabeçalho, presente em todas as telas, fornece acesso às rotas principais (Home, Sobre, Institutos, Programas), além de controles de acessibilidade (*tamanho da fonte* e *modo noturno*). O rodapé replica links institucionais (Sobre, Ajuda, Termos, Créditos). A Home apresenta um painel de cards com indicadores agregados institucionais (ex.: total de produções, ALC durante vínculo, percentual acima da média do campo, excelência em impacto Top 10%, colaboração internacional, Open Access, citações por fonte, distribuição Qualis, ODS, parcerias internacionais, entre outros). Nas Figuras 14, 15 e 16, ilustram-se, respectivamente, o cabeçalho com controles de acessibilidade, o rodapé institucional e o painel de cards da Home.



Figura 14 – Header com os controles de fonte/modo noturno e links de navegação.

Fonte: Elaboração própria.

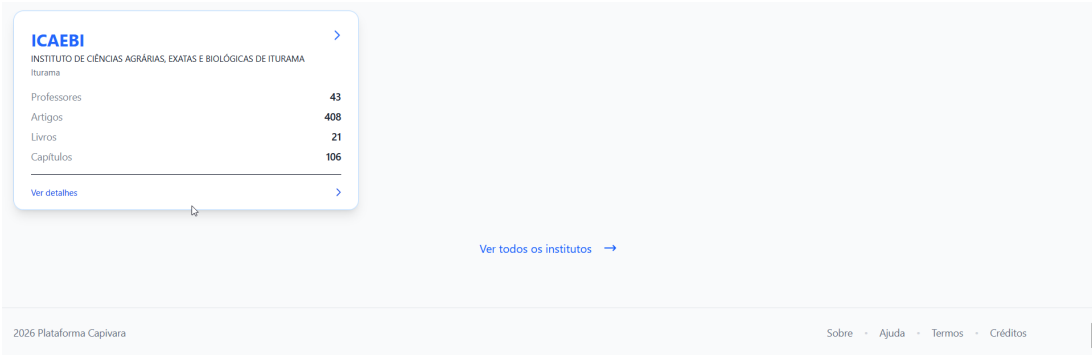


Figura 15 – Footer com links institucionais.
Fonte: Elaboração própria.

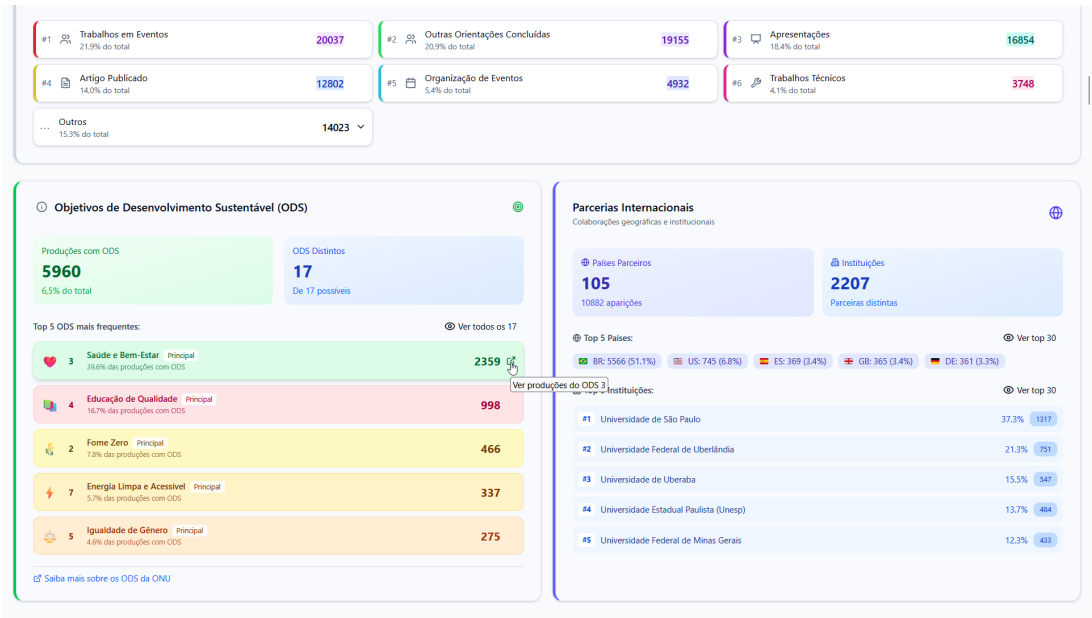


Figura 16 – Pannel de cards de indicadores agregados na Home.
Fonte: Elaboração própria.

4.7.1.2 Núcleo analítico reutilizável: seção de indicadores.

A principal unidade de exploração é a seção de **indicadores gráficos**, presente tanto na Home quanto nas telas institucionais e na tela do professor (em abas específicas). Cada indicador segue um padrão de interação: título/subtítulo, barra de ações, filtros, visualização e recursos de exploração, conforme ilustrado na Figura 17.

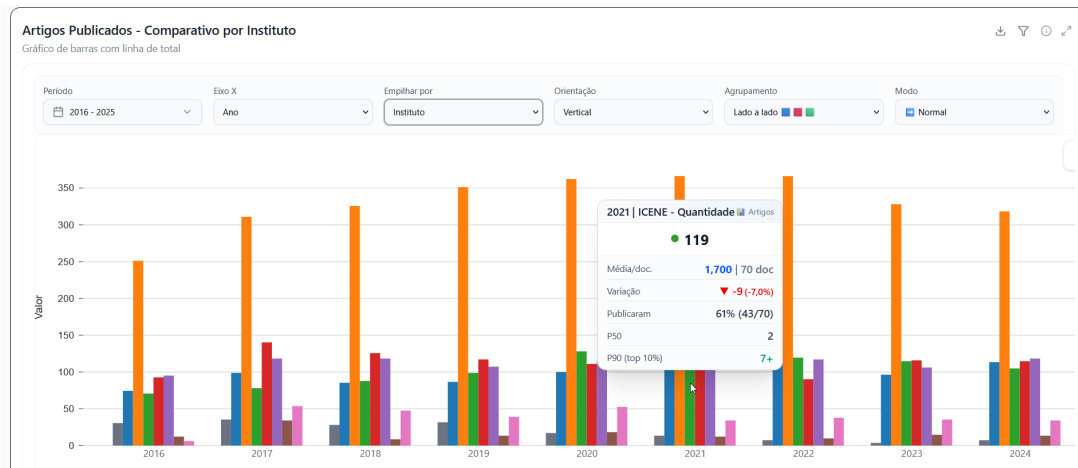


Figura 17 – Um indicador na Home mostrando: título/subtítulo, barra de ações e filtros visíveis.

Fonte: Elaboração própria.

4.7.1.3 Ações e controles padrão dos indicadores.

Os indicadores disponibilizam quatro operações recorrentes: (i) **Download** dos dados (JSON/CSV/XLSX e exportação de imagem), com modos de exportação rápido (*express*) ou configurável (*customizado*), preservando o recorte aplicado; (ii) **mostrar/ocultar filtros**, reduzindo ruído visual sem alterar o recorte; (iii) **informações do indicador** (ícone “i”), abrindo uma janela com definição operacional, fontes, observações e limitações; (iv) **expandir indicador**, abrindo visualização em tela cheia para inspeção. Nas Figuras 18, 19 e 20, ilustram-se, respectivamente, o menu de download, o modal de informações e a expansão em tela cheia.

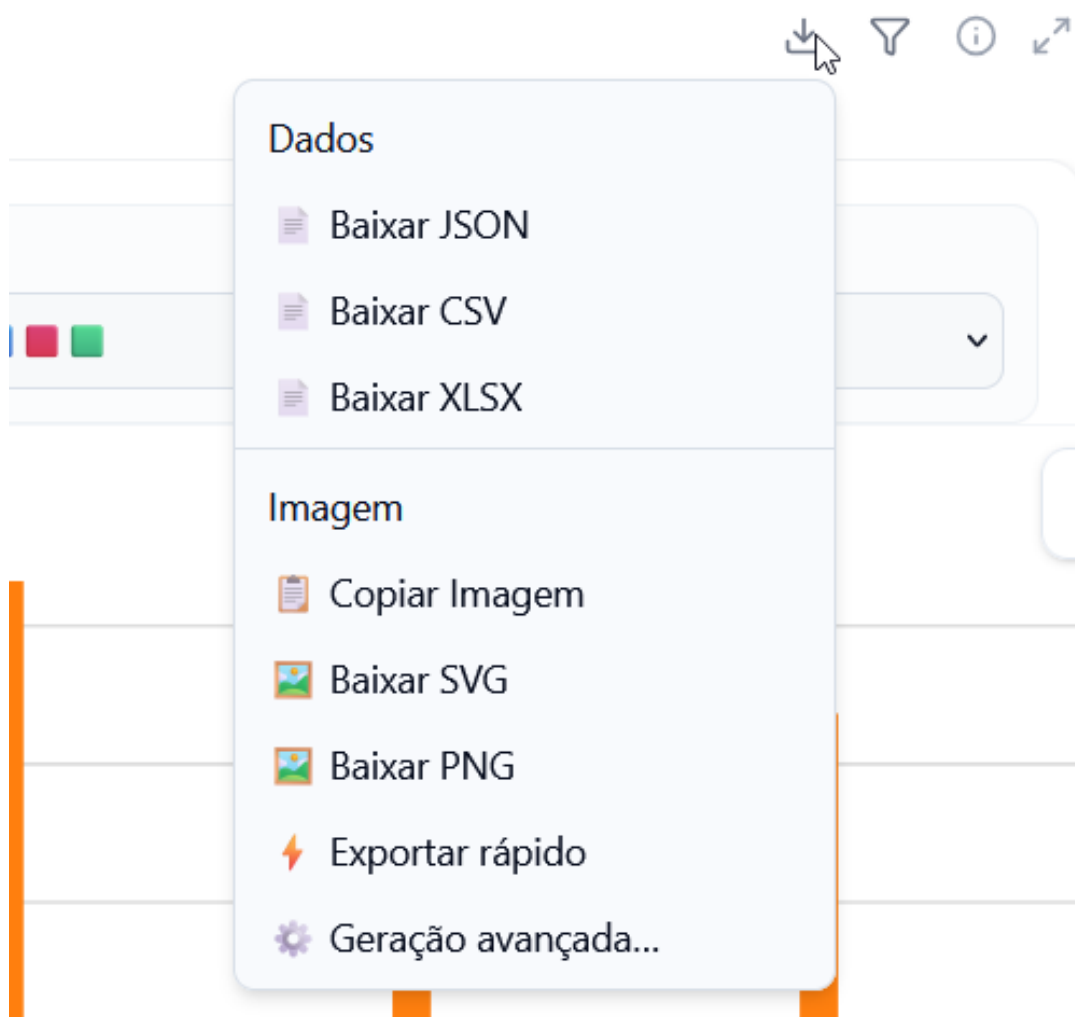


Figura 18 – Popover do Download com opções (JSON/CSV/XLSX/imagem; express/cus-
tom).

Fonte: Elaboração própria.

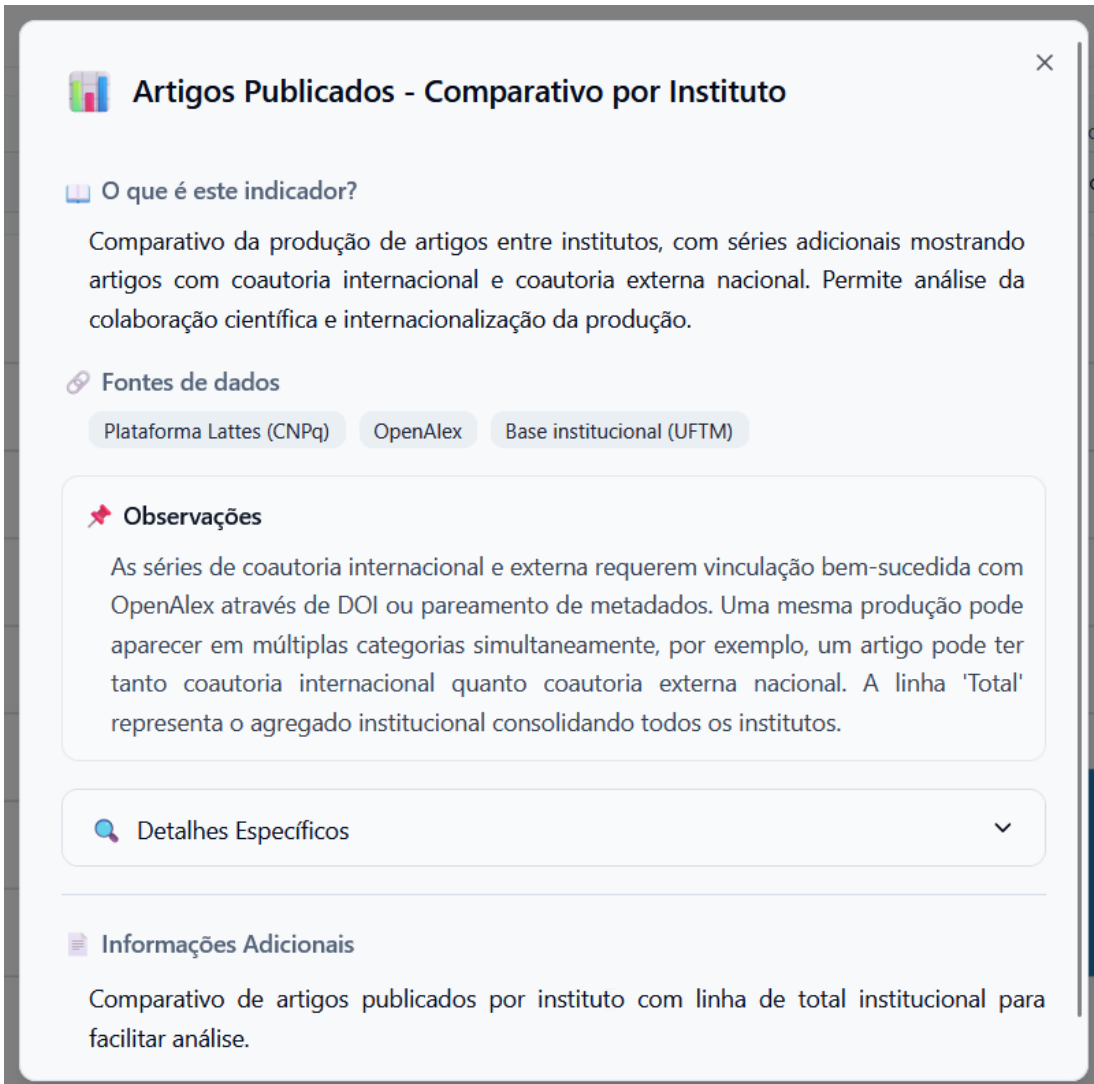


Figura 19 – Modal do "i"(definição + fontes + observações).
Fonte: Elaboração própria.

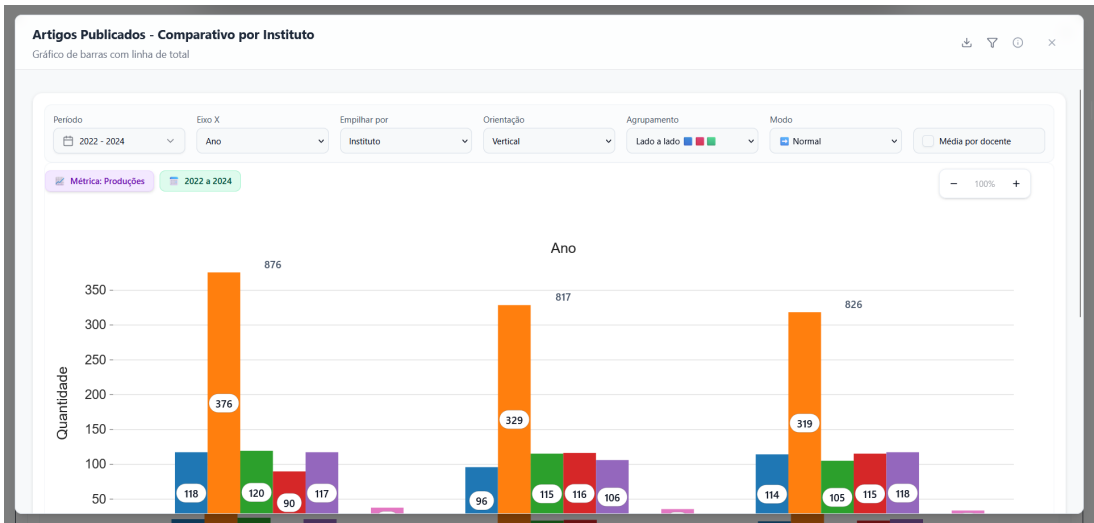


Figura 20 – Indicador expandido (tela cheia).
Fonte: Elaboração própria.

4.7.1.4 Filtros e recortes: mapa de facetas operacionalizadas.

Embora o sistema não apresente uma tela única de “dimensões”, ele operacionaliza um conjunto estável de facetas analíticas que aparecem como filtros e chaves de agregação nos indicadores e na tela de Produções. Na Tabela 21, constam essas facetas como um mapa de verificação (o que pode ser recortado e onde aparece).

Tabela 21 – Mapa de facetas analíticas operacionalizadas no artefato (recortes e onde são aplicados).

Faceta	Recortes usuais	Onde aparece
Temporal	ano, intervalo, últimos N anos; modo normal/cumulativo	Indicadores; Produções
Organizacional	instituto, departamento, programa; hierarquia (pais/filhos)	Telas institucionais; Indicadores; Produções
Tipológica	classe/categoria/tipo (ex.: ALC: artigos/livros/capítulos)	Indicadores; Produções
Colaboração	coautoria externa/internacional; participação discente; redes	Indicadores; Professor (aba Coautorias)
Bibliométrica	citações por fonte/ano; Qualis; impacto OpenAlex (Top 10%/Top 1%)	Indicadores; Professor (aba Impacto)
Base analítica	base declarada <i>versus</i> base deduplicada (consolidada)	Produções; drill-through a partir de indicadores

Nota: ALC = artigos, livros e capítulos (núcleo bibliográfico).

4.7.1.5 Tela de Produções como mecanismo de auditoria.

A tela de Produções funciona como ponto de convergência para verificação: permite busca e filtragem detalhada, alternância entre base declarada e deduplicada, paginação e múltiplos modos de visualização (tabela, cards, grid), além de acesso ao detalhe do registro. Essa alternância permite ao usuário comparar o registro bruto (declarado) com o consolidado (deduplicado), reforçando a verificabilidade do processo descrito na Seção 4.2. Os *drill-throughs* a partir dos indicadores levam diretamente a esta tela com os filtros pré-aplicados, preservando o recorte exibido no gráfico. Nas Figuras 21 e 22, ilustram-se a tela de Produções com seus filtros e resultados paginados.

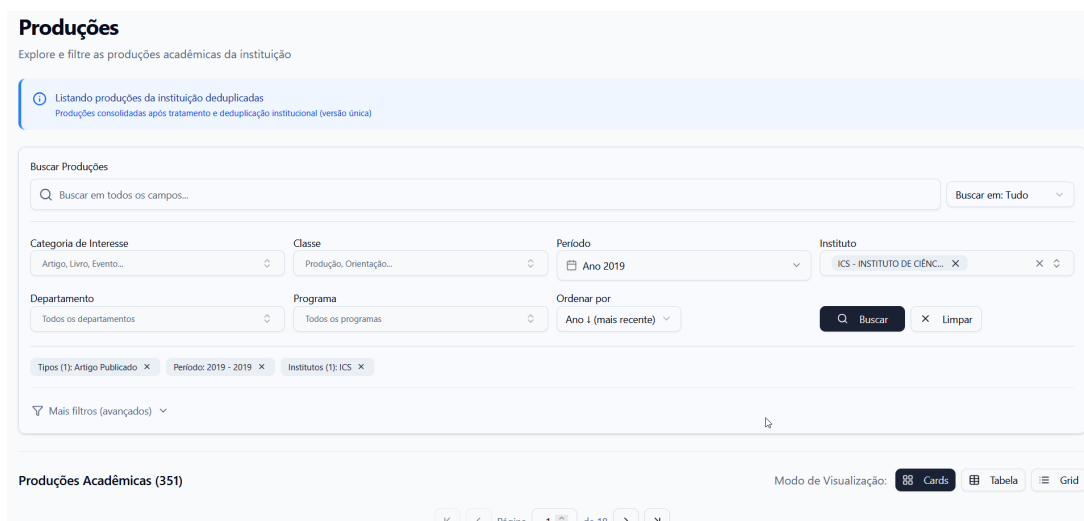


Figura 21 – Tela de Produções com filtros e alternância declarada/deduplicada.

Fonte: Elaboração própria.

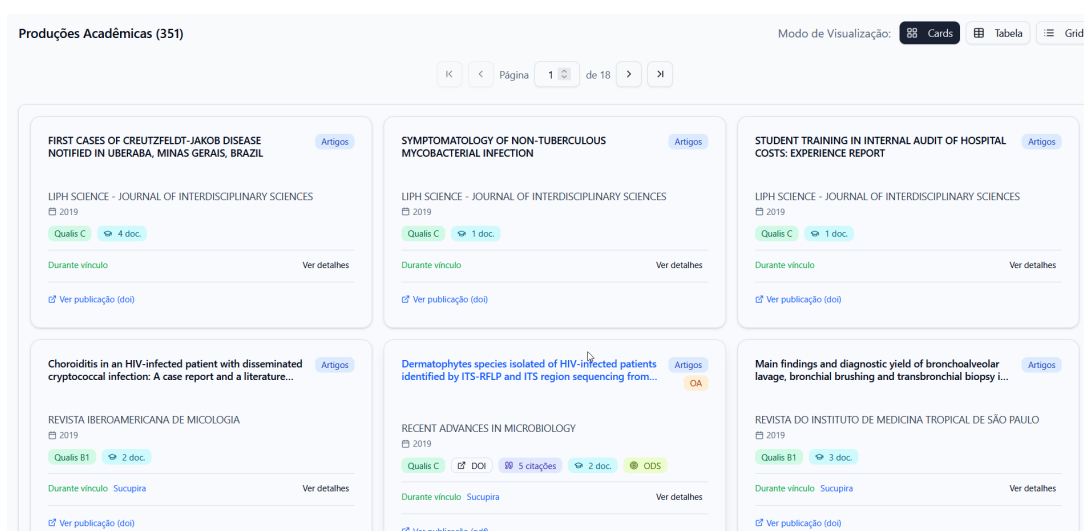


Figura 22 – Tela de Produções: resultado paginado.

Fonte: Elaboração própria.

4.7.1.6 Telas institucionais e composição por blocos.

As telas institucionais (Institutos, Instituto/ID, Departamento/ID, Programas, Programa/ID) seguem um padrão: (i) cabeçalho com descrição, atualização e estrutura (filhos); (ii) cartões de indicadores clicáveis; (iii) tabela/lista de unidades filhas (com navegação hierárquica); (iv) seção de indicadores (mesmo núcleo reutilizável). Nesta seção, o usuário pode selecionar quais indicadores visualizar via um painel lateral de busca/seleção, com botões de acesso também no topo e no final da página (e atalho lateral). Nas Figuras 23 e 24, ilustram-se a tela de instituto com seus componentes.

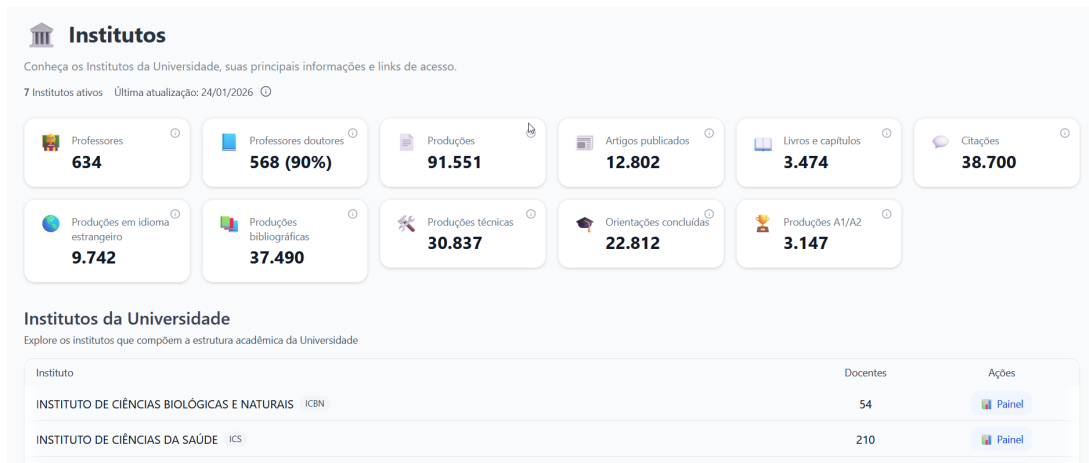


Figura 23 – Tela Instituto/ID: cabeçalho, cartões de indicadores e início da tabela de departamentos.

Fonte: Elaboração própria.

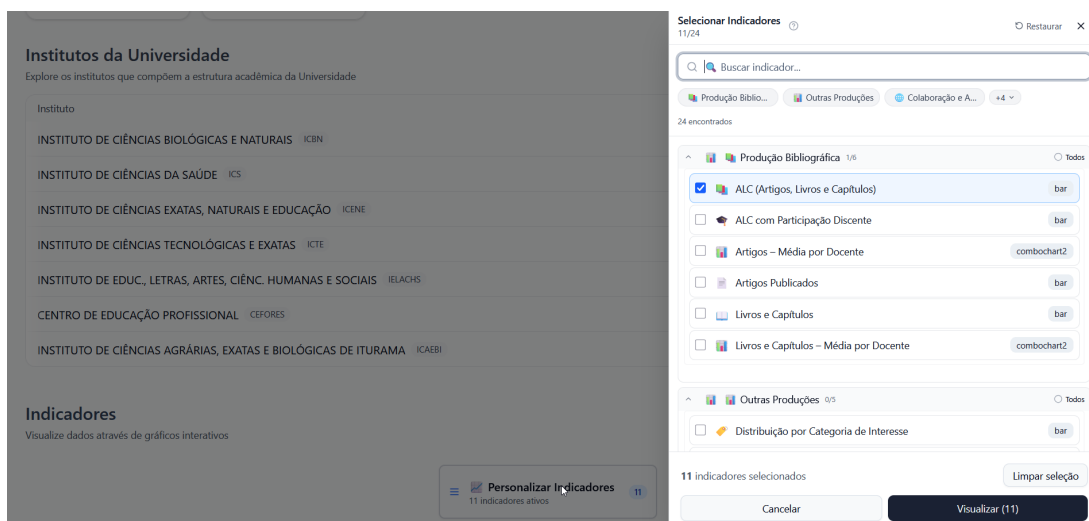


Figura 24 – Tela Instituto/ID: drawer de seleção de indicadores e continuação da tabela.

Fonte: Elaboração própria.

4.7.1.7 Tela do professor e exploração por abas.

A tela do professor funciona como perfil analítico, com cabeçalho (identificação, vínculo organizacional, titulação, links externos) e abas temáticas (Indicadores, Produções, Orientações, Impacto, Coautorias). Nela, os mesmos princípios de verificabilidade se repetem: indicadores com filtros e ações; listas filtradas; e acesso a fontes bibliométricas (OpenAlex, Scholar, OpenCitations etc.) com datas de atualização. Nas Figuras 25 e 26, ilustram-se o cabeçalho do professor e a aba de Impacto.



Figura 25 – Cabeçalho do professor + abas.

Fonte: Elaboração própria.

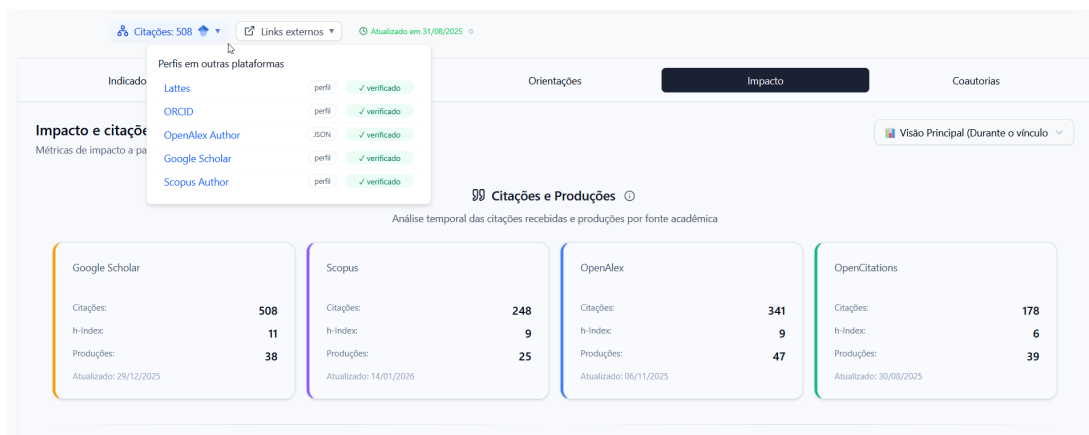


Figura 26 – Aba Impacto com citações por fonte e evolução temporal.

Fonte: Elaboração própria.

4.7.2 Rastreabilidade e validação por consulta

4.7.2.1 Princípio: do agregado ao registro, mantendo o recorte.

A rastreabilidade é operacionalizada pelo ciclo (i) **filtrar** → (ii) **observar** → (iii) **inspecionar** → (iv) **exportar**, preservando o mesmo recorte ao longo das etapas. Na prática, isso se expressa em três mecanismos: **drill-through** do gráfico para listagem, **modal de informação** (contrato de definição/fonte/limitações), e **exportação dos dados** com filtros. A exportação preserva o recorte aplicado, permitindo auditoria externa do mesmo subconjunto observado no gráfico. O mesmo recorte é preservado ao atravessar visualização, inspeção e exportação, reduzindo discrepâncias entre “número visto” e “conjunto auditável” como propriedade de engenharia. Esse desenho favorece a interpretação responsável ao tornar explícitos recortes, denominadores e conjuntos auditáveis, em linha com recomendações de transparência no uso de indicadores (Hicks *et al.*, 2015; Szomszor *et al.*, 2021).

Assim, cada número apresentado no gráfico possui: (a) **recorte explícito** via filtros; (b) **conjunto auditável** via listagem; (c) **contrato operacional** via modal “Info”; e (d) **extração reprodutível** via exportação.

4.7.2.2 Observação sobre filtros passíveis de drill-through.

Nem todas as facetas exibidas em alguns indicadores são necessariamente passadas ao drill-through; o sistema preserva um subconjunto de filtros estáveis (conjunto permitido) para garantir consistência e evitar recortes ambíguos.

4.7.2.3 Fluxo padrão de verificação (5 passos).

O fluxo típico de validação no artefato pode ser descrito em cinco passos: (1) o usuário observa um valor em um indicador agregado (barra/linha/heatmap); (2) aplica ou ajusta filtros (período, unidade, tipo, facetas bibliométricas/colaboração); (3) aciona *drill-through* clicando no elemento do gráfico, abrindo uma modal intermediária com resumo dos dados e botão “Ver Produções”, que direciona à Tela de Produções já filtrada; (4) inspeciona o conjunto retornado (títulos, ano, DOI-chave, metadados e fontes), conferindo a contagem e coerência do recorte; (5) exporta o recorte (CSV/JSON/XLSX ou imagem), de modo a manter alinhamento entre “o que foi baixado” corresponde ao que foi visualizado. Na Figura 27, ilustra-se essa sequência de drill-through.

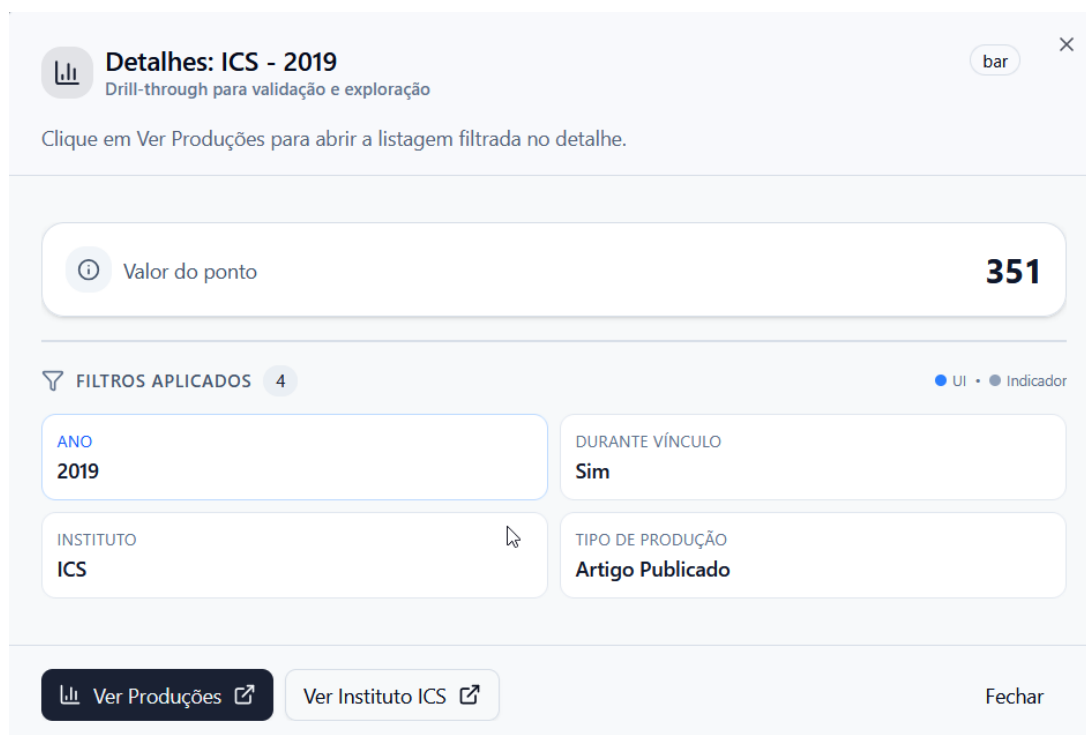


Figura 27 – Sequência de drill-through: gráfico com tooltip e Tela de Produções resultante.

Fonte: Elaboração própria.

4.7.2.4 Exemplo concreto de rastreabilidade (vinculado ao 4.6).

Como exemplo verificável, considere o caso reportado na Seção 4.6.1: **319 artigos publicados no ICS em 2024** (no recorte de participação institucional e período de vínculo). No artefato, o usuário pode: selecionar o instituto ICS e o ano 2024 no indicador temporal; clicar na barra do ano; ser direcionado para a Tela de Produções com filtros equivalentes (instituto = ICS, ano = 2024, tipo = artigoPublicado, base deduplicada, produção durante vínculo); e, então, verificar a presença e metadados dos 319 registros listados, incluindo DOI-chave quando disponível e identificadores de consolidação. O recorte pode ser exportado conforme demonstrado na Figura 18, preservando os filtros aplicados.

4.7.2.5 Modal de “Info” como contrato metodológico.

O ícone “i” de cada indicador fornece documentação operacional do cálculo: definição da métrica, fontes envolvidas (Lattes, OpenAlex, Scholar, OpenCitations, Qualis), observações sobre cobertura e limitações (conforme exemplificado na Figura 19). Esse recurso reduz ambiguidade interpretativa e oferece transparência científica sem exigir que o usuário leia o código ou as queries.

4.7.3 Exemplos de uso e implicações institucionais

Nesta subseção, descrevem-se três cenários representativos de uso, enfatizando: (i) pergunta institucional, (ii) caminho de navegação no sistema, (iii) saída verificável (indicador + drill-through + exportação possível), e (iv) tipo de decisão/encaminhamento que o artefato apoia.

4.7.3.1 Cenário 1: gestão institucional e dinâmica recente de publicações.

Pergunta: há queda recente de artigos publicados em uma unidade específica e qual a magnitude? **Caminho:** Home → Institutos → indicador comparativo por instituto → ICS (detalhe). **Resultado observável:** o ICS apresenta queda de **376 artigos (2022)** para **319 (2024)**, **-15,2%**, enquanto o ICTE exibe crescimento de **90** para **115**, **+27,8%** no mesmo período (conforme Seção 4.6.1 e Figura 1). Ao clicar na barra do ICS em 2024 no indicador temporal do artefato, o drill-through abre a Tela de Produções com 319 registros filtrados (instituto=ICS, ano=2024, tipo=artigoPublicado, base deduplicada), conforme mecanismo demonstrado na Figura 27. **Implicação:** apoio ao monitoramento de variações recentes e à priorização de análises complementares por período e unidade, com recorte rastreável.

4.7.3.2 Cenário 2: coordenação acadêmica e engajamento discente.

Pergunta: a participação discente em artigos e livros/capítulos difere entre unidades e tipos? **Caminho:** Home (ou tela institucional) → indicador de engajamento

discente → seleção de tipo e período (2016–2024). **Resultado observável:** em artigos publicados, o **ICS** apresenta **62,56% com participação discente** (1.875 de 2.997 artigos), enquanto em livros+capítulos, o **IELACHS** exibe apenas **19,39%** (146 de 753), apesar de concentrar **44,1%** de livros+capítulos no ALC institucional (conforme Seção 4.6.2). O mecanismo de drill-through (Figura 27) permite inspecionar os registros subjacentes e confirmar o contraste entre perfis formativos, com exportação dos subconjuntos filtrados conforme Figura 18. **Implicação:** apoio a políticas de formação (IC, TCC, pós-graduação) por tipo e unidade. O contraste entre perfis tipológicos e formativos também ajuda a identificar áreas em que livros/capítulos aparecem mais como produto autoral do que formativo, orientando investigações institucionais.

4.7.3.3 Cenário 3: internacionalização e evidência bibliométrica de alto impacto.

Pergunta: coautoria internacional se associa a maior frequência de alto impacto (Top 10%/Top 1%) no subconjunto com DOI-chave? **Caminho:** Home → indicador bibliométrico (coautoria × impacto) → filtros do núcleo ALC. **Resultado observável:** no subconjunto de **1.755 produções com coautoria internacional**, das quais **1.724 possuem valor normalizado disponível**, observa-se **29,23% no Top 10%** (504/1.724), comparado a **16,97%** para coautoria nacional e **8,37%** sem coautoria externa, resultando em uma razão de proporções **3,5×** (internacional *versus* sem coautoria) e **9,0×** no Top 1% (5,68% *versus* 0,63%) (conforme Tabela 20 na Seção 4.6.3). O drill-through (Figura 27) permite abrir o conjunto subjacente, verificando DOI-chave, instituições parceiras e fonte OpenAlex, com exportação conforme Figura 18. **Implicação:** suporte exploratório a planejamento de internacionalização (monitoramento de colaborações), e comunicação institucional baseada em métricas normalizadas por campo/ano, sem inferir causalidade (associação observável).

4.7.3.4 Síntese: verificabilidade como propriedade de engenharia.

Em conjunto, os mecanismos descritos (filtros, modais de informação, drill-through, exportação consistente e alternância declarada/deduplicada) transformam o artefato em um instrumento de **engenharia de evidência**: resultados agregados podem ser confirmados por inspeção de registros e auditados externamente via exportação, reduzindo a distância entre “visualização” e “prova” no contexto de indicadores institucionais. Esse tipo de deconstrução do agregado em registros é alinhado a recomendações de interpretação responsável de indicadores e transparência de evidência em contextos decisórios (Hicks *et al.*, 2015; Szomszor *et al.*, 2021).

4.8 SÍNTESE E LIMITAÇÕES

4.8.1 Síntese dos principais achados

Esta seção consolida os resultados e discussão do Capítulo 4 em quatro eixos: (i) perfil e consistência do recorte deduplicado; (ii) padrões comparativos (temporais, tipológicos e colaborativos); (iii) padrões com enriquecimento bibliométrico; e (iv) verificabilidade operacional via exploração interativa.

Em síntese, o capítulo mostra que a produção institucional é multifacetada, e que essa diversidade pode ser descrita e auditada por um artefato que integra consolidação, enriquecimento e exploração interativa.

No eixo temporal-organizacional (Seção 4.6.1), observou-se variação entre institutos, com evidência verificável de mudanças recentes no volume de artigos (por exemplo, no ICS, 376→319 entre 2022 e 2024), além de um padrão transversal de ruptura em tipos associados a eventos presenciais. No eixo tipológico-colaborativo (Seção 4.6.2), destacaram-se perfis institucionais distintos (e.g., maior proporção de livros+capítulos no IELACHS no núcleo ALC) e hierarquias de engajamento discente dependentes do tipo de produção. No eixo bibliométrico (Seção 4.6.3), a colaboração internacional mostrou associação robusta com maior frequência de alto impacto (Top 10%/Top 1% normalizados do OpenAlex) no subconjunto com DOI-chave, e padrões consistentes de intensidade de citações por estratos Qualis em anos recentes. Por fim, a Seção 4.7 evidenciou que os resultados são rastreáveis no artefato por filtros, modais de definição, drill-through e exportação consistente.

4.8.2 Limitações e ameaças à validade

As análises deste capítulo são deliberadamente orientadas à **verificabilidade**, porém dependem de escolhas operacionais e da cobertura das fontes. As principais limitações podem ser agrupadas em seis pontos.

4.8.2.1 (L1) Cobertura e viés de DOI-chave.

Resultados que dependem de OpenAlex (impacto normalizado, afiliações e colaboração internacional) exigem um identificador de ligação (DOI-chave). Esse requisito reduz o universo analisado e pode privilegiar áreas e tipos de produção com maior adoção de DOI, afetando comparabilidade com subconjuntos onde DOI é menos frequente (e.g., parte de livros e capítulos, produção técnica e certos campos). Assim, indicadores bibliométricos devem ser lidos como descritores do subconjunto integrável, não do universo total.

4.8.2.2 (L2) Cobertura de afiliações e classificação de colaboração.

A classificação em *sem coautoria externa* / *coautoria nacional* / *coautoria internacional* depende da completude dos registros de afiliação no OpenAlex (*authorships*, instituições e *country_code*). Assim, ausências ou inconsistências de afiliação podem levar a sub/superclassificações, especialmente em produções com metadados incompletos.

4.8.2.3 (L3) Janela temporal e maturação de citações.

Indicadores baseados em citações são sensíveis ao tempo de acumulação: anos recentes tendem a subestimar médias e totais relativos. Por isso, comparações temporais devem privilegiar janelas menos recentes (ou com maior tempo de acumulação), por exemplo excluindo o último ano ou tratando-o separadamente.

4.8.2.4 (L4) Deduplicação e dependência de metadados.

Embora a estratégia de deduplicação seja rastreada e auditável, deduplicações baseadas em heurísticas e metadados estão sujeitas a erro residual: ocorrências distintas podem ser indevidamente agrupadas, e duplicatas podem persistir quando metadados divergem (título, ano, DOI, autoria). O impacto esperado é maior em registros com baixa padronização e em recortes finos (por tipo/ano/unidade) do que em padrões agregados de larga escala. Assim, análises em recortes muito finos (ano×tipo×unidade) são mais sensíveis a erro residual do que tendências agregadas.

4.8.2.5 (L5) Atribuição organizacional e lacunas históricas.

Análises por instituto/departamento dependem da qualidade do vínculo funcional e da identificação de unidades ao longo do tempo. Lacunas históricas (docentes antigos sem departamento registrado, mudanças organizacionais e vínculos incompletos) podem concentrar parte da base de dados em categorias indefinidas ou reduzir precisão da atribuição institucional em subconjuntos específicos. Em consequência, resultados por unidade podem subrepresentar períodos/unidades com cadastro histórico incompleto.

4.8.2.6 (L6) Associação não implica causalidade.

Os padrões comparativos descritos (e.g., colaboração internacional associada a alto impacto) são evidências observacionais na base de dados analisada. Eles não estabelecem causalidade e podem refletir fatores de seleção (áreas, redes, estratégias de publicação) não modelados diretamente. A interpretação causal exigiria controle de variáveis confundidoras (e.g., área de atuação, estratégia de publicação, seleção de parceiros) não disponíveis no recorte analítico.

4.8.3 Implicações para uso e direções de aprimoramento

Apesar das limitações, o conjunto de resultados é apropriado para **diagnóstico exploratório institucional** e **monitoramento rastreável** de perfis e evoluções. Direções plausíveis de aprimoramento incluem: (i) ampliar estratégias de ligação além de DOI (quando aplicável), reduzindo a dependência do subconjunto bibliométrico-integrável; (ii) refinar a atribuição organizacional histórica e o tratamento de lacunas de vínculo; (iii) incorporar análises de sensibilidade por janelas temporais (incluindo maturação de citações e tratamento separado de anos recentes); e (iv) triangular com dados institucionais complementares para sustentar interpretações explicativas mais fortes.

5 CONCLUSÃO

Este trabalho apresentou o desenvolvimento e a aplicação da **Plataforma Capivara**, uma solução tecnológica para curadoria, análise e exploração interativa de produção científica, com foco na verificabilidade e rastreabilidade dos resultados. A partir da consolidação de registros declarados em currículos Lattes, enriquecimento com fontes bibliométricas externas (Crossref, OpenAlex, OpenCitations, Qualis, Scholar, Scopus) e operacionalização de indicadores exploratórios, o artefato materializa o conceito de **engenharia de evidência**: resultados agregados podem ser confirmados por inspeção de registros subjacentes, auditados externamente, por meio da exportação dos dados já recortados e verificados por múltiplas facetas analíticas (temporal, organizacional, tipológica, colaborativa e bibliométrica). A caracterização da produção institucional apresentada no Capítulo 4 evidencia que **não é adequadamente descrita por uma única métrica agregada**, mas exige múltiplas lentes analíticas para capturar heterogeneidade temporal, organizacional, tipológica e bibliométrica.

5.1 CONTRIBUIÇÕES PRINCIPAIS

As contribuições deste trabalho podem ser organizadas em quatro dimensões complementares:

5.1.1 (C1) Metodológica: consolidação rastreável e deduplicação auditável.

A estratégia de deduplicação baseada em DOI-chave e normalização de título/ano/-tipo, aliada à preservação de rastreabilidade (lista de declarantes, identificadores originais e metadados divergentes), viabiliza análises em *base declarada* ou *base deduplicada*, permitindo ao usuário comparar o registro bruto com o consolidado (Seção 4.2). A taxa global de deduplicação (11,29% no recorte de vínculo institucional) quantifica a redução de redundâncias por múltiplas declarações após a consolidação, evidenciando heterogeneidade por tipo de produção e instituto.

5.1.2 (C2) Analítica: padrões observáveis e verificáveis em múltiplas dimensões.

As análises comparativas (Seções 4.6.1–4.6.3) evidenciaram padrões rastreáveis: (i) trajetórias organizacionais heterogêneas entre institutos no período analisado; (ii) perfis tipológicos distintos no núcleo bibliográfico, coerentes com composições disciplinares; (iii) variação do engajamento discente conforme o tipo de produção, indicando articulações diferenciadas entre produção e formação; e (iv) associação entre colaboração internacional e maior presença relativa em faixas de alto impacto normalizado no subconjunto bibliométrico-integrável. Os resultados são descritivos e reproduzíveis na base analisada; interpretações explicativas demandam triangulação adicional.

5.1.3 (C3) Técnica: interoperabilidade e arquitetura modular.

A arquitetura do artefato organiza responsabilidades em camadas: coleta de fontes externas, processamento e consolidação em ambiente analítico, enriquecimento assíncrono com fontes bibliométricas, e interface web para exploração interativa. A modularidade e uso de padrões de interoperabilidade (APIs REST, exportação estruturada, contêineres) facilitam extensões futuras, tanto na integração de novas fontes quanto na adaptação para outros contextos institucionais.

5.1.4 (C4) Institucional: diagnóstico exploratório e suporte à decisão informada.

A plataforma não prescreve uma métrica única de “desempenho”, mas oferece **múltiplas lentes** sobre a produção institucional, permitindo diagnóstico exploratório, comunicação baseada em evidência rastreável e identificação de **unidades setoriais** para análise aprofundada (por exemplo, departamentos e programas). A exploração interativa (Seção 4.7) operacionaliza verificabilidade por meio de filtros, navegação do indicador ao registro (*drill-through*) e exportação consistente, reduzindo discrepâncias entre “número visto” e “conjunto auditável”. Com isso, resultados agregados podem ser **auditados pelo usuário**, aproximando “indicador” e “prova”.

5.2 LIMITAÇÕES E CONTEXTO DE APLICABILIDADE

Conforme discutido na Seção 4.8, as análises dependem de escolhas operacionais e da cobertura das fontes. As principais limitações incluem: (i) cobertura e atualização dos currículos Lattes; (ii) completude de afiliações para classificação de colaboração internacional; (iii) dependência de DOI-chave no enriquecimento bibliométrico; (iv) sensibilidade de citações à janela temporal; (v) erro residual na deduplicação heurística; (vi) lacunas históricas de atribuição organizacional; e (vii) natureza observacional dos padrões.

Essas limitações não comprometem o uso dos resultados para **diagnóstico exploratório** e **monitoramento institucional**, mas delimitam seu escopo de generalização.

5.3 ALINHAMENTO COM OBJETIVOS E PERGUNTA DE PESQUISA

O objetivo geral — *desenvolver uma aplicação para integração e análise de dados institucionais e bibliográficos, voltada à visualização de indicadores estratégicos e ao suporte à tomada de decisão em nível institucional* — foi alcançado pela operacionalização de quatro dimensões verificáveis: (i) integração, consolidação e deduplicação rastreável (Capítulo 3); (ii) enriquecimento bibliométrico com hierarquia de fontes (Seção 4.5); (iii) caracterização e análise comparativa por múltiplas facetas (Seção 4.6.1); e (iv) exploração interativa com verificabilidade operacional (Seção 4.7).

As questões de pesquisa foram respondidas ao: (a) integrar registros dispersos

com rastreabilidade das transformações e documentação das escolhas; (b) fundamentar indicadores úteis à análise institucional por princípios de curadoria, qualidade e interoperabilidade, reconhecendo limitações de cobertura e vieses; e (c) materializar uma solução tecnológica responsiva e reprodutível, com mecanismos de auditoria (*drill-through* e exportação consistente).

5.4 CONSIDERAÇÕES FINAIS

Em um contexto de crescente demanda por transparência, **prestação de contas** e uso de evidência na gestão universitária, este trabalho mostra que é possível construir instrumentos computacionais que conciliam **complexidade analítica** e **acessibilidade operacional**. O artefato desenvolvido constitui um ponto de partida metodológico e técnico para curadoria e exploração sistemática da produção científica, extensível a outros contextos e fontes.

A possibilidade de conferir os resultados reforça a transparência na construção da evidência. Ao explicitar fontes, critérios e limitações e permitir a inspeção dos registros subjacentes, o artefato favorece decisões informadas por dados rastreáveis e passíveis de contestação fundamentada.

6 TRABALHOS FUTUROS

A Plataforma Capivara consolidou um fluxo de curadoria e exploração interativa da produção científica institucional, com ênfase em rastreabilidade e conferência dos resultados. Como evoluções futuras, priorizam-se a ampliação do escopo de dados e o aprofundamento das análises, mantendo o princípio de que todo resultado agregado deve ser conferível nos registros subjacentes.

6.1 AMPLIAÇÃO DO ESCOPO DE DADOS E INDICADORES

6.1.1 (TF1) Expansão para domínios além da produção científica registrada no Lattes.

Incorporar domínios institucionais que complementem a leitura da atividade acadêmica e permitam análises integradas, por exemplo: (i) extensão e ações de impacto social; (ii) iniciação científica e bolsas (IC/IT); (iii) projetos de pesquisa e seus vínculos (financiamento, vigência, equipe); (iv) inovação e transferência (patentes, software, registros tecnológicos); (v) ensino (disciplinas, turmas, participação discente) para análises pesquisa-formação. A expansão demanda modelagem incremental, identificadores consistentes e regras de consolidação específicas por domínio.

6.2 APROFUNDAMENTO ANALÍTICO E EXPLORAÇÃO INTERATIVA

6.2.1 (TF2) Rotulagem temática e taxonomias institucionais.

Aprimorar a análise temática por duas frentes: (i) uso sistemático de classificações já disponíveis em bases externas (e.g., tópicos/áreas do OpenAlex); (ii) definição de taxonomias institucionais (eixos do PDI, vocabulários controlados), aplicadas a produções e novos domínios. O objetivo é viabilizar painéis por “temas institucionais” com critério de rotulagem explicitado.

6.2.2 (TF3) Módulos avançados de análise com rastreabilidade.

Evoluir a interface com módulos especializados, mantendo navegação do agregado ao registro: (i) redes de colaboração/coautoria (estrutura, comunidades e evolução temporal); (ii) análise textual de títulos, palavras-chave e resumos (agrupamentos e similaridade), quando disponíveis; (iii) diagnósticos de qualidade e anomalias (duplicidades residuais, inconsistências) como sinais para investigação.

6.2.3 (TF4) Painéis longitudinais e análises de sensibilidade.

Explorar análises longitudinais que cruzem dimensões organizacionais, tipológicas, formativas e bibliométricas, incluindo análises de sensibilidade (janelas temporais,

critérios de consolidação e cobertura por fonte) para qualificar a robustez dos padrões observados.

6.3 TRANSFERIBILIDADE

6.3.1 (TF5) Adaptação multi-institucional e comparabilidade.

Facilitar a implantação em outras instituições por parametrização de estrutura organizacional, conectores por domínio e padronização do esquema de indicadores. Uma evolução incremental é iniciar por parametrização e, se necessário, avançar para suporte multi-instituição, preservando rastreabilidade.

6.3.2 Síntese.

As direções futuras concentram-se em (i) ampliar o escopo de evidências institucionais e (ii) aprofundar análises e formas de exploração, mantendo o compromisso central do trabalho: produzir indicadores úteis com transparência sobre como os dados foram curados, consolidados e agregados.

7 ORÇAMENTO

A maior parte dos recursos necessários para a execução deste projeto esteve disponível na universidade ou pôde ser acessada gratuitamente. As ferramentas de desenvolvimento (Node.js, React e Next.js) são de código aberto e não exigiram gastos com licenças. As fontes de dados utilizadas incluem bases públicas (por exemplo, Plataforma Lattes e páginas institucionais) e serviços acessíveis no contexto acadêmico (por exemplo, portal CAPES).

Para testes e operação do artefato em ambiente realista, foram utilizados recursos de computação em nuvem. Sempre que possível, adotaram-se *tiers* gratuitos; contudo, houve necessidade de custos temporários com serviços em nuvem (por exemplo, AWS e, após o término do período de testes, ambiente na GCP) para garantir disponibilidade e desempenho durante etapas de validação e uso. Os componentes de infraestrutura (como PostgreSQL e Elasticsearch) foram executados priorizando alternativas gratuitas, institucionais ou de baixo custo.

Adicionalmente, o projeto contou com suporte institucional da **Universidade Federal do Triângulo Mineiro** por meio da **Ação de Desenvolvimento em Serviço para Pós-Graduação Stricto Sensu**, fundamentada no Decreto nº 9.991/2019, no Decreto nº 10.506/2020 e na Instrução Normativa SGP-ENAP/SEDGG/ME nº 21/2021, que viabiliza flexibilização de jornada para dedicação às atividades acadêmicas.

Por fim, destaca-se como insumo relevante o acesso institucional a bases bibliométricas mediante autenticação federada (Rede CAFE) e serviços licenciados pela UFTM, incluindo a viabilização de acesso à API Scopus para fins acadêmicos e não comerciais, dentro de limites estabelecidos pelo provedor.

Dessa forma, o orçamento do projeto considera recursos tecnológicos e eventuais custos financeiros de infraestrutura, bem como investimento institucional indireto e acesso a serviços licenciados necessários à execução.

8 CRONOGRAMA

O projeto foi executado de forma iterativa, com replanejamentos decorrentes de dependências externas (tramitação no CEP, solicitações/limites de APIs e ciclos de carga e reprocessamento de dados). Apresenta-se a seguir a síntese do **cronograma executado** por marcos e entregáveis.

8.1 CRONOGRAMA EXECUTADO (MARCOS)

Tabela 22 – Cronograma executado por marcos e entregáveis

Período	Marco / entregável
T1–T2 (2025)	Revisão de literatura e delimitação do problema/objetivos; definição das fontes e requisitos de rastreabilidade.
T2 (2025)	Prototipação da arquitetura e do pipeline de curadoria; definição de modelos de dados e estratégia de deduplicação.
T3 (2025)	Implementação do processamento e consolidação inicial; primeiros indicadores e validações internas.
T4 (2025)	Infraestrutura de armazenamento/consulta e rotinas de carga (incluindo reprocessamentos); ajustes de qualidade dos dados.
T1 (2026)	Integrações e enriquecimentos bibliométricos (OpenAlex/-Qualis/Scholar) e rotinas de auditoria/exportação.
05/01/2026	Aprovação ética: CEP — CAAE 94775125.7.0000.5154; Parecer nº 8.090.187 (Uberaba–MG).
T2 (2026)	Interface web e exploração interativa (filtros, definições, drill-through); estabilização de fluxos críticos.
T3–T4 (2026)	Consolidação das análises, geração de resultados e redação final (capítulos de resultados, discussão e conclusão).

REFERÊNCIAS

BAR-ILAN, Judit. **Which H-Index?—A Comparison of WoS, Scopus and Google Scholar**. *Scientometrics*, v. 74, p. 257–271, 1 fev. 2008. DOI: 10.1007/s11192-008-0216-y.

BELTER, Christopher W. **Bibliometric Indicators: Opportunities and Limits**. *Journal of the Medical Library Association - JMLA*, v. 103, n. 4, p. 219–221, out. 2015. ISSN 1536-5050, 1558-9439. DOI: 10.3163/1536-5050.103.4.014. Disponível em: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4613388/>. Acesso em: 13 jan. 2026.

BRASIL. **Lei Nº 12.527, de 18 de Novembro de 2011**. Brasília: Presidência da República, nov. 2011. Acesso em: 20 out. 2024.

BREMBS, Björn. **Prestigious Science Journals Struggle to Reach Even Average Reliability**. *Frontiers in Human Neuroscience*, Frontiers, v. 12, 20 fev. 2018. ISSN 1662-5161. DOI: 10.3389/fnhum.2018.00037. Disponível em: <https://www.frontiersin.org/journals/human-neuroscience/articles/10.3389/fnhum.2018.00037/full>. Acesso em: 13 jan. 2026.

CAPES. **Avaliação Quadrienal 2021 - 2024**. Brasília, 2025. Disponível em: <https://www.gov.br/capes/pt-br/acesso-a-informacao/acoes-e-programas/avaliacao/avaliacao-quadrienal/avaliacao-quadrienal-2021-2024>. Acesso em: 2 out. 2025.

CASTRO, Pablo de. **The Role of Current Research Information Systems (CrIS) in Supporting Open Science Implementation: The Case of Strathclyde**. *ITlib. Informačné technológie a knižnice*, p. 21–30, Special Issue 2018 2018. DOI: 10.25610/itlib-2018-0003.

CONSELHO NACIONAL DE DESENVOLVIMENTO CIENTÍFICO E TECNOLÓGICO. **Resolução Normativa nº 01, de 23 de março de 2023: dispõe sobre o Extrator da Plataforma Lattes**. [*S. l.: s. n.*], 2025. Diário Oficial da União: seção 1, Brasília, DF, 26 mar. 2025, p. 9. Dispõe sobre o Extrator da Plataforma Lattes. Disponível em: <https://pesquisa.in.gov.br/imprensa/jsp/visualiza/index.jsp?data=26/03/2025&jornal=515&pagina=9>. Acesso em: 26 jan. 2026.

CONWAY, Paul. **Modeling the Digital Content Landscape in Universities**. Edição: Michael Seadle. *Library Hi Tech*, v. 26, n. 3, p. 342–354, 5 set. 2008. ISSN 0737-8831. DOI: 10.1108/07378830810903283. Disponível em: <http://www.emerald.com/lht/article/26/3/342-354/261404>. Acesso em: 13 jan. 2026.

CULBERT, Jack H. *et al.* **Reference Coverage Analysis of OpenAlex Compared to Web of Science and Scopus**. *Scientometrics*, v. 130, n. 4, p. 2475–2492, 1 abr. 2025. ISSN 1588-2861. DOI: 10.1007/s11192-025-05293-3. Disponível em: <https://doi.org/10.1007/s11192-025-05293-3>. Acesso em: 13 jan. 2026.

DIGIAMPIETRI, Luciano Antonio *et al.* **EXTRAÇÃO, CARACTERIZAÇÃO E ANÁLISES DE DADOS DE CURRÍCULOS LATTES**. *Revista Eletrônica de Sistemas de Informação*, v. 14, n. 2, 31 ago. 2015. ISSN 1677-3071. DOI:

10.21529/RESI.2015.1402001. Disponível em:

<https://www.periodicosibepes.org.br/index.php/reinfo/article/view/1528>.

Acesso em: 14 jan. 2026.

FORTUNATO, Santo *et al.* **Science of Science**. *Science (New York, N.Y.)*, v. 359, n. 6379, eaa0185, 2 mar. 2018. ISSN 1095-9203. DOI: 10.1126/science.aao0185.

PMID: 29496846.

GOUVÊA, Alessandra Lacerda *et al.* **Índice H dos pesquisadores brasileiros: um olhar comparativo entre as bases de dados WoS, Scopus e Google Scholar**.

Research, Society and Development, v. 11, n. 5, e13711527832–e13711527832, 31 mar. 2022. ISSN 2525-3409. DOI: 10.33448/rsd-v11i5.27832. Disponível em:

<https://rsdjournal.org/rsd/article/view/27832>. Acesso em: 13 jan. 2026.

HAUSCHKE, Christian *et al.* **Roadmap to FAIR Research Information in Open Infrastructures**. *Journal of Library Metadata*, v. 21, n. 1–2, p. 45–61, 3 abr. 2021.

ISSN 1938-6389, 1937-5034. DOI: 10.1080/19386389.2021.1999156. Disponível em:

<https://www.tandfonline.com/doi/full/10.1080/19386389.2021.1999156>. Acesso em: 13 jan. 2026.

HICKS, Diana *et al.* **The Leiden Manifesto for Research Metrics**. *Nature*, v. 520, p. 429–431, 23 abr. 2015. DOI: 10.1038/520429a.

KINDLING, Maxi; STRECKER, Dorothea. **Data Quality Assurance at Research Data Repositories**. *Data Science Journal*, v. 21, n. 1, 22 nov. 2022. ISSN

1683-1470. DOI: 10.5334/dsj-2022-018. Disponível em:

<https://datascience.codata.org/articles/dsj-2022-018>. Acesso em: 13 jan. 2026.

LIAW, Siaw-Teng *et al.* **Quality Assessment of Real-World Data Repositories across the Data Life Cycle: A Literature Review**. *Journal of the American Medical Informatics Association : JAMIA*, v. 28, n. 7, p. 1591–1599, 26 jan. 2021.

ISSN 1067-5027. DOI: 10.1093/jamia/ocaa340. PMID: 33496785. Disponível em:

<https://pmc.ncbi.nlm.nih.gov/articles/PMC8475229/>. Acesso em: 13 jan. 2026.

MALTESE, Vincenzo. **Digital Transformation Challenges for Universities: Ensuring Information Consistency Across Digital Services**. *Cataloging & Classification Quarterly*, Routledge, v. 56, n. 7, p. 592–606, 3 out. 2018. ISSN

0163-9374. DOI: 10.1080/01639374.2018.1504847. Disponível em:

<https://doi.org/10.1080/01639374.2018.1504847>. Acesso em: 13 jan. 2026.

MANGHI, Paolo. **Challenges in Building Scholarly Knowledge Graphs for Research Assessment in Open Science**. *Quantitative Science Studies*, v. 5, n. 4, p. 991–1021, 1 nov. 2024. ISSN 2641-3337. DOI: 10.1162/qss_a_00322. Disponível em:

https://doi.org/10.1162/qss_a_00322. Acesso em: 13 jan. 2026.

MAZONI, Alysson Fernandes *et al.* **Exploring Interoperability Between Local and Global Databases in Scientometrics: Lattes, Capes, and OpenAlex.** 21 jul. 2025. DOI: 10.1590/SciELOPreprints.12668. Disponível em: <https://preprints.scielo.org/index.php/scielo/preprint/view/12668>. Acesso em: 13 jan. 2026. Pré-publicado.

MELO, João Henrick Neri de; TRINCA, Tatiane Pacanaro; MARICATO, João de Melo. **Limites dos indicadores bibliométricos de bases de dados internacionais para avaliação da Pós-Graduação brasileira: a cobertura da *Web of Science* nas diferentes áreas do conhecimento.** *Transinformação*, Pontifícia Universidade Católica de Campinas, v. 33, e200071, 2021. ISSN 0103-3786, 2318-0889. DOI: 10.1590/2318-0889202133e200071. Disponível em: <https://www.scielo.br/j/tinf/a/KmfGWMDK8zSvZszWfSX4VnP/?lang=pt>. Acesso em: 14 jan. 2026.

MENA-CHALCO, Jesús Pascual *et al.* **Brazilian Bibliometric Coauthorship Networks.** *Journal of the Association for Information Science and Technology*, v. 65, n. 7, p. 1424–1445, jul. 2014. ISSN 2330-1635, 2330-1643. DOI: 10.1002/asi.23010. Disponível em: <https://asistdl.onlinelibrary.wiley.com/doi/10.1002/asi.23010>. Acesso em: 14 jan. 2026.

MITCHELL, Sonia Natalie *et al.* **FAIR Data Pipeline: Provenance-Driven Data Management for Traceable Scientific Workflows.** *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, v. 380, n. 2233, p. 20210300, 3 out. 2022. ISSN 1364-503X, 1471-2962. DOI: 10.1098/rsta.2021.0300. Disponível em: <https://royalsocietypublishing.org/doi/10.1098/rsta.2021.0300>. Acesso em: 21 jan. 2026.

MUGNAINI, Rogério; QUONIAM, Luc. **Indicadores bibliométricos da produção científica brasileira: uma análise a partir da base Pascal.** *Ciência da Informação*, Brasília, v. 33, n. 2, p. 123–131, 2004.

PIRES, Alause Da Silva *et al.* **Implicações do sistema de classificação de periódicos Qualis em práticas de publicação no Brasil entre 2007 e 2016.** *Education Policy Analysis Archives*, v. 28, p. 25, 17 fev. 2020. ISSN 1068-2341. DOI: 10.14507/epaa.28.4353. Disponível em: <https://epaa.asu.edu/index.php/epaa/article/view/4353>. Acesso em: 14 jan. 2026.

SEGLÉN, Per O. **Citation Frequency and Journal Impact: Valid Indicators of Scientific Quality?** *Journal of Internal Medicine*, v. 229, n. 2, p. 109–111, fev. 1991. ISSN 0954-6820, 1365-2796. DOI: 10.1111/j.1365-2796.1991.tb00316.x. Disponível em: <https://onlinelibrary.wiley.com/doi/10.1111/j.1365-2796.1991.tb00316.x>. Acesso em: 14 jan. 2026.

SÖDERLIND, Johan; GESCHWIND, Lars. **Making Sense of Academic Work: The Influence of Performance Measurement in Swedish Universities. Policy Reviews in Higher Education**, v. 3, n. 1, p. 75–93, 2 jan. 2019. ISSN 2332-2969, 2332-2950. DOI: 10.1080/23322969.2018.1564354. Disponível em: <https://www.tandfonline.com/doi/full/10.1080/23322969.2018.1564354>. Acesso em: 15 jan. 2026.

SZOMSZOR, Martin *et al.* **Interpreting Bibliometric Data. Frontiers in Research Metrics and Analytics**, v. 5, p. 628703, 9 fev. 2021. ISSN 2504-0537. DOI: 10.3389/frma.2020.628703. Disponível em: <https://www.frontiersin.org/articles/10.3389/frma.2020.628703/full>. Acesso em: 13 jan. 2026.

TAŞKIN, Zehra. **Sustaining the “Frozen Footprints” of Scholarly Communication through Open Citations: An Annual Review of Information Science and Technology (ARIST) Paper. Journal of the Association for Information Science and Technology**, v. 77, n. 1, p. 23–39, jan. 2025. ISSN 2330-1635, 2330-1643. DOI: 10.1002/asi.24982. Disponível em: <https://asistdl.onlinelibrary.wiley.com/doi/10.1002/asi.24982>. Acesso em: 21 jan. 2026.

THIEDIG, Christoph; WEGNER, Antje. **Evidence use in higher education decision-making and policy: a scoping review of empirical studies from 2010 to 2022. London Review of Education**, UCL Press, v. 22, n. 1, 6 nov. 2024. ISSN 1474-8479. DOI: 10.14324/LRE.22.1.36. Disponível em: <https://journals.uclpress.co.uk/lre/article/pubid/LRE-22-36/>. Acesso em: 15 jan. 2026.

UNIVERSIDADE FEDERAL DO TRIÂNGULO MINEIRO. **Plano de Desenvolvimento Institucional – PDI 2025–2029**, 2025. Disponível em: <https://www.uftm.edu.br/proplan/planejamento-e-desenvolvimento/planejamento-estrategico/pdi>. Acesso em: 3 out. 2025.

WILKINSON, Mark D. *et al.* **The FAIR Guiding Principles for Scientific Data Management and Stewardship. Scientific Data**, Nature Publishing Group, v. 3, n. 1, p. 160018, 15 mar. 2016. ISSN 2052-4463. DOI: 10.1038/sdata.2016.18. Disponível em: <https://www.nature.com/articles/sdata201618>. Acesso em: 15 jan. 2026.

APÊNDICE A – CONSTRUÇÃO E CATÁLOGO DE INDICADORES

APÊNDICE A.1 OBJETIVO DO APÊNDICE

Este apêndice descreve a lógica metodológica adotada para a construção dos indicadores analíticos do sistema, sem apresentação de resultados empíricos. Os indicadores atuam como camada intermediária entre os dados consolidados e as análises institucionais, organizando informações de forma a permitir interpretações institucionais consistentes.

O detalhamento aqui apresentado complementa a visão sintética da seção principal, explicitando tipologias, métricas, critérios de agregação e a dualidade analítica adotada.

APÊNDICE A.2 TIPOLOGIA DOS INDICADORES

Os indicadores construídos pela plataforma organizam-se em tipologias analíticas não excludentes, permitindo múltiplas leituras do mesmo conjunto de dados conforme o contexto institucional.

De forma conceitual, distinguem-se três naturezas principais:

- a) **Indicadores institucionais:** referem-se ao volume, distribuição e evolução temporal da produção acadêmica, sem considerar métricas externas de impacto. Exemplos incluem contagem de produções por unidade, evolução anual da produção e distribuição por tipo de produção;
- b) **Indicadores bibliométricos:** incorporam métricas de impacto científico, como citações, fatores de impacto de periódicos e classificações Qualis. Esses indicadores contextualizam a produção institucional em seu ambiente científico ampliado;
- c) **Indicadores colaborativos:** analisam padrões de coautoria, internacionalização e participação discente, revelando redes de colaboração e características da produção conjunta.

Essas tipologias não são mutuamente excludentes: um mesmo indicador pode apresentar simultaneamente características institucionais, bibliométricas e colaborativas, conforme as dimensões analíticas aplicadas.

APÊNDICE A.3 MÉTRICAS E OPERAÇÕES FUNDAMENTAIS

As métricas adotadas na construção dos indicadores baseiam-se em operações fundamentais de contagem, agregação e distribuição, priorizando simplicidade, interpretabilidade e reprodutibilidade.

As operações fundamentais incluem:

- a) **Contagens absolutas:** quantificação simples de produções, autores ou eventos, sem normalização;

- b) **Agregações:** soma ou agrupamento de valores segundo critérios definidos (por exemplo, somatório de citações por unidade organizacional);
- c) **Distribuições:** cálculo de proporções, percentuais e participações relativas, permitindo comparações entre categorias de diferentes magnitudes.

Essas operações são aplicadas de forma consistente em todos os indicadores, garantindo comparabilidade metodológica e facilidade de interpretação institucional.

APÊNDICE A.4 CRITÉRIOS DE AGREGAÇÃO

A agregação dos dados é realizada por eixos analíticos recorrentes, que estruturam a organização dos indicadores e definem as dimensões disponíveis para exploração.

Os eixos analíticos principais incluem:

- a) **Tempo:** ano de publicação, períodos agregados ou recortes temporais específicos;
- b) **Unidades organizacionais:** institutos, departamentos ou outras subdivisões da estrutura institucional;
- c) **Tipos de produção:** artigos, capítulos de livro, livros, trabalhos em eventos, patentes, entre outros;
- d) **Pessoas:** docentes, discentes ou autores externos vinculados às produções.

Dimensões temáticas adicionais são incorporadas quando coerentes com o indicador analisado, incluindo idioma de publicação, país de origem, áreas de conhecimento e classificação Qualis dos periódicos. A aplicação dessas dimensões respeita a disponibilidade de dados e a pertinência metodológica de cada indicador.

APÊNDICE A.5 DUALIDADE ANALÍTICA

Um princípio central da metodologia é a manutenção explícita da **dualidade analítica** entre a visão declarativa (preservando as declarações individuais dos currículos) e a visão deduplicada (institucional), evitando sobrecontagens em análises agregadas.

A visão declarativa reflete a produção conforme informada por cada docente em seu currículo Lattes, sem consolidação de coautorias. Essa perspectiva é útil para auditorias curriculares, avaliações individuais e verificação de declarações.

A visão deduplicada consolida produções em coautoria, contabilizando cada produção institucional uma única vez, independentemente do número de autores internos. Essa perspectiva é adequada para análises institucionais agregadas, comparações entre unidades e avaliação de impacto científico.

Essa dualidade não é tratada como mera decisão técnica, mas como escolha metodológica e institucional, que permite atender simultaneamente a demandas de auditoria

curricular e de avaliação institucional, sem comprometer a integridade de nenhuma das perspectivas.

APÊNDICE A.6 SÍNTESE DO APÊNDICE

Os indicadores construídos pela plataforma *Capivara* baseiam-se em tipologias flexíveis, métricas fundamentais e critérios de agregação explícitos. A dualidade analítica entre visões declarativa e deduplicada constitui decisão metodológica central, assegurando que o sistema atenda a demandas institucionais diversas sem comprometer a coerência analítica.

APÊNDICE B – EXPLORAÇÃO ANALÍTICA E VISUALIZAÇÃO

APÊNDICE B.1 OBJETIVO DO APÊNDICE

Este apêndice descreve o espaço metodológico de exploração analítica disponibilizado pelo sistema, entendendo a visualização como extensão operacional da metodologia, e não como etapa interpretativa isolada.

O detalhamento aqui apresentado complementa a visão sintética da seção principal, explicitando dimensões analíticas, filtros, operações de exploração e mecanismos de exportação que asseguram reprodutibilidade e transparência dos resultados.

APÊNDICE B.2 DIMENSÕES ANALÍTICAS

A exploração dos indicadores é organizada a partir de dimensões analíticas explícitas, que estruturam a navegação entre diferentes níveis de agregação sem alteração das regras metodológicas de cálculo.

As dimensões centrais incluem:

- a) **Tempo:** exploração por ano, períodos agregados ou recortes temporais específicos;
- b) **Unidades organizacionais:** navegação hierárquica entre institutos, departamentos e outras subdivisões institucionais;
- c) **Tipo de produção:** segmentação por categorias de produções acadêmicas (artigos, livros, capítulos, trabalhos em eventos, entre outros);
- d) **Recortes temáticos específicos:** dimensões adicionais conforme a natureza de cada indicador, incluindo classificação Qualis, idioma, país de origem e áreas de conhecimento.

Essas dimensões permitem múltiplas leituras do mesmo conjunto de dados, adequando a exploração ao contexto analítico institucional.

APÊNDICE B.3 FILTROS E CONTROLE DE GRANULARIDADE

Os filtros atuam como mecanismos de refinamento do recorte dos dados, permitindo seleção interativa de subconjuntos sem alteração das regras de cálculo dos indicadores.

O sistema permite:

- a) **Navegação hierárquica:** exploração progressiva entre níveis organizacionais (instituição, instituto, departamento);
- b) **Controle de granularidade:** ajuste do nível de detalhe temporal (anual, por período, agregado);

- c) **Seleção dinâmica de categorias:** escolha interativa de tipos de produção, áreas de conhecimento ou outras dimensões disponíveis.

Esses controles asseguram flexibilidade analítica sem comprometer a consistência metodológica dos indicadores.

APÊNDICE B.4 OPERAÇÕES DE EXPLORAÇÃO

O sistema suporta operações de exploração que possibilitam aprofundamento analítico progressivo, do nível institucional agregado a subconjuntos específicos.

As operações incluem:

- a) ***Drill-down***: aprofundamento hierárquico, permitindo transição de visões agregadas para visões detalhadas (por exemplo, de instituto para departamento);
- b) **Troca de eixos analíticos**: reorganização da apresentação dos dados segundo diferentes dimensões, sem alteração dos valores calculados;
- c) **Seleção interativa de categorias**: escolha de subconjuntos específicos para análise comparativa ou focalizada.

Essas operações respeitam os limites metodológicos definidos para cada indicador, garantindo que a flexibilidade exploratória não introduza inconsistências analíticas.

APÊNDICE B.5 MECANISMOS DE EXPORTAÇÃO

As funcionalidades de exportação preservam o recorte analítico aplicado (filtros, dimensões e perspectiva deduplicada ou declarativa), assegurando reprodutibilidade e transparência dos resultados.

Os mecanismos de exportação incluem:

- a) **Preservação do recorte analítico**: filtros, dimensões e perspectiva (deduplicada ou declarativa) são mantidos nos dados exportados;
- b) **Registro de metadados**: informações sobre data de geração, parâmetros aplicados e origem dos dados são incluídas;
- c) **Separação entre dado estruturado e representação visual**: os indicadores podem ser exportados em formatos estruturados (JSON, CSV) ou visuais (gráficos), permitindo reutilização externa.

Essa abordagem assegura que análises realizadas na plataforma possam ser reproduzidas, auditadas e estendidas por usuários institucionais.

APÊNDICE B.6 SÍNTESE DO APÊNDICE

A exploração analítica disponibilizada pela plataforma *Capivara* organiza-se por dimensões explícitas, filtros controlados e operações definidas, permitindo flexibilidade interpretativa sem comprometer a consistência metodológica. Os mecanismos de exportação asseguram reprodutibilidade e transparência, alinhando a visualização aos princípios metodológicos do trabalho.

APÊNDICE C – REGRAS DE INTEGRAÇÃO E DEDUPLICAÇÃO DAS PRODUÇÕES

APÊNDICE C.1 OBJETIVO DO APÊNDICE

Este apêndice descreve, de forma sintética, as regras metodológicas adotadas pela plataforma *Capivara* para a integração, normalização e deduplicação das produções acadêmicas. O foco está nas decisões conceituais e critérios operacionais, não em detalhes algorítmicos de implementação.

As regras aqui apresentadas sustentam a transformação de registros curriculares individuais em produções institucionais consolidadas, preservando simultaneamente a perspectiva declarativa do docente e a visão agregada da instituição.

APÊNDICE C.2 PRINCÍPIOS METODOLÓGICOS

A integração e deduplicação das produções seguem os seguintes princípios fundamentais:

- a) **Primazia institucional do currículo Lattes:** apenas produções declaradas no Lattes são consideradas válidas no corpus institucional;
- b) **Não substituição de dados declarados:** processos automáticos não removem nem alteram produções informadas pelos docentes;
- c) **Rastreabilidade completa:** toda produção consolidada mantém vínculo explícito com suas declarações originais;
- d) **Dualidade analítica:** o sistema preserva visões individual (por docente) e institucional (deduplicada).

APÊNDICE C.3 INTEGRAÇÃO DAS PRODUÇÕES

Inicialmente, cada produção é tratada como pertencente exclusivamente ao docente que a declarou em seu currículo. A integração ocorre somente após a totalidade das produções individuais estar disponível no sistema.

O processo de integração envolve:

- a) Identificação de coautorias institucionais;
- b) Associação das produções às unidades organizacionais correspondentes;
- c) Definição do vínculo temporal da produção com a instituição.

A associação temporal é realizada com base no ano de publicação, considerando-se que a produção pertence à instituição quando ocorre durante o período de vínculo do autor com a universidade.

APÊNDICE C.4 CRITÉRIOS DE DEDUPLICAÇÃO

A deduplicação visa identificar produções idênticas ou equivalentes declaradas por diferentes autores institucionais. Para isso, são aplicados critérios hierárquicos:

1. **Identificadores persistentes:** produções com o mesmo DOI, ISBN, ISSN associado ou número de patente são automaticamente agrupadas;
2. **Correspondência por atributos normalizados:** na ausência de identificadores, utiliza-se a combinação de título normalizado, ano de publicação e tipo de produção.

Cada grupo de produções equivalentes recebe um identificador institucional único, permitindo análises consolidadas sem contagem inflada.

APÊNDICE C.5 VERSÃO PRINCIPAL DA PRODUÇÃO

Dentro de cada grupo deduplicado, é definida uma versão principal, utilizada como referência para análises institucionais e enriquecimento bibliométrico.

A seleção da versão principal segue os seguintes critérios:

- a) Prioridade para produções cujo primeiro autor possua vínculo institucional no ano da publicação;
- b) Na ausência dessa condição, seleção da produção associada ao autor institucional mais bem posicionado na ordem de autoria.

Esse critério busca maximizar a qualidade e completude do registro de referência, mantendo coerência metodológica.

APÊNDICE C.6 PRESERVAÇÃO DA VISÃO DECLARATIVA

Apesar da deduplicação institucional, o sistema preserva integralmente todas as declarações individuais dos docentes. Dessa forma, é possível:

- a) Analisar a produção conforme declarada por cada docente;
- b) Atender a demandas de avaliação individual e progressão funcional;
- c) Comparar visões individuais e institucionais sem perda de informação.

APÊNDICE C.7 CONSIDERAÇÕES FINAIS

As regras de integração e deduplicação adotadas pela plataforma *Capivara* equilibram automação e fidelidade institucional, permitindo análises consistentes sem descaracterizar a natureza declarativa da produção acadêmica. Essas decisões metodológicas são centrais para a confiabilidade dos indicadores gerados pelo sistema.

APÊNDICE D – TAXONOMIAS E REGRAS DE AGREGAÇÃO

Este apêndice descreve as convenções analíticas adotadas no *pipeline* de processamento dos dados do Lattes, incluindo definições de unidades de análise, critérios de classificação e regras de contabilização por unidade organizacional.

APÊNDICE D.1 UNIDADES E NÍVEIS DE ANÁLISE

O *pipeline* distingue duas unidades fundamentais:

Declaração Registro individual extraído do currículo Lattes de um docente. Uma mesma produção pode ser declarada múltiplas vezes por diferentes docentes (ou pelo mesmo docente com variações descritivas).

Produção deduplicada Registro representativo de um grupo de declarações identificadas como referentes à mesma produção (agrupamento por chave composta de atributos como título, ano, DOI, ISBN, etc.). O número de produções deduplicadas corresponde ao número de grupos formados após o processo de consolidação descrito na Seção 4.2. Uma produção deduplicada corresponde, portanto, a um mapeamento *N para 1*, no qual múltiplas declarações são consolidadas em um único registro analítico representativo.

Convenção geral: As análises de volume utilizam produções deduplicadas (um registro por grupo), enquanto análises de coautoria e participação consideram o número de declarações ou docentes associados a cada grupo. Essa distinção é fundamental para evitar a superestimação de volumes quando análises institucionais são realizadas diretamente sobre declarações.

APÊNDICE D.2 CLASSIFICAÇÃO POR CLASSE DE INDICADOR

O campo *classe de indicador* agrupa os registros em três categorias operacionais, utilizadas para fins de filtragem analítica institucional:

producao Registros correspondentes a produtos finalizados e consolidados da atividade acadêmica. Inclui tipos associados, no Lattes, a demais trabalhos, orientações concluídas, produção artística/cultural, produção bibliográfica e produção técnica.

atividadesComplementares Registros de atividades acadêmicas complementares (participações em eventos, bancas, organização de eventos etc.). Inclui tipos classificados, no Lattes, como outra produção, exceto aqueles identificados como em andamento.

emAndamento Orientações e outras atividades ainda em desenvolvimento. Inclui tipos da categoria outra produção cujo nome contém o marcador de atividade em andamento.

Observação metodológica: Trata-se de um *mapeamento operacional*, criado exclusivamente para fins analíticos institucionais, não refletindo necessariamente a classificação original adotada pelo Lattes. O recorte `classe_indicador = 'producao'` é utilizado como filtro preferencial em análises comparativas, por capturar atividades consolidadas e excluir registros em andamento ou de natureza complementar.

APÊNDICE D.3 AGRUPAMENTO POR CATEGORIA DE INTERESSE

O campo *categoria de interesse* agrupa os 70 tipos nativos do Lattes em 10 categorias semânticas, com o objetivo de facilitar leituras agregadas e comparações analíticas. Na Tabela 23, apresenta-se o mapeamento completo.

Tabela 23 – Mapeamento de tipos do Lattes para categoria de interesse

Categoria de interesse	Tipos incluídos
Artigos	artigos aceitos para publicação; artigos publicados
Livros	livros publicados ou organizados
Capítulos	capítulos de livro publicados
Textos bibliográficos	prefácios e posfácios; textos em jornal ou revista; traduções; outras produções bibliográficas
Trabalhos em eventos	trabalhos apresentados em eventos
Participações em eventos	apresentação de trabalho; organização de evento; participação em congresso, encontro, exposição, feira, oficina, olimpíada, seminário e simpósio; outras participações em eventos e congressos
Orientações	orientações em andamento de aperfeiçoamento/especialização, doutorado, graduação, iniciação científica, mestrado e pós-doutorado; orientações concluídas de doutorado, mestrado e pós-doutorado; outras orientações concluídas e em andamento
Bancas	bancas julgadoras para avaliação de cursos, concurso público, livre-docência e professor titular; participações em bancas de aperfeiçoamento/especialização, doutorado, exame de qualificação, graduação e mestrado; outras bancas julgadoras e outras participações em banca
Produção técnica	patente; software; produto tecnológico; processos ou técnicas; trabalho técnico; desenvolvimento de material didático ou instrucional; curso de curta duração ministrado; relatório de pesquisa; programa de rádio ou TV; carta, mapa ou similar; desenho industrial; editoração; maquete; marca; mídia social, website ou blog; manutenção de obra artística; outras produções técnicas
Produção artística/cultural	artes cênicas; artes visuais; música; curso de curta duração; outras produções artísticas/culturais
Outras	demais trabalhos; outra produção; e demais tipos não classificados nas categorias acima

Fonte: Elaboração própria a partir da taxonomia do Lattes.

Núcleo bibliográfico (ALC): Subconjunto da produção bibliográfica formado por **artigos publicados, livros publicados ou organizados e capítulos de livros**

publicados. Este recorte é adotado como referência preferencial em análises de integração bibliométrica devido ao maior potencial de padronização descritiva e uso de identificadores persistentes.

APÊNDICE D.4 MODO PARTICIPAÇÃO: CONTABILIZAÇÃO POR UNIDADE ORGANIZACIONAL

A contabilização de produções por unidade organizacional (instituto, departamento ou programa de pós-graduação) utiliza o *modo participação*, definido pelas seguintes regras:

1. **Definição de participação:** Uma unidade é associada a uma produção deduplicada quando **ao menos um docente com vínculo institucional válido** participa do respectivo grupo de declarações (i.e., ao menos uma das declarações do grupo foi feita por docente vinculado àquela unidade no período analisado).
2. **Contabilização múltipla entre unidades:** Uma mesma produção deduplicada pode ser contabilizada em mais de uma unidade quando houver coautoria interna entre unidades distintas (por exemplo, um artigo coautorado por docentes de institutos diferentes contribui para a contagem de ambos os institutos).
3. **Não duplicação dentro da mesma unidade:** Uma produção deduplicada conta apenas uma vez por unidade, mesmo que múltiplos docentes daquela unidade sejam coautores.
4. **Particularidade de programas de pós-graduação (PPG):** A associação de um docente a um programa considera a vinculação formal do docente ao PPG no período analisado (atuação em disciplinas, comissões ou orientações vinculadas ao programa), não implicando necessariamente que a produção tenha sido desenvolvida no âmbito formal do PPG. Trata-se de uma *proxy* de atuação no programa, não de autoria declarada como pertencente ao PPG. Essa abordagem privilegia consistência institucional e rastreabilidade, em detrimento de uma atribuição autoral formal inexistente no Lattes.

Implicação interpretativa: Totais por unidade representam *participações* e não devem ser somados diretamente para obter o total institucional, pois produções com coautoria interunidades são contabilizadas em múltiplas unidades. O total institucional corresponde ao número de produções deduplicadas distintas (sem repetição entre unidades).

APÊNDICE E – ARQUITETURA DA PLATAFORMA CAPIVARA

APÊNDICE E.1 VISÃO GERAL DA ARQUITETURA

A plataforma Capivara foi concebida como um sistema modular orientado à integração, tratamento e análise de dados de produção acadêmica institucional. Sua arquitetura segue princípios contemporâneos de engenharia de software, priorizando modularidade, escalabilidade, separação de responsabilidades e governança de dados.

O sistema adota uma abordagem baseada em microsserviços e persistência poliglota, na qual diferentes tecnologias de armazenamento e processamento são utilizadas conforme suas características e finalidades específicas.

De forma conceitual, a arquitetura da plataforma organiza-se em quatro camadas principais:

1. Camada de ingestão e preparação de dados;
2. Camada de persistência e governança;
3. Camada de enriquecimento e integração externa;
4. Camada analítica e de visualização.

APÊNDICE E.2 CAMADA DE INGESTÃO E PREPARAÇÃO DE DADOS

A camada de ingestão é responsável pela coleta e preparação inicial dos dados utilizados pela plataforma. Nessa camada são processadas tanto fontes institucionais quanto fontes curriculares declarativas.

Os principais componentes incluem:

- a) Módulos de extração de dados institucionais;
- b) Módulo de extração de currículos Lattes;
- c) Área de preparação (*staging*) para armazenamento temporário dos dados.

Nessa fase, não são realizadas inferências analíticas nem deduplicações globais, priorizando-se a preservação do conteúdo original e normalizações estruturais mínimas.

APÊNDICE E.3 CAMADA DE PERSISTÊNCIA E GOVERNANÇA DE DADOS

A governança dos dados é assegurada por uma base relacional central implementada em PostgreSQL, que atua como base mestra (*source of truth*) do sistema.

Essa camada é responsável por:

- a) Armazenar dados normalizados com integridade referencial;

- b) Manter histórico completo de processamentos;
- c) Preservar versões dos dados curriculares e institucionais;
- d) Registrar metadados de controle e rastreabilidade.

Campos do tipo JSONB são utilizados para preservar dados semiestruturados, conciliando rigor relacional e flexibilidade analítica.

APÊNDICE E.4 CAMADA DE INTEGRAÇÃO E ENRIQUECIMENTO EXTERNO

A camada de integração é responsável pelo enriquecimento bibliométrico e articulação com bases externas, operando de forma desacoplada e incremental.

Entre suas funções destacam-se:

- a) Descoberta e validação de identificadores persistentes;
- b) Integração com bases bibliométricas abertas e proprietárias;
- c) Desambiguação de autores;
- d) Coleta de métricas de impacto e citações.

Todos os dados enriquecidos são registrados com indicação explícita de origem e data de obtenção.

APÊNDICE E.5 CAMADA ANALÍTICA E DE INDEXAÇÃO

A camada analítica utiliza o Elasticsearch para indexação e consulta eficiente dos dados consolidados.

Entre suas principais características estão:

- a) Suporte a filtros multicriteriais;
- b) Agregações rápidas para construção de indicadores;
- c) Busca textual eficiente;
- d) Estruturas adequadas à visualização interativa.

APÊNDICE E.6 CAMADA DE APOIO E CACHE

Como suporte, a arquitetura incorpora o Redis como mecanismo de cache para resolução rápida de informações auxiliares, contribuindo para desempenho e simplificação da modelagem analítica.

APÊNDICE E.7 CONSIDERAÇÕES FINAIS SOBRE A ARQUITETURA

A arquitetura da plataforma Capivara equilibra rigor metodológico, flexibilidade técnica e viabilidade institucional, permitindo evolução incremental e sustentando os objetivos do trabalho.

APÊNDICE F – ESTRATÉGIAS DE ENRIQUECIMENTO BIBLIOMÉTRICO

APÊNDICE F.1 OBJETIVO DO APÊNDICE

Este apêndice apresenta as estratégias adotadas pela plataforma *Capivara* para o enriquecimento bibliométrico das produções acadêmicas consolidadas. O enriquecimento tem como finalidade ampliar o contexto analítico das produções institucionais, incorporando metadados, métricas de impacto e informações de autoria provenientes de bases externas.

As estratégias aqui descritas complementam o corpus institucional validado, sem alterar os critérios de reconhecimento da produção acadêmica estabelecidos na metodologia principal.

APÊNDICE F.2 PRINCÍPIOS DO ENRIQUECIMENTO BIBLIOMÉTRICO

O enriquecimento bibliométrico fundamenta-se nos seguintes princípios:

- a) **Complementaridade:** dados externos atuam como camadas adicionais de informação;
- b) **Não substituição:** informações externas não invalidam nem substituem dados declarados no currículo Lattes;
- c) **Rastreabilidade:** a origem e a data de obtenção de cada dado enriquecido são registradas;
- d) **Incrementalidade:** o enriquecimento pode ser realizado de forma progressiva e assíncrona.

APÊNDICE F.3 IDENTIFICADORES PERSISTENTES

O principal mecanismo de integração bibliométrica é o uso de identificadores persistentes, especialmente o Digital Object Identifier (DOI). Quando informado no currículo Lattes, o DOI é utilizado diretamente como elo de ligação entre bases.

Na ausência de DOI declarado, o sistema realiza tentativas de descoberta a partir de consultas a bases externas, utilizando título normalizado, ano de publicação e tipo de produção. Identificadores descobertos são registrados como complementares, preservando a informação original do currículo.

APÊNDICE F.4 FONTES BIBLIOMÉTRICAS UTILIZADAS

O enriquecimento bibliométrico utiliza múltiplas fontes, cada qual com papel específico:

- a) **OpenAlex**: *hub* central de integração, fornecendo metadados estruturados, perfis de autores, tópicos temáticos e referências cruzadas;
- b) **Crossref**: resolução e validação de DOIs;
- c) **ORCID**: desambiguação de autores e construção de identidade acadêmica;
- d) **OpenCitations**: obtenção de dados de citação em bases abertas;
- e) **Scopus**: complementação de métricas e perfis de autores, quando disponível;
- f) **Google Scholar**: obtenção de métricas alternativas, de forma exploratória.

APÊNDICE F.5 TRATAMENTO DE AMBIGUIDADE

Em situações nas quais identificadores persistentes não estão disponíveis, são aplicadas estratégias heurísticas baseadas em similaridade textual para associação de produções e autores.

Os resultados dessas associações são classificados por grau de confiança, permitindo:

- a) Aplicação automática de associações de alta confiança;
- b) Registro de associações intermediárias para possível validação posterior;
- c) Descarte de associações de baixa confiabilidade.

Essa abordagem busca equilibrar automação e precisão, evitando impactos negativos na integridade do corpus institucional.

APÊNDICE F.6 MÉTRICAS DE IMPACTO E CITAÇÕES

As métricas de impacto incorporadas ao sistema incluem contagem de citações, indicadores derivados e informações de coautoria externa. Essas métricas são utilizadas para análises comparativas e exploratórias, não constituindo critério de validação da produção institucional.

Quando múltiplas fontes de citação estão disponíveis, os dados são mantidos separadamente, permitindo análises de cobertura e consistência entre bases.

APÊNDICE F.7 RECORTE METODOLÓGICO

Para fins deste trabalho, o enriquecimento bibliométrico é considerado em um recorte temporal estático, correspondente ao estado dos dados no momento da consolidação analítica. Atualizações futuras e novas cargas de enriquecimento bibliométrico não integram o escopo desta dissertação, mas constituem uma possibilidade natural de evolução da plataforma *Capivara*.

Trabalhos futuros poderão explorar estratégias de enriquecimento contínuo, ampliação de fontes externas e análises longitudinais de impacto científico a partir de atualizações periódicas dos dados.

APÊNDICE F.8 CONSIDERAÇÕES FINAIS

As estratégias de enriquecimento bibliométrico adotadas pela plataforma *Capivara* ampliam a capacidade analítica do sistema sem comprometer sua coerência metodológica. Ao preservar a distinção entre dados institucionais e dados externos, o sistema garante transparência, confiabilidade e alinhamento com práticas consolidadas de avaliação acadêmica.