



PGDESIGN | Programa de Pós-Graduação
Mestrado | Doutorado



UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL

ESCOLA DE ENGENHARIA

FACULDADE DE ARQUITETURA

PROGRAMA DE PÓS-GRADUAÇÃO EM DESIGN

José Antonio Meira da Rocha

VidLattes – software para captação, tratamento e exibição de dados acadêmicos, como auxílio à autoavaliação de PPG

Tese de Doutorado

Porto Alegre

2023

JOSÉ ANTONIO MEIRA DA ROCHA

VidLattes – software para captação, tratamento e exibição de dados acadêmicos, como auxílio à autoavaliação de PPGs

Tese apresentada ao Programa de Pós-Graduação em Design da Universidade Federal do Rio Grande do Sul, como requisito parcial à obtenção do título de Doutor em Design.

Orientador: Prof. Dr. Fábio Gonçalves
Teixeira

Porto Alegre

2023

CIP - Catalogação na Publicação

Rocha, José Antonio Meira da
VidLattes - software para captação, tratamento e
exibição de dados acadêmicos, como auxílio à
autoavaliação de PPG / José Antonio Meira da Rocha. --
2023.

157 f.

Orientador: Fábio Gonçalves Teixeira.

Tese (Doutorado) -- Universidade Federal do Rio
Grande do Sul, Faculdade de Arquitetura, Programa de
Pós-Graduação em Design, Porto Alegre, BR-RS, 2023.

1. CAPES. 2. Lattes. 3. autoavaliação. 4. Design
Science Research. 5. software livre. I. Gonçalves
Teixeira, Fábio, orient. II. Título.

VidLattes – software para captação, tratamento e exibição de dados acadêmicos, como auxílio à autoavaliação de PPGs

Esta Tese foi julgada adequada para a obtenção do Título de Doutor em Design, e aprovada em sua forma final pelo Programa de Pós-Graduação em Design da Universidade Federal do Rio Grande do Sul - UFRGS.

Porto Alegre, 31 de agosto de 2023.

Prof. Dr. Fabio Pinto da Silva

Coordenador do Programa de Pós-Graduação em Design da Universidade Federal do Rio Grande do Sul – UFRGS

Banca examinadora

Orientador: Prof. Dr. Fábio Gonçalves Teixeira

Programa de Pós-Graduação em Design – PGDesign/Universidade Federal do Rio Grande do Sul - UFRGS

Prof. Dr. Ricardo Triska

Examinador externo, Universidade Federal de Santa Catarina – UFSC

Prof. Dr. Regio Pierre da Silva

Examinador interno, PGDesign/Universidade Federal do Rio Grande do Sul – UFRGS

Prof. Dr. Fernando Batista Bruno

Examinador externo, Universidade Federal do Rio Grande do Sul – UFRGS.

Dedico este trabalho à mulher na minha vida.

Agradecimentos

*Minha gratidão ao prof. dr. Sérgio Endler,
que me iniciou na vida acadêmica,
e ao prof. dr. Fábio Gonçalves Teixeira,
que me levou ao ponto mais alto de minha carreira.*

Resumo

O objetivo deste trabalho foi desenvolver um artefato digital para auxiliar o processo de autoavaliação de Programas de Pós-Graduação no Brasil. Parte das ferramentas que a CAPES usa para coordenar o ensino de pós-graduação, a autoavaliação adquire importância na medida em que os programas crescem em número e aumenta a quantidade de coordenadores, que necessitam conhecer a produção acadêmica tanto de seu corpo docente quanto de candidatos aos quadros dos programas. Uma das principais fontes para esse conhecimento são as informações obtidas a partir dos dados nos currículos Lattes. No entanto, obter estes dados, atualmente, é demorado e trabalhoso. Para resolver este problema, desenvolvemos um artefato, em *software* livre, que extrai e relaciona dados de currículos Lattes, CNPq e outras fontes e exibe-os em diferentes tipos de gráficos e tabelas, como auxílio à preparação de Programas de Pós-Graduação para as avaliações quadriênais da CAPES. O artefato também pode ser usado para consultas mais frequentes, permitindo aos coordenadores terem uma visão ampla de seus programas a qualquer momento.

Palavras-chave: Lattes, autoavaliação, *software* livre, visualização de dados, *Design Science Research*.

Abstract

The objective of this work was to develop a digital artifact to assist the self-evaluation process of Postgraduate Programs in Brazil. As part of the tools used by CAPES to coordinate postgraduate education, self-evaluation becomes important as the number of programs and coordinators increases, who need to be familiar with the academic production of both their faculty and candidates for program positions. One of the main sources for this knowledge is the information obtained from Lattes curriculum data. However, obtaining this data currently takes time and effort. To solve this problem, we developed an artifact, using open-source software, that extracts and relates data from Lattes curricula, CNPq, and other sources, and displays them in different types of graphs and tables, as an aid in the preparation of Postgraduate Programs for CAPES' quadrennial evaluations. The artifact can also be used for more frequent queries, allowing coordinators to have a broad view of their programs at any time.

Keywords: Lattes, self-evaluation, free software, data visualization, Design Science Research.

Índice de figuras

Figura 1 – Crescimento de programas de pós-graduação no Brasil (1995-2021).....	18
Figura 2 – Crescimento do quadro de professores de pós-graduação no Brasil.....	18
Figura 3 – Evolução da quantidade de discentes de pós-graduação no Brasil (1998-2021).....	19
Figura 4 – Principais entidades avaliadoras dos EUA e suas regiões de atuação.....	21
Figura 5 – Sistemas de AI e suas subdivisões.....	41
Figura 6 – Os cinco tipos de gráficos básicos.....	48
Figura 7 – Matriz que relaciona tipos de gráfico com o efeito pretendido com os dados.....	49
Figura 8 – Gráfico de barras verticais compara porcentagem de funcionários públicos em diversos países.....	51
Figura 9 – gráficos de linhas servem para comparações ao longo do tempo.....	52
Figura 10 – Gráfico de área é uma variação do gráfico de linha.....	53
Figura 11 – Gráficos de pizza podem tornar difícil a comparação de dados.....	54
Figura 12 – Os mesmos dados da pizza, mas em gráficos de barras.....	54
Figura 13 – Gráfico de dispersão relaciona PIB com resultado do teste PISA entre países da OCDE.....	56
Figura 14 – Gráficos de Gantt podem ser usado para outros fins, além de projetos.....	57
Figura 15 – Bolhas representam certos tipos de dados em gráficos de dispersão e mapas.....	59
Figura 16 – Mapas de variação usam três variáveis, ocupando os eixos X, Y, e Z (eixo da cor).....	59
Figura 17 – A combinação de gráficos de linha com tabelas destacadas fornece detalhes adicionais.....	60
Figura 18 – Dados georreferenciados podem ser apresentados como mapas.....	61
Figura 19 – <i>Dashboard</i> com mapa e gráficos de bolhas.....	62
Figura 20 – Velocímetro (<i>Gauge</i> , em inglês) é bastante usado.....	63
Figura 21 – Gráficos de <i>bullet</i> comparam uma medida principal com outras medidas.....	64
Figura 22 – o gráfico de árvore hierárquica dá uma ideia do todo.....	65
Figura 23 – Gráfico tipo Radar.....	67
Figura 24 – Histograma mostra a frequência relativa de uma mesma variável.....	68
Figura 25 – Gráfico de Caixas e Traços combinam dois ou mais tipos de gráficos.....	70
Figura 26 – Gráfico combinado de barras com árvore hierárquica.....	70
Figura 27 – Fluxograma de escolha de gráfico conforme operação a ser apresentada.....	72
Figura 28 – Mapa mental do processo de produção de <i>dashboards</i>	77
Figura 29 – Esquema de validação e verificação no ciclo de vida de software.....	90
Figura 30 – Especificações de software.....	92
Figura 31 – Fluxograma da metodologia Design Science Research.....	95
Figura 32 – Tela principal do GeraLattes onde são feitas as escolhas de arquivos.....	100
Figura 33 – O resultado do GeraLattes era um arquivo HTML.....	101
Figura 34 – O software scriptLattes teve mais de 33 mil downloads, principalmente da América Latina.....	103

Figura 35 – Extractor Lattes CNPq lê dados do Lattes e CNPq e cria banco de dados.....	107
Figura 36 – Gráficos e quadros no site do Currículo Lattes.....	108
Figura 37 – <i>Dashboard</i> do Stela Experta.....	109
Figura 38 – A tela inicial do Cientum era um <i>dashboard</i>	110
Figura 39 – O software Somos usa gráficos comerciais Highcharts.com.....	111
Figura 40 – Exemplo de <i>dashboard</i> criado pelo e-Lattes.....	113
Figura 41 – Interfaces gráficas das principais funções do Research.NET.....	114
Figura 42 – Comparativo entre gerenciadores de currículos Lattes.....	115
Figura 43 – Código XML de um currículo Lattes em estado não estruturado.....	123
Figura 44 – Fragmento do código XML de um currículo Lattes estruturado.....	124
Figura 45 – As cinco fases da Arquitetura da Informação.....	127
Figura 46 – Ambiente de desenvolvimento integrado (IDE) do LiveCode.....	131
Figura 48 – Módulo de gerenciamento de pesquisadores.....	133
Figura 47 – Módulo de edição de bibliografia do protótipo VidLattes.....	133
Figura 49 – Lógica do código para apresentar <i>dashboard</i>	135
Figura 50 – Mockup do artefato com o <i>dashboard</i>	136
Figura 51 – No <i>mockup</i> , visão completa do <i>dashboard</i>	137
Figura 52 – Estrutura do VidLattes. Em destaque, o módulo de visualização de dados.....	138
Figura 53 – Estimativa da Lei Orçamentária Anual para a Plataforma Lattes.....	146

Índice de quadros

Quadro 1 – Quesitos e itens avaliados pela CAPES nos PPGs mestrado e doutorado (2010-2020).....	31
Quadro 2 – Quesitos e itens avaliados pela CAPES nos PPGs com mestrado profissional.....	33
Quadro 3 – Modelo simplificado proposto pelo Grupo de Trabalho Ficha de Avaliação da CAPES.....	35
Quadro 4 – Tipos de gráficos conforme o tipo de operação.....	73
Quadro 5 – Categorização de <i>dashboards</i> por Few.....	75
Quadro 6 – Critérios de validação e verificação.....	94
Quadro 7 – Objetivos desta pesquisa e as fases das teorias utilizadas.....	96
Quadro 8 – Tipos de relatórios fornecidos pelo scriptLattes.....	104
Quadro 9 – Linha típica de um currículo Lattes em formato XML.....	125
Quadro 10 – Comparativo de bibliotecas de software para gráficos de dados.....	129
Quadro 11 – Critérios de validação e verificação aplicados.....	140
Quadro 12 – Atendimento de requisitos funcionais do projeto.....	141

Lista de abreviaturas

CAPES	Fundação Coordenação de Aperfeiçoamento de Pessoal do Nível Superior
CAPTCHA	Completely Automated Public Turing test to tell Computers and Humans
CFE	Conselho Federal de Educação
CNPq	Conselho Nacional de Desenvolvimento Científico e Tecnológico.
COC	Commission on Colleges (USA)
DNS	Domain Name System: Sistema de Nomes de Domínio.
ENQA	European Association for Quality Assurance in Higher Education (EU)
EU	European Union
EUA	European University Association (EU)
EURASHE	European Association of Institutions in Higher Education (EU)
FOSS	Free and Open Source Software – Programa Livre e de Código Aberto
FSF	Free Software Foundation: Fundação do Software Livre.
GPL	General Public License: licença livre de copyleft criada pela FSF.
HLC	Higher Learning Commission (USA)
HTML	HyperText Markup Language: Linguagem de Marcação de HiperTexto.
HTTP	HyperText Transfer Protocol: Protocolo de Transferência de HiperTexto.
HTTPS	Secure HyperText Transfer Protocol: Protocolo Seguro de Transferência de HiperTexto.
IBICT	Instituto Brasileiro de Informação em Ciência e Tecnologia.
ID CNPq	Identificador CNPq, código numérico que identifica univocamente um pesquisador na Plataforma Lattes.
IDE	Integrated Development Environment – Ambiente Integrado de Desenvolvimento de software
IEP	Institutional Evaluation Programme (EU)\Zz
IFES	Instituto Federal do Espírito Santo
IFPE	Instituto Federal de Pernambuco
IFPR	Instituto Federal do Paraná
IFSC	Instituto Federal de Santa Catarina
ISSN	International Standard Serial Number – Número de Série Padrão Internacional.
JSON	JavaScript Object Notation: Notação de Objetos JavaScript.
JSONB	JavaScript Object Notation – Binary : JSON em formato binário.
LMPL	Linguagem de Marcação da Plataforma Lattes.
MCT	Ministério da Ciência e Tecnologia.
MSCHE	Middle States Commission on Higher Education (USA),
NEASC	New England Association of Schools and Colleges (USA),
NWCCU	Northwest Commission on Colleges and Universities (USA)
PPG	Programa de Pós-Graduação
PUCRS	Pontifícia Universidade Católica do Rio Grande do Sul
PUCSP	Pontifícia Universidade Católica de São Paulo
QAA	Quality Assurance Agency (UK)
SACS	Southern Association of Colleges and Schools (USA)
SNPG	Sistema Nacional de Pós-Graduação
SOFTEX 2000	Programa de Exportação de Software, projeto do governo brasileiro de fomento à exportação de software

SQL	Structured Query Language: Linguagem de Consulta Estruturada
UCB	Universidade Católica de Brasília
UCDB	Universidade Católica Dom Bosco
UEG	Universidade Estadual de Goiás
ESC	Universidade Estadual de Santa Cruz
UFBA	Universidade Federal da Bahia
UFFS	Universidade Federal da Fronteira Sul
UFOPA	Universidade Federal do Oeste do Pará
UFPE	Universidade Federal de Pernambuco
UFRGS	Universidade Federal do Rio Grande do Sul
UFSCar	Universidade Federal de São Carlos
UNA	Centro Universitário Una (MG, GO)
UNESP	Universidade Estadual Paulista “Júlio de Mesquita Filho”
UNIBH	Centro Universitário de Belo Horizonte
UNICAMP	Universidade Estadual de Campinas
UNIFOR	Universidade de Fortaleza
UNILASALLE	Universidade La Salle Canoas
UNIMONTE	Universidade Estadual de Montes Claros
UNIOESTE	Universidade Estadual do Oeste do Paraná
UNIRIO	Universidade Federal do Estado do Rio de Janeiro
UNISINOS	Universidade do Vale do Rio dos Sinos
UNISOCIESC	Universidade Sociedade Educacional de Santa Catarina
UNIVILLE	Universidade da Região de Joinville
UPF	Universidade de Passo Fundo
USJT	Universidade São Judas Tadeu
USP	Universidade de São Paulo
UTFPR	Universidade Tecnológica Federal do Paraná UTFPR
VPN	Virtual Private Network
WASC	Western Association of Schools and Colleges (USA)
XML	Extensible Markup Language: Linguagem de Marcação Extensível

Sumário

1 Introdução.....	17
1.1 Contextualização.....	17
1.1.1 Avaliação em outros países.....	20
1.1.1.1 Espanha.....	20
1.1.1.2 Estados Unidos.....	21
1.1.1.3 Alemanha.....	22
1.1.1.4 França.....	22
1.1.1.5 Irlanda.....	23
1.1.1.6 Reino Unido.....	23
1.1.1.7 Áustria.....	24
1.1.1.8 <i>European University Association (EUA)</i>	24
1.1.1.9 Hong Kong.....	24
1.1.2 Avaliação no Brasil.....	25
1.1.2.1 Avaliação atual.....	26
1.1.2.2 Autoavaliação.....	27
1.1.2.3 Autoavaliação periódica.....	29
1.2 Delimitação.....	30
1.2.1 Quesitos e itens de avaliação.....	30
1.3 Hipótese.....	36
1.4 Problema de pesquisa.....	36
1.5 Objetivo geral.....	36
1.6 Objetivos específicos.....	36
1.7 Justificativa.....	37
2 Fundamentação teórica.....	40
2.1 Arquitetura da informação.....	40
2.2 Estrutura de um sistema de informação.....	42
2.2.1 Sistemas de Organização.....	43
2.2.2 Sistemas de busca.....	44
2.2.3 Sistema de rotulação.....	44
2.3 Design da informação.....	45
2.3.1 A imagem dos números.....	46
2.3.2 Da taViz, a visualização de dados.....	47
2.3.2.1 A complexidade da apresentação gráfica.....	47
2.3.2.2 Tipos de gráficos.....	48

Gráfico de Barras horizontais ou verticais.....	51
Gráficos de linhas.....	52
Gráficos de área.....	53
Gráfico de pizza ou de torta.....	54
Gráfico de dispersão.....	56
Diagrama de Gantt.....	57
Gráfico de Bolhas.....	58
Mapa de variações.....	58
Tabela de destaque.....	60
Mapa.....	61
Velocímetro (<i>gauge</i>).....	63
<i>Bullet</i> ou marcador.....	64
Mapa de árvore hierárquica.....	65
Gráfico de Radar.....	66
Histograma.....	67
Gráfico de Caixas e Traços ou de Velas.....	69
2.3.3 Como escolher um gráfico.....	71
2.3.4 Dashboards.....	74
2.3.4.1 Categorias de <i>dashboards</i>	74
Painéis para fins estratégicos.....	76
2.3.4.2 Painéis para fins analíticos.....	76
Painéis para fins operacionais.....	77
2.3.4.3 Como planejar um <i>dashboard</i>	78
2.3.5 Colocando em prática um <i>dashboard</i> eficiente.....	79
Passo 1 – levantamento de requisitos.....	79
Passo 2 – Planejamento de Dashboard.....	79
Sobre usabilidade.....	80
Sobre Estrutura de dados.....	80
Sobre visual.....	80
Passo 3 – Configuração técnica.....	80
Configuração da base de dados.....	81
Configuração da interface.....	81
2.4 Software livre.....	82
2.5 Dados abertos.....	83
2.6 Artefato na <i>Design Science Research</i>.....	86
3 Procedimentos metodológicos.....	89
3.1 Avaliação do artefato.....	89
3.2 Delineamento da pesquisa.....	97
3.3 Requisitos do projeto.....	99

4 Apresentação e análise dos resultados.....	100
4.1 Identificação dos artefatos e definição da classe de problemas.....	100
4.1.1 GeralLattes.....	101
4.1.2 scriptLattes.....	103
4.1.3 synclLattes.....	106
4.1.4 Extractor Lattes CNPq.....	107
4.1.5 Plataforma Lattes.....	108
4.1.6 Stela Experta.....	109
4.1.7 Cientum.....	110
4.1.8 Somos.....	111
4.1.9 E-Lattes Web.....	112
4.1.10 Research.NET Web.....	112
4.1.11 VidLattes v0.5.....	114
4.1.12 Comparação com artefatos semelhantes.....	114
4.2 Proposição de artefato.....	116
4.3 Projeto e desenvolvimento do artefato.....	116
4.3.1 Definição da estrutura de dados.....	116
4.3.2 Sistema de Organização.....	121
4.3.3 Sistemas de Navegação, de Rotulagem e de Busca.....	127
4.3.4 As fases do Processo e a Metodologia.....	128
4.3.4.1 Pesquisa.....	128
4.3.4.2 Estratégia.....	129
4.4 Design.....	130
4.4.1 Escolha da biblioteca gráfica.....	130
4.5 Implementação.....	130
4.5.1 Codificação.....	131
4.5.2 Estruturação do banco de dados.....	131
4.5.3 Ambiente de desenvolvimento.....	132
4.5.4 Protótipo.....	135
4.5.5 Tipos de gráficos usados.....	139
4.5.6 Avaliação.....	139
5 Explicitação do aprendizado.....	142
5.1 Documentação do processo de projeto e desenvolvimento do artefato.....	142
5.2 Discussão dos resultados.....	142
5.3 Outras questões.....	144

6 Considerações finais.....	147
6.1 Conclusões.....	147
6.2 Sugestões para trabalhos futuros.....	148

1 Introdução

Neste capítulo, contextualiza-se a avaliação de qualidade dos cursos de pós-graduação, mostrando o crescimento dos cursos no Brasil; faz-se um breve panorama da avaliação em alguns países; detalha-se o processo de avaliação da CAPES, mostrando o peso que a produção acadêmica tem entre os quesitos analisados; delimita-se o escopo da pesquisa; formula-se o problema de pesquisa e a hipótese; define-se os objetivos principal e específicos; e, finalmente, justifica-se a importância do trabalho para a academia e para a sociedade.

Este trabalho inédito está de acordo com a *expertise* do curso de Design da Universidade Federal do Rio Grande do Sul, Área de avaliação Arquitetura, Urbanismo e Design e Grande Área de Ciências Sociais Aplicadas.

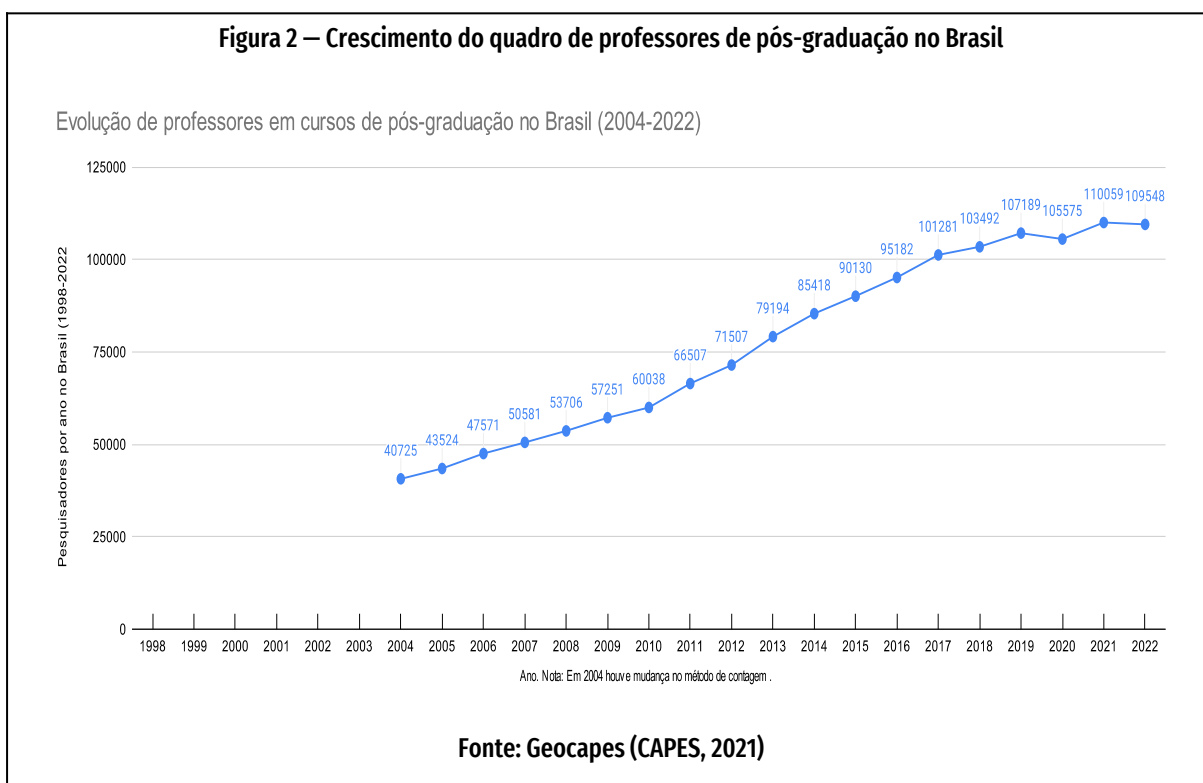
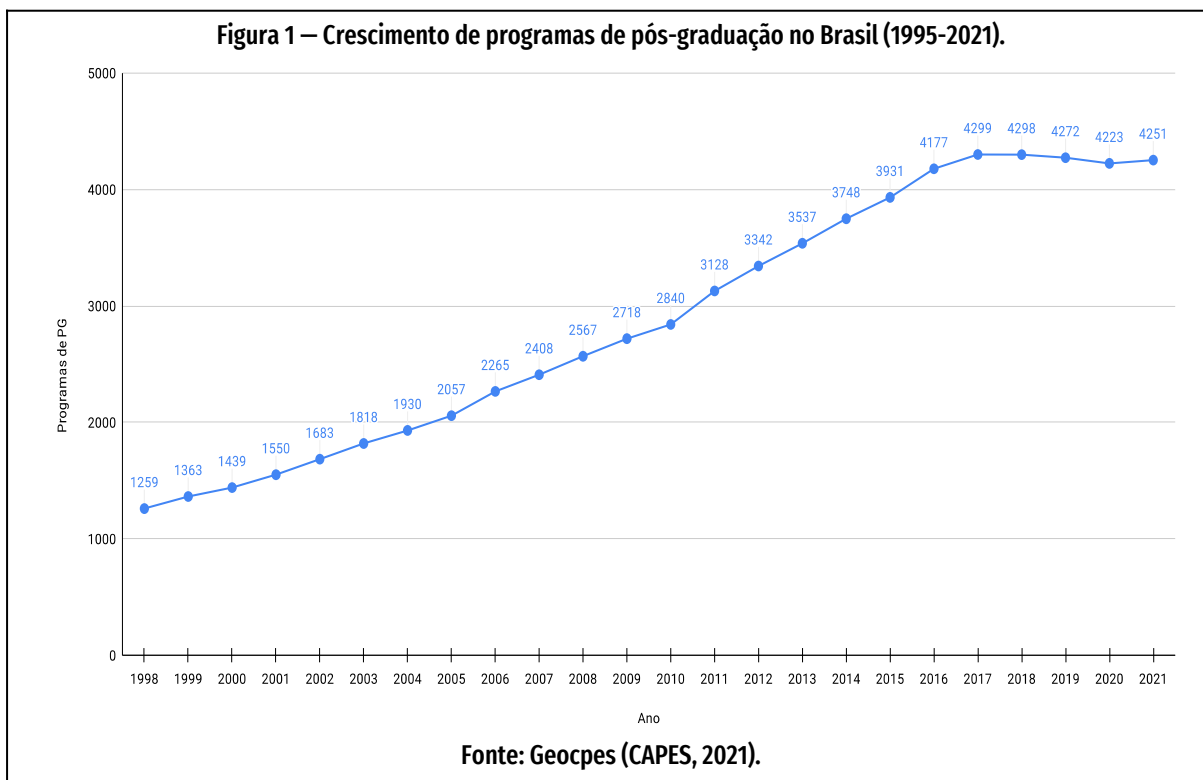
1.1 Contextualização

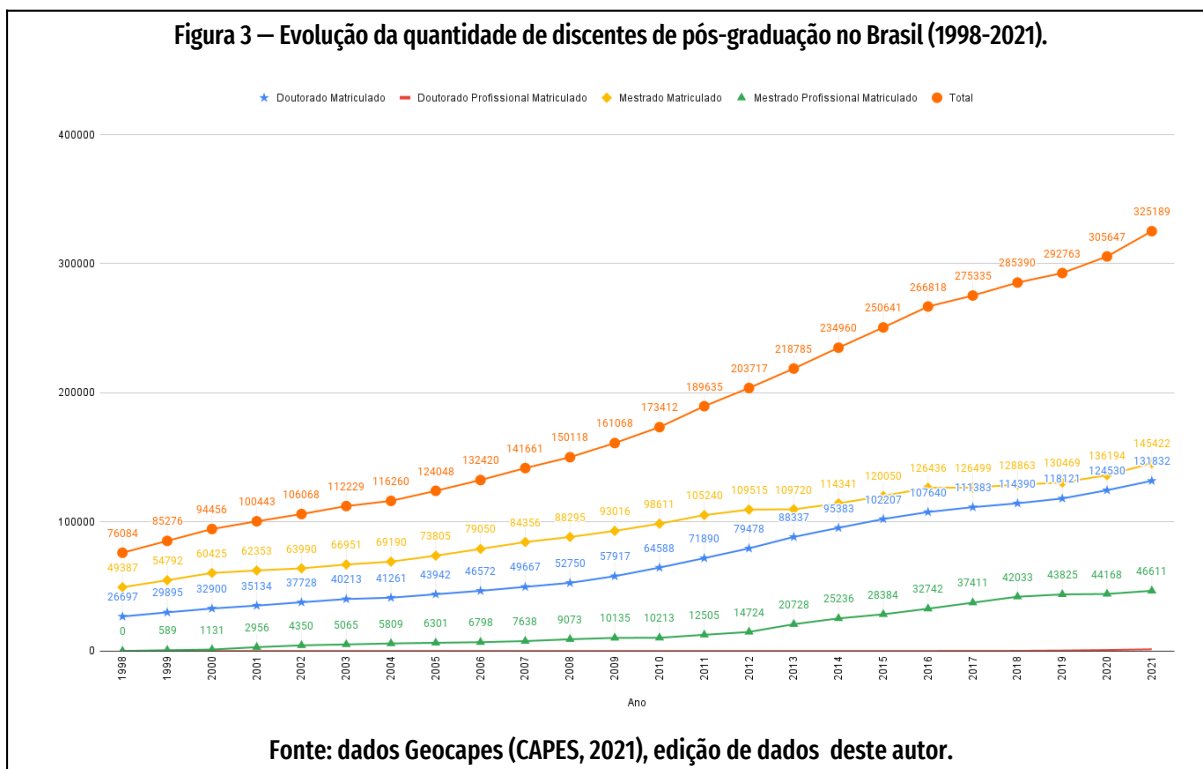
Em 1998, o país tinha 1.259 programas *stricto sensu* (com cursos de doutorado, doutorado profissional, mestrado e mestrado profissional) aprovados com notas entre 3 e 7. Em 2021, já existiam 4.251 Programas de Pós-Graduação (PPGs) no Brasil aprovados com notas entre 3 e 7. A Figura 1, página 18, mostra a evolução da quantidade de programas de 1998 a 2021. Fonte: (CAPES, 2021).

O corpo docente, que era de 40.725 pesquisadores em 2004¹, chegou a 109.548 em 2022, um aumento de 269%, conforme se pode ver na Figura 2, página 18. O número de alunos matriculados nos cursos de mestrado, que era de 49.387 em 1998, em 2021 atingiu 136.194, um crescimento de 275%.

Em cursos de doutorado, o número de estudantes era de 26.697 em 1998 e em 2021 alcança 124.530 matriculados, uma expansão de 466% (CAPES).

1 Em 2004 mudou o método de contagem de docentes em programas de pós-graduação no Brasil para o método atual. Os dados de 1998 a 2004 não podem ser usados para este fim.





No mestrado profissional, o número de alunos, que era de 2.956 em 2011, em 2021 chega aos 44.168, um aumento de 353% (CAPES, 2021).

O aumento da procura por cursos de pós-graduação pode estar relacionado à percepção da oportunidade de maiores remunerações profissionais. A Pesquisa Nacional por Amostra de Domicílios de 2015, do IBGE (IBGE, 2016), levantou dados indicativos de que trabalhadores pós-graduados registravam uma renda média de R\$ 7.400, mais que o dobro do que o verificado para a média de R\$ 3.600 de quem tinha apenas graduação. Já Lima, em sua tese sobre os efeitos da pós-graduação nos salários, não chegou a números tão expressivos, mas observou que

“o doutorado incrementa a renda de seus egressos em aproximadamente 17,7% (em relação ao mestrado), enquanto o mestrado acadêmico e o mestrado profissional têm efeitos, em média, de 13,1% e 9,0% (em relação à graduação), todos já no primeiro ano após a titulação” (Lima, 2016, p. 112).

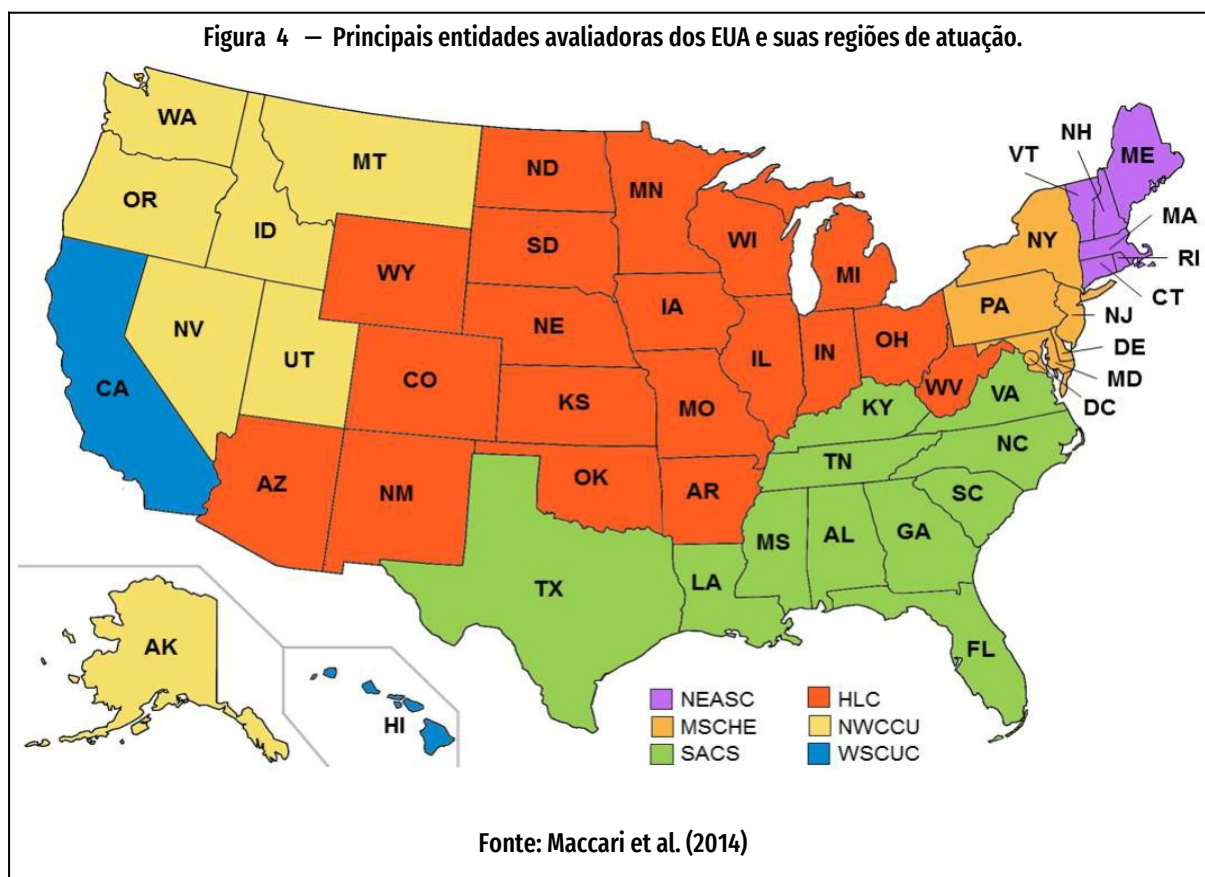
Com o aumento do número de cursos professores e alunos, a atual metodologia de avaliação esgotou-se ante o crescimento do número de cursos e de estudantes, e deve ser “repensado e aprimorado” (CAPES, 2017c, p. 16), justamente em um momento em que se torna também essencial determinar a qualificação dos cursos oferecidos. É o que pensam os pesquisadores que estão buscando novas formas de avaliação, conforme se apreende da apresentação elaborada pela Comissão Especial de Acompanhamento do Programa Nacional de Pós-Graduação (PNPG) a partir de propostas de diversas entidades envolvidas. A avaliação foi um dos primeiros itens listados como necessário aperfeiçoar (CAPES, 2017c).

1.1.1 Avaliação em outros países

A fim de situar a educação superior brasileira no panorama mundial, vai-se comentar brevemente como é feita a avaliação dos cursos de graduação e pós-graduação em alguns países. O tema tem sido uma preocupação dos governos do mundo todo nas últimas décadas. Na União Europeia, desde o **Processo de Bolonha** e a criação da *European Higher Education Area*, a métrica de qualidade do ensino superior tem ganhado importância (Ríos, 2015, p. 25). Em geral, as agências europeias avaliam o que no Brasil se chama “graduação” e “pós-graduação”, e em outros países, respectivamente, “*undergraduate*” e “*graduate*”. Diferentemente do Brasil, onde a pós-graduação é regulada por uma instituição específica, a CAPES, na maioria dos países as entidades analisam a qualidade de ambos os níveis: graduação e pós-graduação.

1.1.1.1 Espanha

No país ibérico, a avaliação do Ensino Superior (graduação e pós-graduação, como muitos países europeus) é responsabilidade da *Agencia Nacional de Evaluación de la Calidad y Acreditación* (ANECA), entidade governamental que se articula com 10



agências de acreditação ligadas a governos regionais (Ríos, 2015, p. 31). A unidade de avaliação é o curso e os docentes titulares ou catedráticos. O método é o de *peer-review* (autoavaliação) e produção dos docentes, na sua acreditação. A periodicidade é de 4 anos para o mestrado e 6 anos para o doutorado – ou demanda espontânea – e participam estudantes, acadêmicos e profissionais, além do *staff* da agência, empresários, gestores, especialistas estrangeiro .

1.1.1.2 Estados Unidos

No país americano, existem sete principais agências regionais acreditadoras não-governamentais atuando em seis regiões (Ryan, 2015, p. 3). As unidades de avaliação são as universidades e os PPGs. As entidades acreditadoras são:

- *Accrediting Commission for Senior Colleges and Universities*
- *Higher Learning Commission (HLC)*

- *Middle States Commission on Higher Education (MSCHE)*
- *New England Association of Schools and Colleges (NEASC)*
- *Northwest Commission on Colleges and Universities (NWCCU)*.
- *Southern Association of Colleges and Schools (SACS) Commission on Colleges (COC)*
- *Western Association of Schools and Colleges (WASC)*

As áreas de atuação aparecem na Figura 4, página 21.

Além destas entidades, existem outras dedicadas a áreas específicas, como a *Association to Advance Collegiate Schools of Business (AACSB)*, entidade sem fins lucrativos que congrega instituições educacionais ligadas à área de Administração e Negócios (Maccari *et al.*, 2014, p. 373).

1.1.1.3 Alemanha

O sistema de avaliação alemão procura credenciar todos os programas de estudo, o que é considerado uma tarefa gigantesca e questionável do ponto de vista didático e científico. “Adicionalmente, o sistema é muito centrado em agências de avaliação, o que pode **não incentivar** as Instituições de Ensino Superior (IES) a fazerem melhorias por **conta própria**. No entanto, reformas recentes levaram o sistema a uma abordagem menos orientada a agências e mais formativa”, diz Weber (2010, p. 2).

1.1.1.4 França

O sistema francês examina os sistemas de governança, gestão e garantia de qualidade das instituições, com bastante abrangência; inclui o desempenho de unidades de pesquisa (departamentos, institutos, laboratórios, centros de pesquisa) e examina a qualidade de todos os programas de ensino e aprendizagem (Weber, 2010, p. 2). Desde 2014, a entidade responsável pela avaliação é a agência inde-

pendente HCERES. A avaliação é obrigatória e acontece a cada 5 anos. A metodologia utiliza **autoavaliação**, métricas, *peer review* e relatórios de comitês (CAPES, 2010, p. 9; Weber, 2010, p. 2) e envolve professores pesquisadores, estudantes, servidores públicos e privados.

1.1.1.5 Irlanda

O sistema de acreditação irlandês é, em grande parte, orientado pelas características da instituição avaliada. As instituições são obrigadas por lei a desenvolver rigorosos **sistemas internos** de avaliação e promover melhoria do ensino e aprendizagem, pesquisa, divulgação no nível de departamentos, além de eficientes serviços de graduação (*undergraduate*). A lei, no entanto, também exige avaliações regulares dessas medidas internas por uma agência externa de garantia de qualidade. Para Weber,

“o sistema irlandês apresenta uma interessante e eficiente divisão de responsabilidades pela garantia da qualidade, responsabilizando as instituições pela qualidade de todos os seus serviços, mas também garantindo que elas levem a sério suas responsabilidades” (2010, p. 3, tradução deste autor).

1.1.1.6 Reino Unido

O ensino superior e a pesquisa, nas quatro nações da ilha, é avaliada anualmente através do *Quality Assurance Agency for Higher Education (QAA)*, entidade governamental independente, em metodologia que envolve **autoavaliação** e *peer review*. Participam, além da agência, estudantes, professores, profissionais das áreas (ALDERMAN; BROWN, 2005).

“A avaliação externa da educação superior, denominada Auditoria (Audit), focaliza, quase exclusivamente, as estruturas e mecanismos adotados pela instituição para assegurar sua qualidade. Os auditores julgam o nível de confiança que eles têm na gestão da qualidade da IES e na integridade das informações por ela publicadas” (CAPES, 2019, p. 36)

O resultado da visita são três possíveis conclusões: “confiança completa”, “confiança limitada” e “nenhuma conclusão”.

1.1.1.7 Áustria

Os austríacos têm um sistema parcial de acreditação de universidades. Contribui para criar mais transparência no setor privado, determinando quais universidades devem ser chamadas de “universidades privadas”. O sistema é, portanto, claramente somativo e conduzido por agências. Contribui indiretamente para melhorar a qualidade, fixando os padrões mínimos de qualidade para a acreditação; no entanto, os **processos internos** de qualidade impostos como critérios de acreditação permanecem **muito superficiais**, segundo Weber (2010, p. 3).

1.1.1.8 *European University Association (EUA)*

O programa de avaliação institucional da *European University Association (EUA)* foi criado para servir as IES membros da entidade e ajudá-las, numa base voluntária, a avaliar como estão em relação à governança, capacidade de mudança e garantia de qualidade. As universidades redigem relatórios de **autoavaliação**, incluindo *análises SWOT* (ou análise FOFA = Força, Oportunidade, Fraqueza, Ameaça). Grupos de cinco especialistas (incluindo um estudante) avaliam os relatórios de autoavaliação e fazem recomendações. O impacto das avaliações da EUA, no entanto, depende muito das instituições, que são livres para ignorar os relatórios ou seguir ativamente as recomendações dos especialistas. “O melhor aspecto do sistema EUA é que ele é muito bom em promover uma dinâmica de mudança e melhoria de qualidade”, diz Weber (2010, p. 3).

1.1.1.9 Hong Kong

O sistema de Hong Kong dá ênfase à **responsabilidade das próprias instituições** na promoção da qualidade do ensino/aprendizado e ao papel e apoio da agência acreditadora nacional, o Conselho de Garantia de Qualidade do Comitê de Subsídios Universitários. Com influências da Austrália, Nova Zelândia e Estados Unidos,

a abordagem de Hong Kong considera que a qualidade advém do reconhecimento de que as IES de Hong Kong têm missões distintas e variadas. Segundo Weber, “o sistema não tenta colocar as instituições em uma ‘camisa de força’ através de um único conjunto de padrões ou objetivos de qualidade, mas reconhece que cada instituição possui seus próprios objetivos apropriados à sua missão” (Weber, 2010, p. 3).

1.1.2 Avaliação no Brasil

O Sistema Nacional de Pós-Graduação (SNPG) foi implantado no Brasil a partir dos anos 1970, como um dos resultados da reforma universitária de 1968, modelada conforme as universidades dos Estados Unidos. O Sistema teve por objetivo a formação pós-graduada de docentes para todos os níveis de ensino; a formação de recursos humanos para o mercado de trabalho não acadêmico; e o fortalecimento da área tecnológica, científica e de inovação do país.

Para assegurar a qualidade dos programas de pós-graduação, a CAPES considera fundamental realizar a atividade de **avaliação**, através de sua Diretoria de Avaliação (DAV). Esta, além do objetivo de certificar a qualidade da pós-graduação do Brasil, também identifica assimetrias regionais e de áreas estratégicas do conhecimento no SNPG “para orientar ações de indução na criação e expansão de programas de pós-graduação no território nacional” (CAPES, 2018).

O sistema de avaliação dos programas de pós-graduação (PPGs) é dividido em dois processos distintos: a avaliação de propostas de cursos novos, e a avaliação para permanência dos cursos de pós-graduação já existentes. Os fundamentos que conduzem ambos os processos são os mesmos: **reconhecimento e confiabilidade** ancorados na qualidade garantida pela análise dos pares; **critérios discutidos e atualizados** pelas comunidades acadêmica e científica a cada avaliação; e

transparência, resultado de ampla divulgação de decisões, ações e resultados (CAPES, 2018).

Para designar indicadores ao resultado da avaliação, inicialmente, no período 1976-1997, a CAPES usou uma classificação conceitual alfabética de “A” a “E”, sendo o conceito “A” a melhor classificação, recebida por PPGs de padrão internacional. A partir de 1997, foi introduzida a escala numérica de 1 a 7, sendo 7 a maior nota (CAPES, 2010, p. 125). Este sistema continua em uso até os dias atuais.

1.1.2.1 Avaliação atual

A avaliação do Sistema Nacional de Pós-Graduação (SNPG) se caracteriza por três eixos (CAPES, 2010, p. 125) :

- É feita por pares com reputação intelectual reconhecida;
- É meritocrática;
- Associa reconhecimento e fomento. Define políticas e estabelece critérios para o financiamento dos programas.

A avaliação é baseada numa combinação de parâmetros acadêmicos – como solidariedade e Nucleação, com o objetivo de fazer com que os melhores programas ajudem os programas com dificuldades (CAPES, 2010, p. 125).

A avaliação é realizada na sede da CAPES em Brasília, a partir de dados colhidos pelos formulários na Web, respondidos pelos PPGs, com os critérios a serem observados. Os quesitos avaliados são os seguintes (CAPES, 2017b, p. 1):

- 1) Proposta do Programa de Pós-Graduação
- 2) Corpo Docente
- 3) Corpo Discente, Teses e Dissertações;
- 4) Produção Intelectual; e
- 5) Inserção Social.

Cada quesito tem uma série de itens de avaliação. Cada item recebe conceitos: Muito Bom, Bom, Regular, Fraco e Insuficiente (CAPES, 2017b, p. 11). O desempenho do PPG no primeiro quesito determina a nota máxima que poderá receber na avaliação, da qual se ocupam os quatro quesitos seguintes. Se avaliado como INSUFICIENTE no quesito 1, o curso poderá receber uma nota máxima 3 e será descredenciado. Se FRACO, poderá receber até 3. Se tiver na avaliação o conceito REGULAR, o PPG poderá receber até a nota 4. Se obtiver conceito MUITO BOM em pelo menos 4 dos 5 quesitos, entre os quais devem estar os quesitos 3 e 4 – centrais, na avaliação da CAPES – o PPG terá nota 5 e será candidato a ter atribuídas as notas 6 ou 7 – ou continuar com nota 5, dependendo de sua inserção internacional (CAPES, 2017b, p. 11).

Este modelo se mostrou eficiente, e a CAPES deve seu sucesso, especialmente, a este processo da avaliação, que concomitantemente,

“orienta a formação de recursos humanos pós-graduada stricto sensu para a produção intelectual de impacto científico dos grupos de pesquisa, para o alto desempenho de processos econômicos inovadores e para a necessária e obrigatória ampliação da qualidade do trabalho acadêmico e para a elevação da dinâmica produtiva do País.” (CAPES, 2020, p. 2)(CAPES, 2020, p. 2).

1.1.2.2 Autoavaliação

A autoavaliação de uma instituição, também chamada avaliação interna ou avaliação institucional, é o processo de uma organização avaliar a si própria. Tem como principal objetivo a aprendizagem. A autoavaliação possibilita uma reflexão sobre contexto e políticas adotadas pela instituição, além da sistematização dos dados que levam à tomada de decisão, já que é planejada, conduzida, implementada e analisada pelos próprios formuladoras e agentes das ações a serem avaliadas (CAPES, 2019, p. 7).

Em 2018, a CAPES estabeleceu uma comissão com a tarefa de Implantar uma sistemática de **autoavaliação** dos programas de pós-graduação que pudesse também ser componente importante para a avaliação realizada pela própria instituição (CAPES, 2019, p. 4)..

O Grupo de Trabalho Autoavaliação de Programas de Pós-Graduação sustentava q/ue, na época, a meta avaliação realizada pela CAPES mostrava que a pós-graduação *stricto sensu* avançara em quantidade de programas e no desenvolvimento de um sistema de avaliação que recebeu aprovação da comunidade acadêmica nacional e internacional. Em relação aos objetivos do “Parecer Sucupira” (Sucupira *et al.*, 1965)², o GT Autoavaliação considerava que fora atingida a meta do objetivo “Formação de pesquisadores e aumento da produção científica brasileira”, embora tenha reconhecido uma lacuna em outros objetivos do parecer, como na formação de professores para o sistema de educação brasileiro e na qualificação de técnicos e trabalhadores intelectuais (CAPES, 2019, p. 4).

O sistema privilegiou a avaliação *post-facto*, incluiu a avaliação por pares; vinculou avaliação com classificação e recursos; e definiu um formato avaliativo que favoreceu os rankings, buscando alcançar um padrão de qualidade estabelecido e conhecido antecipadamente por todos. O padrão, na CAPES, foi negociado em comum acordo com a comunidade acadêmica (CAPES, 2019, p. 4).

Uma vez amadurecido o seu processo de avaliação, a CAPES viu necessidade de ampliar o foco, reconhecendo que é a **autoavaliação**, hoje muito utilizada nas experiências internacionais, que pode trazer mais subsídios para o aprimoramento do Sistema com qualidade (CAPES, 2019, p. 4–5).

2 O relatório 977/1965 do Conselho Federal da Educação, que definiu o formato da pós-graduação brasileira, é conhecido como “Parecer Sucupira”, depois do trabalho de Newton Sucupira, relator da comissão constituída por educadores como Alceu Amoroso Lima e Anísio Teixeira.

A autoavaliação não é novidade na área da educação superior brasileira:

“No País, a graduação tem sido objeto de autoavaliação desde os tempos do Programa de Avaliação Institucional das Universidades Brasileiras (PAIUB), no início da década de 1990, até mais recentemente, na década de 2000, com o Sistema Nacional de Avaliação da Educação Superior e seu componente de autoavaliação Institucional. Portanto, ao adotar a autoavaliação, a CAPES está aproximando sua concepção daquela já adotada pela avaliação da graduação, contribuindo para diminuir a distância hoje existente entre os dois processos nacionais avaliativos.” (CAPES, 2019, p. 6)

A avaliação externa assegura padrões básicos, mas tem limitações. Os avaliadores externos não se envolvem na solução dos problemas identificados, diferentemente do pessoal interno do PPG. Como é feita internamente, a autoavaliação pode favorecer a construção da identidade, heterogeneidade e envolvimento dos programas avaliados, superando os padrões mínimos obtidos pela avaliação externa (CAPES, 2019, p. 5).

“Assim, a autoavaliação, desenvolvida de forma sistemática e contínua, é a abordagem a ser enfatizada, pois garante proximidade entre avaliador e avaliado e leva a aprofundamentos de natureza qualitativa e contextualizada.” (CAPES, 2019, p. 5).

A autoavaliação dos programas de pós-graduação tem resultado na melhoria dos próprios programas, segundo a experiência internacional. A valorização da autoavaliação em todo e qualquer programa avaliativo segue a tendência mundial. (CAPES, 2019, p. 5).

1.1.2.3 Autoavaliação periódica

O documento “Relatório Final 2016 da Comissão Especial de Acompanhamento do PNPGE-2011-2020” (CAPES, 2017c) fez diversas recomendações sobre a pós-graduação brasileira, algumas das quais foram implantadas ao longo da década. Por exemplo, em 2013/2014, foi implementada a mudança mais importante do Plano

Nacional de Pós-Graduação (PNPG), a alteração do período avaliativo dos programas. Em vez de trienal, a avaliação passou a **quadrienal**.

Mas falta aplicar algumas recomendações do PNPG-2011-2020 como a que propôs o “monitoramento parcial periódico com o objetivo de medir os indicadores exigidos para a melhoria do conceito, com caráter obrigatório para todas as áreas” (CAPES, 2017c, p. 10).

É justamente nesta tarefa de monitoramento ou **autoavaliação periódica**, que se dirige esta pesquisa: ela se propõe a colaborar na melhoria das ferramentas que dispõem os avaliadores para qualificarem os programas de pós-graduação (PPGs).

1.2 Delimitação

Esta pesquisa vai se delimitar ao processo de autoavaliação dos programas de pós-graduação brasileiros a partir da produção acadêmica de seus próprios pesquisadores. A fonte dos dados são os currículos Lattes dos professores dos programas em avaliação, mas se projetará a previsão de incluir outras bases de dados como a plataforma Sucupira, a base Qualis e bases de dados sobre grupos de pesquisa do CNPq.

Dos currículos Lattes, extrairemos a produção acadêmica constante nos quesitos exigidos pela da CAPES na avaliação dos PPGs.

1.2.1 Quesitos e itens de avaliação

A seguir, listamos os quesitos e seus pesos relativos e globais exigidos para os PPGs de Design durante a primeira metade do decênio **2010-2020**. Na coluna “Peso

no quesito” estão as proporções (medidas em porcentagem) de cada item dentro do mesmo quesito de avaliação; Na coluna “Peso global” estão as proporções de cada item em relação ao peso de todos os quesitos. No Quadro 1, página 31, podemos ver que os itens mais importantes e de maior peso foram os itens dos quesitos 3 e 4. Em destaque, com bordas tracejadas em vermelho, estão os itens objetos deste estudo.

Quadro 1 – Quesitos e itens avaliados pela CAPES nos PPGs mestrado e doutorado (2010-2020)

Mestrado/Doutorado		
Quesitos/Itens	Peso no quesito (%)	Peso global (%)
1 – Proposta do Programa		
1.1. Coerência, consistência, abrangência e atualização das áreas de concentração, linhas de pesquisa, projetos em andamento e proposta curricular.	50%	Não aplicável
1.2. Planejamento do programa com vistas a seu desenvolvimento futuro, contemplando os desafios internacionais da área na produção do conhecimento, seus propósitos na melhor formação de seus alunos, suas metas quanto à inserção social mais rica dos seus egressos, conforme os parâmetros da área.	30%	n.a.
1.3. Infraestrutura para ensino, pesquisa e, se for o caso, extensão.	20%	n.a.
2 – Corpo Docente		
2.1. Perfil do corpo docente, consideradas titulação, diversificação na origem de formação, aprimoramento e experiência, e sua compatibilidade e adequação à Proposta do Programa.	30%	6%
2.2. Adequação e dedicação dos docentes permanentes em relação às atividades de pesquisa e de formação do programa	30%	6%
2.3. Distribuição das atividades de pesquisa e de formação entre os docentes do programa.	30%	6%
2.4. Contribuição dos docentes para atividades de ensino e/ou de pesquisa na graduação.	10%	2%
3 – Corpo Discente, Teses e Dissertações		
3.1. Quantidade de teses e dissertações defendidas no período de avaliação, em relação ao corpo docente permanente e à dimensão do corpo discente.	20%	6%

Continua

Conclusão Quadro 1 – Quesitos e itens avaliados pela CAPES nos PPGs mestrado e doutorado (2010-2020)

Mestrado/Doutorado		
3.2. Distribuição das orientações das teses e dissertações defendidas no período de avaliação, em relação aos docentes do programa.	20%	6%
3.3. Qualidade das Teses e Dissertações e da produção de discentes autores da pós-graduação e da graduação (no caso de IES com curso de graduação na área) na produção científica do programa, aferida por publicações e outros indicadores pertinentes à área.	40%	12%
3.4. Eficiência do Programa na formação de mestres e doutores bolsistas: Tempo de formação de mestres e doutores e percentual de bolsistas titulados.	15%	5%
3.5. Envolvimento do corpo discente com o ensino da graduação	5%	2%
4 – Produção Intelectual	40%	Global
4.1. Publicações qualificadas do Programa por docente permanente.	45%	18%
4.2. Distribuição de publicações qualificadas em relação ao corpo docente permanente do Programa.	30%	12%
4.3. Produção técnica, patentes e outras produções consideradas relevantes.	20%	8%
4.4. Produção artística, nas áreas em que tal tipo de produção for pertinente.	5%	2%
5 – Inserção Social	10%	Global
5.1. Inserção e impacto regional e (ou) nacional do programa.	40%	4%
5.2. Integração e cooperação com outros programas e centros de pesquisa e desenvolvimento profissional relacionado à área de conhecimento do programa, com vistas ao desenvolvimento da pesquisa e da pós-graduação.	40%	4%
5.3. Divulgação e transparência das atividades e da atuação do Programa	20%	2%

Fonte: Quadro por este autor com dados da CAPES

Nos programas com mestrado/doutorado, o item mais importante de toda avaliação CAPES era o **“4.1. Publicações qualificadas do Programa por docente permanente”**, com o peso de **45%** em relação ao quesito e **18%** em relação à pontuação total (**45%/18%**). Seguiam em importância os itens **“4.2. Distribuição de publicações qualificadas em relação ao corpo docente permanente do Programa”** (40%/12%) e **“3.3. Qualidade das Teses e Dissertações e da produção de discentes autores da pós-graduação e da graduação”**, (30%/12%), cada um com peso de 12% do total. Também importante era o item **“4.3. Produção técnica, patentes e outras**

produções consideradas relevantes”, com peso global de **8%**. Estes quatro itens somavam o peso de **50%** do total da avaliação e os **quesitos 3 e 4 completos** respondiam por **70%** da avaliação.

Os quesitos e itens exigidos de mestrado profissional aparecem no Quadro 2, página 33.

Quadro 2 – Quesitos e itens avaliados pela CAPES nos PPGs com mestrado profissional.

Mestrados Profissionais		
Quesitos / Itens	Peso no quesito (%)	Peso global (%)
1 – Proposta do Programa		
1.1 Coerência, consistência, abrangência e atualização da(s) área(s) de concentração, linha(s) de atuação, projetos em andamento, proposta curricular com os objetivos do Programa	40%	Não aplicável
1.2. Coerência, consistência e abrangência dos mecanismos de interação efetiva com outras instituições, atendendo a demandas sociais, organizacionais ou profissionais.	30%	n.a.
1.3. Infraestrutura para ensino, pesquisa e administração.	10%	n.a.
1.4. Planejamento do Programa visando ao atendimento de demandas atuais ou futuras de desenvolvimento nacional, regional ou local, por meio da formação de profissionais capacitados para a solução de problemas e práticas de forma inovadora.	20%	n.a.
2 – Corpo Docente	20%	Global
2.1. Perfil do corpo docente, considerando experiência como pesquisador e/ou profissional, titulação e sua adequação à Proposta do Programa.	50%	10%
2.2. Adequação da dimensão, composição e dedicação dos docentes permanentes para o desenvolvimento das atividades de pesquisa e formação do Programa.	25%	5%
2.3. Distribuição das atividades de pesquisa, projetos de desenvolvimento e inovação e de formação entre os docentes do Programa.	25%	5%
3 – Corpo Discente e Trabalho de Conclusão	30%	Global
3.1. Quantidade de trabalhos de conclusão (MP) aprovados no período e sua distribuição em relação ao corpo discente titulado e ao corpo docente do programa.	30%	9%

Continua

Conclusão Quadro 2 – Quesitos e itens avaliados pela CAPES nos PPGs com mestrado profissional.

Mestrados Profissionais		
3.2. Qualidade dos trabalhos de conclusão produzidos por discentes e egressos.	40%	12%
3.3. Aplicabilidade dos trabalhos produzidos.	30%	9%
4 – Produção Intelectual	30%	Global
4.1. Publicações qualificadas do Programa por docente permanente.	30%	9%
4.2. Produção artística, técnica, patentes, inovações e outras produções consideradas relevantes.	30%	9%
4.3. Distribuição da produção científica e técnica ou artística em relação ao corpo docente permanente do programa.	30%	9%
4.4. Articulação da produção artística, técnica e científica entre si e com a proposta do programa.	10%	3%
5 – Inserção Social	20%	Global
5.1. Impacto do Programa.	30%	6%
5.2. Integração e cooperação com outros Cursos/Programas com vistas ao desenvolvimento da pós-graduação.	25%	5%
5.3. Integração e cooperação com organizações e/ou instituições setoriais relacionados à área de conhecimento do Programa, com vistas ao desenvolvimento de novas soluções, práticas, produtos ou serviços nos ambientes profissional e/ou acadêmico.	25%	5%
5.4. Divulgação e transparência das atividades e da atuação do Programa.	20%	4%

Fonte: quadro e cálculos deste autor com dados da CAPES.

Nos cursos com mestrado profissional, o item mais importante era “**3.2. Qualidade dos trabalhos de conclusão produzidos** por discentes e egressos”, com 40% do peso do quesito e 12% da nota ponderada da avaliação (40%/12%). Este item está fora do escopo deste projeto, mas se incluem os itens 3.1, 4.1, 4.2 e 4.3, cada um com 30% do peso do quesito e 9% da nota ponderada da avaliação. Estes quatro itens somavam 36%, mais de um terço da pontuação total.

Para o quadriênio 2017-2020, foi aprovada pela CAPES a proposta de aperfeiçoamento da ficha de avaliação dos PPGs com o objetivo de simplificar o processo. A ficha ficou como no Quadro 3, a seguir.

Quadro 3 – Modelo simplificado proposto pelo Grupo de Trabalho Ficha de Avaliação da CAPES.

Quesitos/Itens	Peso maior ou igual a:
1 - Programa	70,00%
1.1. Articulação, aderência e atualização das áreas de concentração, linhas de pesquisa, projetos em andamento e estrutura curricular, bem como a infraestrutura disponível, em relação aos objetivos, missão e modalidade do programa.	25,00%
1.2. Perfil do corpo docente, e sua compatibilidade e adequação à Proposta do Programa	25,00%
1.3. Planejamento estratégico do programa, considerando também articulações com o planejamento estratégico da instituição, com vistas à gestão do seu desenvolvimento futuro, adequação e melhorias da infraestrutura e melhor formação de seus alunos, vinculada à produção intelectual – bibliográfica, técnica e/ou artística.	10,00%
1.4. Os processos, procedimentos e resultados da autoavaliação do programa, com foco na formação discente e produção intelectual.	10,00%
2 - Formação	65,00%
2.1. Qualidade e adequação das teses, dissertações ou equivalente em relação às áreas de concentração e linhas de pesquisa do programa.	15,00%
2.2. Qualidade da produção intelectual de discentes e egressos.	15,00%
2.3. Destino, atuação e avaliação dos egressos do programa em relação à formação recebida.	10,00%
2.4. Qualidade das atividades de pesquisa e da produção intelectual do corpo docente no programa	15,00%
2.5. Qualidade e envolvimento do corpo docente em relação às atividades de formação no programa.	10,00%
3 - Impacto na Sociedade	30,00%
3.1. Impacto e caráter inovador da produção intelectual em função da natureza do programa.	10,00%
3.2. Impacto econômico, social e cultural do programa.	10,00%
3.3. Internacionalização e visibilidade do programa.	10,00%

Fonte: quadro deste autor com dados da CAPES.

Neste novo modelo de ficha, a **autoavaliação** surge com peso idêntico a oito de outros 11 itens. Marcados com linhas vermelhas pontilhadas estão os itens do es-

copo deste trabalho. São itens destacados do processo de autoavaliação e foram escolhidos como objetos de estudo pois, além da importância, são variáveis que temos condições de gerenciar com esta pesquisa.

1.3 Hipótese

Um artefato digital para extrair e combinar informações acadêmicas estruturadas e organizadas da Plataforma Lattes pode subsidiar o processo de autoavaliação de Programas de Pós-Graduação, amparado pela Arquitetura da Informação e pelo Design da Informação e com o uso de técnicas de visualização de dados.

1.4 Problema de pesquisa

Como consultar, estruturar e organizar as informações da produção acadêmica de Programa de Pós-Graduação para fins de subsidiar o processo de autoavaliação dos PPGs, no contexto do Sistema Nacional de Pós-Graduação?

1.5 Objetivo geral

O objetivo deste trabalho é desenvolver um **artefato digital**³ que extraia e relacione dados de currículos Lattes, CNPq e outras fontes e exiba-os em um painel com diferentes tipos de gráficos e tabelas.

1.6 Objetivos específicos

Como objetivos específicos, pretende-se:

³ Este trabalho se desenvolve no âmbito da *Design Science Research*, que utiliza a nomenclatura de **artefato** para referir-se a objetos feitos pelo ser humano para resolver problemas práticos (Johannesson; Perjons, 2014, p. 20). No caso desta pesquisa, adota-se a nomenclatura de **artefato digital** para designar programas de computador (*software*).

1. Levantar os requisitos da CAPES aos gestores de programas de Pós-Graduação em relação à produção acadêmica para definir as características, recursos e funcionalidades do artefato digital.
2. Compreender a estrutura de dados do currículo Lattes para definir a estrutura do banco de dados necessário ao gerenciamento das informações sobre a produção docente e discente a ser analisada.
3. Analisar as características e recursos de artefatos semelhantes a fim de definir os requisitos de projeto do artefato proposto.
4. Desenhar a interface com o usuário do aplicativo principal e determinar os tipos de gráficos apresentados conforme as necessidades de PPGs, a partir dos preceitos da Arquitetura da Informação, do Design da Informação e da Visualização de Dados Numéricos.
5. Programar o artefato com ferramenta de *Rapid Application Development* (RAD).
6. Verificar o artefato segundo as especificações de requisitos de software e conforme os paradigmas da *Design Science Research*.

1.7 Justificativa

A Pós-Graduação brasileira atualiza-se constantemente através do trabalho de toda uma estrutura da CAPES da qual é parte importante a Comissão Especial de Acompanhamento do PNPG.

Os membros desta Comissão e todas as instituições interessadas que contribuíram com a reflexão sobre o aprimoramento do Modelo de Avaliação da Pós-Graduação brasileira, neste momento, consideraram que o modelo atual funcionou no passa-

do, mas esgotou-se ante o crescimento do número de cursos e de estudantes, e deve ser “repensado e aprimorado” (CAPES, 2017c, p. 16).

Os pesquisadores refletiram: “[...] as Universidades devem ser mais protagonistas de seu próprio planejamento e **avaliação da PG**, inserida em seu contexto e no seu território de atuação (**autoavaliação** conectada com plano estratégico institucional de PG)” (CAPES, 2017c, p. 16, negrito deste autor). A Comissão de Acompanhamento também listou pontos convergentes e recorrentes entre as colaborações:

- a) **autoavaliação** [negrito nosso] (PDI ou Plano Estratégico Institucional de PG)
- b) Impacto (no desenvolvimento econômico e social, regional e nacional)
- c) Modelo único de avaliação (multidimensional)
- d) Produções indicadas pelos autores (cinco mais relevantes)
- e) Relevância social e econômica
- f) Acompanhamento de egressos
- g) Balanço entre indicadores quantitativos e qualitativos
- h) Mudanças no Qualis
- i) Internacionalização
- j) Inovação
- k) Redefinição das áreas de conhecimento

Este diagnóstico de 2018 indica que, no sistema brasileiro de Pós-graduação, a autoavaliação é um dos principais pontos a serem aperfeiçoados no final do decênio.

Vimos que os quesitos envolvendo métrica da produção docente e discente são responsáveis por 70% da nota CAPES, em alguns PPGs. Mas as informações que respondem a estes quesitos estão espalhadas pelos currículos Lattes docentes e

discentes e por outras bases de dados. Consolidar toda esta informação é um trabalho que, se feito manualmente, é árduo, demorado e sujeito a falhas (Mena-Chalco; Cesar Jr., 2009, p. 31).

Coletando manualmente a produção do corpo docente, a coordenação de PPG acessa cada currículo acadêmico de seus pesquisadores no banco de dados Lattes, através da interface Web, faz a contagem das diversas produções e registra numa planilha. Este método manual é um desperdício de tempo e de trabalho. Um tempo caro, pois, não raro, é um trabalho feito pelas coordenações de PPG com emprego de recursos humanos sobre-qualificados para a tarefa.

Para gerar maior produtividade nestas tarefas, uma ferramenta que realize uma rápida recuperação de dados do currículo Lattes, (com variedade de análises e apresentação gráfica de dados numéricos) permitirá que os programas de pós-graduação façam autoavaliações mais frequentemente – mensais ou semestrais, por exemplo – tendo um panorama visual de variáveis como a quantidade de pesquisas, de produção nas diversas categorias da CAPES, as relações desta produção com as pesquisas.

Com autoavaliações frequentes e boa visualização da produção acadêmica, pretende-se obter mais facilidade e rapidez na descoberta dos pontos fracos do PPG, facilitando, conseqüentemente, a busca de soluções.

Com o uso de software livre para este projeto, se obterá gratuitamente ferramentas de qualidade; se poderá modificar o código fonte para adaptar o artefato digital às necessidades específicas do projeto; não se terá gastos monetários com o sistema, uma grande vantagem em tempos de poucos recursos, tanto para instituições públicas quanto para escolas privadas; e se poderá compartilhar com os pares o resultado do trabalho, criando-se mais oportunidades de melhoramento do artefato digital.

2 Fundamentação teórica

Esta pesquisa se desenvolve com base na metodologia **Design Science Research** e recebe aportes teóricos da **Arquitetura da Informação**, do **Design da Informação**, da **Representação Visual de Dados** e dos movimentos **Software Livre** e **Dados Abertos**.

Neste capítulo, são apresentados os conceitos básicos destas disciplinas e movimentos, e de que maneira estes conceitos são apropriados pelo projeto. As disciplinas estão distribuídas pelo capítulo em 4 seções:

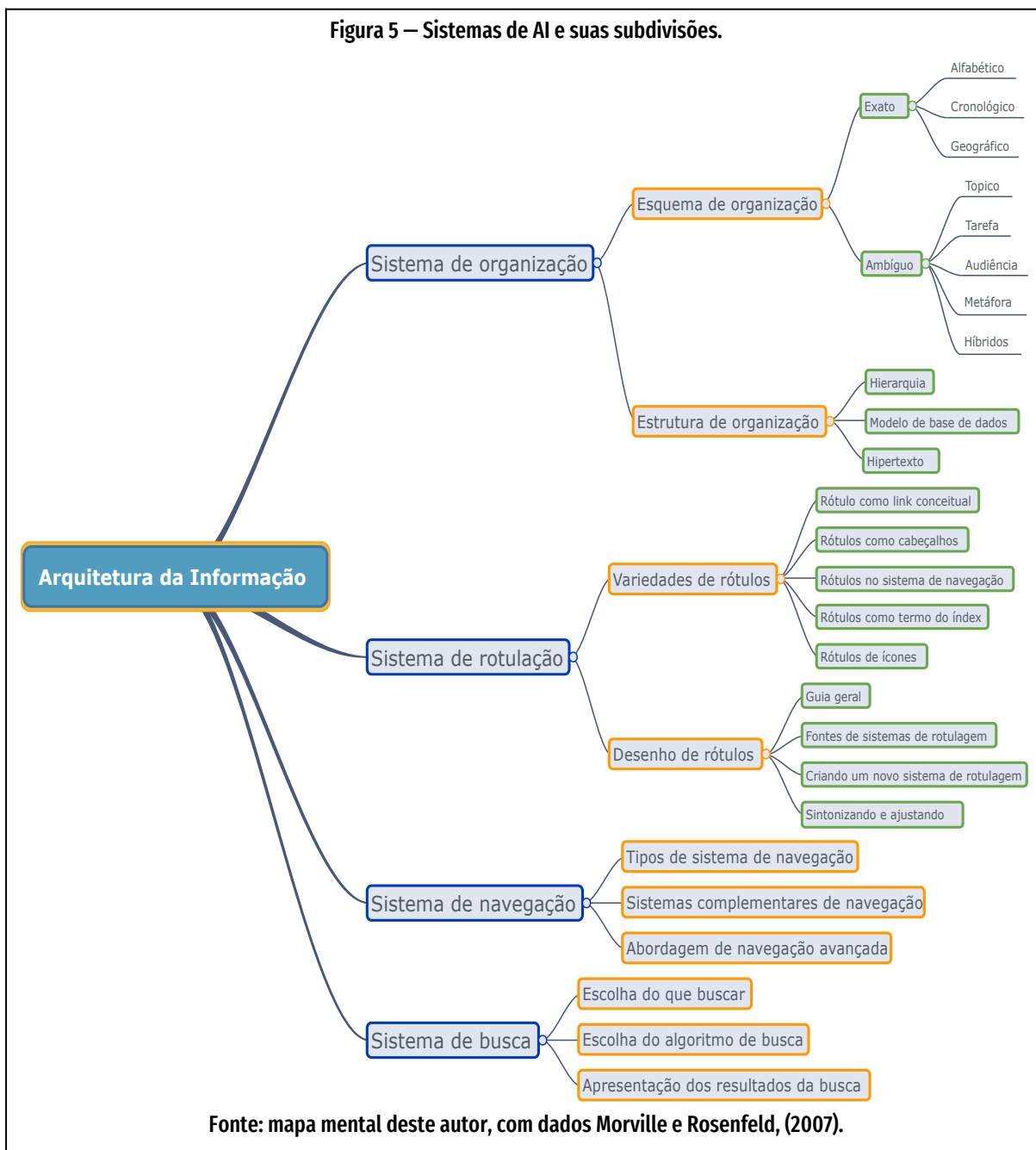
- 1) A **Arquitetura da Informação**, que define a construção da macroestrutura teórica básica do artefato.
- 2) O **Design da Informação**, disciplina que vem complementar os preceitos definidos pela Arquitetura da Informação;
- 3) **Software Livre e Aberto**, e
- 4) **Dados Abertos** e sua importância para o desenvolvimento de um ambiente informacional sadio.

2.1 Arquitetura da informação

A **Arquitetura da Informação** pode ser definida como a disciplina do Design que organiza a informação digital e seu uso (“navegação”) de forma que os usuários possam adquirir conhecimento fácil e eficientemente (Hennum *et al.*, 2005, p. 2).

Outras definições foram listadas por Morville e Rosenfeld:

- 1) “A combinação de organização, rotulagem e esquemas de navegação dentro de um sistema de informação.
- 2) “O projeto estrutural de um espaço informativo para facilitar a conclusão de tarefas e o acesso intuitivo a conteúdo.



- 3) “A arte e a ciência de estruturar e classificar *sites* e *intranets* para ajudar as pessoas a encontrar e gerenciar informações.
- 4) “Uma disciplina emergente e uma comunidade de prática focada em trazer princípios de design e arquitetura para o cenário digital” (Morville; Rosenfeld, 2007, p. 4).

Central a todas estas definições está a concepção de estruturar o espaço da informação para gerenciamento e uso. Isto pode ser interpretado de diversas maneiras: desde a simples preocupação em rotular trechos da informação até intenções mais amplas como facilitar a interação (Morville; Rosenfeld, 2007, p. 4).

O termo “arquitetura da informação” foi cunhado pelo arquiteto Richard Wurman em 1975 para descrever a necessidade de transformar dados em informações que façam sentido para o usuário, segundo Dillon e Turnbull. Eles apontam que não é uma ideia de todo original, mas que foi a primeira vez que se juntaram as duas palavras em um rótulo: Arquitetura da Informação, ou AI (Dillon; Turnbull, 2009, p. 1).

Para Hennem e colegas, uma arquitetura da informação bem-feita passa despercebida, mas a falta dela é gritante: o usuário não consegue achar a informação que precisa ou não consegue assimilar a informação que porventura encontre. O resultado é um usuário exausto que logo abandona a fonte de consulta e busca o que precisa em outro lugar (Hennem *et al.*, 2005, p. 2).

2.2 Estrutura de um sistema de informação

Morville e Rosenfeld (2007, p. 44) afirmam que uma forma básica de estruturar a arquitetura da informação é em quatro categorias:

a) **Sistema de organização**

São as maneiras de apresentar um conjunto de informação.

b) **Sistema de navegação**

São os instrumentos de ajuda ao usuário para se mover pelo sistema.

c) **Sistema de busca**

Mecanismo que permite ao usuário procurar pelo conteúdo rapidamente.

d) **Sistema de rotulação**

Descreve categorias, opções e links em linguagem inteligível para o usuário.

A Figura 5, página 41, mostra os 4 sistemas da Arquitetura da Informação e suas subdivisões. As subseções a seguir detalham cada um destes sistemas.

2.2.1.1 Sistemas de Organização

Esta categoria determina as maneiras de se estruturar a informação.

Morville e Rosenfeld (2007, p. 58) dividem os sistemas de organização em **esquemas de organização** e **estruturas de organização**.

Os esquemas de organização podem ser *Exatos* ou *Ambíguos*.

- Os **esquemas de organização exatos** ou **objetivos** dividem a informação em seções bem definidas e mutualmente exclusivas, como uma agenda telefônica ou listagem de arquivos de uma pasta, no computador. Alguns dos esquemas exatos são:
 - a) a ordenação alfabética,
 - b) a ordem cronológica e
 - c) a disposição geográfica.
- Já os **esquemas de organização ambíguos** ou **subjetivos** dividem as informações em categorias sem definição exata. “Eles estão atolados na ambiguidade da linguagem e organização, para não mencionar a subjetividade humana. Eles são difíceis de projetar e manter. Eles podem ser difíceis de usar”, sustentam Morville e Rosenfeld (2007, p. 61).

Entre os esquemas ambíguos estão:

- 1) **Tópicos**. Uma lista de assuntos. Muito comum em menus ou telas iniciais de softwares.
- 2) **Tarefas**. Uma lista de ações a tomar-se. Mais comum em e-Commerce, nos passos para compra.
- 3) **Audiência**. Organiza a informação conforme o público. Pode ser aberto – permite que uma audiência acesse o material de outra – ou fechado – que não permite a visualização do material por outra audiência.
- 4) **Metáfora**. Ajudam o usuário a conhecer o novo relacionando-o ao já conhecido. Em sistema de informação digital, por exemplo, temos “pastas”, “arquivos”, “lata de lixo” e outras metáforas com objetos de escritório.

- 5) **Híbrido.** É um esquema de organização ambíguo que mistura esquemas diferentes. Por exemplo, num mesmo menu, utilizar verbos e substantivos. É recomendado não se usar este tipo de esquema. Se um item de menu é um verbo, sugerindo que o usuário tome ações, todos os outros elementos do menu devem ser verbos. Se for substantivo, orientando uma atitude mais passiva, todos os outros elementos devem ser substantivos.

Nesta categoria **Sistemas de Organização**, esta pesquisa desenvolveu as estruturas de bancos de dados e da hierarquização das informações a serem apresentadas aos usuários, conforme a organização dos dados da Plataforma Lattes. O procedimento está descrito na subseção “4.3.2 Sistema de Organização”, página 121.

2.2.1.2 Sistemas de busca

Em **Sistema de Busca**, a terceira categoria proposta por Morville e Rosenfeld, o artefato digital proporcionará a busca “*full text*”, ou seja, por todo o conteúdo do banco de dados, através dos recursos do próprio *engine* SQL utilizado. No caso do SQLite, a busca será feita por uma *query* com “*coringas*” (o caractere “%”), que procuram uma palavra-chave, em um campo do banco de dados, com quaisquer caracteres antes e depois da *string* procurada. Ou seja: procura uma palavra em *meio* ao conteúdo do campo de dados, e não o campo *inteiro*.

2.2.1.3 Sistema de rotulação

Na categoria **Sistema de Rotulação** se tomarão cuidados com os termos aplicados a rótulos, botões e outros elementos da interface gráfica, principalmente para que façam parte do repertório de coordenadores de PPGs, mas sem serem específicos da área do Design, para que o artefato digital possa ser usado por pesquisadores de qualquer área.

Estas três categorias são aplicadas na seção 4.3.3, Sistemas de Navegação, de Rotulação e de Busca.

2.3 Design da informação

Enquanto a Arquitetura da Informação cuida da construção de uma estrutura de dados ideal, o **Design da Informação** garante que o uso desta estrutura será eficiente e efetivo.

Design da Informação tem foco na pesquisa sobre princípios para análise, planejamento, apresentação e entendimento de mensagens. Para Petterson (2003, p. 13), “não importa o meio selecionado, um conjunto de informação bem projetado satisfará as exigências estéticas, econômicas, ergonômicas, bem como questões do próprio objeto”.

“Muitas disciplinas entram no escopo de *information design* e cada uma tem sua própria expressão para determinada parte do processo”, dizem Engelhardt e colegas. Eles sustentam que arquitetos que lidam com sinalização, designers de mídia interativa, tipógrafos, designers gráficos, “todos podem ser agrupados sob o denominador comum ‘Design da Informação’” (Engelhardt *et al.*, 2016, p. 12).

Horn tem definições mais objetivas sobre o que é esta disciplina (1999, p. 15–16). Ele diz que *Information Design* é “a arte e ciência de preparar a informação de forma que ela possa ser usada por seres humanos com eficiência e efetividade”.

Para Horn, os objetivos primários da *Information Design* são:

- 1) Desenvolver documentos que sejam compreendidos com rapidez e achados com precisão, além de se poder traduzi-los em ação efetiva.
- 2) Projetar interações com equipamentos que sejam naturais, fáceis e tão prazerosas quanto possíveis, o que envolve resolver muitos problemas no projeto da interface humano-computador.
- 3) Habilitar pessoas a acharem seu caminho no espaço tridimensional com conforto e facilidade tanto no espaço urbano quanto no ciberespaço.

Nesta pesquisa, os preceitos do Design da Informação serão usados para a criação da interface gráfica do artefato digital e para definir as formas de apresentação

dos gráficos criados a partir dos dados dos currículos Lattes e dados do CNPq sobre projetos de pesquisa. Para cada tipo de dado ou comparação de grandezas, se procurará a mais eficiente maneira de mostrá-las visualmente, buscando para isto, elementos da **Representação Visual de Dados**. Esta disciplina também é referida como **DataViz**, contração da expressão em inglês *Data Visualization*.

2.3.1 A imagem dos números

A representação visual de dados numéricos permite o exame de dados, de informação e de conhecimento em **formas gráficas**, de modo que o usuário possa utilizar sua percepção visual natural para analisar e compreender grandes quantidades de dados com mais rapidez e eficiência, para descobrir características escondidas, padrões e tendências (Freitas *et al.*, 2001, p. 144; Gershon; Eick; Card, 1998, p. 9).

Para Edward Tufte, a excelência em gráficos estatísticos consiste em ideias complexas comunicadas com clareza, precisão e eficiência (Tufte, 2001, p. 10). Segundo ele, apresentações gráficas devem:

- 1) “Apresentar os dados
- 2) “Levar o usuário a pensar sobre o conteúdo dos dados, em vez de se ater à metodologia, tecnologia, design ou outro fator.
- 3) “Não distorcer o que os dados revelam.
- 4) “Mostrar muitos números em pouco espaço.
- 5) “Buscar coerência entre vários conjuntos de dados.
- 6) “Encorajar o olhar a comparar diferentes trechos de dados.
- 7) “Mostrar os dados em diversos níveis de detalhes, da visão geral à microestrutura.
- 8) “Servir a um propósito claro: descrição, exploração, tabulação ou decoração.
- 9) “Estar intimamente integrado com as descrições estatísticas e verbais de um conjunto de dados” (Tufte, 2001, p. 10, tradução deste autor).

2.3.2 DataViz, a visualização de dados

Atualmente, é crescente a quantidade de **dados** que precisam fazer sentido, incluindo o que se convencionou chamar de “*big data*”. Por isto, **dados** exigem cuidados especiais em sua apresentação para que sejam interpretados rápida e efetivamente. Para resolver questões sobre a interpretação de conjuntos de dados numéricos complexos, foi criada a disciplina **Visualização de Dados**. Também chamada informalmente de **DataViz**, Visualização de Dados é o processo de transformação de dados, informações e conhecimento em uma forma audiovisual, explorando as capacidades cognitivas visuais do ser humano (Gershon; Eick; Card, 1998, p. 9). É uma área de aplicação de técnicas de computação gráfica que tem o objetivo de “ajudar o processo de análise e compreensão de um conjunto de dados, por meio de representações gráficas manipuláveis” (Freitas *et al.*, 2001, p. 143)

As técnicas de visualização de informações, para Freitas e colegas, buscam apresentar graficamente dados de uma determinada área de modo que a representação visual gerada explore a capacidade de percepção do ser humano e este interprete e compreenda as informações apresentadas, a partir das relações espaciais exibidas, e, finalmente, deduza novos conhecimentos (Freitas *et al.*, 2001, p. 144).

2.3.2.1 A complexidade da apresentação gráfica

O desenvolvimento de sistemas para visualização de informações pode se tornar bastante complexo em função da necessidade de:

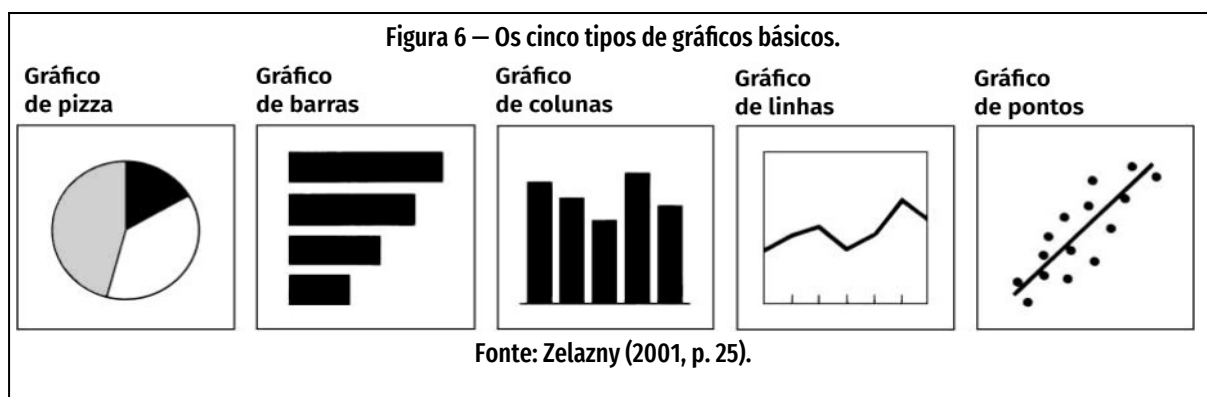
- a) **Criação de uma metáfora visual** que leve à codificação visual do conjunto de informações com o grau de fidelidade necessário ao uso;
- b) **Mecanismos de interação** necessários para manipular os eventualmente volumosos ou complexos conjuntos de dados, e

c) **Implementação de algoritmos geométricos complexos** tanto para a criação da representação visual quanto para sua manipulação (Freitas *et al.*, 2001, p. 145).

Se o item “a” envolve **criação**, uma atividade humana ainda não automatizada, os outros dois itens são de trivial implementação computacional graças à existência de bibliotecas gráficas em linguagens de programação como JavaScript, presente em todos os aplicativos navegadores Web contemporâneos. Tornou-se relativamente simples o desenvolvimento de visualizações complexas, reaproveitando-se de funções desenvolvidas por uma comunidade de programadores (ver Quadro 10, página 129) e colocam seus *softwares* em regime de código livre e aberto (*Free and Open Source Software – FOSS*). Com este mecanismo de licenciamento, o código de um programador pode ser compartilhado pela comunidade, e ele mesmo pode usar dezenas de soluções comunitárias, numa via de mão dupla superavitária para cada profissional. O programador doa um pequeno software e pode usar milhares de softwares de outros programadores. O software livre e aberto foi tratado com mais profundidade na seção “Software livre”, página 2.4 .

2.3.2.2 Tipos de gráficos




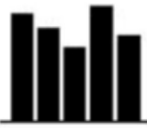
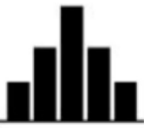
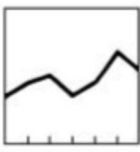


Conforme a Arquitetura da Informação, o estágio de **desenvolvimento** da interface de usuário é o momento de se escolher o tipo de gráfico mais adequado ao que se quer informar com os dados. Cada tipo de finalidade ou de público pode exigir uma metáfora visual, uma forma de expressão gráfica específica.



Zelazny (2001) afirma que existem apenas cinco tipos básicos de gráficos, apresentados na Figura 6, página 48. As centenas de outros tipos são combinações destes básicos. São eles: Gráfico de pizza, gráfico de barras, gráfico de colunas, gráfico de linhas e gráficos de pontos.

A matriz na Figura 7, página 49, resume o uso adequado de gráfico conforme o tipo de comparação de dados desejado. Nas legendas à esquerda da matriz, no eixo vertical, estão os tipos básicos de gráficos propostos por Zelazny (2001, p. 27) dispostos em linhas: pizza, barras, colunas, linhas e pontos. Nas legendas superiores, eixo horizontal, estão os tipos de comparação desejadas: componentes de um todo, itens, série temporal, frequência e correlação. Na junção de linhas com colu-

Figura 7 – Matriz que relaciona tipos de gráfico com o efeito pretendido com os dados.

		TIPOS DE COMPARAÇÃO				
		COMPONENTE	ITEM	SÉRIE TEMPORAL	FREQUÊNCIA	CORRELAÇÃO
TIPOS BÁSICOS DE GRÁFICOS	PIZZA					
	BARRA					
	COLUNA					
	LINHA					
	PONTO					

Fonte: Zelazny, 2001, p.27. Tradução deste autor.

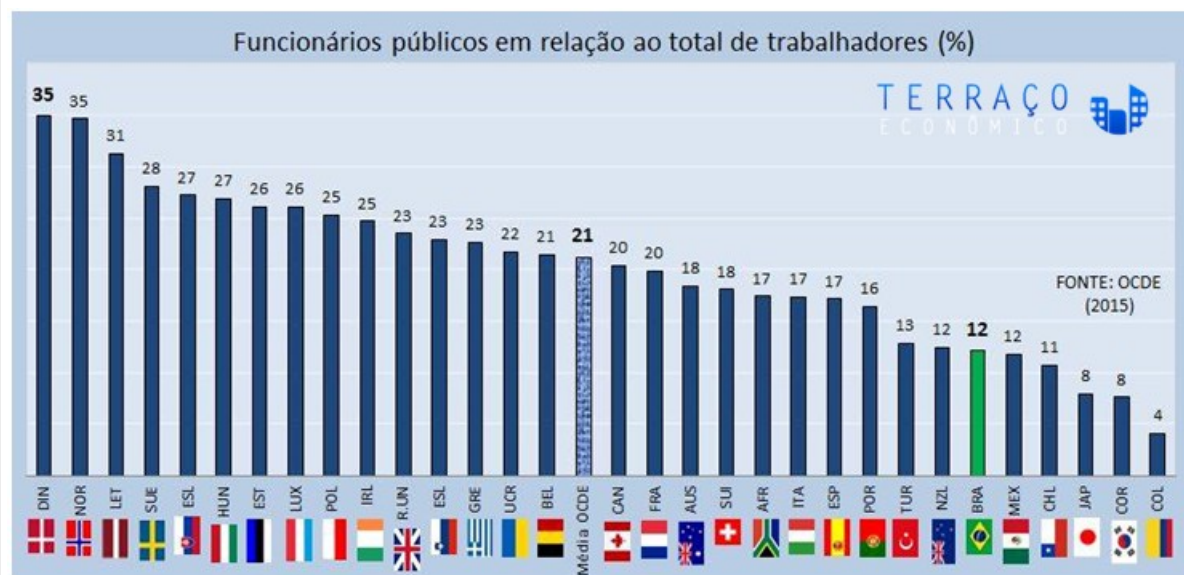
nas estão os ícones correspondentes ao tipo de gráfico adequado ao tipo de comparação desejada. Por exemplo, gráficos de coluna são melhor empregados em **frequências** e em **séries temporais**. Gráficos de barra fazem melhor papel em **comparação entre itens** e em **correlações**.

Pino (2017), Hardin e otros (2012) acrescentam a estes tipos outros gráficos básicos. Na seção seguinte, compilou-se estes aportes e se mostra o uso dos gráficos mais comuns. Os tipos de gráficos têm tópicos chamados de **Função** (o tipo de comparação usada), **Legibilidade** (considera a experiência do leitor com interpretação de dados, ou seja, se os gráficos são fáceis, de dificuldade média ou difíceis de ler), **Uso** (em que contexto usar o gráfico), **Exemplo** (como usar o gráfico) e **Cuidado** (recomendações para usar ou evitar no gráfico).

São estes os gráficos:

- 1) Gráfico de barras horizontais ou verticais
- 2) Gráfico de linhas
- 3) Gráficos de áreas
- 4) Gráfico de pizza
- 5) Gráfico de dispersão
- 6) Diagrama de Gantt
- 7) Gráfico de bolhas
- 8) Histograma
- 9) Marcador
- 10) Mapa
- 11) Velocímetro
- 12) Gráfico de marcador ou *bullet*
- 13) Mapa de árvore hierárquica
- 14) Gráfico de radar

Figura 8 – Gráfico de barras verticais compara porcentagem de funcionários públicos em diversos países.



Fonte: Terraço Econômico, (2016). <https://i0.wp.com/terraoeconomico.com.br/wp-content/uploads/2016/08/Graf.png?quality=75&strip=all>

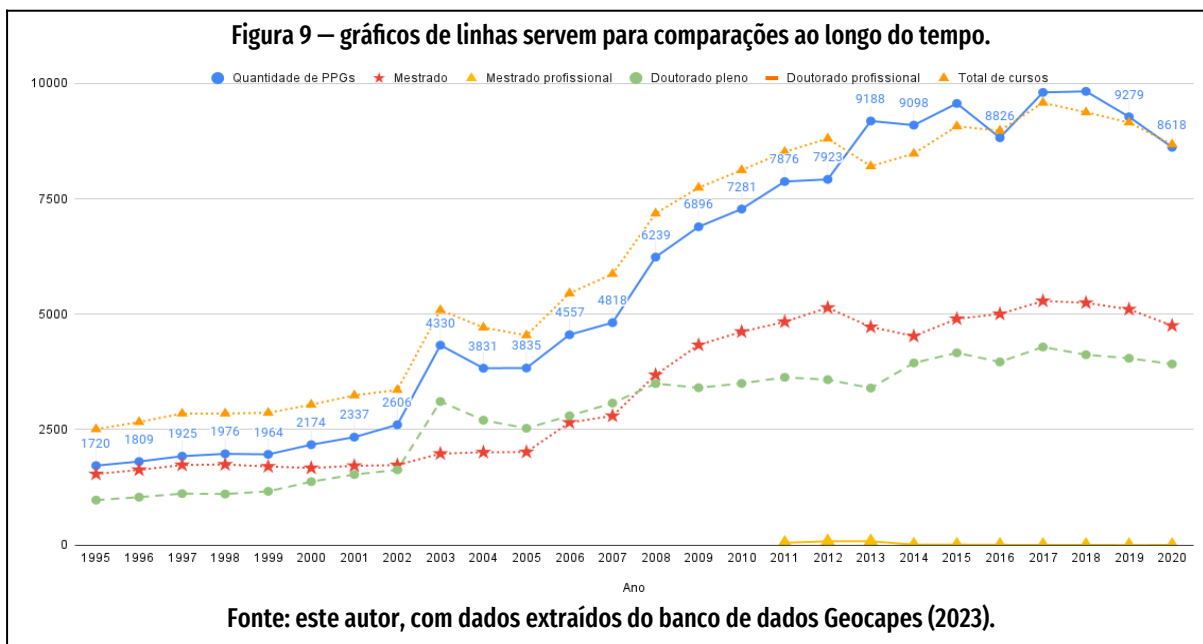
- 15) Histograma
- 16) Gráfico de caixas e traços ou de velas
- 17) Mapa de variação
- 18) Tabela de destaque

Gráfico de Barras horizontais ou verticais

Função: Permite a rápida comparação de grandezas (Figura 8, página 51).

Legibilidade: fácil.

Usos: É um dos formatos mais simples e mais usados para visualização de dados. Gráficos de barras são especialmente eficazes quando se tem “informações numéricas que se dividem muito bem em duas categorias diferentes, possibilitando a visualização rápida das tendências nos dados” (Hardin *et al.*, 2012, p. 2). Os gráficos de barras horizontais geralmente são mais fáceis de ler que os gráficos de barras verticais (Pino, 2017).



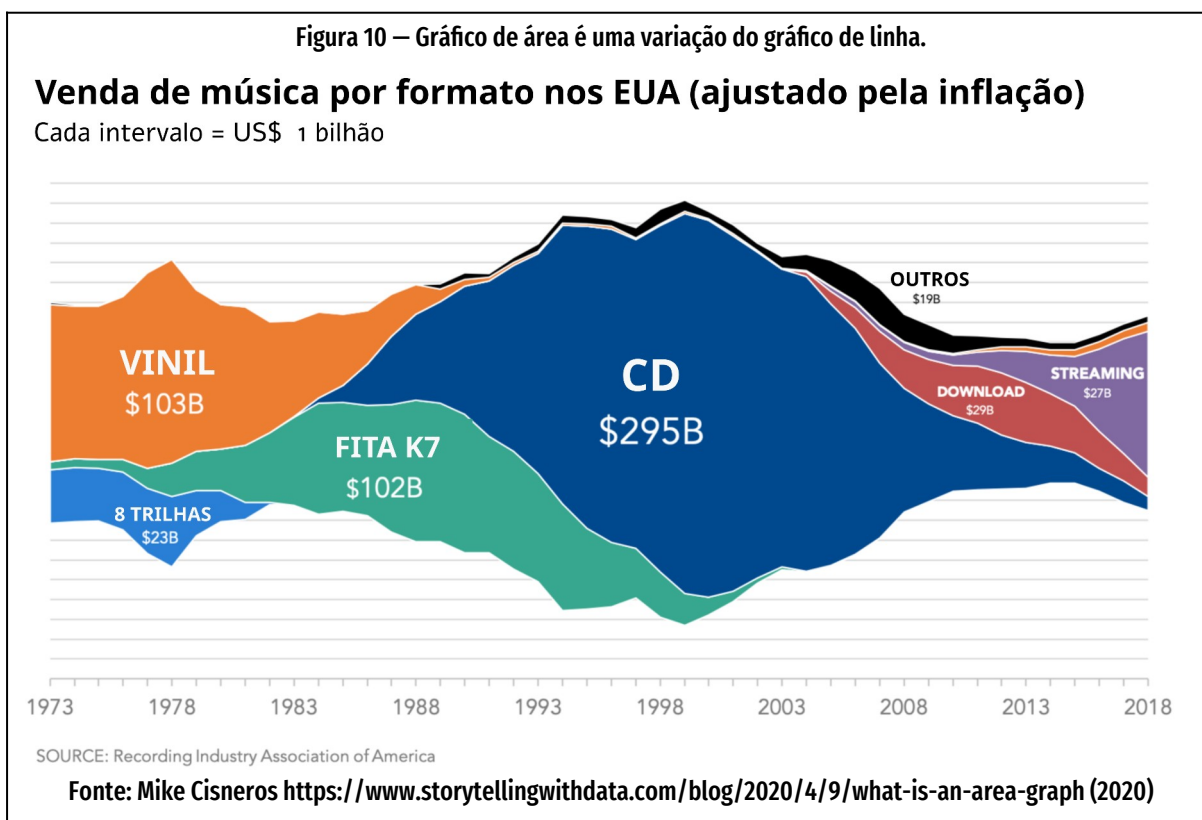
Cuidados: Hardin e colegas recomendam usar gráficos de barras na comparação de dados entre categorias, como percentual de gastos por departamento ou tráfego do site por site de origem. Também sugerem testar ambos os formatos (horizontais e verticais) para decidir qual o melhor conforme o tipo de dado. Aconselham ainda o uso de cores para dar mais impacto ao gráfico e revelar mais informações úteis; variar a orientação das barras (usar barras verticais e horizontais); combinar gráfico de barras com mapas, e usar barras dos dois lados de um eixo – dados positivos de um lado e negativos de outro (Hardin *et al.*, 2012, p. 3).

Gráficos de linhas

Função: comparação.

Legibilidade: fácil de ler e compreender.

Uso: Os gráficos de linha (Figura 9, página 52) conectam pontos que representam dados individuais, resultando numa maneira simples e direta de visualizar uma sequência de valores. Em geral, são usados em percentuais, contagem, frequência relativa, para exibir tendências ao longo do tempo (Hardin *et al.*, 2012, p. 3), em



comparações, dados em muitos períodos não cíclicos ou em poucos períodos e muitas categorias (Pino, 2017). O gráfico de linha é melhor que o gráfico de colunas para enfatizar as mudanças dos dados ao longo do tempo.

Exemplo: cotações de ações na bolsa durante uma semana; temperaturas globais durante um século; produção acadêmica nos últimos 25 anos.

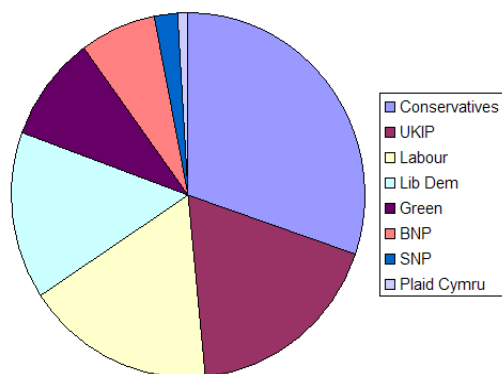
Cuidados: o excesso de linhas pode prejudicar a identificação das legendas. Linhas sobrepostas dificultam a leitura. Pino (2017) recomenda manter no gráfico somente as informações relevantes ou destacar as informações mais importantes.

Gráficos de área

Função: composição.

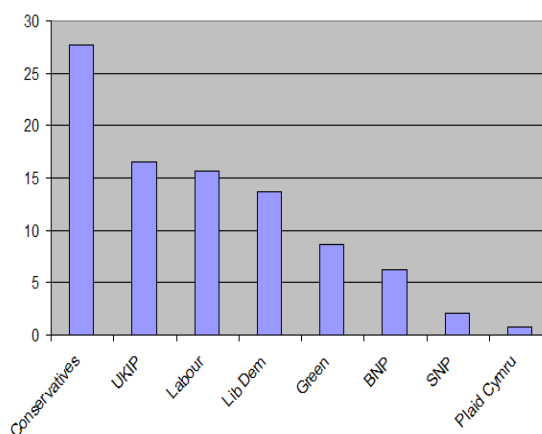
Legibilidade: média.

Figura 11 – Gráficos de pizza podem tornar difícil a comparação de dados.



Qual a diferença entre os três partidos britânicos das fatias de baixo da pizza? Fonte: Fenton (2009).

Figura 12 – Os mesmos dados da pizza, mas em gráficos de barras.



A diferença entre os três partidos com resultados parecidos é facilmente visível. Fonte: Fenton (2009).

Uso: São uma variação dos gráficos de linhas, onde o espaço abaixo da linha é preenchido geralmente com cores. (Figura 10, página 53)⁴. Usa-se em composições variáveis ao longo do tempo, com muitos períodos ou tipos.

Exemplos: Visitas, receitas, vendas, faturamento da indústria da música conforme a mídia, ao longo dos anos.

Cuidados: quando as áreas das séries são empilhadas, pode haver distorção na visão de cada série, pela dificuldade de identificação de padrões (Pino, 2017).

Gráfico de pizza ou de torta

Função: Composição estática, que mostre uma porção simples do total (Pino, 2017).

Legibilidade: fácil.

Cuidados: frequentemente este tipo de gráfico é usado de maneira inapropriada. Fenton (2009) alinha três argumentos contra o uso indiscriminado de gráficos de pizza. Primeiro, gráficos de pizza dificultam a comparação de dados. Segundo, o

4 <https://www.storytellingwithdata.com/blog/2020/4/9/what-is-an-area-graph>

gráfico de pizza não fica mais legível quando é aumentado, ao contrário do gráfico de barras. E terceiro, se o objetivo é comparar dados, barras verticais ou horizontais fazem melhor a tarefa (Figura 11 e Figura 12, página 54). “Não peça ao usuário para traduzir fatias de pizza em dados relevantes ou para comparar uma pizza com outra. Os pontos principais de seus dados serão perdidos, e o usuário terá que trabalhar muito”, argumentam Hardin *et al* (2012, p. 5).

“A razão pela qual você está mostrando informação em um gráfico é que você quer fazê-la fácil de interpretar e comparar informação. Gráfico de pizza faz exatamente o contrário”, diz Fenton (2009).

Usos: Para Hardin e outros (2012, p. 5), “gráficos de pizza devem ser usados para mostrar proporções relativas – ou porcentagens – de informações”. Ou seja: quando a soma dos diversos elementos do gráfico for 100%.

No entanto, apesar de suas desvantagens, os gráficos de pizza podem ser bem usados se limitarem a quantidade de fatias e se forem combinados, por exemplo, com um mapa (Hardin *et al.*, 2012, p. 6). Ou, como Fung sugere (2020), seguir algumas convenções: usar um número razoável de fatias – 5 ou 6 fatias – e uma última fatia agrupando itens restantes, se houver, com o nome de “Outros”; Ordenar as fatias da maior para a menor; Colocar a fatia “Outros” em último lugar; Posicionar a maior fatia iniciando no raio vertical superior, na posição das 12 horas, e arranjar as fatias restantes no sentido horário; Variar a cor das fatias apenas se a cor está codificada nos dados. Neste caso, usar um tom mais claro na última fatia, “Outros”, indicando que a fatia é um grupo de itens e a menos importante do gráfico.

Cuidados: Não usar quando a soma das fatias não for 100%. Não usar mais de seis fatias. Não usar efeitos tridimensionais nem de perspectiva.

Figura 13 – Gráfico de dispersão relaciona PIB com resultado do teste PISA entre países da OCDE.

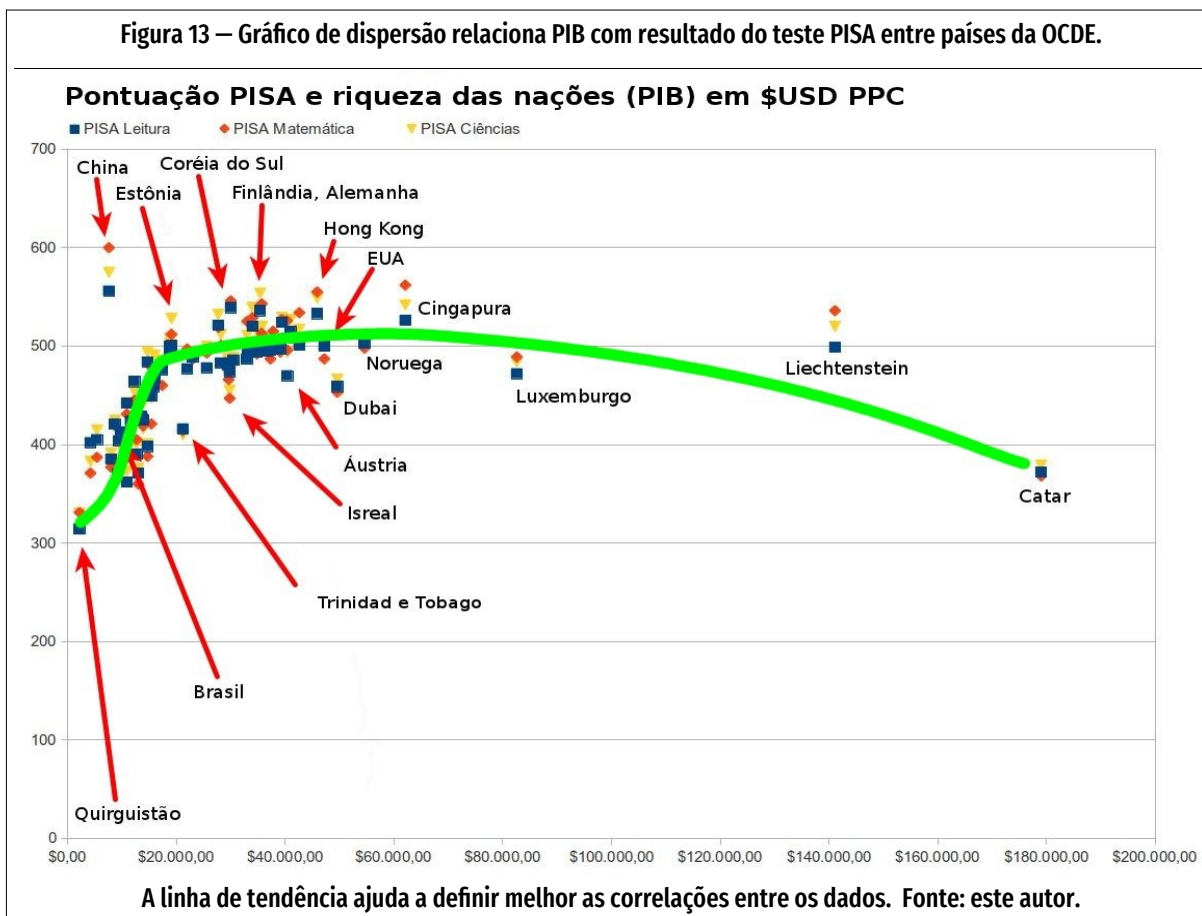
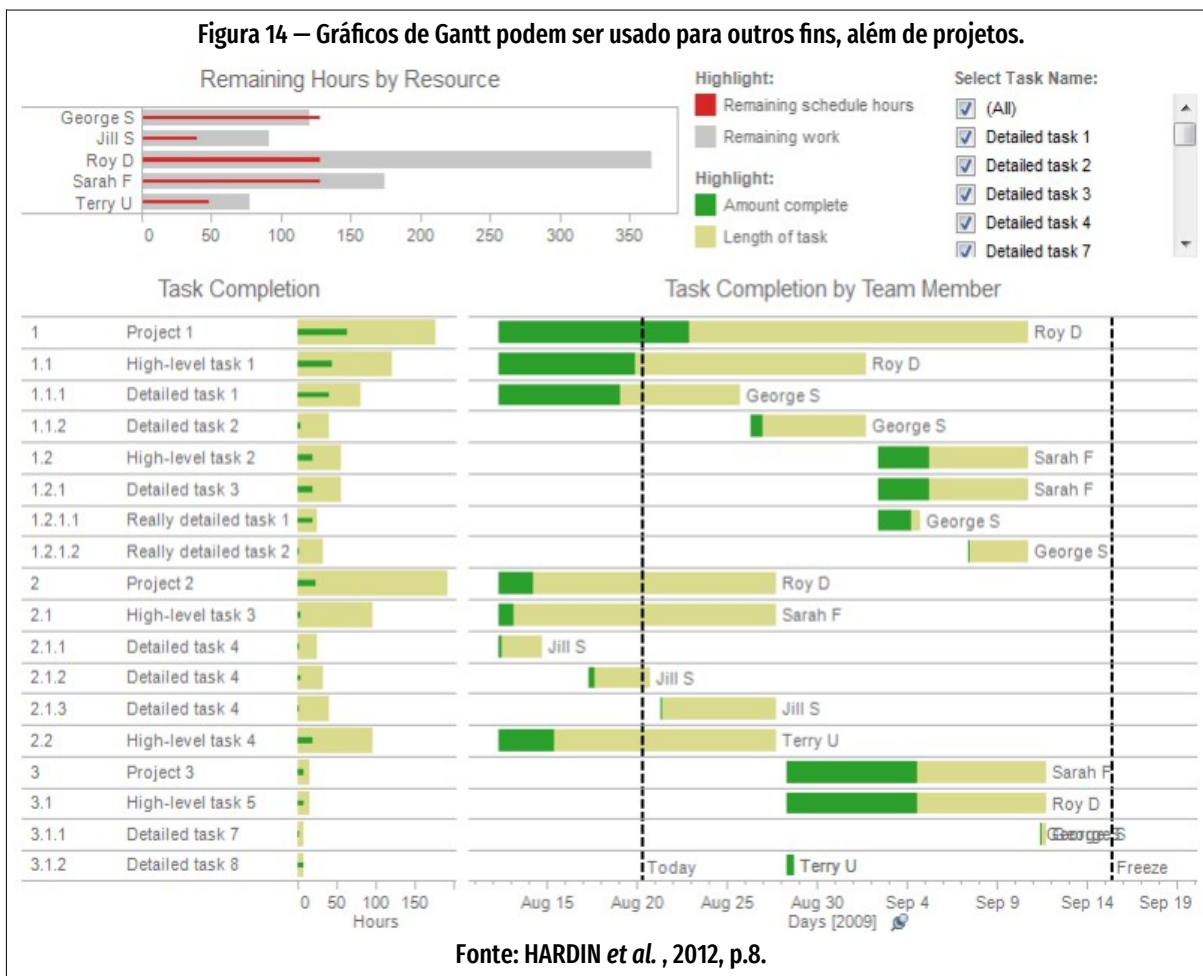


Gráfico de dispersão

Função: distribuição ou relação

Legibilidade: média.

Uso: usa-se o gráfico de dispersão para analisar dados com mais profundidade e verificar se e como diferentes informações estão relacionadas. Eles são eficazes em dar senso de tendências, de concentrações e de exceções (Figura 13), direcionando para onde se quer concentrar mais esforços de investigação (Hardin *et al.*, 2012, p. 8). Este tipo de gráfico serve para se observar o que acontece com uma variável quando outra se altera. Com isto, pode-se identificar relações de causa e efeitos (Pino, 2017).



Alguns autores recomendam usar-se uma linha de tendência a fim de tornar a correlação entre os dados mais claramente definida (Hardin *et al.*, 2012, p. 8).*

Exemplo: correlação entre notas do teste PISA e PIB dos países (Figura 13, página 56).

Cuidados: a correlação nem sempre indica relação direta de causalidade.

Diagrama de Gantt

Função: são mais associados ao gerenciamento de projetos. No entanto, também podem ser usados para entender como outros elementos, como pessoas ou máquinas, variam ao longo do tempo.

Legibilidade: difícil.

Exemplo: pode-se usar Gantt para fazer o planejamento de recursos e ver quanto tempo as pessoas levarão para atingir determinados marcos e como isso foi distribuído ao longo do tempo (Hardin *et al.*, 2012, p. 8). Ou mostrar a disponibilidade de participantes de uma equipe ou duração de uso de máquinas ou recursos (Figura 14, página 57).

Cuidados: o uso de cor nos gráficos de Gantt e a sua combinação com outros tipos de gráficos, em um *dashboard*, pode ajudar a filtrar e analisar dados para ampliar as informações úteis disponíveis.

Gráfico de Bolhas

Função: As bolhas não são propriamente um tipo de visualização. Elas são um recurso usado para representar dados em mapas ou gráficos de dispersão, criando “eixos” ou “dimensões” a mais (Figura 15, página 59).

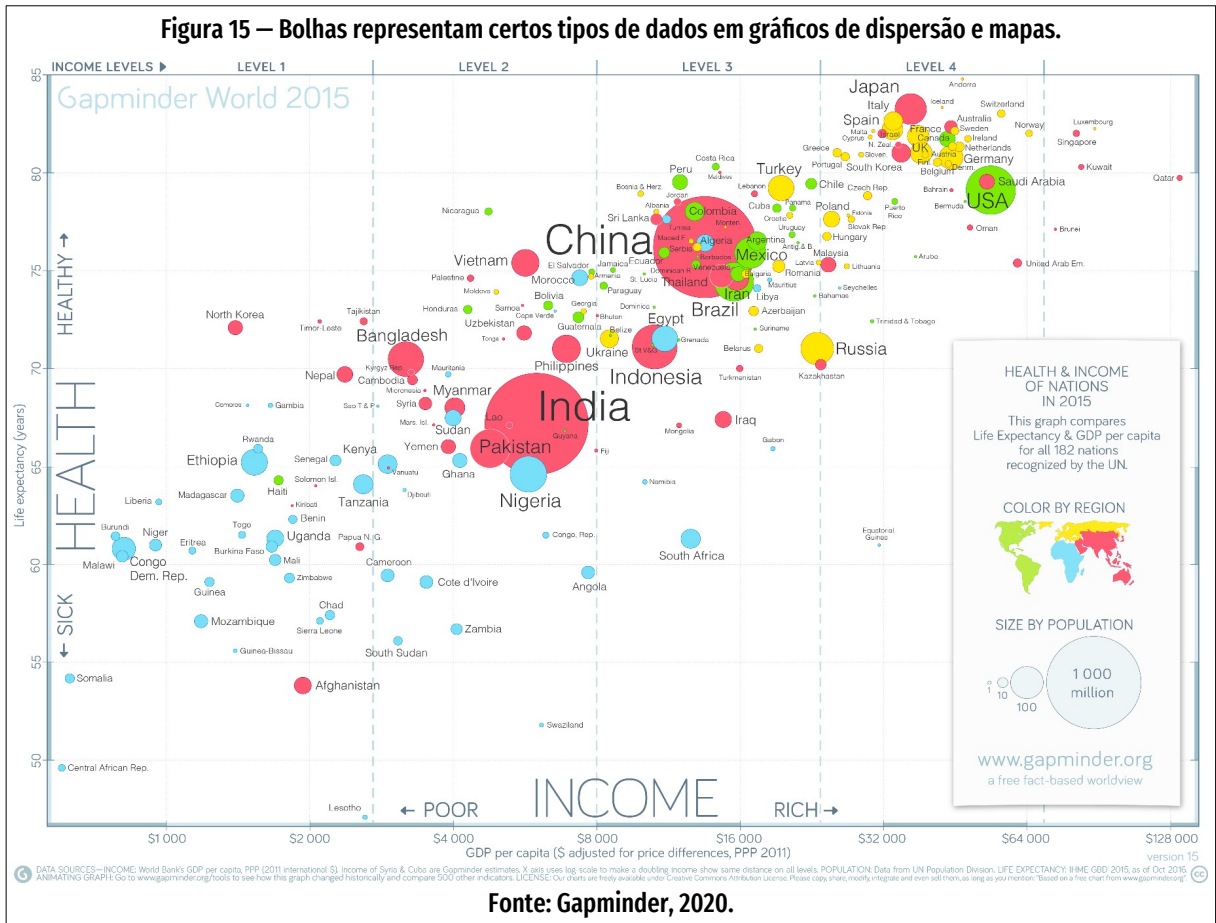
Uso: Usam-se as bolhas para mostrar a concentração de dados ao longo de dois eixos. Por exemplo, a taxa de mortalidade infantil em relação ao PIB dos países. As bolhas também podem ser sobrepostas a mapas, contextualizando rapidamente o usuário sobre a qual localidade pertencem os dados (Hardin *et al.*, 2012, p. 9).

Cuidados: “Quando as bolhas variam por cor e tamanho, o impacto é ampliado.” (Hardin *et al.*, 2012, p. 9)

Mapa de variações

Função: comparação.

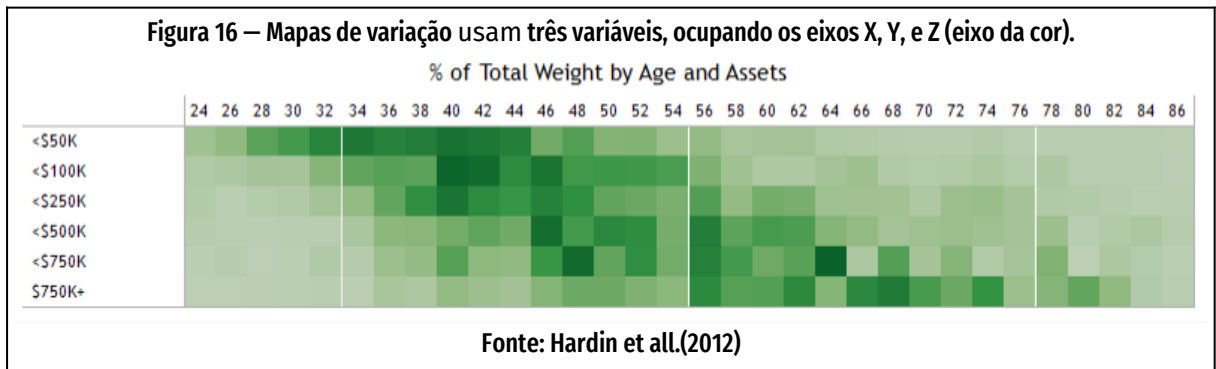
Figura 15 – Bolhas representam certos tipos de dados em gráficos de dispersão e mapas.



Uso: os mapas de variações comparam dados em duas categorias com o uso de cores (Figura 16, página 59), permitindo que se veja rapidamente onde a interseção das categorias é mais forte e mais fraca (Hardin *et al.*, 2012, p. 12).

Pode-se variar o tamanho dos quadrados ou elementos usados. Assim, obtêm-se uma terceira dimensão, além dos eixos X e Y do gráfico, a dimensão da cor ou do tom. (Hardin *et al.*, 2012, p. 12).

Figura 16 – Mapas de variação usam três variáveis, ocupando os eixos X, Y, e Z (eixo da cor).



Fonte: Hardin et al.(2012)

Figura 17 – A combinação de gráficos de linha com tabelas destacadas fornece detalhes adicionais.

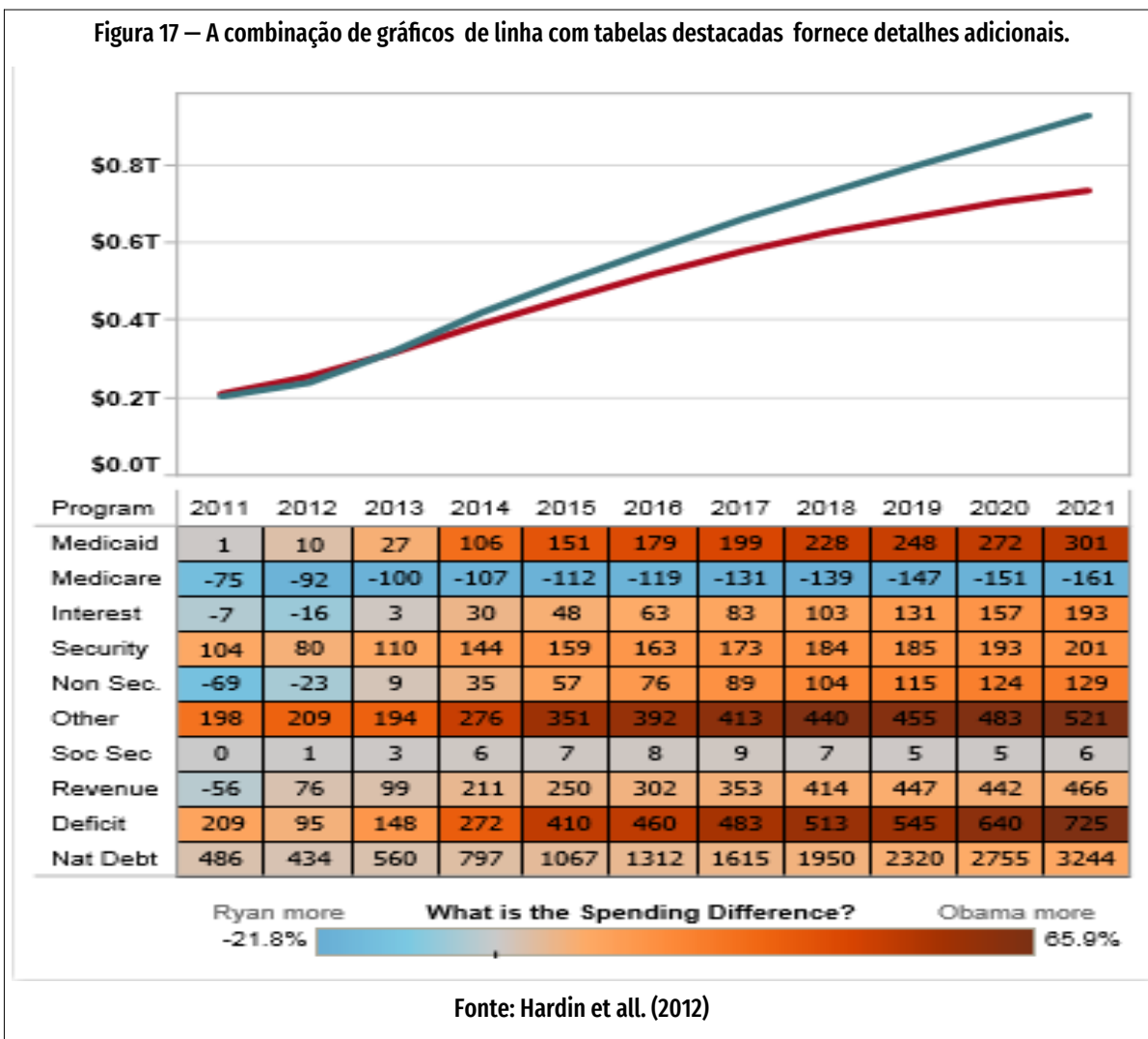
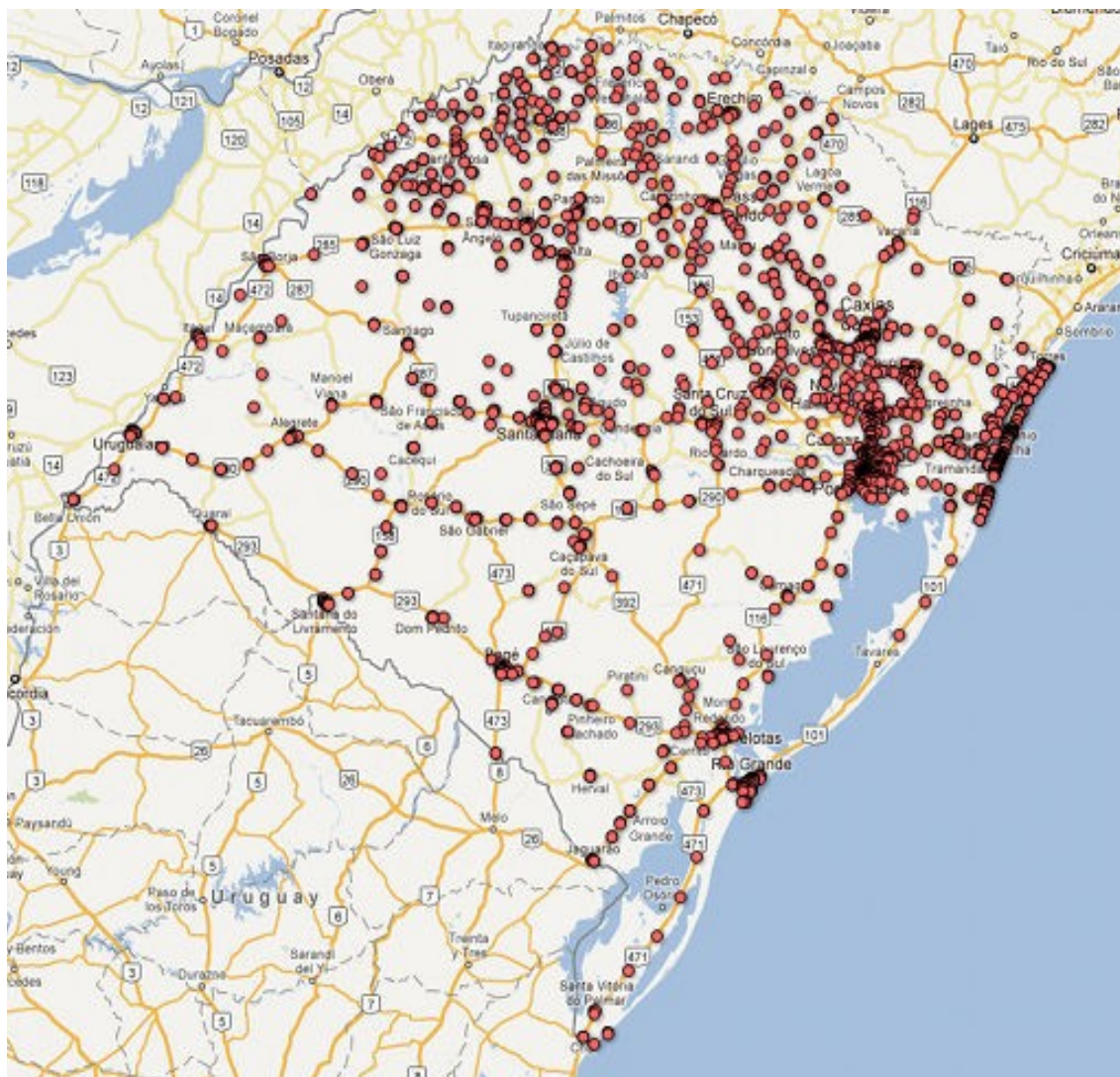


Tabela de destaque

Função: Tabelas de destaque são uma evolução dos mapas de variação (Hardin et al., 2012, p. 12)

Uso: Tabelas de destaque são usadas ao se fornecer informações detalhadas, em algarismos arábicos, em mapas de variações (Figura 17, página 60). Além da cor – elemento icônico – comunica-se também valores numéricos decimais – elementos simbólicos.

Figura 18 – Dados georreferenciados podem ser apresentados como mapas.



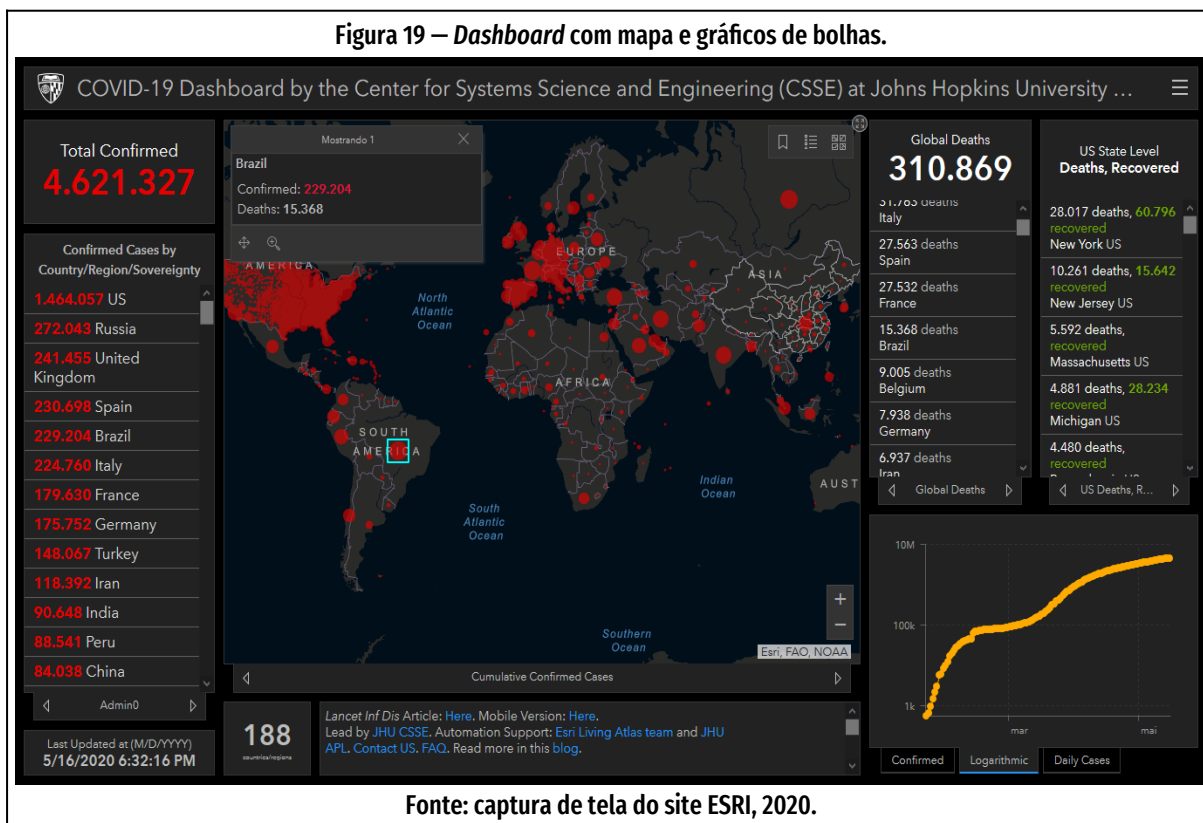
Localização de torres de telefonia celular no Rio Grande do Sul.

Fonte: este autor com dados da Anatel.

Cuidado: Combinar tabelas de destaque com outros tipos de gráficos, como um gráfico de linhas, “permite que um usuário entenda tendências gerais assim como análise rapidamente em uma seção cruzada específica de dados” (Hardin *et al.*, 2012, p. 15).

Mapa

Função: distribuição geográfica.



Legibilidade: fácil.

Uso: Dados geográficos e de local (códigos postais, abreviações de estado, nomes de países ou suas codificações geográficas) são candidatos à apresentação em forma de mapas (Figura 18, página 61).

Exemplo: Mapas do coronavírus nas entidades da área da saúde e imprensa, Figura 19, página 62.

Cuidados: Mapas podem ser usados com países, estados ou municípios em cores diferentes, com a tonalidade proporcional aos valores atribuídos à entidade, ou pode fazer dos elementos do mapa um índice ou menu com links para detalhamento da informação. Alguns exemplos: quantidade de estudantes por município, solicitações de seguro por estado, destinos de exportação de produto por país, acidentes de carro por CEP (Hardin *et al.*, 2012, p. 6).

Figura 20 – Velocímetro (*Gauge*, em inglês) é bastante usado.



Fonte: Pino e Galindo, 2017.

Esses autores sugerem também que se distribua gráficos de bolhas sobre mapas representando a concentração de dados (Figura 27). Seus tamanhos variados são uma maneira rápida de comparar dados. Também pode-se distribuir gráficos de pizza sobre os mapas, uma maneira interessante de mostrar tendências regionais. Deve-se manter pequeno o número de fatias de pizza, neste caso (Hardin *et al.*, 2012, p. 6).

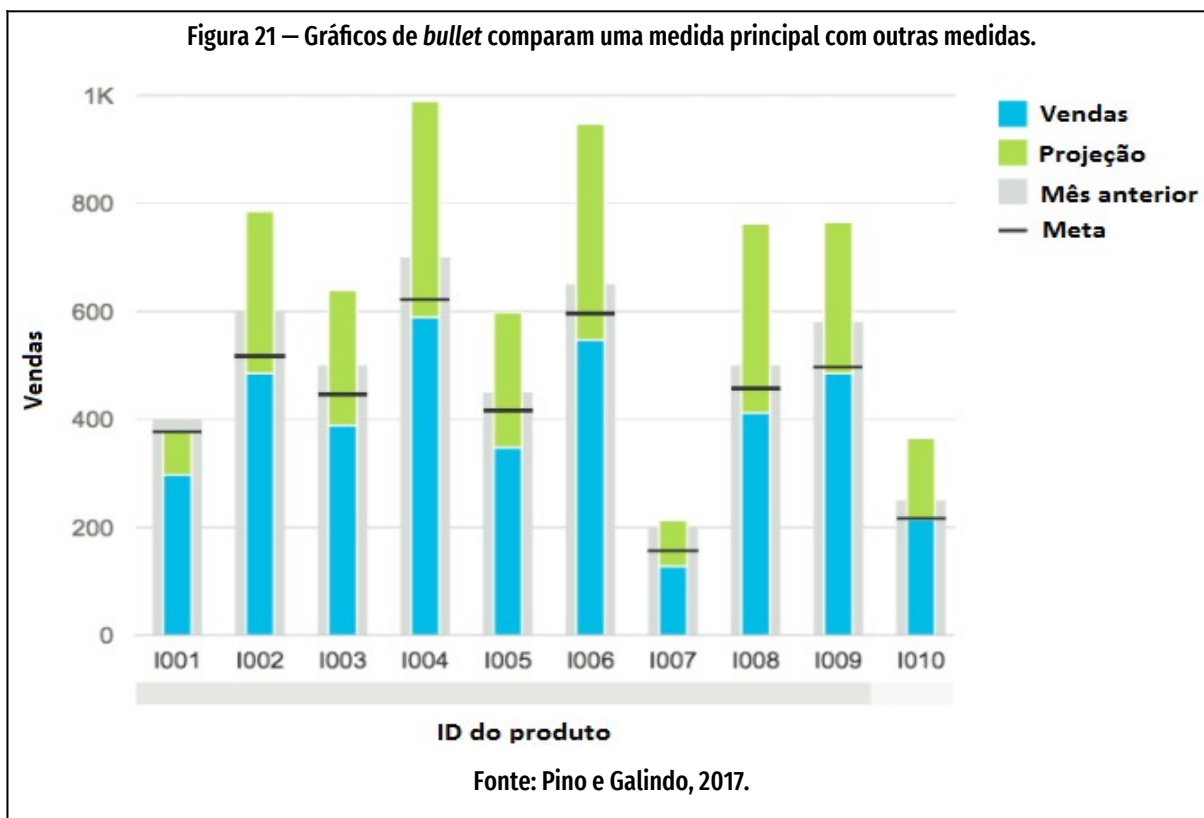
Velocímetro (*gauge*)

Função: comparação.

Legibilidade: fácil (Figura 20, página 63).

Uso: acompanhamento de metas, alertas para instabilidades.

Cuidado: a percepção visual de comprimento por humanos é melhor que a percepção de áreas, fazendo com que gráficos lineares (colunas, barras, *bullet*) sejam mais



fáceis de ler que gráficos radiais (pizza, radar, velocímetro). Embora seja considerada uma escolha atraente e muito usada, deve-se considerar o gráfico *bullet* mais adequado para este tipo de dados (Pino; Galindo, 2017b).

Bullet ou marcador

Função: comparação.

Legibilidade: média.

Cuidado: Pino e Galindo consideram que o gráfico *bullet* (Figura 21, página 64) não é tão comum quanto o velocímetro, embora mostre o cumprimento da meta de forma mais precisa que o *gauge*.

Figura 22 – o gráfico de árvore hierárquica dá uma ideia do todo.



Uso: Gráficos de marcador são uma variação dos gráficos de barras. Eles comparam uma medida principal com outras medidas, geralmente metas a serem atingidas ou preço de ações. São usados para avaliar desempenho de uma métrica em relação a um objetivo. Este tipo de gráfico foi criado para substituir níveis, medidores, velocímetros e termômetros de *dashboards*, porque estas imagens geralmente não exibem informações suficientes e exigem um espaço valioso em painéis (Hardin *et al.*, 2012, p. 10).

Exemplos: meta de vendas, objetivos de atividades.

Cuidado: o gráfico de marcador é mais raro que o velocímetro, mas mostra o cumprimento de metas mais precisamente que este tipo (Pino; Galindo, 2017b).

Mapa de árvore hierárquica

Função: composição.

Legibilidade: difícil.

Uso: Esses gráficos (Figura 22, página 65) usam retângulos aninhados em outros retângulos, para mostrar dados hierárquicos – por isso se chama de “árvore hierárquica”, em referência a “galhos” grandes que se subdividem em galhos menores – como uma proporção do todo (Hardin *et al.*, 2012, p. 13). Os retângulos podem ainda ser de cores diferentes conforme a categoria dos dados, aumentando a carga de informação, como mostra a Figura 22.

Exemplo: Mapas de árvore podem ser usados na representação de armazenamento de dados em computadores, gerenciamento do número e da prioridade de casos de suporte técnico, comparação de orçamentos fiscais de cada ano, área de aplicação dos impostos. Pode-se combinar gráficos de barra com gráficos de árvore, como na Figura 26, página 70, que mostra a evolução do PIB mundial ao longo do tempo (Hardin *et al.*, 2012, p. 13).

Gráfico de Radar

Função: comparação.

Legibilidade: difícil.

Uso: dados com muitas variáveis no mesmo eixo ou sobre períodos cíclicos. Também usado em avaliações qualitativas conforme critérios passíveis de serem quantificados (Figura 23, página 67).

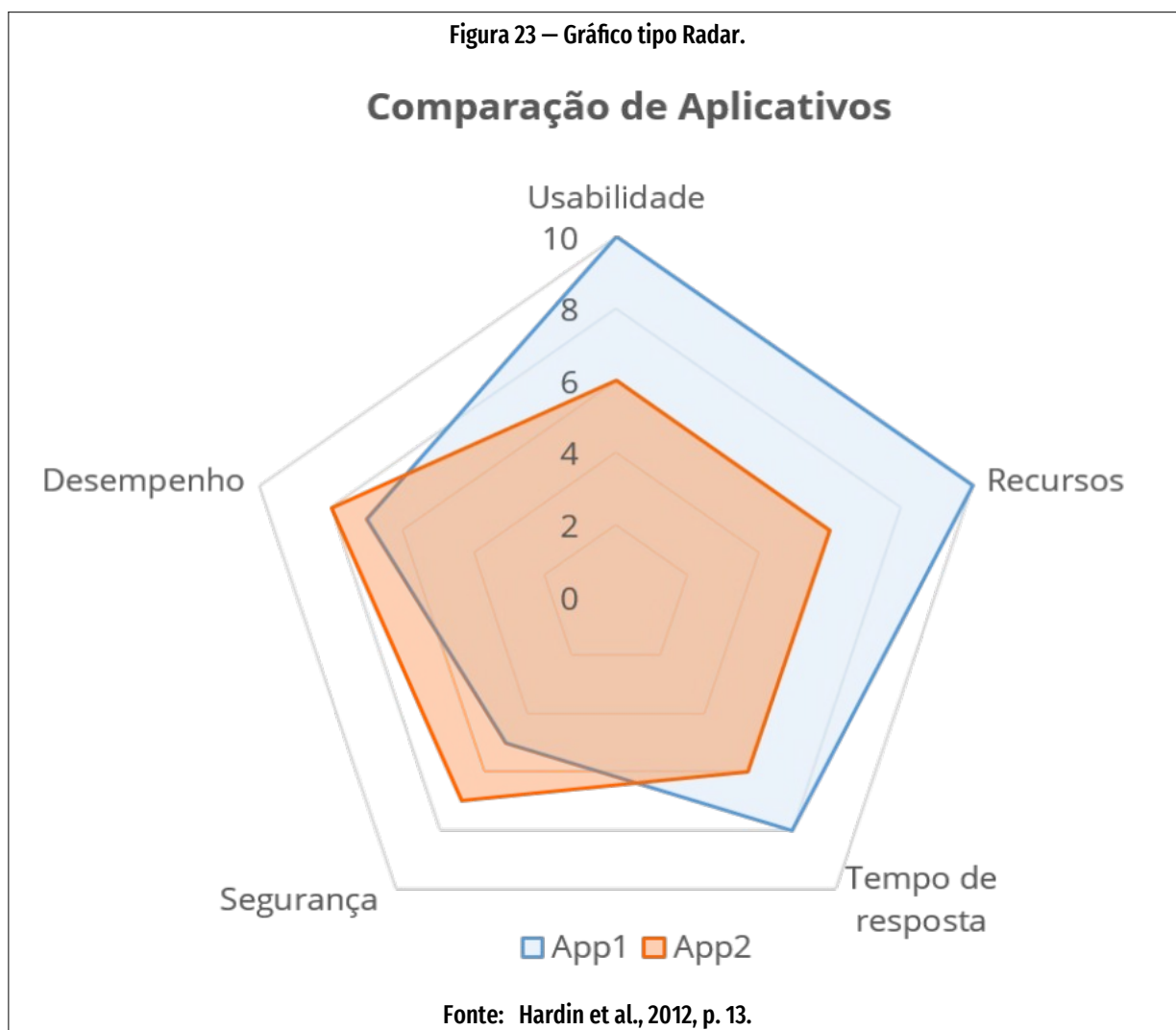
Cuidado: a ordem dos eixos é importante. Sua alteração pode influenciar a figura e modificar a interpretação dos dados.

Histograma

Função: distribuição.

Legibilidade: fácil.

Uso: quando se quer ver como dados são distribuídos entre grupos. Cria-se um histograma ao agrupar dados em alguma categoria e, em seguida, plotá-los com barras verticais ao longo de um eixo (Figura 24, página 68). Se verá a distribuição dos itens de acordo com suas propriedades.



Cuidado: frequentemente, não necessariamente se saberá qual abordagem de categorização faz sentido para determinados dados. “Pode usar histogramas para testar diferentes abordagens e verificar se os grupos de dados estão equilibrados em tamanho e em relevância para análise” (Hardin *et al.*, 2012, p. 10).

Quando o objetivo for testar diferentes agrupamentos de dados, é conveniente que o histograma seja interativo e conte com filtros, isto é, com meios que permitam apresentar e comparar rapidamente diferentes valores. “Quando você está explorando seus dados e procurando agrupamentos ou ‘compartimentos’ que fazem sentido, criar uma variedade de histogramas pode ajudá-lo a determinar o conjunto de dados mais útil” (Hardin *et al.*, 2012, p. 10).

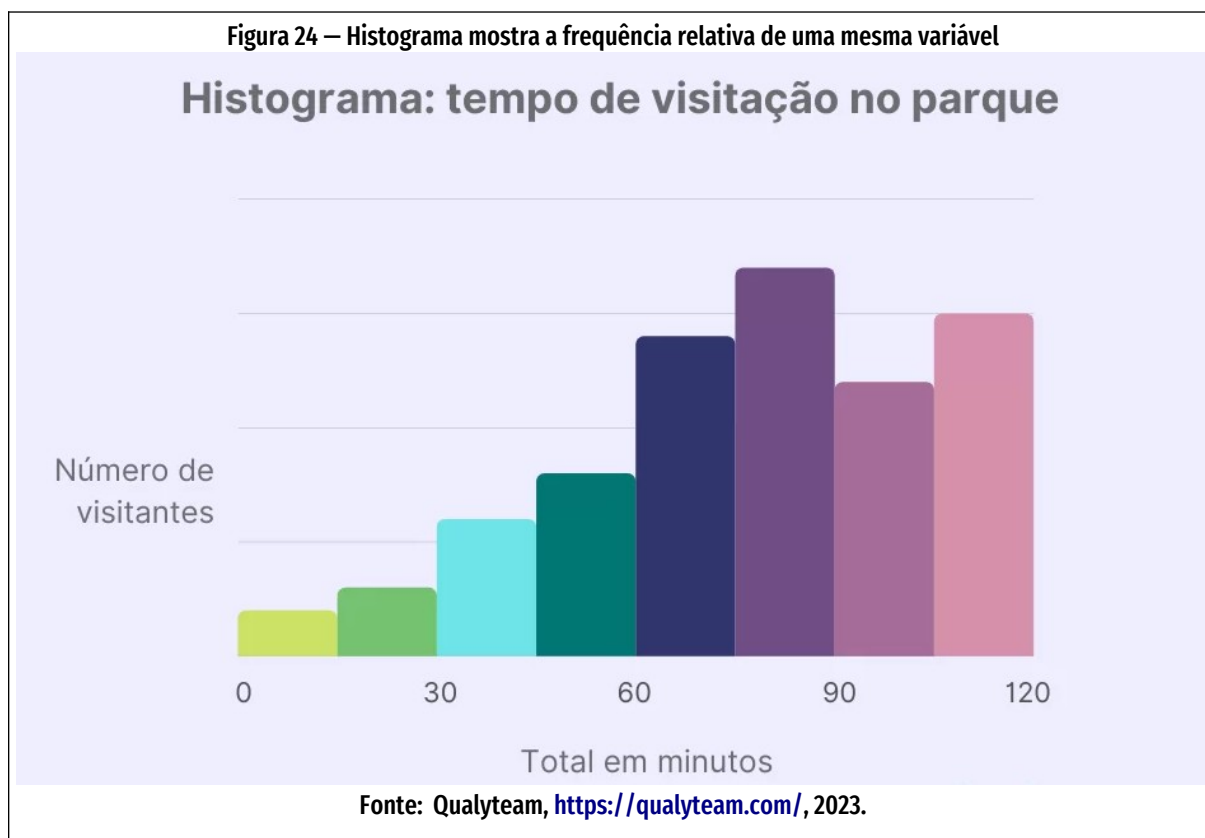


Gráfico de Caixas e Traços ou de Velas

Função: distribuição.

Legibilidade: difícil.

Uso: os gráficos Caixa e Traços, ou de velas ou *boxplots* (Figura 25, página 70), são uma combinação de outros formatos: gráficos de barras e gráficos de dispersão. É uma importante forma de mostrar as distribuições de dados em conjuntos estatísticos. A compreensão dos conceitos estatísticos de **mediana**⁵ e de **quartis**⁶ é necessária para interpretar gráficos de Vela.

O tipo de gráfico de caixa tem este nome porque se constitui de duas partes: a caixa, que contém a mediana dos dados junto com o 1º e o 3º quartis (25% maior e 25% menor do que a mediana); e os traços, que representam os dados uma vez e meia a amplitude interquartilica (a diferença entre o 1º e o 3º quartis). Os traços também podem ser usados para mostrar os pontos mínimos e máximos dos dados (Hardin *et al.*, 2012, p. 14). Em outras palavras, “os gráficos Caixas e Traços são usados para mostrar onde estão localizados os 50% dos valores mais prováveis, a e os valores extremos de um conjunto” (Pino; Galindo, 2017a). Eles permitem compreender-se rapidamente os dados, ver como os eles estão distorcidos em uma extremidade e identificar exceções no conjunto (Hardin *et al.*, 2012, p. 14).

5 “**Mediana** é o valor que separa a metade maior e a metade menor de uma amostra, uma população ou uma distribuição de probabilidade. Em termos mais simples, mediana pode ser o valor do meio de um conjunto de dados” (Mediana (estatística), 2019)

6 “Na estatística descritiva, um **quartil** é qualquer um dos três valores que divide o conjunto ordenado de dados em quatro partes iguais, e assim cada parte representa 1/4 da amostra ou população” (Quartil, 2019)

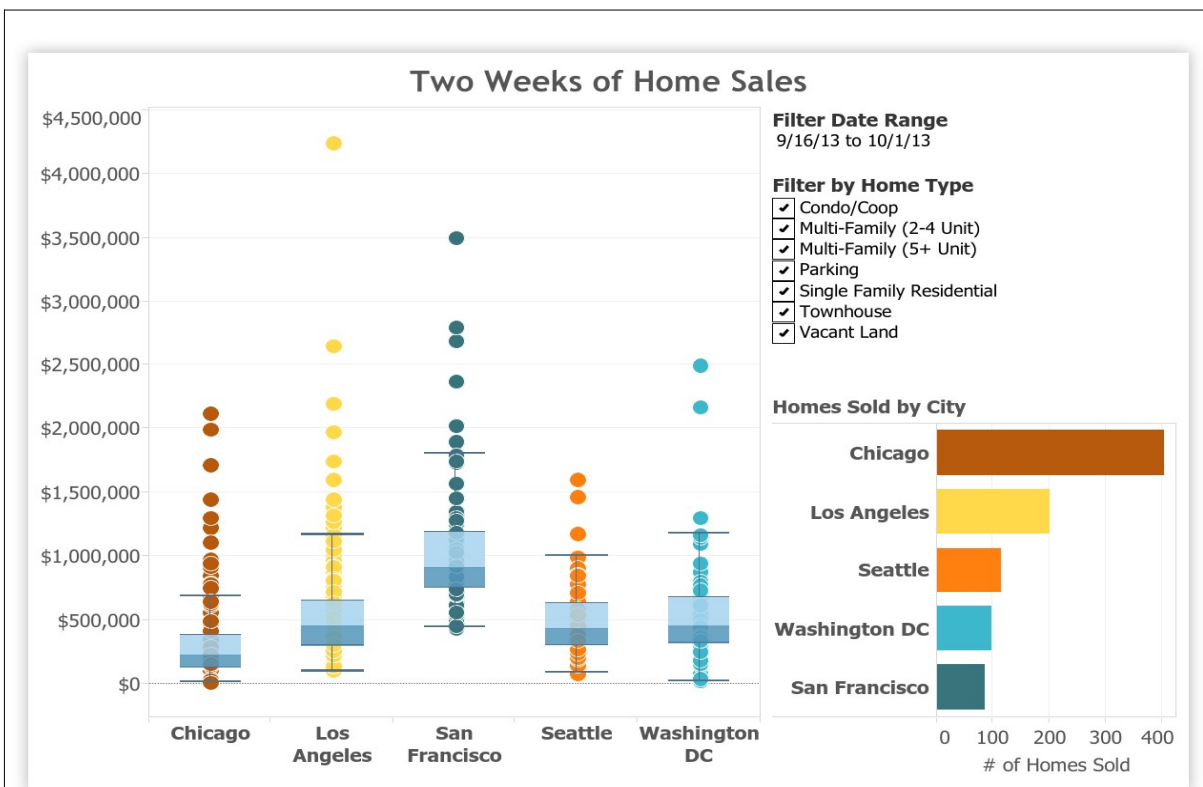
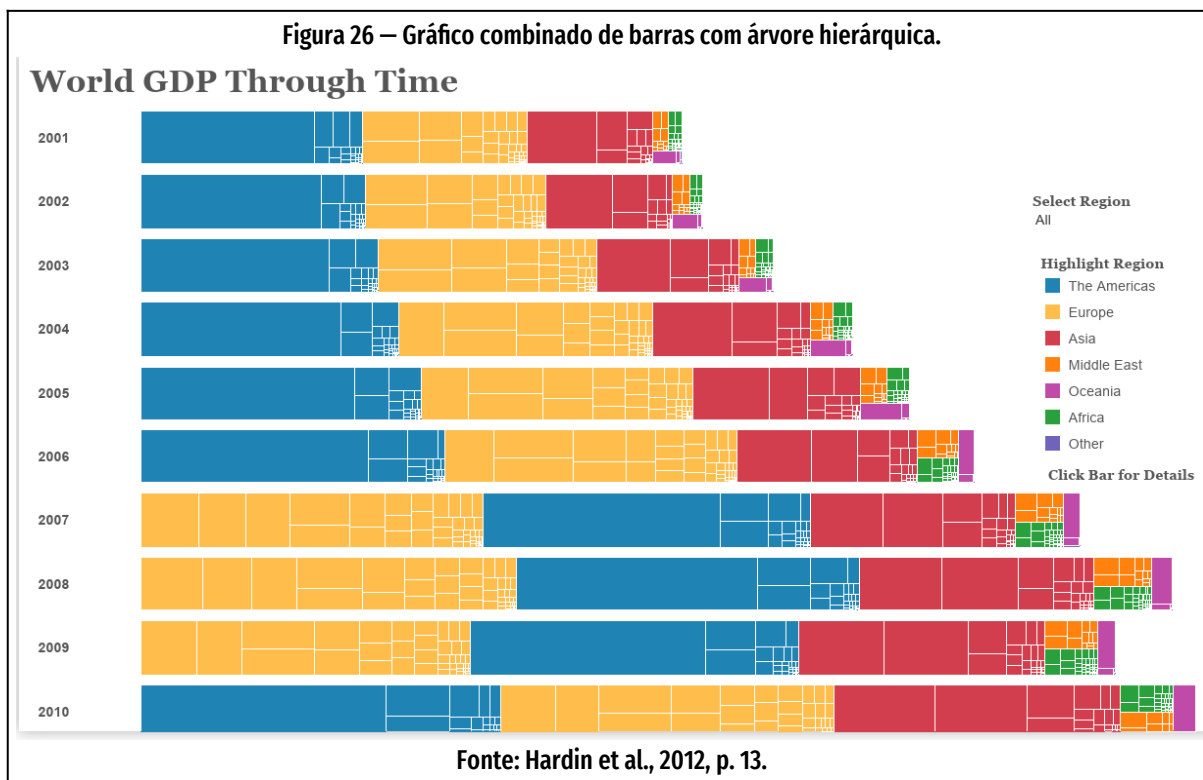


Figura 25 – Gráfico de Caixas e Traços combinam dois ou mais tipos de gráficos.
Fonte: (HARDIN et al., 2012, p. 14)



Fonte: Hardin et al., 2012, p. 13.

2.3.3 Como escolher um gráfico

A numerosa quantidade de gráficos diferentes pode dificultar a escolha da melhor opção conforme a intenção de seu autor.

A escolha de um gráfico depende do tipo de dado a ser representado, e este depende da intenção do autor da apresentação, que depende do público a que se destina. A partir deste ponto de vista, Abela (2009) cria um fluxograma com quatro dos principais motivos usados para a apresentação de dados, as variáveis disponíveis e o tipo de gráfico mais adequado para apresentar estes dados (Figura 27, página 72).

O objetivo de uma apresentação tem, segundo Abela, principalmente, uma destas operações:

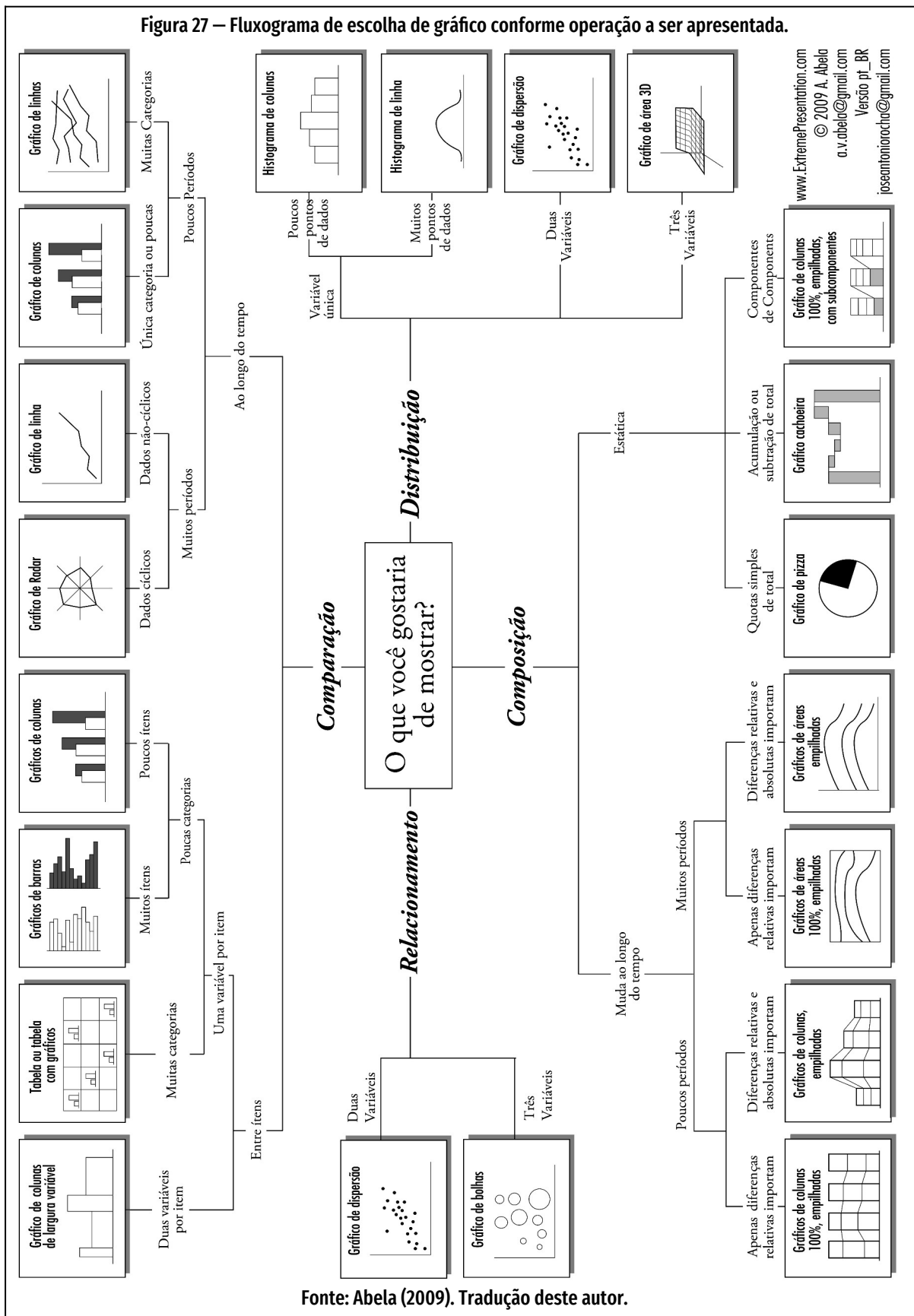
- 1) **comparação** de dados,
- 2) **composição** de dados,
- 3) **distribuição** de dados e
- 4) **relacionamento** de dados.

Cada operação tem características próprias que vão definir o tipo de gráfico adequado. Para melhor legibilidade, os dados da Figura 27, página 72, estão também no Quadro 4, página 73.

Depois de definidos os tipos adequados de gráficos, pode-se combiná-los em um painel (*dashboard*), criando-se um ambiente onde pode-se não apenas perguntar, mas obter-se respostas em tempo real (HARDIN et al., 2012, p. 15).

Essas relações ficam mais claras através do quadro construído por Abela, apresentado na Figura 28 — Mapa mental do processo de produção de dashboards.

Figura 27 – Fluxograma de escolha de gráfico conforme operação a ser apresentada.



Quadro 4 – Tipos de gráficos conforme o tipo de operação

Operação	Detalhe 1	Detalhe 2	Detalhe 3	Detalhe 4	Gráfico
Comparação	Entre itens	Duas variáveis por item			Gráfico de coluna de largura variável
		Uma variável por item	Muitas categorias		Tabela simples ou tabela com gráficos
			Poucas categorias	Muitos itens	Gráfico de barras
				Poucos itens	Gráfico de colunas
	Ao longo do tempo	Muitos períodos	Dados cíclicos		Gráfico Radar
			Dados não cíclicos		Gráfico de linhas
		Poucos períodos	Única categoria, ou poucas		Gráfico de colunas
			Muitas categorias		Gráfico de linhas
Composição	Muda ao longo do tempo	Muitos períodos	Apenas diferenças relativas importam		Gráfico de áreas empilhadas 100%
			Diferenças relativas e absolutas importam		Gráfico de áreas empilhadas
		Poucos períodos	Apenas diferenças relativas importam		Gráfico de áreas empilhadas 100%,
			Diferenças relativas e absolutas importam		Gráfico de áreas empilhadas
	Estática	Quotas simples de total (100%)			Gráfico de pizza
		Acumulação ou subtração de total			Gráfico cachoeira
		Componentes de componentes			Gráfico de colunas empilhadas 100%, com subcomponentes
	Distribuição	Variável única	Poucos pontos de dados		
Muitos pontos de dados			Histograma de linha		
Duas variáveis			Gráfico de dispersão		
Três variáveis			Gráfico de área 3D		
Relacionamento	Duas variáveis			Gráfico de dispersão	
	Três variáveis			Gráfico de bolhas	
Fonte: Abela (2009)					

2.3.4 Dashboards

Várias definições de **dashboard** (painel) foram compiladas por Stephen Few em um enunciado que define este dispositivo como “um painel das informações mais importantes necessárias para alcançar alguns objetivos, consolidadas e organizadas em uma única tela para que as informações possam ser monitoradas rapidamente” (Few, 2006a, p. 26). Para Few, o ideal é que o painel ocupe uma única tela que monitore todas as informações necessárias para alcançar seu objetivo específico. As informações devem ser principalmente visuais, pois graficamente pode-se comunicar “com mais eficiência que a informação apenas textual” (Few, 2006a, p. 26).

Os *dashboards*, se interativos, devem ter mecanismos de exibição pequenos, concisos, claros e intuitivos. Os painéis devem ser personalizáveis, isto é, devem permitir filtros de informações para que possibilitem a rápida identificação de tendências, padrões e anomalias e possam ajudar a orientar decisões efetivas (Few, 2006a, p. 27).

2.3.4.1 Categorias de dashboards

Painéis podem ser categorizados de diversas maneiras. Mesmo reconhecendo que categorizar é um esforço limitado e falho, Few (2006a, p. 30) propõe uma categorização dos *dashboards*, vista no Quadro 5, página 75.

Quadro 5 – Categorização de *dashboards* por Few

Categorias	Variáveis
Quanto a seu Papel	Operacional
	Analítico
	Estratégico
Quanto ao Tipo de Dados	Quantitativo
	Não quantitativo
Quanto ao Domínio dos Dados	Vendas
	Finanças
	Marketing
	Fabricação
	Recursos humanos
Quanto ao Tipo de Métricas	<i>Balanced Scorecard</i> (p. ex., KPIs)
	Seis Sigma
	Sem avaliação de desempenho
Quanto ao Alcance dos Dados	Toda a empresa
	Departamental
	Individual
Quanto à Frequência de Atualização	Mensalmente
	Semanalmente
	Diariamente
	Por hora
	Em tempo real ou em tempo quase real
Quanto à Interatividade	Exibição estática
	Exibição interativa
Quanto aos Mecanismos de Exibição	Principalmente gráficos
	Principalmente texto
	Integração de gráficos e texto
Quanto à Funcionalidade de portal	Condutor para dados adicionais
	Nenhuma funcionalidade de portal

Fonte: Few, 2006, p. 31.

Destas categorias, a mais importante, para o autor, é a que **define o papel de um dashboard**. Few reconhece que a divisão do painel em três papéis não é a única maneira de expressar os tipos de negócios com que um *dashboard* pode lidar. No entanto, ele diz que é a única categoria, entre as listadas no Quadro 5, que envolve significativas diferenças entre o design visual dos painéis (Few, 2006a, p. 31).

Painéis para fins estratégicos

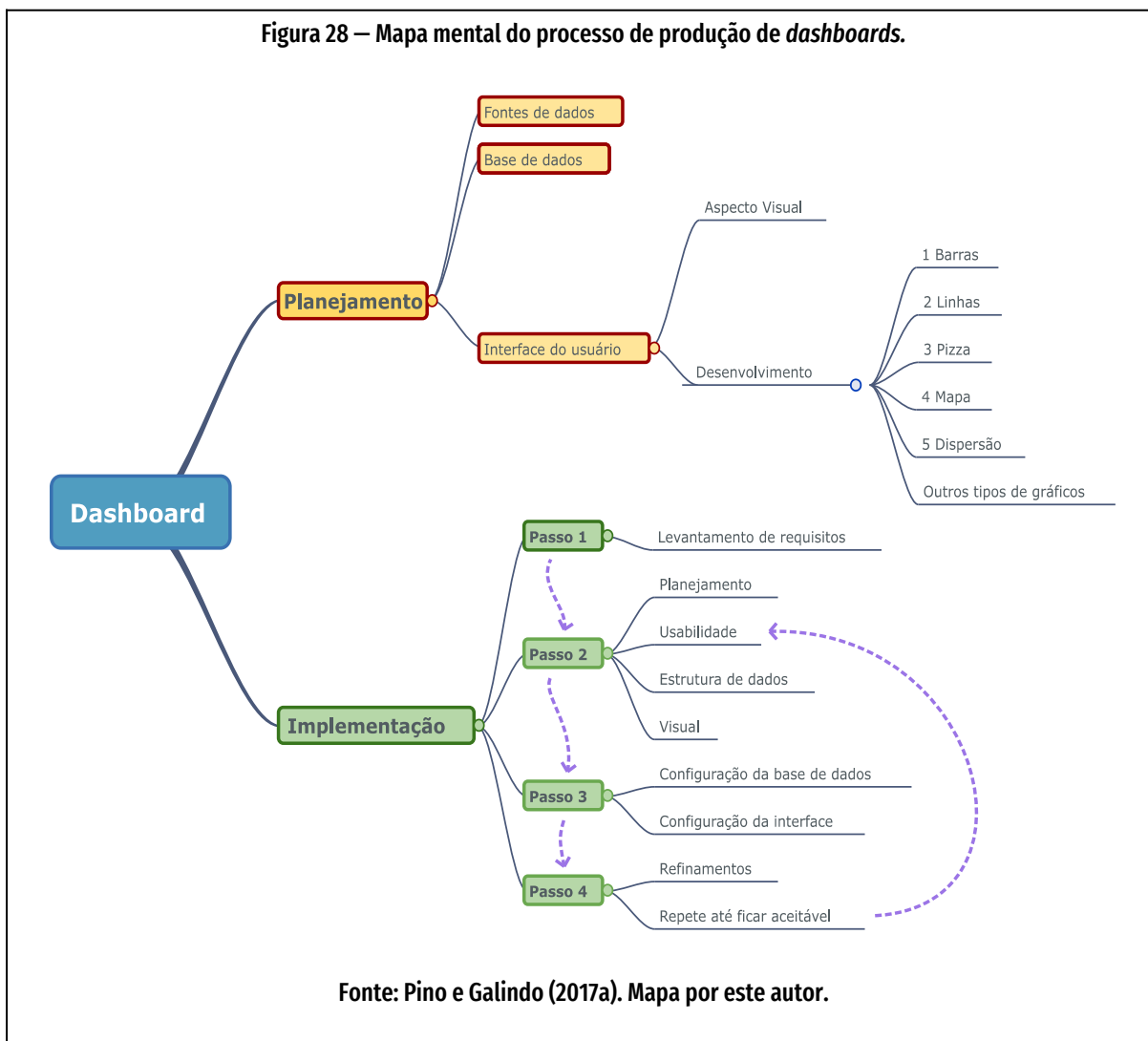
Atualmente, o uso primário de *dashboards* é para **fins estratégicos**, como suporte à gestão. Os painéis fornecem uma visão rápida que os tomadores de decisões necessitam para monitorar a saúde da empresa e oportunidades de negócios. *Dashboards* estratégicos trabalham medindo desde desempenho atual do empreendimento até previsões para o futuro (Few, 2006a).

Painéis deste tipo funcionam melhor com mecanismos de apresentação extremamente simples, pois excesso de contextualizações, históricos, comparações complexas e gradações sutis de informação podem distrair o gestor de sua meta imediata, que é a tomada de decisões (Few, 2006b).

Esses painéis não precisam de dados em tempo real, mas “fotografias” estáticas obtidas mensalmente, semanalmente ou diariamente. São, geralmente, apresentações unidirecionais que se limitam a apresentar o que está acontecendo. Não são desenhadas para interação necessária a análises mais profundas da situação, pois raramente isto é função do gestor estratégico (Few, 2006a, p. 32).

2.3.4.2 Painéis para fins analíticos

Dashboards para análise de dados requerem uma abordagem diferente dos painéis estratégicos. Painéis analíticos necessitam de informações altamente contextualizadas, históricos, avaliadores de performance sutis. Também se beneficiam de “fotografias” estáticas, que não mudam de um momento para outro, como os pai-



néis analíticos, mas sua apresentação é mais sofisticada para permitir ao analista que examine a complexidade dos dados, suas reações e suas causas.

Painéis para fins operacionais

Quando os painéis são usados para monitorar operações, eles devem ser projetados de maneira diferente daqueles que suportam a tomada de decisão estratégica ou a análise de dados. A característica das operações que influenciam exclusivamente o design dos painéis é a sua natureza dinâmica e imediata. Ao monitorar as operações, deve-se manter a consciência das atividades e eventos que estão

constantemente mudando e podem exigir atenção e resposta a qualquer momento. Se o braço robótico na linha de montagem de fabricação que prende a porta do carro ao chassi ficar sem parafusos, não se pode esperar até o dia seguinte para tomar conhecimento do problema e tomar uma ação. Da mesma forma, se o tráfego em um site cair repentinamente para metade do nível normal, a pessoa responsável será notificada imediatamente.

2.3.4.3 Como planejar um *dashboard*

Pino e Galindo dividem o planejamento de um *dashboard* em três camadas: **Fonte de dados**, **Base de dados** e **Interface com o usuário** (Pino; Galindo, 2017a).

Fonte de dados é a origem dos dados captados, ainda brutos. Trabalhar com fontes primárias de dados exige conhecimentos em “raspagem”⁷ de dados, processamento de texto, busca-e-troca de letras e palavras, expressões regulares⁸, locais onde obter dados primários.

Base de dados é onde os dados estão armazenados para serem utilizados em gráficos. A manipulação de dados nesta camada exige conhecimentos em armazenamento em bancos de dados, linguagens de programação como SQL, planejamento de métricas, técnicas de cruzamento de dados.

Interface de usuário é onde os usuários visualizam e interagem com os dados representados graficamente. Exige conhecimentos de ferramentas de apresentação de dados, design gráfico e design de informação. O plano de **Interface de usuário** pode ser dividido, segundo Pino e Galindo (2017a), em **Desenvolvimento** e **Aspecto Visual**.

7 “**Raspagem de dados**” é a técnica de obter, programaticamente, através da rede, com softwares especializados, acesso a dados desestruturados ou semiestruturados, publicados em diversos formatos (CSV, PDF, DOC, ODT, TXT, etc). Os programas podem ser locais ou serviços Web em rede.

8 “**Expressões regulares**” são padrões de letras e símbolos tipográficos que atuam como filtro ou “coringas” e auxiliam na procura por sequências de caracteres específicas em grandes quantidades de texto.

Desenvolvimento da interface de usuário envolve a identificação do tipo de gráfico mais eficiente conforme o tipo de dado. Para isto, pode-se usar algum dos tipos padrão de gráficos, como barras, linhas e pizzas, ou usar tipos mais elaborados.

Aspecto visual da interface de usuário tem a função de não apenas deixar os gráficos mais bonitos, mas imediatamente inteligíveis pelo usuário.

2.3.5 Colocando em prática um *dashboard* eficiente

Para criar um bom *dashboard*, Pino e Galindo (2017a) consideram que é preciso executar alguns passos essenciais, que se relacionam conforme a Figura 28, página 77.

Segundo as autoras, a produção de um *dashboard* é constituída das etapas de planejamento e da implantação. No planejamento, se definem as características técnicas (tecnologia usada) e os tipos de gráficos em relação aos tipos de dados, entre outros fatores. A implementação do *dashboard* é dividido pelas autoras em 4 passos, descritos a seguir.

Passo 1 – levantamento de requisitos

O que o negócio precisa e que pode ser resolvido ou auxiliado com o *dashboard*. Compreensão dos objetivos e das estratégias do negócio. Identificação dos principais indicadores de desempenho (*KPIs* — *Key Performance Indicators*).” (Pino; Galindo, 2017a)

Passo 2 – Planejamento de Dashboard

Reunião com todos os envolvidos na construção do *dashboard* e com quem receberá os dados, para acertar métricas, dimensões e compreender as particularidades e necessidades da equipe que vai consumir os dados. Pino e Galindo observam: “nesta fase é importante encontrar a melhor ferramenta para a criação do painel com base nos dados que serão extraídos” (Pino; Galindo, 2017a).

Pino e Galindo dividem o planejamento em três áreas: “**Usabilidade**”, “**Estrutura de dados**” e “**Visual**” (Pino; Galindo, 2017a). Em cada área, as autoras listam as principais questões que o planejamento deve responder.

Sobre usabilidade

- 1) Quais áreas da empresa vão acessar esses dados e de quais funcionalidades precisam?
- 2) Quais são as ferramentas ou plataformas mais adequadas para as necessidades desse projeto?

Sobre Estrutura de dados

- a) Quais são as métricas e dimensões que é preciso incluir no *dashboard*?
- b) Como esses dados se cruzam?
- c) De quais fontes vêm?
- d) Como devem ser estruturadas as bases de dados?

Sobre visual

- a) Quais aspectos visuais podem facilitar a interpretação dos dados? (Pino; Galindo, 2017a)

Passo 3 – Configuração técnica

Nesta etapa, é necessário entender as possibilidades técnicas para construir o *dashboard*. Em geral, há a participação de pessoal da área de Tecnologia da Informação com conhecimento sobre a ferramenta que será utilizada, além de conhecimentos sobre extração e importação de dados (Pino; Galindo, 2017a).

As autoras dividem a **configuração técnica** em duas frentes: **configuração da base de dados** e **configuração da interface** (2017a).

Configuração da base de dados

Consolidação dos dados em uma única base de dados ou armazém de dados (*data warehouse*). Configura o nível intermediário entre as fontes de dados e a interface que será acessada pelos usuários finais (Pino; Galindo, 2017a).

Configuração da interface

É a criação visual dos gráficos na ferramenta de visualização, com esquema de cores, avisos, legendas. Também envolve o desenvolvimento avançado de gráficos com fórmulas ou como as informações que aparecem nos painéis. Geralmente são fragmentadas de duas maneiras:

- Separados em telas discretas para as quais é necessário navegar;
- Separados em diferentes instâncias de uma única tela que são acessadas através de alguma forma de interação (Few, 2006a, p. 40).

Few argumenta que não é, em geral, uma prática ruim permitir que os usuários naveguem para telas diferentes ou a instâncias diferentes de uma única tela para acessar informações adicionais. Pode ser um recurso poderoso do painel permitir que a navegação tenha mais informação ou um grupo diferente de informações que atinja seu objetivo melhor sozinho. No entanto, essa fragmentação penaliza as vantagens exclusivas de um painel quando todas as informações devem ser vistas ao mesmo tempo para obter as ideias desejadas. “Fragmentar dados que devem ser vistos juntos é um erro” (Few, 2006a, p. 40).

2.4 Software livre

Praticamente todo o desenvolvimento do artefato digital proposto neste projeto será feito com *software* livre e aberto, com exceção do sistema operacional Windows, que é proprietário e fechado. Este conceito de *free and open source code* foi criado nos anos 1980 como reação aos *softwares* que começavam a ser vendidos com técnicas de marketing⁹, como um produto físico e com o código fonte¹⁰ fechado. Ou seja, não se podia descobrir como o *software* funcionava nem melhorar suas características.

Programadores da academia, como Richard Stallman, acostumados a desenvolverem e a trocarem trechos de software, não gostaram da situação e desenvolveram o conceito de software livre. Stallman criou em 1985 a Free Software Foundation, entidade sem fins lucrativos para promover a liberdade de usuários de computador. Esta liberdade de uso está expressa em quatro liberdades essenciais do software livre:

- 0) **A liberdade de executar o programa como o usuário desejar**, com qualquer objetivo. Esta é a Liberdade 0 (zero).
- 1) **A liberdade do usuário de estudar como o programa funciona**, e adaptá-lo às suas necessidades. Esta é a liberdade 1. Para usufruir desta liberdade, é necessário que o usuário tenha acesso ao código-fonte do software.
- 2) **A liberdade do usuário de redistribuir cópias** de modo que possa ajudar outros. Esta é a liberdade 2.
- 3) **A liberdade do usuário de distribuir cópias de suas versões modificadas** a outros. Esta é a liberdade 3. Assim, o usuário ou programador pode dar a

9 Software proprietário é software comercial cuja propriedade não é vendida, apenas licenciada.

10 Código fonte é o texto que determina as operações a serem realizadas por um computador.

toda comunidade a oportunidade de aproveitar suas mudanças. Para isto, também é necessário acesso ao código-fonte (O que é software livre?, 2018).

Esta política levou ao aparecimento de dezenas de milhares de programas desenvolvidos comunitariamente. O mais conhecido é o sistema operacional GNU/Linux, que funciona em diferentes dispositivos, desde os menores equipamentos para a “Internet das Coisas” (IoT) como lâmpadas, roteadores internet, modems, relógios, telefones celulares, até 498 dos 500 maiores supercomputadores do mundo (Prakash, 2019). Atualmente, o GNU/Linux é o sistema operacional mais usado no mundo, embora não tenha grande participação no mercado de computadores pessoais *desktop*.

Em função das licenças de software – especificamente a licença do LiveCode – aplicativos desenvolvidos com a versão comunitária do LiveCode só podem ser distribuídas, igualmente, de forma livre e aberta. Se a intenção for produzir software comercial, vendido, é necessário se usar a versão corporativa do LiveCode, que é comercializada por uma taxa anual.

2.5 Dados abertos

O conceito de dados abertos é, em alguns aspectos, similar ao de software livre. Segundo o Portal dos Dados Abertos (Brasil, [s. d.]), “dados são abertos quando qualquer pessoa pode livremente acessá-los, utilizá-los, modificá-los e compartilhá-los para qualquer finalidade”, sujeitando-se a exigências que visem preservar a proveniência e abertura destes dados.

Este conceito foi popularizado depois que um grupo de ativistas do movimento norte-americano Governo Aberto, entre eles Lawrence Lessig, Tim O'Reilly, Ethan Zuckerman e Aaron Swartz, reuniu-se em 2007, na Califórnia, para definir os conceitos de Dados Abertos. Eles listaram 8 princípios segundo os quais os dados do governo serão considerados abertos. Para receber esta classificação, os dados precisam ser:

1. **“Completos.** Todos os dados públicos são disponibilizados. Dados públicos são dados que não estão sujeitos a limitações válidas de privacidade, segurança ou privilégio.
2. **“Primários.** Os dados são coletados na origem, com o nível mais alto possível de granularidade, não em formas agregadas ou modificadas.
3. **“Oportunos.** Os dados são disponibilizados o mais rápido possível para preservar seu valor.
4. **“Acessíveis.** Os dados estão disponíveis para o maior número de usuários para a maior variedade de propósitos.
5. **“Processáveis por máquinas.** Os dados são razoavelmente estruturados para permitir o processamento automatizado.
6. **“Não discriminatórios.** Os dados estão disponíveis para qualquer pessoa, sem exigência de registro.
7. **“Não proprietários.** Os dados estão disponíveis em um formato sobre o qual nenhuma entidade possui controle exclusivo.
8. **“Livre de licenças.** Os dados não estão sujeitos a nenhum regulamento de direitos autorais, patente, marca comercial ou segredo comercial. Privacidade razoável, restrições de segurança e privilégio podem ser permitidas” (The 8 Principles of Open Government Data, [s. d.]).

Esta pesquisa depende inteiramente de dados abertos, ou seja, depende das informações disponíveis no banco de dados Currículo Lattes e outros, mantidos pelo CNPq. Estes dados possuem praticamente todas as características listadas como definidoras de dados abertos. Estão em formato XML, o que, por definição deste formato, os caracterizam como **Não-Proprietários** e **Processáveis por Máquinas**. Embora não sejam dados **Completos**, são **Primários, Livres de Licenças** e atualizados. Portanto, **Oportunos**.

Os dados Lattes só possuem alguma limitação quanto à **acessibilidade**. No Plano Diretor de Tecnologia da Informação e Comunicação (PDTIC) 2017-2019, a CAPES explicita suas diretrizes, as linhas pelas quais pretende alcançar seus objetivos estratégicos. Na segunda diretriz, D2, o órgão de tecnologia da entidade se propõe a “Fomentar a disponibilização e o uso de dados abertos” (CAPES, 2017a, p. 16) . Este objetivo da CAPES parece não encontrar eco no CNPq.

Apesar dos dados do currículo Lattes serem públicos, que pelos princípios acima e pelo decreto 8.777/2016 (Brasil, 2016) deveriam estar disponíveis não apenas a humanos mas também a máquinas, o CNPq filtra entradas e permite acesso apenas a humanos através do recurso chamado de “CAPTCHA”¹¹: a exigência de entrada manual de um código numérico que aparece na tela. Com isto, não é qualquer pessoa que pode automatizar o acesso aos dados XML dos currículos através de programas de computador. O CNPq alega que softwares que não navegadores Web fazem múltiplas consultas simultâneas, o que sobrecarregaria o servidor do banco de dados (Não sou um robô, 2016, p. 8–9). Assim, o acesso total a dados não pesso-

11 “*Completely Automated Public Turing test to tell Computers and Humans Apart*: Teste de Turing público completamente automatizado para diferenciação entre computadores e humanos, é um teste bastante utilizado na Internet para assegurar que não sejam realizados acessos automatizados a servidores, e geralmente é composto por imagens distorcidas de letras e números ou sons de difícil compreensão.” (Matias, 2015, p. 56)

ais dos currículos é liberado apenas a instituições que tenham convênio com o CNPq e, via formulário, expliquem qual o uso que vão dar aos dados.

Essas exigências vão contra as políticas de dados do Decreto nº 8.777, que diz textualmente em seu artigo 3º, inciso II, que “a Política de Dados Abertos do Executivo Federal será regida pelos seguintes princípios e diretrizes: [...] II – garantia de acesso irrestrito às bases de dados, as quais devem ser legíveis por máquina e estar disponíveis em formato aberto” (Brasil, 2016).

Estas limitações terão um impacto negativo na produtividade proporcionada pelo resultado desta pesquisa. Se uma instituição não tiver acordo com o CNPq, as coordenações de PPG terão de baixar os currículos em linguagem XML um a um, enfrentando dois *CAPTCHAs* para cada documento.

2.6 Artefato na *Design Science Research*

Esta pesquisa se desenvolve com base na metodologia ***Design Science Research***, a chamada “Ciência do artificial”. Numa curta definição, Roel Wieringa entende *Design Science Research* como o projeto e investigação de um artefato em um contexto (2014, p. 3). Johannesson e Perjons ampliam um pouco mais esta definição: *Design Science Research* é “o estudo científico e a criação de artefatos conforme eles são desenvolvidos e usados por pessoas, com o objetivo de resolver problemas práticos de interesse geral” (2014, p. 7). Hevner e Chavetear têm uma definição um pouco mais extensa: *Design Science Research* “é um paradigma de pesquisa no qual um designer, através da criação de artefatos inovadores, responde a questões relevantes para problemas humanos, criando, assim, novo conhecimento”. Eles completam: “O artefato planejado é tanto útil quanto fundamental para o entendimento do problema” (Hevner; Chatterjee, 2010, p. 5). Dresch, Lacerda e

Antunes Jr. sintetizam, em um conceito mais completo, definições de diversos autores: *Design Science Research* é a ciência que consolida conhecimentos sobre o projeto e desenvolve soluções para melhorar sistemas existentes, para resolver problemas e para criar outros artefatos (Dresch; Lacerda; Antunes Jr., 2014, p. xiii,57).

Artefato, neste contexto, é algo construído pelo ser humano, algo artificial. Por isto, *Design Science* também é chamada de “a ciência do artificial”. É a organização dos componentes de um ambiente interno para atingir os objetivos de determinado ambiente externo (Dresch; Lacerda; Antunes Jr., 2014, p. 57, 108). Comum a todo artefato está a intenção de ser uma resposta a um problema prático (Johannesson; Perjons, 2014, p. 3).

O artefato pode ser organizado em categorias, elas próprias também artefatos:

- **Constructos.** São os conceitos utilizados para se descreverem os problemas dentro de um domínio, um vocabulário para se especificarem as soluções apropriadas.
- **Modelos.** São conjuntos de proposições ou declarações que expressam as relações entre os constructos.
- **Métodos.** São conjuntos de passos necessários para desempenhar determinada tarefa. Etapas do método podem usar partes dos modelos. Favorecem a construção e a representação das necessidades de melhoria de um sistema.
- **Instanciações.** Execução do artefato em seu ambiente. Artefatos que operacionalizam outros artefatos (constructos, modelos e métodos). Conjunto coerente de regras que regulam o uso dos artefatos em um ambiente real (Dresch; Lacerda; Antunes Jr., 2014, p. 111–112).
- **Design propositions.** Este 5º artefato envolve as contribuições teóricas feitas por meio da aplicação da *Design Science Research*. Estas devem permitir generalizações de uma solução para determinada classe de problema, tornando-se conhecimento que pode ser aplicado em situações similares.

É importante notar que a *Design Science*, mesmo ocupando-se da resolução de problemas, não busca uma solução ótima em um modelo simplificado – solução que dificilmente será ótima no mundo real –, mas um resultado satisfatório no contexto em que se encontra o problema (Dresch; Lacerda; Antunes Jr., 2014, p. 57).

3 Procedimentos metodológicos

3.1 Avaliação do artefato

No contexto da engenharia de software, “**verificação**” — derivada do latim *veritas* (verdade) — é o processo de determinar se os produtos de uma fase do ciclo de desenvolvimento de software atende ou não aos requisitos estabelecido no planejamento. Já “**validação**” — derivada do latim *valorem*¹² (o que importa, o que tem valor) —, é o ato de avaliação do software no final do processo de desenvolvimento, para garantir a conformidade com requisitos iniciais do produto (Boehm, 1984, p. 75).

Para Boehm, informalmente pode-se definir as duas ações com as perguntas:

- a) **Verificação**: “Eu estou construindo corretamente este dispositivo?”
- b) **Validação**: “Eu estou construindo o dispositivo correto?”

Verificação e validação de requisitos de software e especificações de design têm os objetivos básicos de descobrir e resolver problemas de programação, de interface e questões de alto risco no início do ciclo de vida do software. Encontrando e corrigindo problemas no início do ciclo de desenvolvimento, grandes projetos podem economizar bastantes recursos (Boehm, 1984).

Na Figura 29, página 90, um “gráfico-V” mostra os momentos em que a verificação e validação são aplicadas à fase de desenvolvimento de artefatos (Boehm, 1979). A haste descendente do “V”, à esquerda, e as áreas coloridas, representam a fase de desenvolvimento do software, enquanto a haste ascendente à direita representa as fases de testes de software e sua integração a sistemas já instalados.

12 Boehm fala em “valere”, que é um termo italiano. Em latim, “o que tem valor” é “valorem”.



A chave para a diferenciação entre atividades de verificação e atividades de validação é a Linha Base de Requisitos, que são as especificações desenvolvidas e validadas durante a fase Planos e Requisitos, acima da Linha Base de Requisitos. Abaixo da Linha, as ações são de verificação, que é a comparação entre a Linha Base de Requisitos e sucessivos refinamentos (Projeto de Produto, Projeto Detalhado, Código, Base de Dados e Documentação). O processo de codificação (programação) iterativo. A cada iteração, foi realizada a verificação dos módulos codificados, para confirmar que estavam dentro das especificações.

Boehm (1984) argumenta que o processo de desenvolvimento de software tem custo mais efetivo se sua especificação é produzida na seguinte ordem:

1. **Especificação de requisitos.** Estabelecem as funções que o software deve executar, o nível desejado de desempenho (velocidade, acurácia) e a natureza das interfaces requeridas entre o produto de software e seu ambiente.
2. **Especificações de projeto de produto.** Descreve a arquitetura geral do produto de software e seus componentes.
3. **Especificações detalhadas de projeto.** Identifica entradas e saídas, lógica de controle e estrutura de dados de cada componente de baixo nível do *software* (Boehm, 1984, p. 76) .

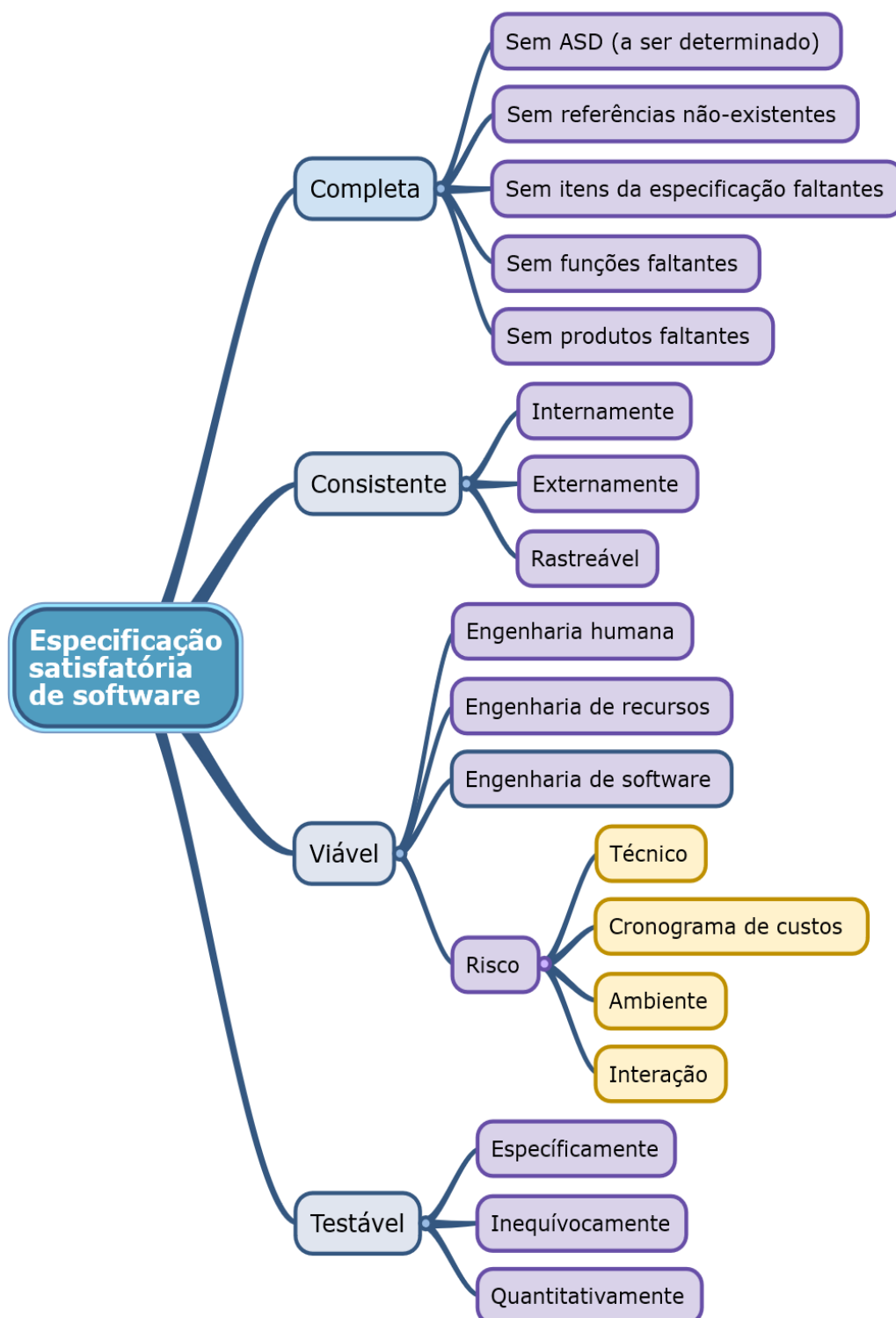
Validação e verificação têm quatro critérios básicos para requisitos e especificações de projeto: **integridade, consistência, viabilidade e testabilidade**. A taxonomia destes componentes aparece na Figura 30, página 92 e está desdobrada no Quadro 6, página 94.

Integridade. Uma especificação é íntegra quando todas as suas partes estão presentes e cada parte está totalmente desenvolvida. Não deve haver: elemento “ASD” (“A Ser Determinado”); referência a elementos não definidos na especificação (funções, entradas e saídas etc); Sem itens da especificação ausentes (itens que normalmente devem fazer parte das especificações); Sem funções ausentes (que devem fazer parte do produto mas não estão nas especificações); Sem produtos ausentes (produtos que deveriam fazer parte da distribuição do software mas não estão nas especificações).

Consistência. Uma especificação é consistente quando seus elementos não conflituam uns com os outros ou com especificações e objetivos. A consistência pode ser: **Interna**, quando não há conflito entre elementos internos do software; **Externa**, quando não há conflito entre elementos internos e externos, como as especificações; **Rastreável**, quando os itens na especificação

“têm antecedentes claros em especificações anteriores ou declarações de objetivos do sistema. Particularmente em especificações grandes, cada item deve indicar o item ou itens em especificações anteriores das quais é derivado” (Boehm, 1984, p. 77–79).

Figura 30 – Especificações de software



Fonte: Boehm (1979), ilustração deste autor.

A rastreabilidade tenta evitar má interpretações e “embelezamentos” (a inclusão de módulos de software que não são necessários ao sistema).

Viabilidade. Os **benefícios** do ciclo de vida devem ser maiores que o **custo** do ciclo de vida. A verificação de viabilidade envolve mais do que conferir se o sistema cumpre os requisitos funcionais e de desempenho. Envolve também validar que o sistema tenha manutenção e segurança capazes de assegurar um ciclo de vida sem prejuízo. Do ponto de vista da **engenharia humana**, verificação e validação podem ser definidas pelas respostas às perguntas: o sistema pode prover ao usuário uma maneira satisfatória de operação? O sistema satisfaz as necessidades humanas em vários níveis? O sistema ajudará as pessoas atingirem seu potencial? Do ponto de vista da **engenharia de recursos**, as perguntas poderiam ser estas: O sistema pode ser desenvolvido de maneira a cumprir, a um aceitável custo, todas as especificações requeridas? Em sua vida útil e a um custo-benefício favorável, o sistema poderá acomodar o crescimento esperado de requisitos operacionais? (Boehm, 1984, p. 79). Ou seja, seria **escalável**?

Do ponto de vista da **engenharia de programa**, as perguntas seriam relativas ao custo-benefício da manutenção e da portabilidade, e se o sistema seria suficientemente preciso, confiável e disponível.

Risco. Se os riscos não forem identificados e resolvidos antecipadamente, “há uma forte probabilidade de desastre se ou quando esse aspecto do sistema não for realizado conforme o esperado” (BOEHM, 1984, p. 79).

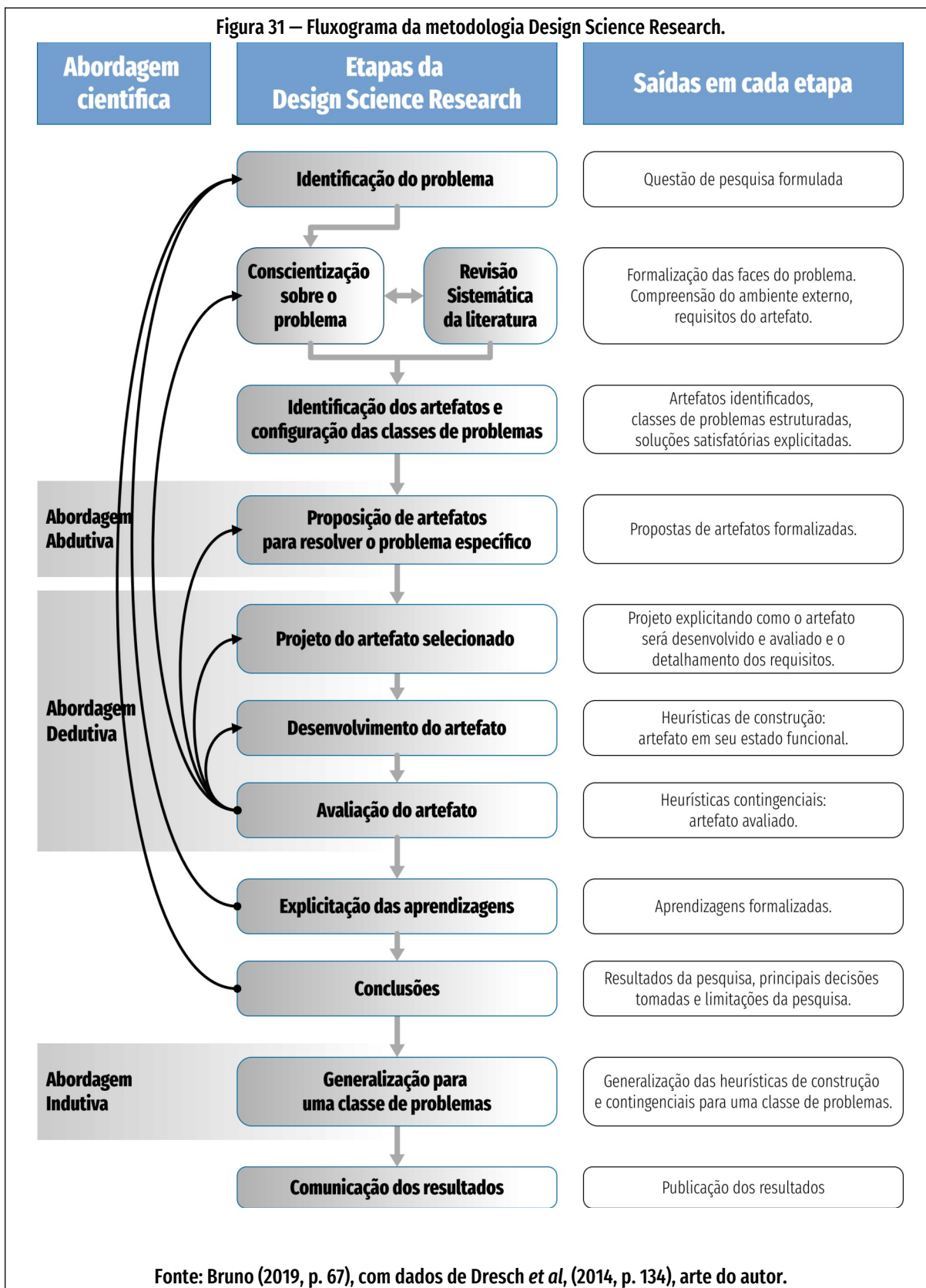
Testabilidade. “Uma especificação é testável na medida em que se pode identificar uma técnica economicamente viável para determinar se o software desenvolvido satisfará ou não a especificação” (Boehm, 1984, p. 79).

O Quadro 6 descreve mais as principais especificações satisfatórias de software. Nem todo critério listado por Boehm foi usado nesta pesquisa por não ser aplicável ou por não ser útil ao objetivo.

Quadro 6– Critérios de validação e verificação

Critério	Propriedade	Descrição
Compleitude	Sem “ASD”	Não deve haver elemento “ASD” (“A Ser Definido”) ou locais onde haja lacunas nas especificações.
	Sem referências não-existentes	Não deve haver referência a elementos não definidos nas especificações.
	Sem itens da especificação faltantes	Todos os itens especificados devem estar presentes.
	Sem funções faltantes	Funções que deveriam fazer parte do sistema mas não estão nas especificações.
	Sem produtos faltantes	Produtos que deveriam fazer parte da distribuição do software mas não estão nas especificações.
Consistência	Interna	Quando não há conflito entre elementos internos do software.
	Externa	Quando não há conflito entre elementos internos e externos, como as especificações
	Rastreável	Quando os itens na especificação “têm antecedentes claros em especificações anteriores ou declarações de objetivos do sistema. Particularmente em especificações grandes, cada item deve indicar o item ou itens em especificações anteriores das quais é derivado” (BOEHM, 1984, p.77-79) .
	Erros de digitação	Erros de escrita e de digitação (<i>typos</i>), em geral, levam a problemas de consistência.
Viabilidade	Engenharia humana	Verificação e validação podem ser definidas pelas respostas às perguntas: <ul style="list-style-type: none"> • O sistema pode prover ao usuário uma maneira satisfatória de operação? • O sistema satisfaz as necessidades humanas em vários níveis? • O sistema ajudará as pessoas atingirem seu potencial?
	Engenharia de recursos	<ul style="list-style-type: none"> • O sistema pode ser desenvolvido de maneira a cumprir, a um aceitável custo, todas as especificações requeridas? • Em sua vida útil e a um custo-benefício favorável, o sistema poderá acomodar o crescimento esperado de requisitos operacionais? Ou seja, seria escalável?
	Engenharia de software	<ul style="list-style-type: none"> • A manutenção do sistema terá custo-benefício vantajoso? • A portabilidade terá custo-benefício vantajoso? • O sistema terá suficiente precisão, confiança e disponibilidade?
	Risco	As quatro principais fontes de risco em requisitos de software e especificações de projeto são: <ul style="list-style-type: none"> • Efeitos técnicos; • Cronograma de custos; • Riscos ambientais; e • Riscos de interação.
Testabilidade		Para serem testáveis, as especificações devem ser, sempre que possível: <ul style="list-style-type: none"> • Específicas, • Inequivocas e • Quantitativas.

Fonte: (Boehm, 1984)



Quadro 7 – Objetivos desta pesquisa e as fases das teorias utilizadas.

Objetivo geral da pesquisa	Objetivos específicos da pesquisa	Etapas da Design Science Research	Fases da Arquitetura da Informação	Design da informação	Seções desta pesquisa
<p>O objetivo deste trabalho é desenvolver um artefato digital de auxílio à gestão acadêmica. Um dispositivo que extraia e relacione dados de currículos Lattes, CNPq e outras fontes e exiba-os em um painel com diferentes tipos de gráficos e tabelas</p>	<p>Levantar os requisitos da CAPES aos gestores de programas de Pós-Graduação em relação à produção acadêmica para definir as características, recursos e funcionalidades do artefato digital</p> <p>Compreender a estrutura de dados do currículo Lattes para definir a estrutura do banco de dados necessário ao gerenciamento das informações sobre a produção docente e discente a ser analisada</p> <p>Analisar as características e recursos de artefatos semelhantes a fim de definir os requisitos de projeto do artefato proposto</p> <p>Desenhar a interface com o usuário do aplicativo principal e determinar os tipos de gráficos apresentados conforme as necessidades de PPGs</p> <p>Programar o artefato com ferramenta de <i>Rapid Application Development</i> (RAD).</p> <p>Verificar o artefato segundo as especificações de requisitos de software e conforme os paradigmas da Design Science Research.</p>	<p>1 Identificação do problema</p> <p>2.1 Conscientização do problema</p> <p>2.2 Revisão sistemática da literatura</p> <p>3 Identificação do artefato e configuração das classes de problemas</p> <p>4 Proposição de artefatos para resolver o problema específico</p> <p>5 Projeto do artefato</p> <p>6 Desenvolvimento do artefato</p> <p>7 Avaliação do artefato. Iteração: retorna às etapas 4, 5 ou 6 até completar toda a codificação planejada</p> <p>8 Explicitação das aprendizagens</p> <p>9 Conclusões</p> <p>10 Generalização para uma classe de problemas</p> <p>11 Comunicação dos resultados</p>	<p>Pesquisa</p> <p>Design</p> <p>Implementação</p>	<p>Contextualização</p> <p>Problematização, Delimitação do tema</p> <p>Revisão bibliográfica</p> <p>Definição do escopo e dos requisitos de projeto</p> <p>Processo iterativo de codificação</p> <p>Verificação dos algoritmos, validação do artefato</p> <p>Documentação do processo, discussão dos resultados</p> <p>Conclusão</p> <p>Publicação online</p>	

Fontes: este autor, baseado em quadro de Bruno (2019), Dresch, Lacerda e Antunes Jr. (2015)

3.2 Delineamento da pesquisa

Dresch, Lacerda e Antunes Jr. compilaram a produção de mais de duas décadas dos principais autores sobre *Design Science* e consolidaram a metodologia em um modelo mais completo de etapas, como se pode ver no fluxograma da Figura 31, página 95. Com este modelo, se relacionaram os objetivos desta pesquisa com as fases correspondentes da *Design Science Research*, da Arquitetura da Informação, e do Design da Informação (Quadro 7, página 96).

Na fase “Pesquisa” da DSR, também presente nas fases propostas pela Arquitetura da Informação, mapeou-se a estrutura completa dos arquivos XML do Currículo Lattes. Os dados Lattes estão disponíveis à consulta nas formas HTML e XML zipado (Arquivo “comprimido”). Em HTML, as informações são formatadas para visualização por seres humanos, embora alguns pesquisadores tenham desenvolvido *scripts* para extrair informações lendo o HTML e descobrindo os campos de dados através dos atributos das *tags* HTML. É o que se chama “raspar dados”. Ferramentas de software como o *scriptLattes* usam esta abordagem (Mena-Chalco; Cesar Jr., 2009). Este artefato aqui proposto, no entanto, usa os arquivos de currículos em formato XML disponíveis no Lattes e baixados localmente, pois é mais estruturado e confiável que dados resultantes da “raspagem”. O documento DTD, consultado para a formatação do XML, tem referências às tabelas e campos do banco de dados, assegurando, assim, a correta estrutura dos dados locais.

As modernas linguagens de programação têm bibliotecas apropriadas para manipular XML com facilidade. É o caso da linguagem *script* do *software* LiveCode, bem como da linguagem JavaScript, que serão usadas no desenvolvimento do artefato.

Essas linguagens também possuem bibliotecas que manipulam arquivos zipados, tornando possível ao usuário usar os arquivos de currículos diretamente no sistema, sem necessidade de descomprimir os dados manualmente. O artefato digital fará a descompressão automaticamente. Este detalhe poupa algum tempo, em um

estágio que já é consumidor de tempo em função dos dois CAPTCHAs que o CNPq impõe para acesso aos currículos em XML. Se a instituição do PPG tiver convênio com o CNPq para extração de conteúdo da plataforma Lattes, os dados poderão ser acessados e baixados ao PC da instituição dedicado à tarefa. Caso a instituição não tenha convênio com o CNPq, o usuário — geralmente o coordenador do PPG — deve baixar manualmente os currículos individuais, a partir do link para os dados em formato XML. Para cada currículo, o operador deve digitar dois CAPTCHAs.

No Quadro 7, página 86, estão dispostos os objetivos específicos deste estudo em relação às fases do design. O primeiro objetivo específico deste trabalho,

“Levantar as necessidades da gestão do Programa de Pós-Graduação em Design (PGDesign) em relação à produção acadêmica para definir as características e recursos do artefato digital”,

é tratado na seção “3.3 Requisitos do projeto”, página 99.

O segundo objetivo específico,

“Compreender a estrutura de dados do currículo Lattes para definir a estrutura do banco de dados necessário ao gerenciamento das informações sobre a produção docente e discente a ser analisada”,

é tratado na seção “4.3.1 Definição da estrutura de dados”, página 116.

O terceiro objetivo específico,

Analisar as características e recursos de artefatos semelhantes a fim de definir os requisitos de projeto do artefato proposto,

é abordado na seção “4.1 Identificação dos artefatos e definição da classe de problemas”, página 100.

O quarto objetivo específico,

Desenhar a interface com o usuário do aplicativo principal e determinar os tipos de gráficos apresentados conforme as necessidades de PPGs, a partir dos preceitos da Arquitetura da Informação, do Design da Informação e da Visualização de Dados Numéricos,

é coberto pelas seções “4.3.2 Sistema de Organização”, página 121, e “4.3.3 Sistemas de Navegação, de Rotulagem e de Busca”, página 132.

Já o quinto objetivo específico,

Programar o artefato com ferramenta de *Rapid Application Development (RAD)*,

é abordado na seção 4.3.4 As fases do Processo e a Metodologia , página 128, que descreve os procedimentos de programação do dispositivo.

Finalmente, o sexto objetivo específico, a verificação do artefato digital é tratada na seção “4.5.6 Avaliação”, página 139.

3.3 Requisitos do projeto

Para se descobrir as necessidades da gestão de PPGs da área do Design, em relação à produção acadêmica, foi feita pesquisa bibliográfica nos documentos oficiais da CAPES publicados no decênio 2010-2020 e se fez entrevista não-estruturada com coordenadores do PGDesign da UFRGS, definindo-se as características e recursos desejados do artefato digital.

Assim, identificaram-se as necessidades da coordenação do PPG quanto às exigências da CAPES para avaliação do curso de Pós-Graduação em Design. Os requisitos listados são:

- Levantamento rápido da produção docente e discente do PPG registrada no banco de dados do Currículo Lattes.
- Geração de relatórios com a produção docente e discente.
- Análise quantitativa da produção docente e discente ao longo de vários anos, tendo como *default* o intervalo de anos entre as avaliações, determinado pela CAPES (atualmente, quadrienal).
- Capacidade de gerar gráficos a partir dos dados.
- Capacidade de fazer busca por todo o conteúdo.
- Capacidade de imprimir os resultados.
- Capacidade de exportar os resultados para outro sistema de representação gráfica, caso necessário, em formato CSV.
- Capacidade multiplataforma. O sistema deve funcionar pelo menos nos ambientes mais usados por usuários de computador *desktop*: Windows , Linux e MacOS.

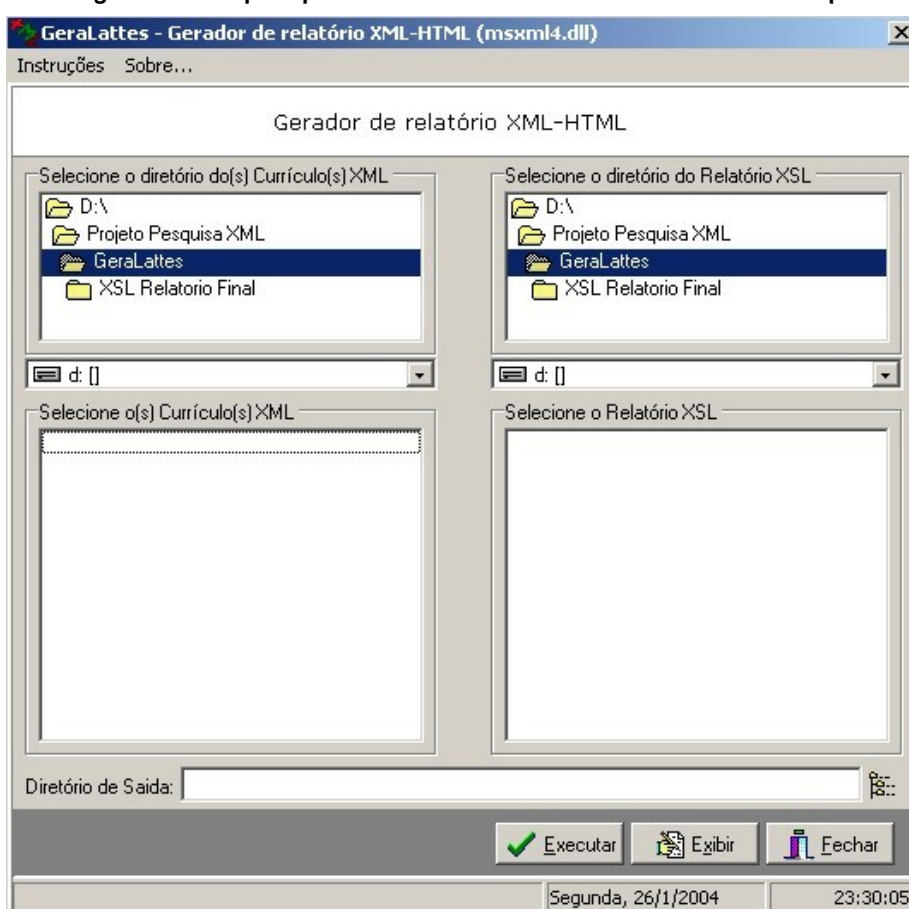
4 Apresentação e análise dos resultados

Com o objetivo de resolver o problema de pesquisa, nesta seção se analisam as características de sistemas similares ao artefato aqui proposto.

4.1 Identificação dos artefatos e definição da classe de problemas

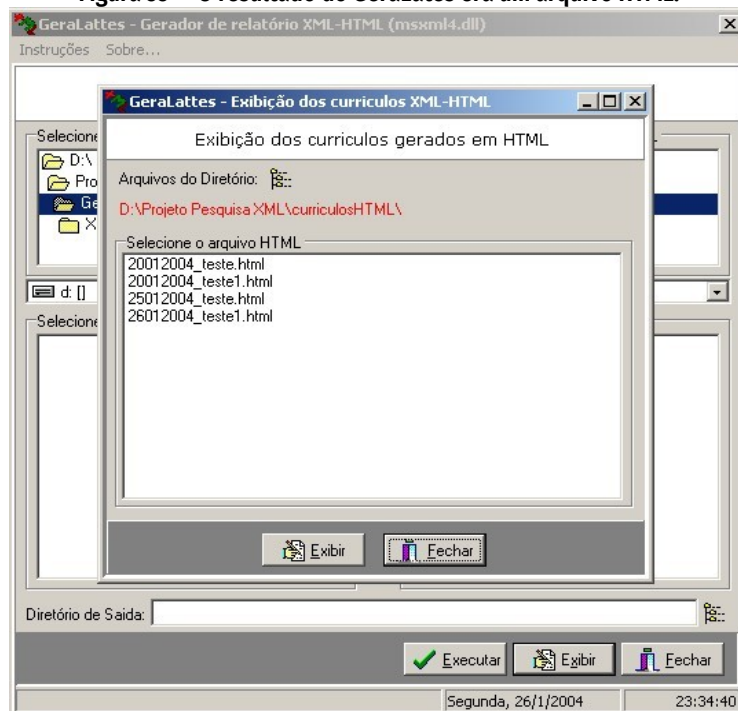
Na fase da pesquisa bibliográfica, encontraram-se sistemas prontos para instalação e descrições de ferramentas de gerenciamento dos dados da Plataforma Lattes similares, em algum aspecto, ao artefato que se propõe desenvolver nesta pesquisa.

Figura 32 – Tela principal do GeraLattes onde são feitas as escolhas de arquivos



Fonte: (Oliveira; Bermejo; Kern, 2004)

Figura 33 – O resultado do Geralattes era um arquivo HTML.



Fonte: (Oliveira; Bermejo; Kern, 2004)

4.1.1 GeraLattes

A ferramenta pioneira em análise de dados do Currículo Lattes, em nossa pesquisa bibliográfica, foi o GeraLattes, desenvolvido por integrantes do grupo Stela, ligado à criação do Currículo Lattes. Está descrito em um artigo de 2004, com ilustrações da interface gráfica do programa, vista na Figura 32, página 100.

O GeraLattes era uma interface escrita em *Object Pascal/Delphi* entre a biblioteca MSXML.DLL, que manipulava o arquivo de *schema* do Lattes (que define os campos do banco de dados Lattes) e o arquivo de estilos XSL. Este software não pode ser testado, pois não foram encontradas as fontes ou binários do programa na rede. Mas, segundo seus autores (Oliveira; Bermejo; Kern, 2004), o aplicativo extraía o conteúdo do Lattes e gerava um relatório formatado em HTML com dados que poderiam ser transformados em gráficos através de outros softwares. No entanto, a escolha do tipo de relatório, neste método, devia ser feita programando-se o

arquivo de estilo XSL, o que é uma tarefa demorada e complexa, não visual, exigindo conhecimento de XML.

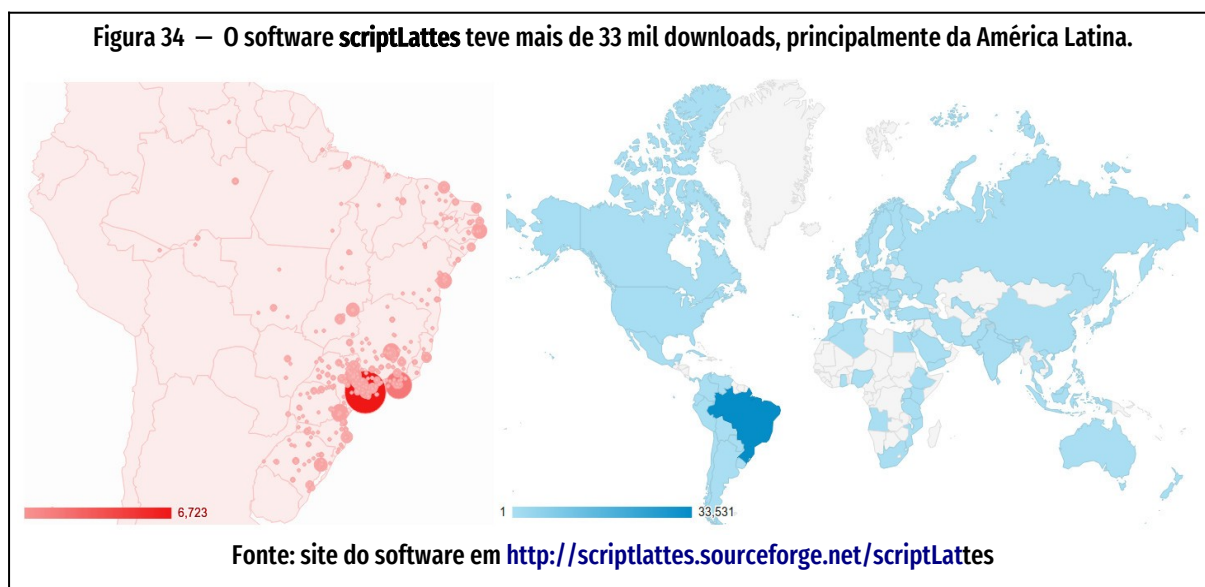
A interface gráfica (Figura 32, página 100 e Figura 33, página 101) servia para se escolher os arquivos XML com um currículo Lattes, um arquivo XSL de estilo que determinava o layout de página e o nome do arquivo HTML de saída dos dados.

O GeraLattes produzia 4 relatórios de informação operacional e gerencial:

- 1) Contabilização da produção acadêmica por ano,
- 2) Contabilização da produção acadêmica por idioma,
- 3) Relatório de orientações concluídas e
- 4) Relatório de participações em bancas de conclusão.

Os itens do banco de dados Lattes extraídos pelo GeraLattes eram:

- 1) Artigo completo em periódico internacional
 - 2) Artigo completo em periódico nacional
 - 3) Resumo em periódico internacional
 - 4) Resumo em periódico nacional
 - 5) Artigo completo em evento internacional
 - 6) Artigo completo em evento nacional
 - 7) Resumo em evento internacional
 - 8) Resumo em evento nacional
 - 9) Livro publicado ou organizado
 - 10) Capítulo de livro
 - 11) Orientação concluída (doutorado)
 - 12) Orientação concluída (mestrado)
 - 13) Orientação concluída (especialização)
 - 14) Orientação concluída (graduação)
 - 15) Orientação concluída (iniciação científica)
 - 16) Software em geral
 - 17) Produtos tecnológicos
 - 18) Processos ou técnicas
 - 19) Texto em jornal ou revista (magazine) internacional
 - 20) Texto em jornal ou revista (magazine) nacional
- (Oliveira; Bermejo; Kern, 2004)



4.1.2 scriptLattes

Outro pioneiro na área de tratamento de currículos Lattes foi o **scriptLattes**, desenvolvido a partir de 2005 e mantido até a data atual. É uma coleção de scripts em linguagem Python e Perl, inicialmente – depois totalmente em Python – acessa os currículos Lattes em formato HTML, extrai as informações pertinentes e gera grafos de colaborações, mapa de geolocalização (Figura 34, página 103) e de internacionalização (Mena-Chalco; Cesar Jr., 2009).

Foi desenvolvido por equipe de pesquisadores da Universidade Federal do Grande ABC (UFABC) e da Universidade de São Paulo (USP), e está disponibilizado em licença livre no site de desenvolvimento de software Sourceforge. Quanto à portabilidade, tem a característica de ser desenvolvido no ambiente GNU/Linux. Para ser usado em ambiente Windows ou MacOS, é necessária a instalação não trivial de requisitos como módulos Python que não estão na distribuição oficial da linguagem.

Originalmente, o scriptLattes foi escrito em linguagem Perl, posteriormente sendo vertido totalmente para linguagem Python. Como software livre, pode ser baixado

pela rede e testado, modificado. No entanto, a tentativa deste autor em realizar instalação em ambiente Windows e Linux fracassou em função de bibliotecas¹³ de Python que mudaram de nome ao longo dos anos (o software foi publicado em 2005). O código fonte do scriptLattes está em <http://scriptlattes.sourceforge.net>.

Alguns módulos do scriptLattes foram usados por outras soluções, como o syncLattes, evidenciando uma das vantagens do software livre: a liberdade de copiar, modificar e passar adiante livremente o código de programas.

Quadro 8 – Tipos de relatórios fornecidos pelo scriptLattes

Nº	Tipo de produção	Subtipo de produção
1	Produções bibliográficas	Artigos completos publicados em periódicos
		Livros publicados/organizados ou edições
		Capítulos de livros publicados
		Textos em jornais de notícias/revistas
		Trabalhos completos publicados em anais de congressos
		Resumos expandidos publicados em anais de congressos
		Resumos publicados em anais de congressos
		Artigos aceitos para publicação
		Apresentações de trabalho
		Demais tipos de produção bibliográfica
2	Produções técnicas	Softwares com registro de patente
		Softwares sem registro de patente
		Produtos tecnológicos
		Processos ou técnicas
		Trabalhos técnicos
		Demais tipos de produção técnica

Continua

13 Bibliotecas de software são conjuntos de funções ou trechos de programas reutilizáveis por outros programas, agrupadas em um único arquivo. Exemplo: os arquivos *.DLL do Windows.

Conclusão

Quadro 8 – Tipos de relatórios fornecidos pelo scriptLattes

Nº	Tipo de produção	Subtipo de produção
3	artísticas/culturais	
4	Projetos de pesquisa	Supervisão de pós-doutorado
		Tese de doutorado
		Dissertação de mestrado
		Monografias de conclusão de curso de aperfeiçoamento/especialização
		Trabalho de conclusão de curso de graduação
		Iniciação científica
	Orientações de outra natureza	
5	Projetos de pesquisa	
6	Prêmios e títulos	
7	Grafo de colaborações (redes de coautoria)	As produções com título igual ou similar (dentro do mesmo tipo e ano de publicação) são consideradas como colaborações entre pesquisadores.
8	Mapa de geolocalização	Baseados nos CEPs correspondentes aos endereços profissionais cadastrados nos currículos Lattes, são calculadas automaticamente as latitudes e longitudes de cada pesquisador.
9	Internacionalização	Coautoria e internacionalização
Fonte: (Mena-Chalco; Cesar Jr., 2009)		

Merece menção o módulo de consolidar a bibliografia repetida em uma única entrada no banco de dados, operação necessária para eliminação de duplicatas de obras. Estes passos, chamados de **deduplicação**, são necessários para obtenção da contagem real da produção acadêmica. Os registros extraídos do banco de dados Lattes ficam duplicados quando a produção tem mais de um autor e cada um escreve de maneiras levemente diferentes o nome da obra. Enquanto o ORCID¹⁴ e o DOI não forem obrigatórios para todo pesquisador e toda produção científica, essa normalização será necessária.

14 ORCID é um sistema numérico de identificação única, de âmbito mundial, para pesquisadores. DOI é um sistema idêntico, mas para conteúdo digital, como livros, publicações, documentos em geral e artigos eletrônicos. (Digital Object Identifier, 2022; ORCID, 2022)

Atualmente (2023), o **scriptLattes** está apto a fornecer relatórios a partir de dados de tabelas homônimas do banco de dados Lattes, conforme o Quadro 8, página 104.

O artefato funciona através de interface por linha de comandos (CLI), mas a versão 8 do software tem uma interface gráfica desenvolvida por colaboradores. É interessante o módulo de grafos (visualização de redes de colaboração) feito a partir das parcerias entre pesquisadores. No entanto, atualmente o resultado geral parece um tanto ultrapassado, com gráficos com baixa interatividade e pouco atrativos.

4.1.3 syncLattes

Este artefato, desenvolvido com software livre, aprimorou alguns scripts do scriptLattes. É mantido por pesquisadores na Universidade Federal de São Carlos, SP (Matias, 2015) e usado em diversas produções de áreas de engenharia de produção e bibliometria.

O syncLattes foi programado totalmente em linguagem Python e faz uso do acesso via VPN à Plataforma Lattes, desde que a instituição de pesquisa tenha convênio com o CNPq. A programadora do syncLattes desenvolveu também um proxy-cache em linguagem Lua para superar a limitação imposta pelo CNPq de dar acesso ao banco de dados Lattes a apenas um computador e um único número IP fixo por instituição. Com o proxy-cache, vários computadores dentro da instituição podem acessar o banco Lattes completo sem sobrecarregar os servidores da plataforma.

O código fonte do syncLattes está no Github (<https://github.com/nitmateriais/synclattes>).

Figura 35 – Extractor Lattes CNPq lê dados do Lattes e CNPq e cria banco de dados.

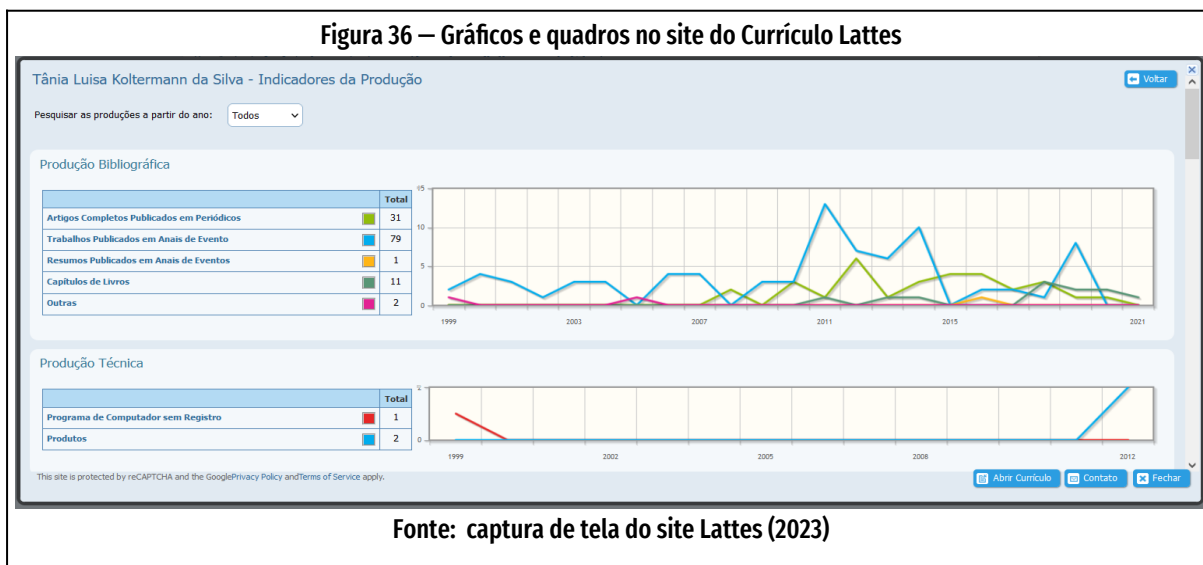


4.1.4 Extractor Lattes CNPq

É o resultado do trabalho de conclusão de curso de graduação (TCC) de Abreu (2020) na Universidade de Joinville, SC (Univille), visto na Figura 35, página 107. Desenvolvido em linguagem C# (C-sharp), é disponibilizado como software livre e aberto no site de desenvolvimento de programas Github. É uma ferramenta para extrair e guardar em banco de dados SQL as informações sobre currículos Lattes e produção acadêmica. Não tem módulo para apresentação. Em vez disto, gera dados em formato *Comma Separated Values (CSV)*, o padrão mais comum para troca de dados tabulares entre sistemas. Para visualização dos dados, os autores usam o **Tableau**, software de “*Business Intelligence*” (BI) proprietário e comercial para mostrar os dados, mas, como a saída é em *Comma Separated Values (CSV)* – formato universal de transferência de bancos de dados relacionais –, os dados podem ser importados por diferentes sistemas.

O Extractor Lattes CNPq funciona em ambiente Windows, com a linguagem C# (C-Sharp), e em Linux, com o compilador Mono, compatível com a tecnologia da linguagem C#.

Figura 36 – Gráficos e quadros no site do Currículo Lattes



O sistema inclui banco de dados da classificação Qualis de periódicos. O código do Extractor Lattes CNPq é aberto e o repositório está em <https://fabricadesoftwareuniville.github.io/ExtratorLattesCNPq/>.

4.1.5 Plataforma Lattes

Depois de produtores de software independentes lançarem ferramentas para análise de seu banco de dados, a própria Plataforma Lattes passou a publicar gráficos da acadêmica. A opção de ver gráficos com a produção de um pesquisador específico está na pré-apresentação do currículo, no link “Indicadores da produção”, e apresenta gráficos como os da Figura 36, página 108.

A plataforma também mostra gráficos de seus dados gerais — estes, a partir da capa do site (<https://lattes.cnpq.br/>).

Mas o Lattes, em 2023, apresenta gráficos fixos, sem interatividade, não é possível se usar filtros ou se modificar parâmetros como faixa de tempo — períodos quadrienais, por exemplo —, não podem ser comparados com outros e não podem ser salvos ou impressos, a não ser por “*print screen*” (instantâneo da tela do computador).

4.1.6 Stela Experta

Uma das principais empresas na área de currículos Lattes é a **Stela Experta**, serviço Web desenvolvido pelos implementadores da plataforma Lattes, o grupo Stela, que surgiu na Universidade Federal de Santa Catarina (UFSC). Desenvolveu a plataforma Lattes e tornou-se empresa privada de consultoria nesta plataforma, de nome **Stelatek** (<https://www.stelatek.com.br/>). Em 2019, conforme o site da empresa, tinha cerca de 35 clientes, entre eles ANIMAEDUCACAO, EBMS, EGC, ESTACIO, FIOCRUZ, FUMEC, IFPE, IFPR, IFSC, MACKENZIE, PUCRS, PUCSP, UCB, UCDB, UEG, UESC, UFBA, UFFS, UFOPA UFPE, UNA, UNIBH, UNIFOR, UNILASALLE, UNIMONTE, UNIOESTE, UNIRIO, UNISINOS, UNISOCIESC, UNIVILLE, UPF, USJT e UTFPR (“STELA EXPERTA”, [s.d.]).

A empresa oferece os serviços de relatórios para análise do desempenho das instituições de graduação e pós-graduação, com base na Plataforma Lattes, Plataforma Sucupira e Bancos de dados do CNPq.

O seu forte são as possibilidades de comparações. Por exemplo, pode se comparar um curso com a média de cursos nacionais e instituições de

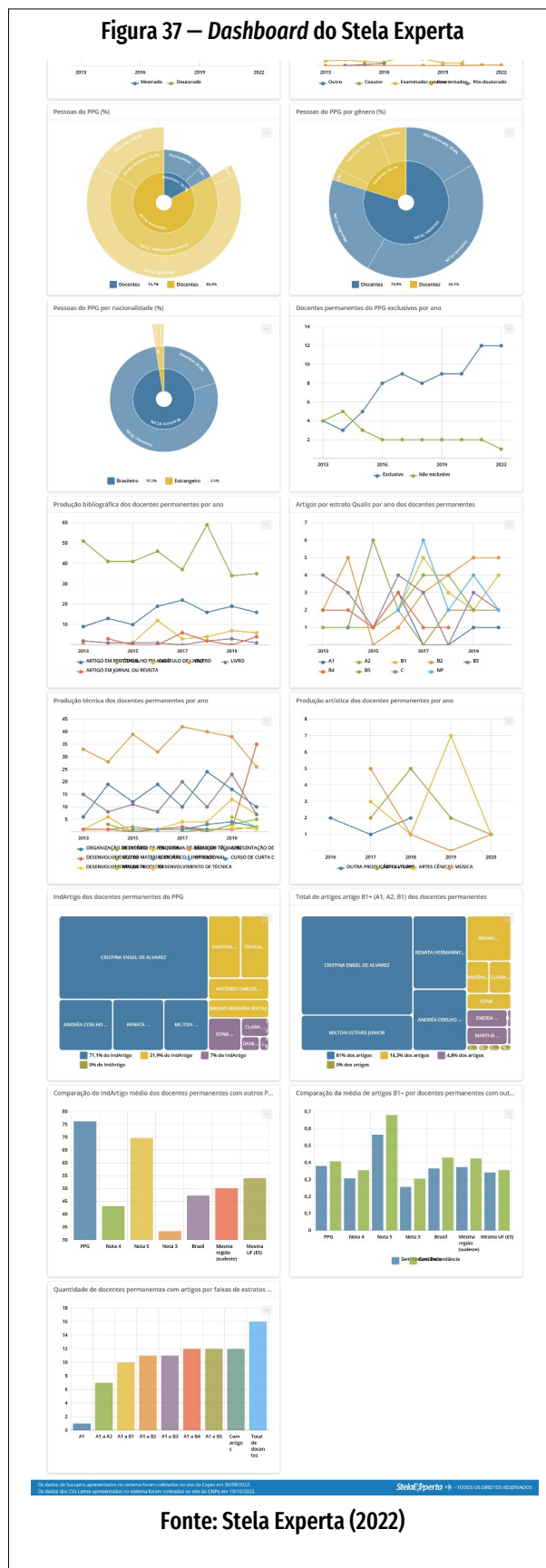
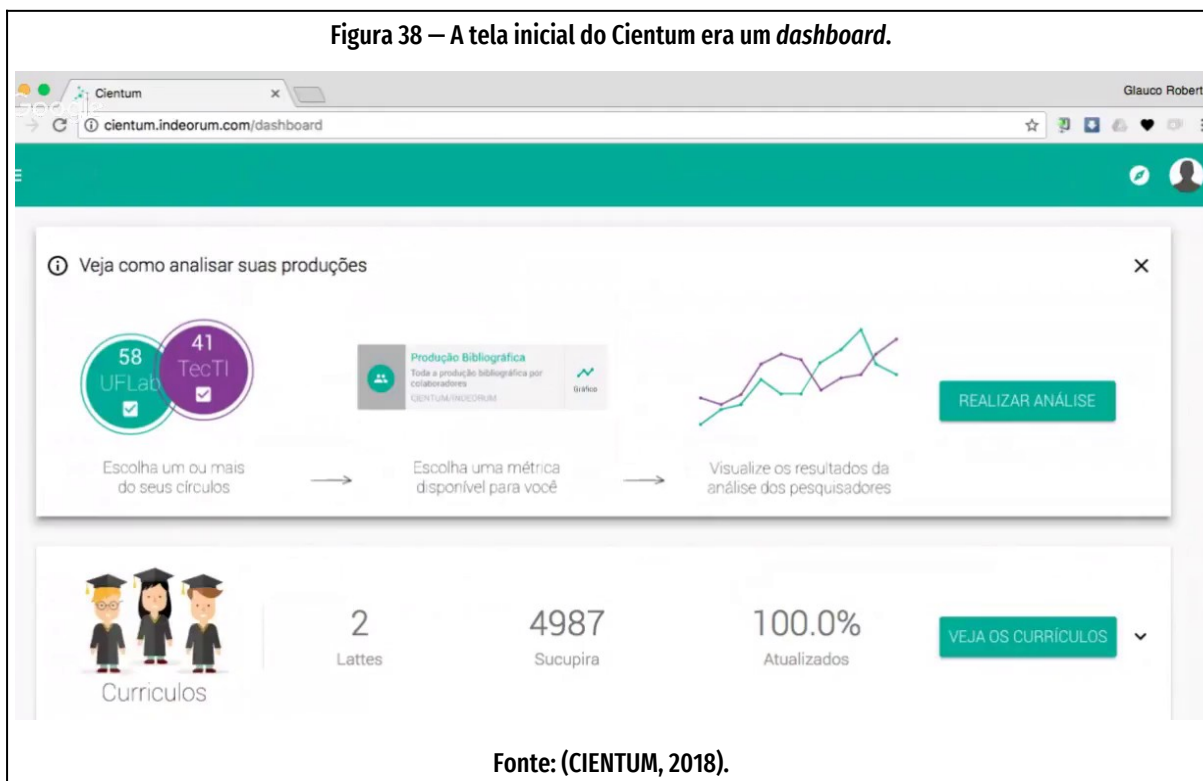


Figura 38 – A tela inicial do Cientum era um *dashboard*.



mesma área e mesma nota CAPES. Ou comparar os indicadores de desempenho dos cursos de uma instituição com os similares de outras instituições de todo o país. A Figura 37, página 109, mostra um *dashboard* gerado pelo Stela Experta.

4.1.7 Cientum

Outro produto comercial destacado da área de extração de bancos de dados acadêmicos era o **Cientum** (Figura 38, página 110), desenvolvido pela empresa **Indeorum**, *start up* nascida na incubadora empresarial da Universidade Federal de Pelotas (UFPel), RS, Brasil. O Cientum se definia como um sistema de gestão de currículos:

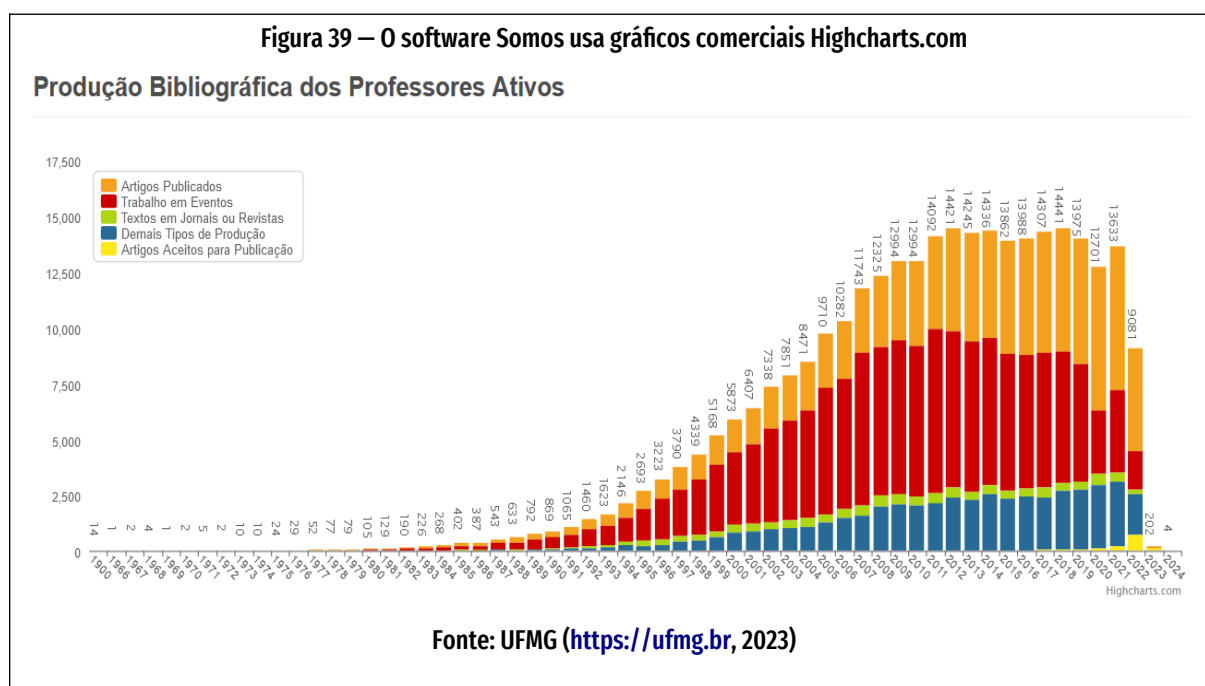
“O Cientum é um sistema de gestão de pesquisadores cujo objetivo é fornecer análises gráficas de forma qualitativa, quantitativa e temporal, sendo uma ferramenta de auxílio à tomada de decisão e alocação de recursos” (“CIENTUM”, 2018)

O serviço Web era organizado em torno do conceito de *círculos*, que podiam ser grupos de pesquisadores ou instituições. A partir destes círculos, podia-se extrair diversos tipos de gráficos analíticos e comparativos. A interface do software era via Web.

O sistema Cientum ficou em primeiro lugar no Expolattes (CAPES, 2018a), na categoria Produção de informações estruturadas por meio de rotinas dinâmicas para subsidiar processos de tomada de decisão, à frente de sistemas como o e-Lattes (software livre, segundo colocado no ExpoLattes) e a plataforma Stela Experta (comercial, terceiro lugar). No entanto, desde 2022 as páginas Web da empresa não existem mais, o Google Maps marca a sede como “permanentemente fechada” e não se encontram informações sobre o desaparecimento da empresa. Uma reclamação séria de cliente no site Reclame Aqui jamais foi respondida. A Indeorum atendia pelo endereço <https://cientum.indeorum.com>.

4.1.8 Somos

O Somos foi desenvolvido na FUNDEP, incubadora de inovação da Universidade Federal de Minas Gerais (UFMG), que o utiliza para publicar o portalq “Somos



UFMG”, portal web com os dados de bancos acadêmicos do seu quadro de pesquisadores. Além do uso próprio, o Somos é licenciado para outras instituições, onde adota o nome da instituição, como “Somos UNIFESP”. É um sistema simples de consulta a dados extraídos do Currículo Lattes. Usa uma abordagem de “site de busca”, onde existe um destacado campo de busca. O conteúdo do banco de dados vai sendo filtrado enquanto se digita cada tecla, o que facilita a busca. Assim que o fragmento de texto procurado é encontrado, não é necessário continuar a digitar o resto da frase ou palavra.

Na prática, o Somos é menos um instrumento de avaliação e mais um sistema de passeio digital pela produção acadêmica da instituição.

O sistema tem o *front-end* (código do lado do cliente) baseado na linguagem JavaScript e as bibliotecas de gráficos em JavaScript são do pacote comercial *Highcharts JS v2.2.5* (Figura 39, página 111). O Somos tem sua home page em <https://www.fundep.ufmg.br/solucoes/sistema-somos>.

4.1.9 E-Lattes Web

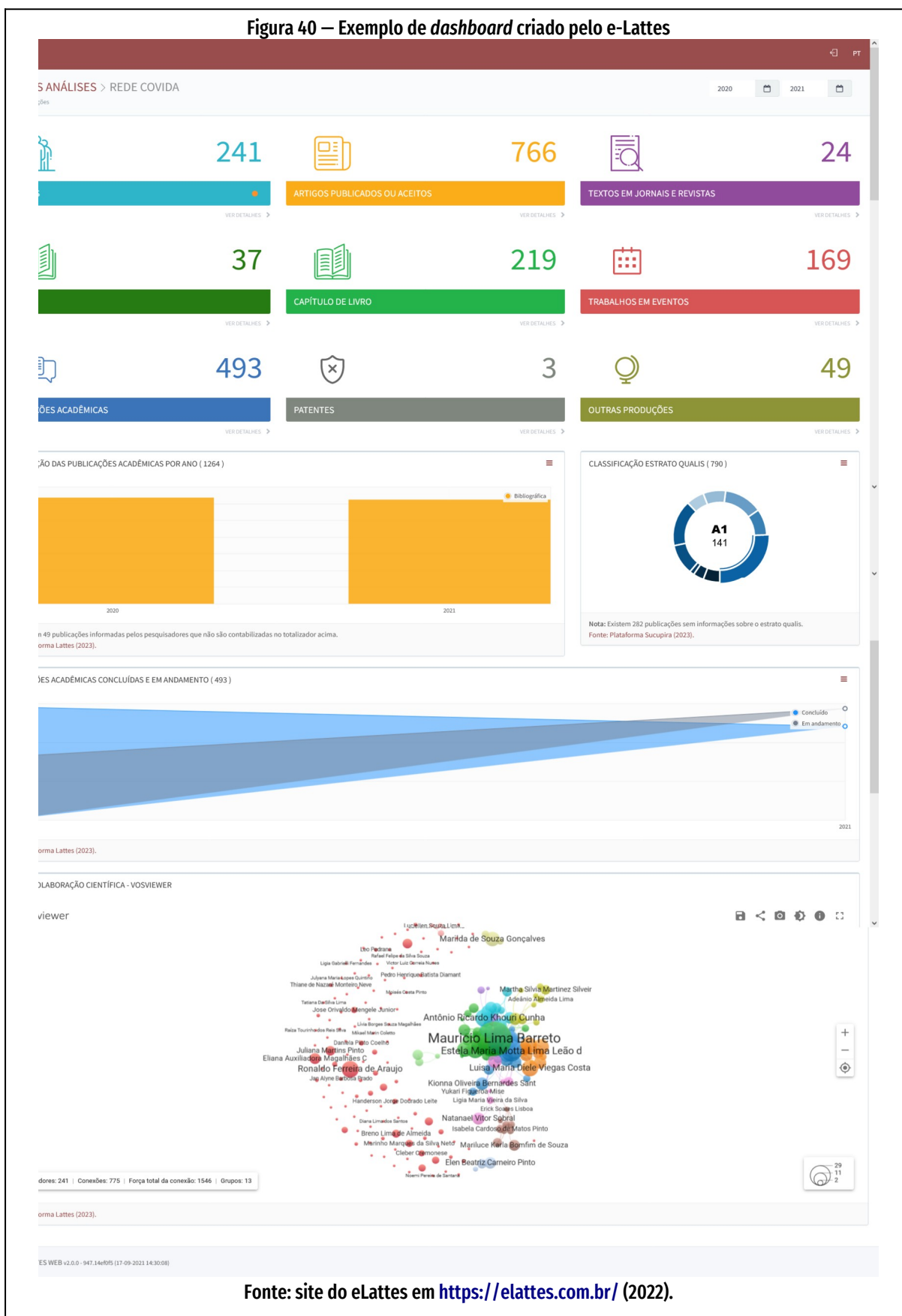
O e-Lattes Web foi criado por alguns dos autores do scriptLattes, como Mena-Chalk. Funciona com linguagem estatística “R” e apresenta-se com um visual mais contemporâneo, fazendo uso de tecnologias mais atualizadas que o scriptLattes. Uma das novidades é o módulo de visualização de redes de pesquisadores usando o servidor de grafos VOSviewer (Figura 40, página 113).

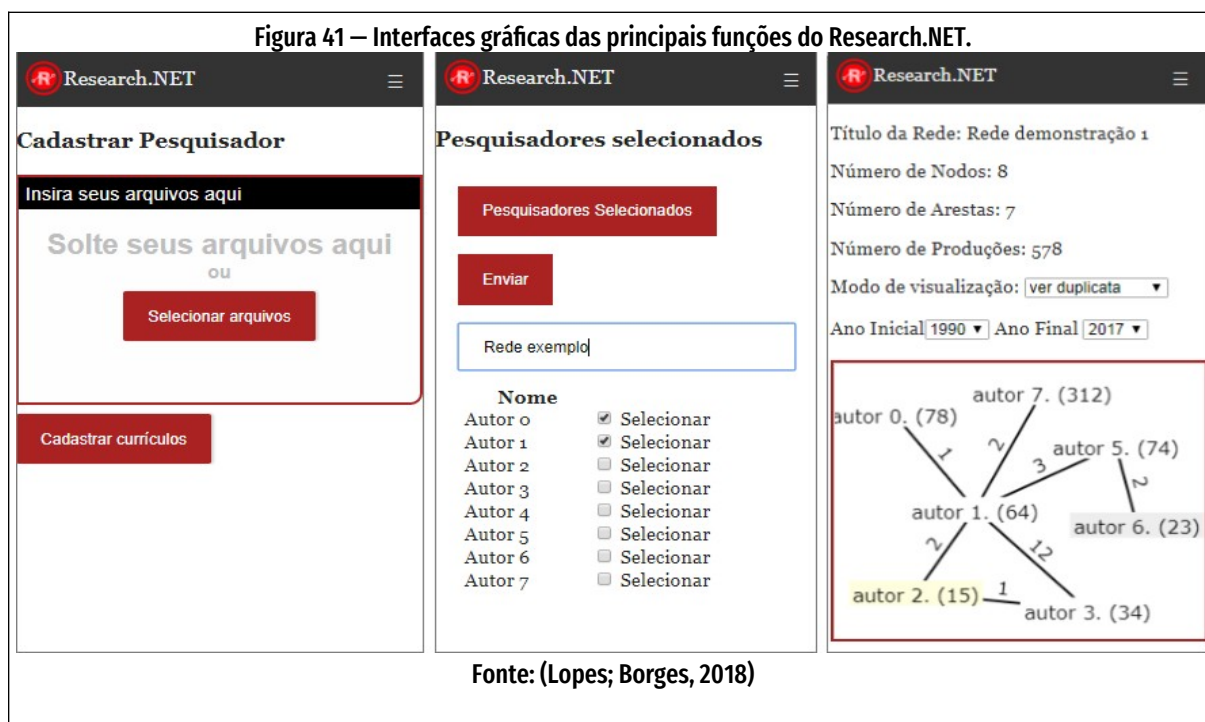
Em seu site de demonstração (<https://elattes.com.br/user/login>), é possível se inscrever e testar as funcionalidades.

4.1.10 Research.NET Web

O Research.NET Web é uma ferramenta de visualização de redes de colaboração acadêmicas feita para ser executada em servidor Web com informações do banco de dados da Plataforma Lattes (Lopes; Borges, 2018).

É constituído de três módulos: o recebimento de currículos, o construtor de redes





e o visualizador de redes. O módulo de recebimento acessa arquivos de currículo Lattes em formato XML e cadastra o currículo, se ainda não estava registrado. O módulo construtor de redes recebe os pesquisadores e, depois de identificar as duplicatas, constrói a rede, salvando o resultado em formato XML. Finalmente, o terceiro módulo usa o arquivo XML gerado para apresentar graficamente a rede de colaborações (Figura 41, página 114).

4.1.11 VidLattes v0.5

Esse é o protótipo que serviu de base ao artefato proposto por esta pesquisa. Mais detalhes estão na seção “4.5.4 Protótipo”, página 135.

4.1.12 Comparação com artefatos semelhantes

Analisando os sistemas concorrentes, elaborou-se o quadro apresentado na Figura 42, página 115, que salienta os recursos semelhantes ou exclusivos de cada sistema e seus pontos fortes ou problemáticos.

Figura 42 – Comparativo entre gerenciadores de currículos Lattes

	Recursos	Currículo Lattes	GeraLattes	StelaExperta	scriptLattes	synclattes	Extractor Lattes CNPq	Cientum	Somos	e-Lattes	Research.NET Web	VidLattes 0.5
Características	URL	http://lattes.cnpq.br		http://site.stelaexperta.com.br	http://scriptlattes.sourceforge.net	https://github.com/nimaterais/synclattes	https://fabricadesoftwareuiville.github.io/ExtractorLattesCNPq/	https://cientum.indeorum.com	https://www.fundep.ufmg.br/solucoes/sistema-somos	https://elattes.com.br/user/login	https://github.com/erick-luiz/research.net-web/	PGDesign/UFRGS
	Ano de lançamento		2001		2005		2007	2016		2018	2013, 2018	2016
	Sistema operacional compatível	Web	Desktop Windows	Web	Linux (instalação complexa em MacOS ou Windows)	Web e desktop	Windows e Linux	Web	Web	Web	Desktop/Web	Linux, Windows, MacOS e Android
	Tecnologia de desenvolvimento		Delphi/Object Pascal, XML, XSL		Python	Python	C#	JavaScript		Linguagem R	Java/JavaScript, PHP	LiveCode
	Interface	gráfica (web)	gráfica, só para escolha de arquivos de dados.	gráfica	Por linha de comando e gráfica	Por linha de comandos	Por linha de comandos	Gráfica	Web	Gráfica, Web	Gráfica, desktop (2013), Web (2018)	Gráfica (desktop)
	Modelo de negócios	Público e gratuito	Público e gratuito, mas indisponível	Comercial	Free and Open Source Software (FOSS)	FOSS	FOSS	Comercial		FOSS	FOSS	FOSS
	Documentação, ajuda	Não Aplicável (N/A)					Boa documentação no Github			Sim		Comentários no código.
Forma grupos de pesquisadores/ Instituições			Sim			Sim	Sim				Não	
Entrada de dados (input)	Tipo de dados	Qualquer produção Lattes e dados da Plataforma Sucupira	Qualquer produção Lattes	Qualquer produção Lattes	Qualquer produção Lattes	Qualquer produção Lattes	Produção Lattes, Qualis e CNPq	Produção Lattes, Qualis, Sucupira e CNPq	Qualquer produção Lattes	Qualquer produção Lattes	Só "trabalhos em eventos"	Qualquer produção Lattes
	Busca dados automaticamente	Não se aplica (N/A), pois é o servidor Web que outros programas clientes acessam	Não	Busca automática dos currículos em XML no site Lattes, se a instituição tiver convênio com o CNPq.	Carrega automaticamente os currículos Lattes em HTML.		Busca automática dos currículos em XML no site Lattes, se a instituição tiver convênio com o CNPq.			Busca automática dos currículos em XML no site Lattes, se a instituição tiver convênio com o CNPq	Deve-se baixar manualmente os currículos em XML	Deve-se baixar manualmente os currículos em XML
	Importação de dados	Não Aplicável (N/A)	Currículo em formato XML		Carrega automaticamente os currículos Lattes em HTML.	Módulo próprio em Python lê currículos em XML					Usa módulo do scriptLattes para ler XML zipado do Currículo Lattes	Módulo próprio lê XML zipado do Lattes
	Carrega dados em batelada (batch)	Não Aplicável (N/A)	Sim		Carrega automaticamente os currículos Lattes em HTML.							Sim, carrega todos os currículos XML de uma pasta do PC
Processamento dos dados	Filtros de consultas	N/A	Programável através de edição em arquivo XSL.	Sim	Carrega automaticamente os currículos Lattes em HTML.	Sim	Sim	Sim	Sim	Sim	Sim	Por autor, por tipo de produção, por ano, por quatriênio ou qualquer intervalo anual
	Geração de gráficos	Sim, estáticos.	Não	Sim, gráficos dinâmicos	Carrega automaticamente os currículos Lattes em HTML.	Não	Não. Autores usam Tableau, software fechado comercial	Sim, gráficos dinâmicos	Sim, gráficos comerciais highchart.com	Sim, integrado ao gerador de gráficos de rede VOSviewer Online	Módulos do scriptLattes	Sim. Plugin do LiveCode. Apenas gráficos de linhas
	Compara pesquisadores?	Não		Sim				Sim	Não	Sim	Sim	Não
	Compara cursos/programas?	Não	Não	Sim					Não	Sim	Sim	Não
	Análises ou Indicadores prontos ou predefinidos	Não	Não	Mais de cem indicadores pré-definidos					Não	Sim. Usuário pode gravar análises customizadas	?	Não
	Duplicação de artigos			Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Não
	Mostra redes de pesquisadores	Não	Não						Sim	Sim	Sim	Sim
Mostra produção de pesquisador	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Não	
Saída (output)	Exportação de dados	Sim	Sim	Sim			Sim	Sim	Não	Sim	Sim	Sim
	Formatos dos dados exportados	HTML, XML, bibliografia em formato RIS	Formatos para Web (HTML, PNG, JPEG)		RIS		Formato CSV		--	CSV e JSON	Através de cópia e colagem, imagens em PNG, ou salvar HTML	BIBtex, RIS, CSV, PNG
	Comparação entre: pesquisadores, cursos, programas, instituições.	Não	Não				?	Sim	Não	Sim	?	Não
	Saída pronta para impressão	Não, só apresenta gráficos e tabelas na tela sem opção de imprimir	Não	Sim					Não	Gera HTML		Não
Saída para arquivo	Não, só apresenta na tela sem opção de gravar arquivo	Sim, em formato HTML						Não			Sim	
Avaliação	Pontos fortes	É a fonte primária e oficial de dados Lattes	Segue padrões e usa-os adequadamente	Feito pelos criadores da plataforma Lattes	É pioneiro, é bastante difundido, é FOSS	Melhorou alguns scripts do scriptLattes, criou proxy-cache		Premiado à frente do e-Lattes e Stela Expert	Simple e funcional para o que se propõe	Muitos recursos		Captura todos os dados do CL
	Pontos fracos	Não pode ser acessado programaticamente. Tem CAPCHA que exige intervenção humana	Programar em XSL é complexo	Não é FOSS	Não pode baixar automaticamente os CL se o CNPq colocar captchas	É um front end que exige um back end, complicando a instalação	Talhado para Windows. Pode rodar em Linux com esforço.	Empresa fechou	Pouca interatividade dos gráficos	Exige instalação em servidor Web	Análise se limita a "trabalhos em eventos", na versão avaliada	Exige baixar manualmente os CL

Fonte: este autor.

4.2 Proposição de artefato

A partir da análise das ferramentas similares, pela leitura da documentação e análise do código fonte — onde foi possível, pois não se teve acesso os códigos de alguns sistemas — observou-se que os artefatos apresentam recursos semelhantes. Alguns artefatos têm, até mesmo, trechos do código anterior melhorados, como prega o movimento do software livre. Todos os artefatos precisam:

- 1) Obter os CV em XML zipado.
- 2) Alimentar um banco de dados.
- 3) Fazer a deduplicação.
- 4) Fornecer consultas ao banco de dados com opções de filtros.
- 5) Gerar visualizações com os dados.
- 6) Editar os dados.
- 7) Gerar banco de dados com produções bibliográficas.

Nem todos os artefatos possuem todas estas funções. Alguns, por exemplo, apenas geram dados em formatos como CSV ou JSON para plotagem de gráficos e grafos em outros sistemas.

4.3 Projeto e desenvolvimento do artefato

Nesta seção, explicitamos os passos e decisões tomadas no projeto do artefato. Inicialmente, descreve-se a estrutura de dados na sub-seção 4.3.1 . A seguir, na sub-seção 4.3.2 , discute-se o sistema de organização; e na sub-seção 4.3.3 , discute-se sobre os sistemas de navegação, de rotulagem e de busca.

4.3.1 Definição da estrutura de dados

A estrutura hierárquica de dados do Lattes fica evidente quando se carrega um currículo Lattes em XML no editor XML Notepad. A partir destas visualizações, criou-se uma série de tabelas em banco de dados relacional SQLite.

O elemento raiz do XML, “CURRICULO-VITAE”, possui quatro atributos: SISTEMA-ORIGEM-XML, DATA-ATUALIZAÇÃO, HORA-ATUALIZAÇÃO E NUMERO-IDENTIFICADOR. Foi criada então uma tabela SQLite com o nome “curriculo_vitae” e com os campos “sistema_origem_xml”, “data_atualização”, “hora_atualização” e “numero_identificador”. O atributo mais importante é o número identificador, que é único para cada pesquisador. Isto o torna o candidato ideal à chave primária da estrutura de dados SQLite. O Código SQL para criação desta tabela é o seguinte:

```
CREATE TABLE IF NOT EXISTS curriculo_vitae (  
id INTEGER PRIMARY KEY, # identificador do pesquisador no CNPq.  
sistema_origem_xml text,  
data_atualizacao TEXT,  
hora_atualizacao TEXT  
);
```

O número identificador foi o indexador de todas as outras tabelas do banco de dados.

O elemento raiz, além dos atributos, também é “pai” de quatro outros elementos: DADOS-GERAIS, PRODUCAO-TECNICA, OUTRA-PRODUCAO e DADOS COMPLEMENTARES. Cada um destes elementos gerará uma tabela própria, todas com o número identificador CNPq como índice.

As características do desenvolvimento do formato Currículo Lattes Markup Language (CLML) resultaram em um formato com apêndices, campos introduzidos em novas versões do padrão, segundo o histórico de desenvolvimento em comentários na documentação do formato. Para não se gerar um número excessivo de tabelas relacionais, optou-se por reunir numa única tabela os campos que, por evolução histórica do formato CLML, estivessem distribuídos em “galhos” diferentes da árvore hierárquica mas que pertencessem ao mesmo pesquisador. Por exemplo, os elementos de produções avaliadas pela Capes têm a seguinte estrutura, segundo o DTD¹⁵:

15 Neste trabalho, código de programação ou arquivos de dados XML estruturados são apresentados em fonte tipográfica monoespacia.

```
<!ELEMENT TRABALHOS-EM-EVENTOS (TRABALHO-EM-EVENTOS*)>
<!ATTLIST TRABALHOS-EM-EVENTOS
    SEQUENCIA-PRODUCAO CDATA #IMPLIED
>
<!ELEMENT TRABALHO-EM-EVENTOS (DADOS-BASICOS-DO-TRABALHO?,
DETALHAMENTO-DO-TRABALHO?, AUTORES*, PALAVRAS-CHAVE?, AREAS-DO-
CONHECIMENTO?, SETORES-DE-ATIVIDADE?, INFORMACOES-ADICIONAIS?)>
<!ELEMENT DADOS-BASICOS-DO-TRABALHO EMPTY>
<!ATTLIST DADOS-BASICOS-DO-TRABALHO
    NATUREZA (COMPLETO | RESUMO | RESUMO_EXPANDIDO) #IMPLIED
    TITULO-DO-TRABALHO CDATA #IMPLIED
    ANO-DO-TRABALHO CDATA #IMPLIED
    PAIS-DO-EVENTO DATA #IMPLIED
    IDIOMA CDATA #IMPLIED
    MEIO-DE-DIVULGACAO (IMPRESSO | WEB | MEIO_MAGNETICO |
MEIO_DIGITAL | FILME | HIPERTEXTO | OUTRO | VARIOS |
NAO_INFORMADO) "NAO_INFORMADO"
    HOME-PAGE-DO-TRABALHO CDATA #IMPLIED
    FLAG-RELEVANCIA (SIM | NAO) "NAO"
    DOI CDATA #IMPLIED
    TITULO-DO-TRABALHO-INGLES CDATA #IMPLIED
    FLAG-DIVULGACAO-CIENTIFICA (SIM | NAO) "NAO"
>
<!ELEMENT DETALHAMENTO-DO-TRABALHO EMPTY>
<!ATTLIST DETALHAMENTO-DO-TRABALHO
    CLASSIFICACAO-DO-EVENTO (INTERNACIONAL | NACIONAL | REGIONAL |
LOCAL | NAO_INFORMADO) "NAO_INFORMADO"
    NOME-DO-EVENTO CDATA #IMPLIED
    CIDADE-DO-EVENTO CDATA #IMPLIED
    ANO-DE-REALIZACAO CDATA #IMPLIED
    TITULO-DOS-ANAIS-OU-PROCEEDINGS CDATA #IMPLIED
    VOLUME CDATA #IMPLIED
    FASCICULO CDATA #IMPLIED
    SERIE CDATA #IMPLIED
    PAGINA-INICIAL CDATA #IMPLIED
    PAGINA-FINAL CDATA #IMPLIED
    DADOS-BASICOS-DO-TRABALHO?
    DETALHAMENTO-DO-TRABALHO?
    AUTORES*
    PALAVRAS-CHAVE?
    AREAS-DO-CONHECIMENTO?
    SETORES-DE-ATIVIDADE?
    INFORMACOES-ADICIONAIS?
    ISBN CDATA #IMPLIED
    NOME-DA-EDITORIA CDATA #IMPLIED
    CIDADE-DA-EDITORIA CDATA #IMPLIED
    NOME-DO-EVENTO-INGLES CDATA #IMPLIED >
```

Este fragmento da DTD estabelece atributos de 6 elementos:

- DADOS-BASICOS-DO-TRABALHO,
- DETALHAMENTO-DO-TRABALHO,
- AUTORES, PALAVRAS-CHAVE,
- AREAS-DO-CONHECIMENTO,
- SETORES-DE-ATIVIDADE,
- INFORMACOES-ADICIONAIS.

Praticamente todo tipo de produção requerido pela CAPES tem **dados básicos** e **detalhamento** em entidades separadas. Essa separação é desnecessária nas tabelas SQL, pois as informações referem-se a um mesmo pesquisador ou equipe de pesquisadores, e não são repartidos com nenhum outro. Então, aqui foram reunidas em uma única tabela.

Outros dados como AUTORES, PALAVRAS-CHAVE, AREAS-DO-CONHECIMENTO, SETORES-DE-ATIVIDADE são compartilhados com outros pesquisadores da base de dados. A tabela de produções acadêmicas, então, tem campos com IDs apontando para tabelas específicas com estas informações compartilhadas.

INFORMACOES-ADICIONAIS, como refere-se a um ou alguns pesquisadores determinadas e não é uma informação compartilhada, não necessita de uma tabela própria. É apenas um campo de dados na tabela de produtos de um determinado pesquisador, não um ID para outra tabela.

Na elaboração das tabelas SQL, os dados de produção acadêmica foram consolidados em uma única tabela por produto. Por exemplo, o *schema* da tabela de produto “Trabalho em Eventos” ficou assim:

```
CREATE TABLE IF NOT EXISTS trabalho_em_eventos (  
id INTEGER PRIMARY KEY AUTOINCREMENT  
, natureza TEXT  
, titulo_do_trabalho TEXT  
, ano_do_trabalho TEXT
```



```
, pais_do_evento TEXT
, idioma TEXT
, meio_de_divulgacao TEXT
, home_page_do_trabalho TEXT
, flag_relevancia TEXT
, doi TEXT
, titulo_do_trabalho_ingles TEXT
, flag_divulgacao_cientifica TEXT
, classificacao_do_evento TEXT
, nome_do_evento TEXT
, cidade_do_evento TEXT
, ano_de_realizacao TEXT
, titulo_dos_anais_ou_proceedings TEXT
, volume TEXT
, fasciculo TEXT
, serie TEXT
, pagina_inicial TEXT
, pagina_final TEXT
, isbn TEXT
, nome_da_editora TEXT
, cidade_da_editora TEXT
, nome_do_evento_ingles TEXT
-- campos que apontam para IDs em outras tabelas-----
, palavra_chave_1 TEXT
, palavra_chave_2 TEXT
, palavra_chave_3 TEXT
, palavra_chave_4 TEXT
, palavra_chave_5 TEXT
, palavra_chave_6 TEXT
, area_do_conhecimento_1 INTEGER
, area_do_conhecimento_2 INTEGER
, area_do_conhecimento_3 INTEGER
, setor_de_atividade_1 TEXT
, setor_de_atividade_2 TEXT
, setor_de_atividade_3 TEXT
--Fim dos Ids para outras tabelas-----
, descricao_informacoes_adicionais TEXT
, descricao_informacoes_adicionais_ingles TEXT
);
```

As informações sobre autoria e coautoria estão em duas tabelas à parte, pois, em geral, há mais de uma autoria em um produto acadêmico.

O código SQLite para a criação desta tabela é o seguinte:

```
CREATE TABLE IF NOT EXISTS autorias_produtos (
```

```
id INTEGER PRIMARY KEY AUTOINCREMENT
, id_atoria INT
, id_coatoria INT
, tabela_produto TEXT
, id_produto INT
, sequencia_producao INT
, ordem_importancia INT
, ordem_de_atoria INT
, flag_relevancia BOOLEAN
, fonte TEXT
, FOREIGN KEY(id_atoria) REFERENCES aorias(id) ON DELETE
CASCADE
, FOREIGN KEY(id_coatoria) REFERENCES coatorias(id) ON DELETE
CASCADE
);
```

As tabelas de produtos mereceram uma atenção especial, já que dois ou mais coautores de um mesmo trabalho poderão registrar a obra com um nome levemente diferente. Se comparados nome a nome, um único caractere diferente já basta para que as obras sejam consideradas obras diferentes. Então, foi necessário o desenvolvimento de um algoritmo que procure por obras de nomes grafados com modificações. Ao se registrarem as produções acadêmicas, deverá se fazer a **deduplicação**: procurar as equivalências usando-se algoritmos e vários outros critérios além dos nomes, como comparação dos campos únicos como ISSN, ISBN, DOI e outros que possam ser coincidentes.

4.3.2 Sistema de Organização

Nesta pesquisa, as estruturas dos arquivos XML dos currículos Lattes e o banco de dados SQL (através da implementação SQLite) com que o artefato digital proposto manipulará estas informações se enquadram na primeira categoria da Arquitetura da Informação, **Sistema de Organização**. Nesta categoria também se definem o relacionamento e a hierarquia dos dados.

Nesta fase da pesquisa, buscou-se compreender a estrutura de dados do currículo Lattes disponível ao público em formato XML (chamada de Linguagem de Marcação da Plataforma Lattes – LMPL), para definir a estrutura do banco de dados com o qual o artefato digital vai operar. As informações sobre esta estrutura estão à disposição no site da **Comunidade para Ontologias em Ciência, Tecnologia e Informações de Aperfeiçoamento de Nível Superior (CONSCIENTIAS)**, grupo criado pelo CNPq “para desenvolver ontologias que se prestem ao intercâmbio de informações entre agências de fomento e instituições ligadas aos temas ‘Ciência’, ‘Tecnologia’, ‘Inovação’ e ‘Informações de Aprimoramento de Nível Superior’” (Comunidade Conscientias-LMPL, [s.d.]).

A comunidade define “ontologia”, no contexto do Currículo Lattes, como “uma descrição formal dos conceitos e relacionamentos existentes dentro de um domínio” (idem, ibidem). Estes conceitos e relacionamentos são os campos dos bancos de dados Lattes e os valores permitidos em cada um deles, que estão disponíveis em formato XML.

As informações sobre a estrutura dos dados Lattes é definida em um arquivo DTD (*Document Type Definition*), um tipo de documento XML que especifica o formato de outro documento XML através de regras que definem o que pode ou não pode constar neste segundo documento. Os DTDs são usados para se verificar se um arquivo XML está bem formatado e tem dados válidos (DTD Introduction, [s. d.]). Neste caso, determina o nome, a hierarquia e o conteúdo permitido para os campos do banco de dados Lattes. Este DTD está disponível no portal da Comunidade CONSCIENTIAS e no Github^{16 17} sob o nome **LMPLCurrículo.DTD**.

16 O DTD está em <https://github.com/antoanne/LMPLCurrículo> .

17 GitHub é um site Web que funciona como repositório de projetos, sistema de controle de versões dos arquivos e gerenciador de trabalho em grupo.

Figura 43 – Código XML de um currículo Lattes em estado não estruturado.

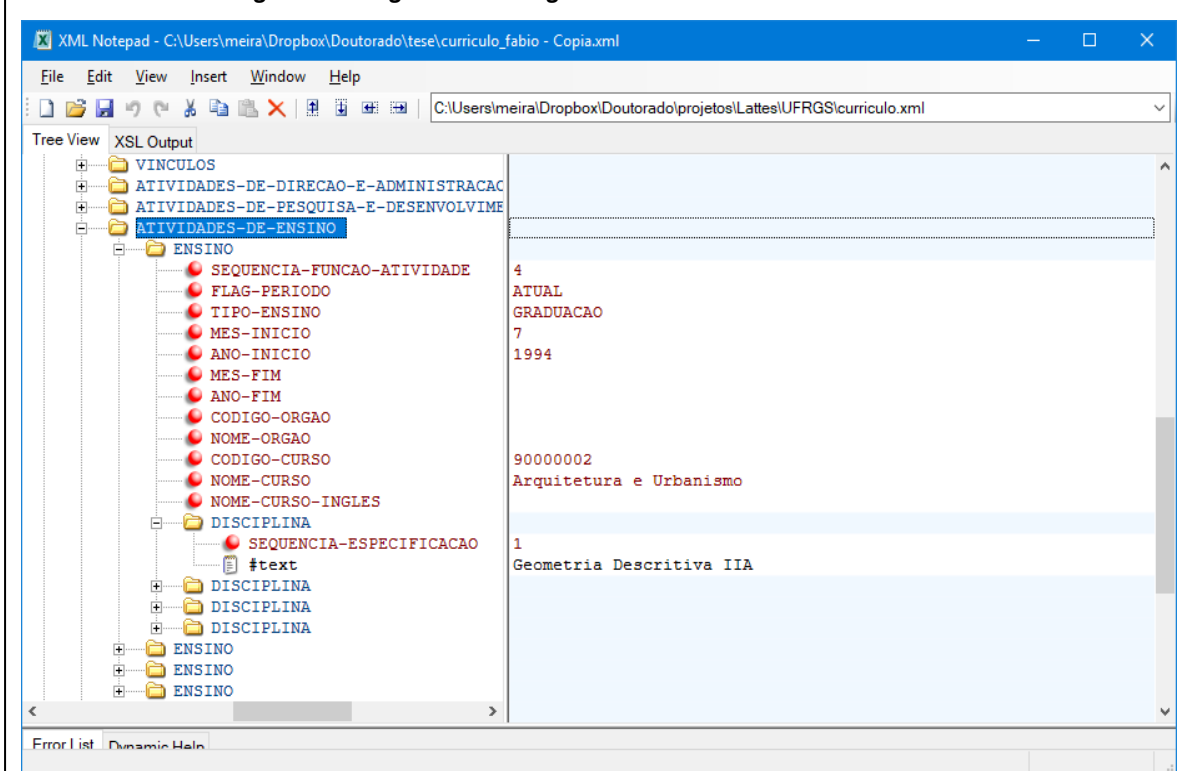
```
FLAG-LINHA-DE-PESQUISA-ATIVA="NAO" OBJETIVOS-LINHA-DE-PESQUISA=""
TITULO-DA-LINHA-DE-PESQUISA-INGLES="" OBJETIVOS-LINHA-DE-PESQUISA-INGLES=""
<<AREAS-DO-CONHECIMENTO/></LINHA-DE-PESQUISA><LINHA-DE-PESQUISA
SEQUENCIA-LINHA="2" TITULO-DA-LINHA-DE-PESQUISA="Design Virtual"
FLAG-LINHA-DE-PESQUISA-ATIVA="NAO" OBJETIVOS-LINHA-DE-PESQUISA=""
TITULO-DA-LINHA-DE-PESQUISA-INGLES="" OBJETIVOS-LINHA-DE-PESQUISA-INGLES=""
<<AREAS-DO-CONHECIMENTO/>
</LINHA-DE-PESQUISA></PESQUISA-E-DESENVOLVIMENTO></ATIVIDADES-DE-PESQUISA-E-D
ESENVOVIMENTO><ATIVIDADES-DE-ENSINO><ENSINO SEQUENCIA-FUNCAO-ATIVIDADE="4"
FLAG-PERODO="ATUAL" TIPO-ENSINO="GRADUACAO" MES-INICIO="7" ANO-INICIO=
"1994" MES-FIM="" ANO-FIM="" CODIGO-ORGAO="" NOME-ORGAO="" CODIGO-CURSO=
"90000002" NOME-CURSO="Arquitetura e Urbanismo" NOME-CURSO-INGLES=""
><DISCIPLINA SEQUENCIA-ESPECIFICACAO="1">Geometria Descritiva IIA
</DISCIPLINA><DISCIPLINA SEQUENCIA-ESPECIFICACAO="2">Geometria Descritiva
III</DISCIPLINA><DISCIPLINA SEQUENCIA-ESPECIFICACAO="3">Informática
Aplicada à Arquitetura I</DISCIPLINA><DISCIPLINA SEQUENCIA-ESPECIFICACAO="4"
>Informática Aplicada à Arquitetura II</DISCIPLINA></ENSINO><ENSINO
SEQUENCIA-FUNCAO-ATIVIDADE="9" FLAG-PERODO="ANTERIOR" TIPO-ENSINO=
"GRADUACAO" MES-INICIO="7" ANO-INICIO="1994" MES-FIM="6" ANO-FIM="1997"
CODIGO-ORGAO="" NOME-ORGAO="" CODIGO-CURSO="90000002" NOME-CURSO=
"Arquitetura e Urbanismo" NOME-CURSO-INGLES=""><DISCIPLINA
SEQUENCIA-ESPECIFICACAO="1">Introdução ao Projeto Assistido por Computador
</DISCIPLINA></ENSINO><ENSINO SEQUENCIA-FUNCAO-ATIVIDADE="10" FLAG-PERODO=
"ANTERIOR" TIPO-ENSINO="GRADUACAO" MES-INICIO="3" ANO-INICIO="1996" MES-FIM=
"12" ANO-FIM="1996" CODIGO-ORGAO="" NOME-ORGAO="" CODIGO-CURSO="90000002"
NOME-CURSO="Arquitetura e Urbanismo" NOME-CURSO-INGLES=""><DISCIPLINA
```

Fonte: este autor (2018).

Através da leitura do DTD do currículo Lattes, foi construída a estrutura das tabelas do banco de dados relacional SQLite, com ajuda de um software de modelagem de dados relacionais, e se definiu as funções de programação que extrairão as informações dos currículos XML. Esta etapa já foi realizada ao se desenvolver o protótipo funcional do artefato.

Os dados de origem do artefato digital proposto são os currículos Lattes em formato XML, obtidos a partir de um link disponível na visualização Web dos currículos, no site Lattes, no local salientado na Erro: Origem da referência não encontrada, página Erro: Origem da referência não encontrada:

Figura 44 – Fragmento do código XML de um currículo Lattes estruturado.



Fonte: este autor (2018).

O link retorna um arquivo com dados comprimidos¹⁸ para baixar, nomeado com o número do currículo Lattes do pesquisador e contendo um arquivo chamado “curriculo.xml”. Para carregar este arquivo no artefato digital, não é necessário descomprimi-lo. O artefato aqui proposto fará isto automaticamente. A Figura 46, página 131, tela do editor de texto “Notepad++”, mostra um trecho deste arquivo com sintaxe colorida.

Vê-se os três elementos principais do XML: as marcações (*tags*, “<” e “>”), os parâmetros das marcações (chamados de “atributos”, dentro dos caracteres “<” e “>”, logo depois do nome da *tag*, no formato “parâmetro=valor”) e o conteúdo do “nó” XML, entre a abertura e o fechamento da marcação. Este conjunto de marcação e conteúdo é chamado de “elemento”. O Quadro 9 mostra a estrutura de um elemento XML.

18 Arquivo comprimido, chamado também de “zipado”, em função da extensão “.ZIP” no nome do arquivo)

Quadro 9 – Linha típica de um currículo Lattes em formato XML.

<code><DISCIPLINA SEQUENCIA-ESPECIFICACAO="1">Geometria Descritiva IIA</DISCIPLINA></code>			
Início do Elemento	Atributo e valor (dentro das marcas de elemento '<' e '>')	Conteúdo do elemento	Fim do elemento
<code><DISCIPLINA</code>	<code>SEQUENCIA-ESPECIFICACAO="1"></code>	Geometria Descritiva IIA	<code></DISCIPLINA></code>

Fonte: fragmento de código de currículo gerado pela plataforma Lattes.

Verificou-se, pela análise de seu DTD, que a estrutura dos currículos Lattes é hierárquica, isto é, pode ser expressa e manipulada por linguagens que possuam a estrutura de dados tipo “matriz” (*array*) multidimensional. No entanto, a armazenagem destes dados, no contexto do artefato digital proposto, é feita em banco de dados relacional, que não trabalha com dados hierárquicos, mas em tabelas (dados tabulares). Dessa forma, será necessário construir tabelas SQLite que compo-tem as informações hierárquicas.

Na Figura 44, página 124, o mesmo fragmento de código da Figura 43, página 123, é visto em um editor de XML¹⁹ que apresenta os dados em estrutura hierárquica, também chamada de “árvore” hierárquica. Ficam destacados com um ícone de bolinha vermelha os atributos dos elementos “Ensino” e “Disciplina”. No painel à direita, ficam os conteúdos dos atributos e dos elementos.

Estes dados XML são importados pelo artefato digital e transformados por uma função em variáveis “array” (matriz) no formato nativo da linguagem *script Live-Code*. Por exemplo, dado o fragmento de código XML de um currículo Lattes (simplificado):

```
<CURRICULO-VITAE>
  <ATIVIDADES-DE-ENSINO>
    <ENSINO>
      <DISCIPLINA SEQUENCIA-ESPECIFICACAO="1">
        Geometria Descritiva IIA
      </DISCIPLINA>
    </ENSINO>
  </ATIVIDADES-DE-ENSINO>
</CURRICULO-VITAE>
```

¹⁹ O editor é o XMLNotepad, software gratuito da Microsoft.

Serão geradas as *arrays* multidimensionais no formato do LiveCode²⁰:

```
myArrayA["CURRICULO-VITAE"] ["ATIVIDADES-DE-ENSINO[1]"]  
["ENSINO[1]"] ["DISCIPLINA[1]"]["@value"] = "Geometria Descritiva  
IIA"
```

Logo depois do nome da *array*, aparecem expressões entre aspas e colchetes. São os chamados “subscritos” ou “índices”. Cada subscrito define uma “dimensão” da matriz. A linguagem *LiveCode Script* permite *arrays* com número ilimitado de dimensões, chamadas de multidimensionais. Nesta pesquisa, usou-se como subscritos das matrizes os “galhos” da árvore hierárquica dos currículos Lattes. Os valores de um elemento são armazenados com o subscrito “@value” e os atributos de um elemento são guardados numa *array* com subscrito “@attributes”, por escolha do criador da função de tradução entre o formato de árvore hierárquica XML e uma matriz LiveCode.

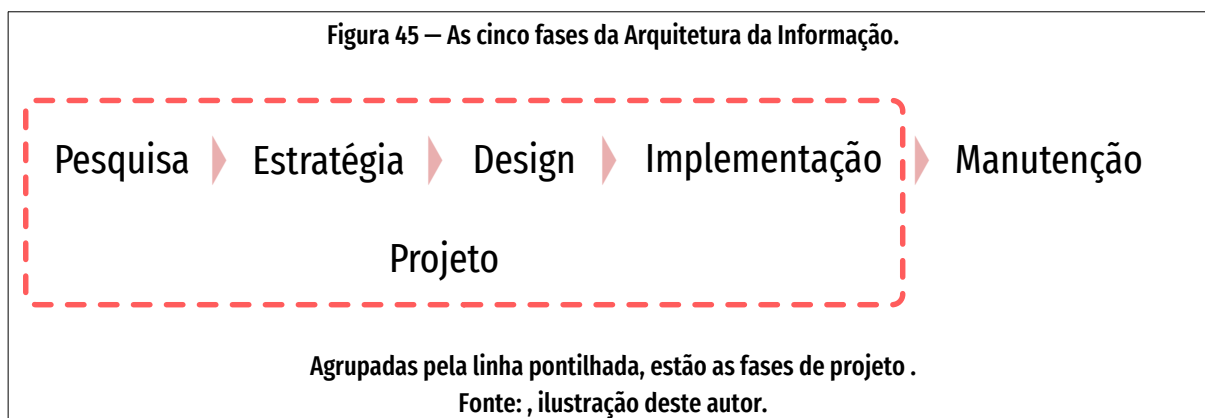
Esta *array* acima, por exemplo, guarda o valor – o conteúdo – do elemento XML, no caso, a *string* “Geometria Descritiva IIA”.

Outro exemplo é a *array* a seguir, que guarda um atributo da tag “DISCIPLINA[1]”, no caso, o número de ordem na sequência de itens:

```
myArrayA["CURRICULO-VITAE"] ["ATIVIDADES-DE-ENSINO[1]"]  
["ENSINO[1]"] ["DISCIPLINA[1]"]["@attributes"] ["SEQUENCIA-  
ESPECIFICACAO"] = 1
```

Elementos que se repetem iterativamente, como “DISCIPLINA” – entram várias disciplinas “filhas” neste “elemento-pai” –, recebem um número sequencial, anexo ao final de seu nome, entre chaves para facilitar sua identificação e guardar sua ordem dentro da estrutura de dados, que frequentemente é uma informação importante. No exemplo abaixo, vê-se o número de ordem das *arrays* armazenado na *string* de nome do subscrito.

20 A matriz ocupa apenas uma linha. É apresentada aqui em muitas linhas por ser muito longa.



```
myArrayA["CURRICULO-VITAE"] ["ATIVIDADES-DE-ENSINO[1]"]
myArrayA["CURRICULO-VITAE"] ["ATIVIDADES-DE-ENSINO[2]"]
myArrayA["CURRICULO-VITAE"] ["ATIVIDADES-DE-ENSINO[3]"]
| | |
# Parte do string subscrito+-----+--+|
# String subscrito da matriz+-----+
```

Estes números não devem ser confundidos com os subscritos da *array*, embora tenham os delimitadores “[” e “]”. Eles fazem parte da *string* de nome do subscrito da *array*. No entanto, podem ser usados para ordenar e classificar *arrays*, graças à flexibilidade da linguagem LiveCode no tratamento de *strings*.

4.3.3 Sistemas de Navegação, de Rotulagem e de Busca

A segunda categoria da Arquitetura da Informação, **Sistema de Navegação**, foi aplicada na construção de um conjunto de opções de consulta de dados disponível ao usuário, como menus, botões, listas, guias e assistentes de tarefas.

A terceira categoria, **Sistema de rotulagem**, ajudou a definir o texto de botões, menus, links, títulos e outros elementos.

Estas categorias foram desenvolvidas com ajuda dos preceitos do Design da Informação contemporâneo.

Finalmente, a categoria **Sistema de Busca** definiu a maneira pela qual são feitas as buscas por informações, em que campos do banco de dados serão feitas as buscas e de que forma se apresentarão os resultados. Este sistema ganha relevância à medida que é necessário um robusto algoritmo de **deduplicação**, a eliminação de produtos acadêmicos duplicados (quando dois ou mais coautores registram a obra com leves diferenças de nome, por exemplo).

4.3.4 As fases do Processo e a Metodologia

O processo de desenvolvimento proposto pela Arquitetura da Informação passa por 5 fases, sendo 4 delas atividades de projeto (Morville; Rosenfeld, 2007, p. 232) e uma atividade de manutenção do sistema. Estas fases aparecem na Figura 45, página 127: Das cinco fases, este trabalho aplicou as quatro primeiras – Pesquisa, Estratégia, Design e Implementação – que constituem o projeto, propriamente dito.

Já quanto aos procedimentos metodológicos, a aplicação dos processos de desenvolvimento está descrita nas subseções “3.2 Delineamento da pesquisa” (página 97), “3.3 Requisitos do projeto” (página 99), e “3.1 Avaliação do artefato” (página 89).

4.3.4.1 Pesquisa

A primeira fase da Arquitetura da Informação é a **pesquisa**. Neste trabalho, se fez a atualização da pesquisa do material disponível já encontrado para desenvolvimento do protótipo. Buscou-se pelas palavras-chave: CAPES, Lattes, autoavaliação, *software* livre, *design* da informação, visualização de dados, *Design Science Research* nos sites de busca acadêmica *ScienceDirect* (<http://www.sciencedirect.com>), IEEE Xplore (<https://ieeexplore.ieee.org>), Google Acadêmico (<https://scholar.google.com>) – para buscas e “alertas” diários por *email* – e *Web of Science* (<http://www.webofknowledge.com>).

Quadro 10 – Comparativo de bibliotecas de software para gráficos de dados.

JavaScript Charting Library - comparison			Supported Chart Types													Supported Bar Chart Types					Other	Interactivity			Rendering Technologies				
	License	Total features	Árvore	Linha	Linha de tempo	Dispersão	Área	Pizza	Rosquinha	Bala	Radar	Funnel	Gantt	Network	Grouped	Stacked	Negative	Discrete	Horizontal	3D	Legends	Mouse Over	onClick	HTML5 Canvas	SVG	VML	jQuery	WebGL rendering	
Google Charts	Free	0	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Não	Não	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	
D3.js	BSD-3	0	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Não	Sim	Não	Sim		
TreeChart JS	MIT	0	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Não	Sim	Não	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Não		
chart.js	MIT	0	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Não	Sim	Não	Sim		
RGraph	MIT	0	Sim	Sim	Não	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Não	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim		
Echarts	BSD-3	0	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Não	Não	Sim	Sim	Sim	Sim	Não	Sim	Sim	Sim	Sim	Sim	Sim	Não	Não	Sim	
JenScript	BSD-3	0	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Não	Não	Não	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Não	Sim	Não	Sim		
billboard.js	MIT	0	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Não	Sim	Não	Não	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim		
ReactiveChart	MIT	0	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Não	Não	Não	Não	Sim	Sim	Sim	Sim	Sim	Sim	Não	Sim	Sim	Sim	Sim	Sim	Não	Sim		
ApexCharts	MIT	0	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Não	Sim	Não	Não	Não	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Não	Sim	Não		
C3.js	MIT	0	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Não	Não	Não	Não	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Não	Sim	Não	Sim	
Chart.js	MIT	0	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Não	Sim	Não	Não	Não	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Não	Não	Não		
DHTMLX Charts	GPL	0	Sim	Sim	Não	Sim	Sim	Sim	Sim	Não	Sim	Não	Sim	Não	Sim	Sim	Sim	Sim	Não	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim		
dimple	MIT	0	Sim	Sim	Não	Sim	Sim	Sim	Sim	Não	Não	Não	Não	Não	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Não	Sim	Não	Sim	
NVD3	Apache 2.0	0	Sim	Sim	Não	Sim	Sim	Sim	Sim	Sim	Sim	Não	Não	Não	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Não	Sim	Não	Sim		
Dojo Charting	BSD or AFL	0	Sim	Sim	Não	Sim	Sim	Sim	Sim	Não	Não	Não	Não	Não	Sim	Sim	Sim	Sim	Não	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim		
VisJS	Apache 2.0 and f	0	Sim	Sim	Sim	Sim	Sim	Sim	Não	Não	Não	Não	Sim	Sim	Sim	Sim	Sim	Sim	Não	Não	Sim	Sim	Sim	Sim	Sim	Não	Não		
Chartist	WTFPL or MIT	0	Sim	Sim	Não	Sim	Sim	Sim	Sim	Não	Não	Não	Não	Não	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Não	Sim	Sim	Não	Sim	Não		
jqPlot	MIT or GPLv2	0	Sim	Sim	Não	Sim	Sim	Sim	Sim	Não	Não	Não	Não	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Não		
MuzeJS	MIT	0	Sim	Não	Sim	Sim	Sim	Sim	Sim	Não	Não	Não	Não	Não	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Não	Sim	Não	Sim			
canvasXpress	GPLv3, name/en	0	Sim	Sim	Não	Sim	Sim	Sim	Sim	Não	Não	Não	Não	Sim	Sim	Sim	Sim	Não	Não	Sim	Sim	Sim	Sim	Sim	Sim	Não	Sim		
Webix JS Charts	GPL	0	Sim	Sim	Não	Sim	Sim	Sim	Sim	Não	Sim	Não	Não	Não	Sim	Sim	Sim	Sim	Não	Sim	Sim	Sim	Sim	Sim	Sim	Não	Sim	Sim	
NextCharts	Apache 2.0	0	Sim	Sim	Não	Não	Sim	Sim	Sim	Não	Não	Não	Não	Não	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Não	Não	Sim	
Chart Builder By Livegap	Free	0	Sim	Sim	Não	Não	Sim	Sim	Sim	Não	Sim	Não	Não	Não	Sim	Sim	Sim	Sim	Não	Sim	Sim	Sim	Sim	Sim	Sim	Não	Não		
Dygraphs	MIT	0	Sim	Sim	Sim	Não	Sim	Não	Não	Não	Não	Não	Não	Não	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Não	Não		
Flotr2	MIT	0	Sim	Sim	Não	Sim	Não	Sim	Não	Não	Sim	Não	Não	Não	Sim	Sim	Sim	Sim	Não	Sim	Sim	Sim	Sim	Sim	Sim	Não	Não		
dc	Apache 2.0	0	Sim	Sim	Não	Sim	Sim	Sim	Sim	Sim	Não	Não	Não	Não	Sim	Sim	Sim	Sim	Não	Não	Não	Sim	Sim	Sim	Sim	Não	Sim		
Flot Charts	MIT	0	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Não	Não	Não	Não	Não	Sim	Sim	Sim	Não	Não	Não	Sim	Sim	Sim	Sim	Sim	Não	Não		
PlusCharts	Free	0	Sim	Sim	Não	Não	Sim	Sim	Sim	Não	Não	Não	Não	Não	Sim	Sim	Sim	Não	Não	Sim	Sim	Sim	Sim	Sim	Não	Sim	Sim		
Webcharts	MIT	0	Sim	Sim	Sim	Sim	Não	Não	Não	Sim	Não	Não	Não	Não	Sim	Sim	Sim	Não	Não	Sim	Sim	Sim	Sim	Sim	Não	Sim	Não		
YUI Charts	BSD-3	0	Sim	Sim	Não	Não	Sim	Sim	Sim	Não	Não	Não	Não	Não	Sim	Sim	Sim	Não	Não	Não	Sim	Sim	Sim	Não	Sim	Sim	Sim		
gRaphael	MIT	0	Sim	Sim	Não	Sim	Não	Sim	Sim	Não	Não	Não	Não	Não	Sim	Sim	Sim	Não	Não	Sim	Sim	Sim	Não	Não	Sim	Sim	Não		
rickshaw	MIT	0	Sim	Sim	Não	Sim	Sim	Sim	Sim	Não	Não	Não	Não	Não	Sim	Sim	Sim	Não	Não	Não	Sim	Sim	Sim	Não	Não	Sim	Não		
Cytoscape.js	MIT	0	Sim	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Sim	Não	Não	Não	Não	Não	Não	Não	Sim	Sim	Sim	Não	Não	Sim		

Fonte dos dados : Wikipedia (2023). Edição de dados: Este autor.

4.3.4.2 Estratégia

A fase da pesquisa permitiu um entendimento contextual que levou a uma **estratégia** da arquitetura da informação (Morville; Rosenfeld, 2007, p. 232), a segunda fase. Dentro de uma perspectiva *top-down*, a estratégia permitiu definir dois ou três níveis mais altos em que se organizam as informações e características do artefato digital e suas estruturas de navegação. De uma perspectiva *bottom-up*, a estratégia definiu formatos de arquivos e de dados que foram usados no artefato digital. Por exemplo, quais formatos de bibliografia estão disponíveis para exportação de dados do artefato digital para outros sistemas (BibTex, EndNotes, RIS, JSON, Zotero).

4.4 Design

A fase de **Design** de **Sistemas de Informação** transformou a estratégia em uma arquitetura de informação, envolvendo design gráfico, programação e conteúdo (textos para telas de ajuda, tutoriais, texto da licença de uso do software, ficha técnica etc).

4.4.1 Escolha da biblioteca gráfica

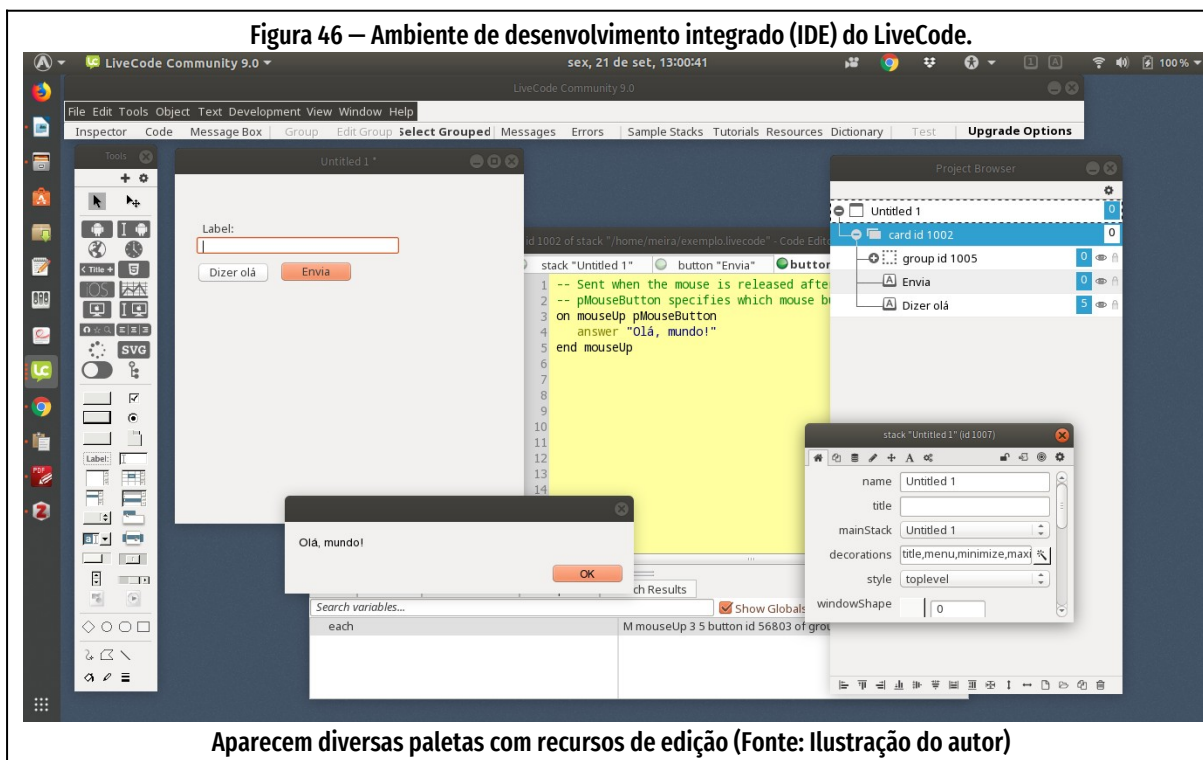
A biblioteca de funções gráficas foi escolhida pelos seguintes critérios:

- 1) Quanto à licença, deveria ser *open source*.
- 2) Quanto aos recursos, deveria ter o maior número possível de tipos de gráficos.
- 3) Quanto à popularidade, deveria ter uma grande base de usuários.

O conjunto de análise (Quadro Quadro 10, página 129) foi obtido de uma listagem com as principais bibliotecas de apresentações em JavaScript, comerciais ou abertas. Decidiu-se pela biblioteca **Plotly.js**, em função da facilidade de programação e por usar a biblioteca **D3.js**, sendo, portanto, mais simples e avançado que esta que é uma das bibliotecas mais completas para a tarefa.

4.5 Implementação

A fase de Implementação realiza, de forma prática, tudo o que foi encontrado na pesquisa, organizado na estratégia e planejado no design. Envolve a programação em linguagens de computador, programação do banco de dados, layout e aspecto visual do artefato digital, tipos de gráficos conforme o tipo de dados usados, testes e depuração de erros.



4.5.1 Codificação

Esta fase se ocupa de implementar o produto proposto. Para prototipagem e geração final do artefato digital, utilizou-se o sistema LiveCode, que conta com uma linguagem de programação com sintaxe semelhante à língua inglesa. Os *scripts* desta linguagem podem ser atribuídos a qualquer objeto da interface gráfica com o usuário.

4.5.2 Estruturação do banco de dados

O banco de dados do Currículo Lattes contava com mais de sete milhões de registros em 2023. Esta quantidade excede o limite de planilhas como LibreOffice Calc ou Microsoft Excel, que carregam no máximo 1.048.576 registros ou linhas de da-

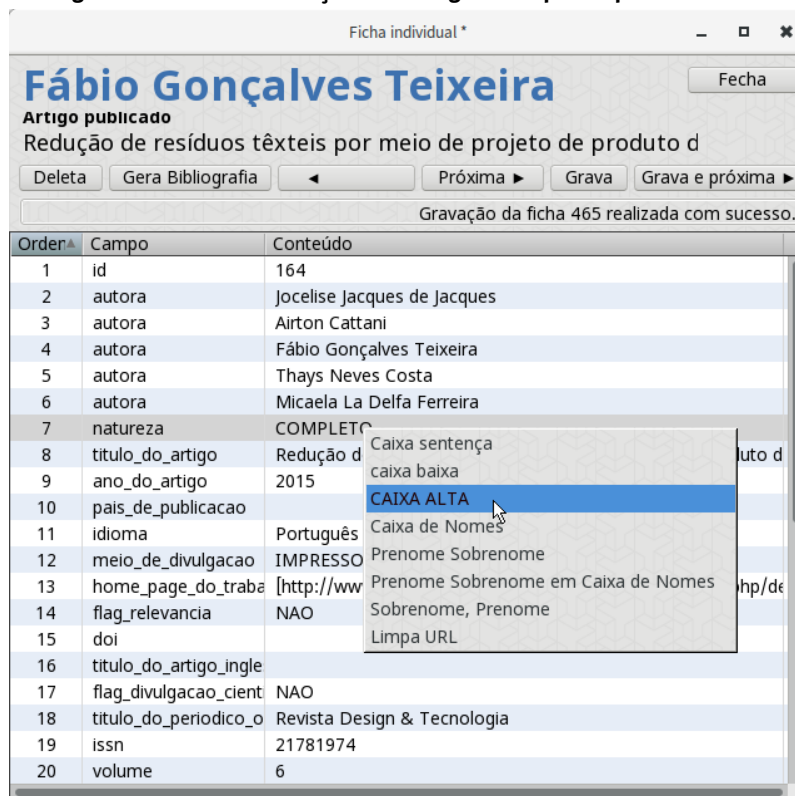
dos. Embora esteja previsto para ser usado em grupos de até 100 pesquisadores, o artefato digital aqui proposto é dimensionado para trabalhar com qualquer volume razoável de dados. Na prática, o artefato deve trabalhar com cerca de 20 pesquisadores, no caso de um único PPG, a alguns milhares de pesquisadores, no caso de análise de uma instituição com dezenas de PPGs. Para guardar e gerenciar o grande volume de informação do Lattes, foi usado o banco de dados de código aberto SQLite, desenhado para integração a outros aplicativos e adequado a acessos por um único cliente por vez. Ele é rápido, confiável e largamente usado no mundo, o que garante mais chances de se resolver eventuais problemas ou dúvidas de operação, através da ajuda da comunidade de usuários. É o banco de dados padrão do sistema operacional Android, por exemplo. O SQLite está bem documentado e integrado ao LiveCode. Foi usado também para guardar diversas configurações do artefato digital, além de análise e testes nos dados acadêmicos da CAPES e CNPq.

Para o projeto das consultas SQL ao banco de dados pelo artefato, além do gerenciamento e manipulação dos dados, foi usado o **DB Browser for SQLite**, versão 3.10.99 (DB4S, 2018), artefato digital de código aberto e multiplataforma: existem versões para os sistemas operacionais GNU/Linux, Windows e Macintosh. Com esta ferramenta, definiu-se as *queries* (consultas) em linguagem SQL, que posteriormente foram usadas pela linguagem de programação escolhida.

4.5.3 Ambiente de desenvolvimento

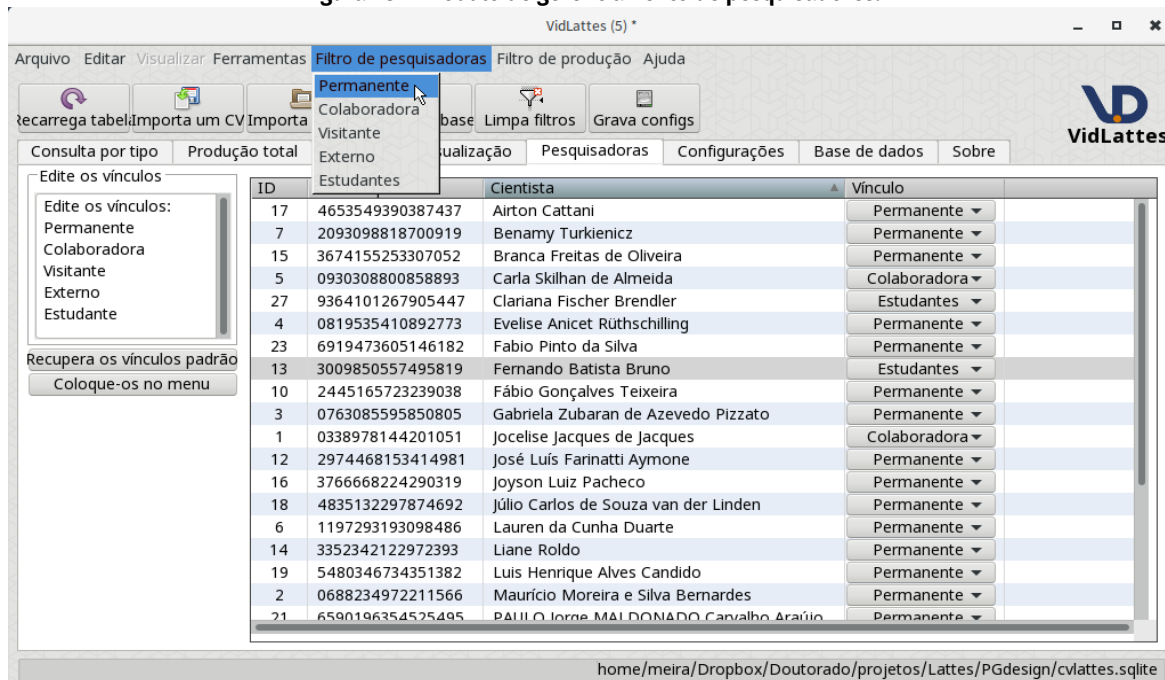
Para a codificação do artefato digital foi usado o **LiveCode**, um Ambiente de Desenvolvimento integrado (IDE) e linguagem de programação *script*, com versão livre e

Figura 47 – Módulo de edição de bibliografia do protótipo VidLattes.



Fonte: este autor (2023).

Figura 48 – Módulo de gerenciamento de pesquisadores.



Fonte: este autor (2023).

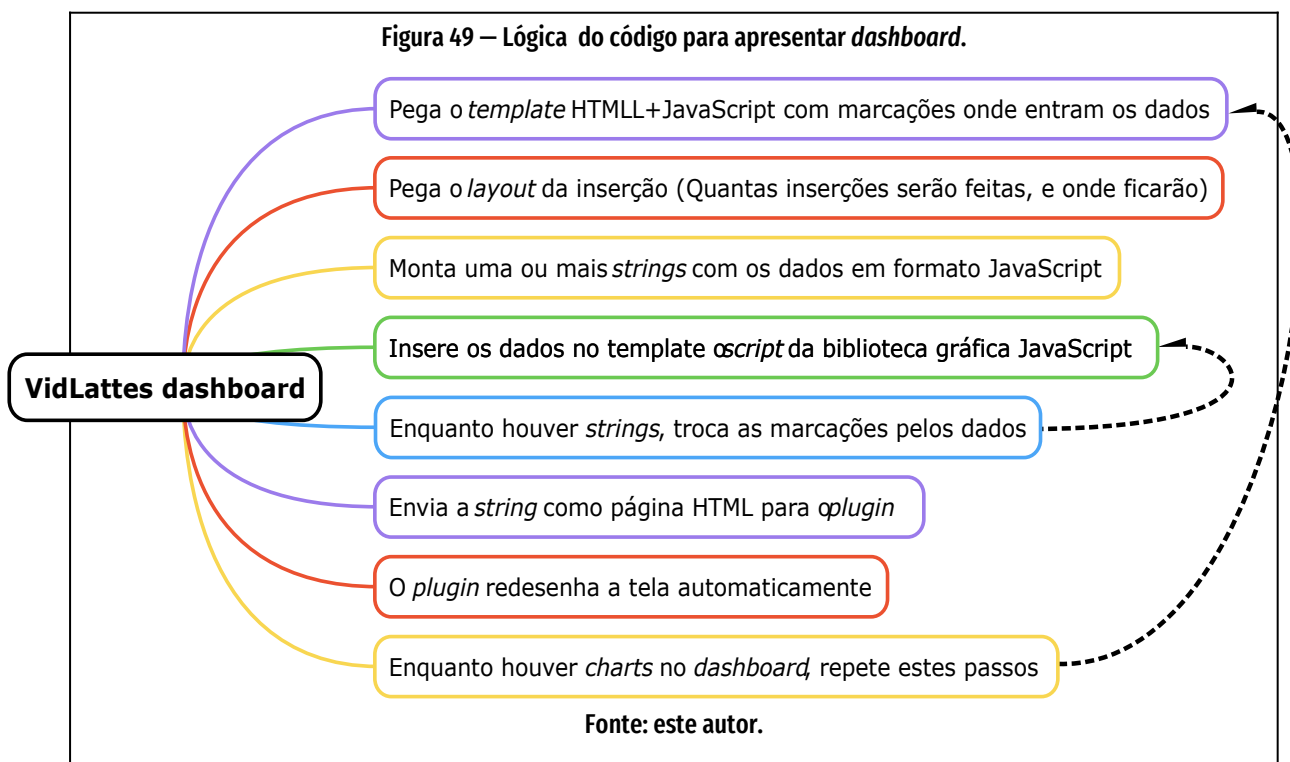
aberta que permite, a partir de um único código, a geração de aplicativos para várias plataformas: Windows, Macintosh, Linux, iOS, Android e Web (html5). A arquitetura do LiveCode, similar à do *HyperCard* (software embutido nas primeiras gerações de computadores Apple Macintosh), permite o desenvolvimento rápido de protótipos e aplicativos (*Rapid Application Development—RAD*). Trabalha através de um Ambiente de Desenvolvimento Integrado (*Integrated Development Environment—IDE*) com **dois modos principais de operação: modo edição e modo aplicação**, selecionáveis por um botão correspondente na paleta de ferramentas da IDE. Uma tela com o artefato digital em funcionamento aparece na Figura 46, página 131.

No **modo edição**, edita-se a interface gráfica e os *scripts* que cada elemento gráfico pode conter. No **modo aplicação**, pode-se executar o programa com opção de se fazer o *tracing*, isto é, a execução de uma instrução por vez, para efeitos de depuração. Qualquer mudança na estrutura gráfica da interface ou na lógica do programa pode ser imediatamente testada. Esta característica resulta em maior produtividade do programador, se comparado com o modo tradicional de desenvolvimento de software, que envolve ciclos de edição-compilação-teste.

O **LiveCode** possui um objeto do tipo navegador Web baseado no *Chromium Embedded Framework (CEF)*, versão do navegador *open source Chromium* especialmente desenvolvida para ser embutida em outros aplicativos. Este navegador, como todos os programas desta classe, utiliza um interpretador de código em linguagem de programação JavaScript. Desta forma, se teve acesso às bibliotecas de *software JavaScript*, para apresentação visual dos dados.

O artefato digital extrai os dados do currículo Lattes em formato XML e os gerencia através de um banco de dados relacional **SQLite**. No capítulo a seguir, descreve-se em mais detalhes estes processos e as tecnologias empregadas.

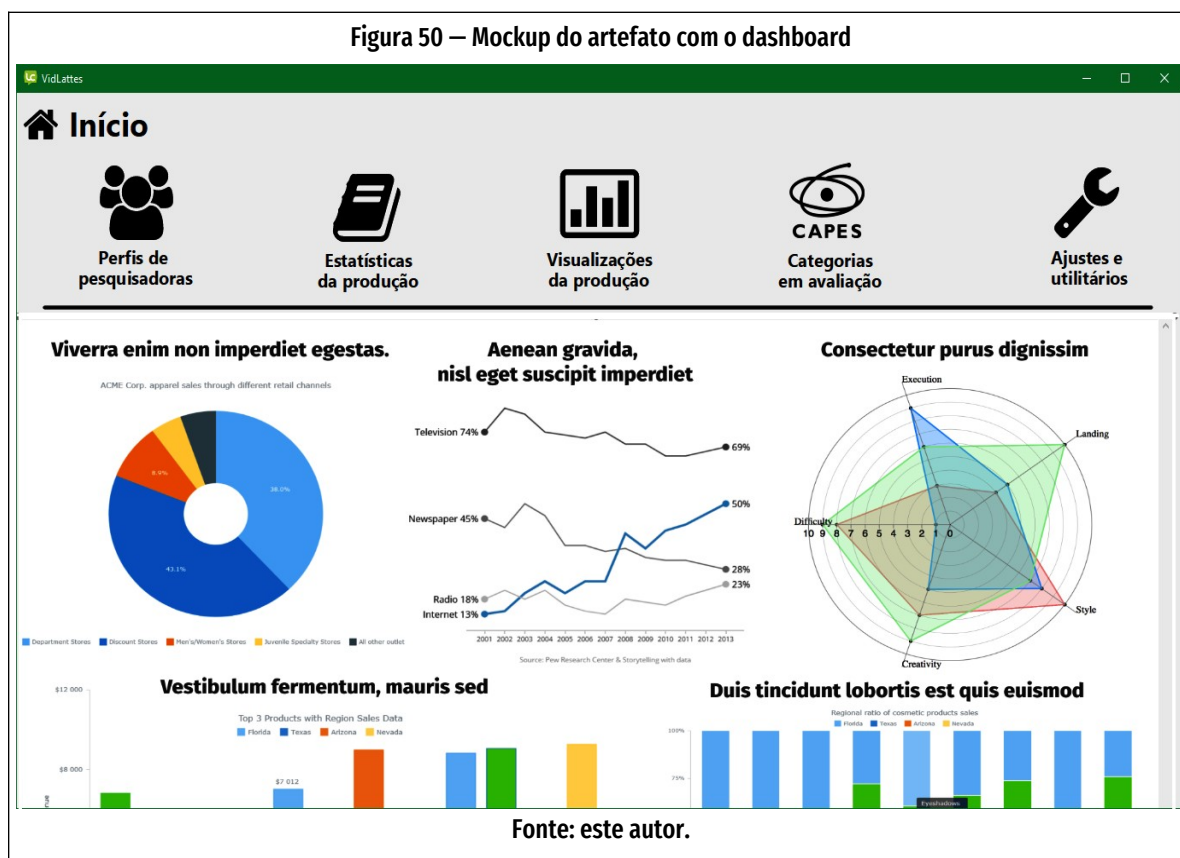
O **LiveCode** usado foi a versão 9.6.3 *Open Source*, também conhecida como “*Community*”, cuja licença proíbe o uso comercial dos aplicativos desenvolvidos. Ou seja, se produzirá software livre, assim como a ferramenta geradora. Poderá ser usado gratuitamente pelos milhares de PPGs do país.



4.5.4 Protótipo

O artefato aqui proposto foi desenvolvido a partir de um protótipo funcional com o mecanismo básico de operação: leitura de currículos em formato XML zipado, compilação e apresentação de tabelas numéricas da produção acadêmica. As funcionalidades elementares (Figura 48, página 133) foram definidas em entrevista informal com coordenadores do PPG em Design da UFRGS. Nas páginas seguintes aparecem algumas das telas do protótipo. O protótipo possui ainda um editor de bibliografia com alguns utilitários – nomes, padronização da URL – que facilitam a padronização do banco de dados (Figura 47, página 133).

A partir deste protótipo/prova de conceito, pretendeu-se melhorar a interface gráfica, desenvolver novos módulos de funções e integrar o módulo *Chrome Embedded* (CEF) de navegação Web. Isto permitirá o uso de bibliotecas gráficas em JavaScript, linguagem embutida no CEF e padrão em todo navegador Web. A Figura 52, página



138, mostra o mapa das funções do VidLattes, e salienta o trecho de que trata esta pesquisa, o módulo de visualização de dados.

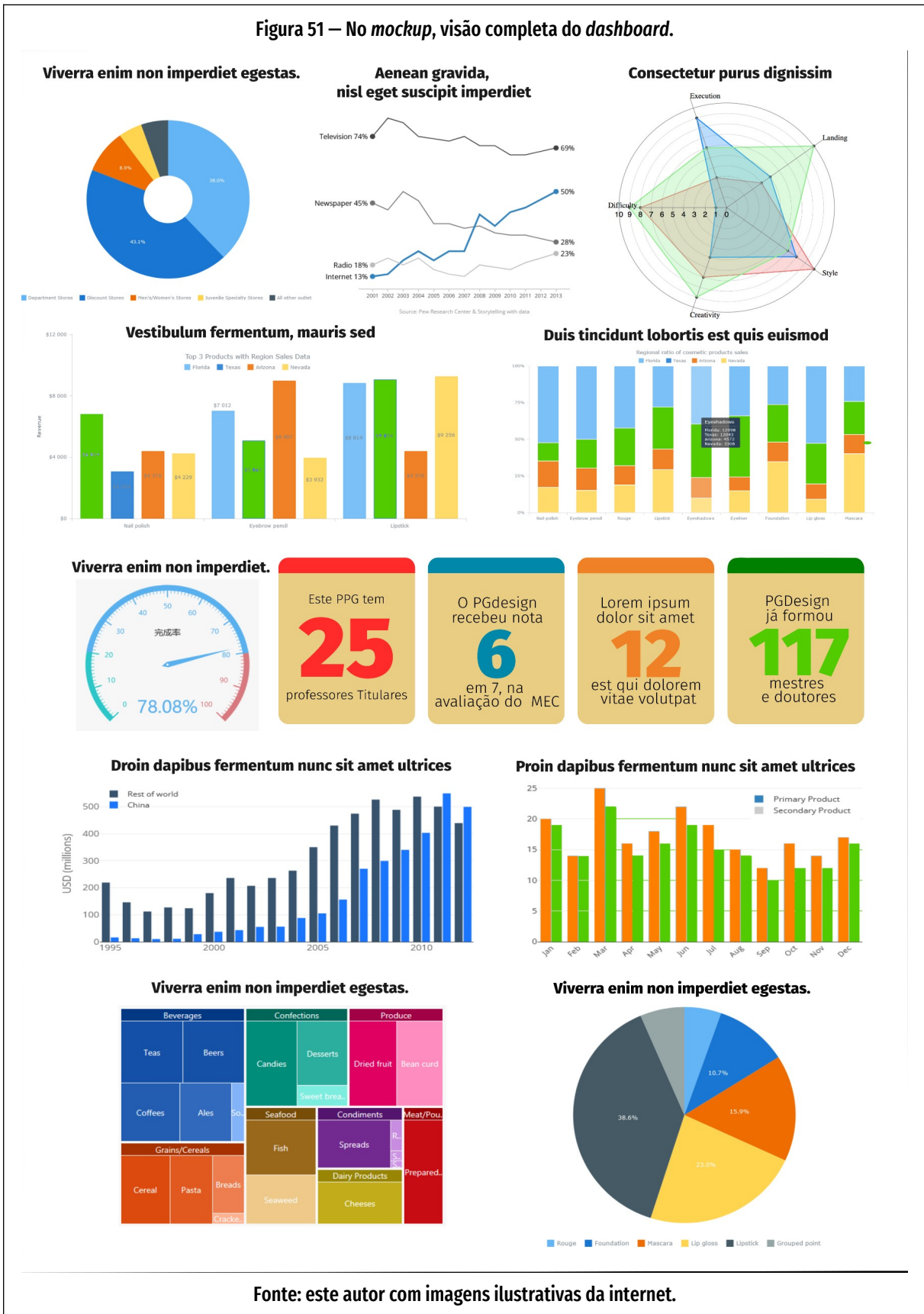
A Figura 49, página 135, mostra a lógica da função de apresentar um *dashboard*.

A interface do artefato com o código JavaScript em arquivos HTML é baseada em estruturas de dados em arquivos de configuração no formato TXT para facilitar a edição. Para cada tipo de gráfico, há um arquivo com as particularidades de cada biblioteca de apresentação. Isto permitirá se usar qualquer biblioteca de gráficos em JavaScript. Neste trabalho, optou-se pela biblioteca **plotly.js**²¹.

A Figura 50, página 136, mostra um *mockup* do artefato, com o *dashboard* em uma área “rolável”. A Figura 51, página 137, apresenta um *mockup* de todo o painel com diversos tipos de gráficos.

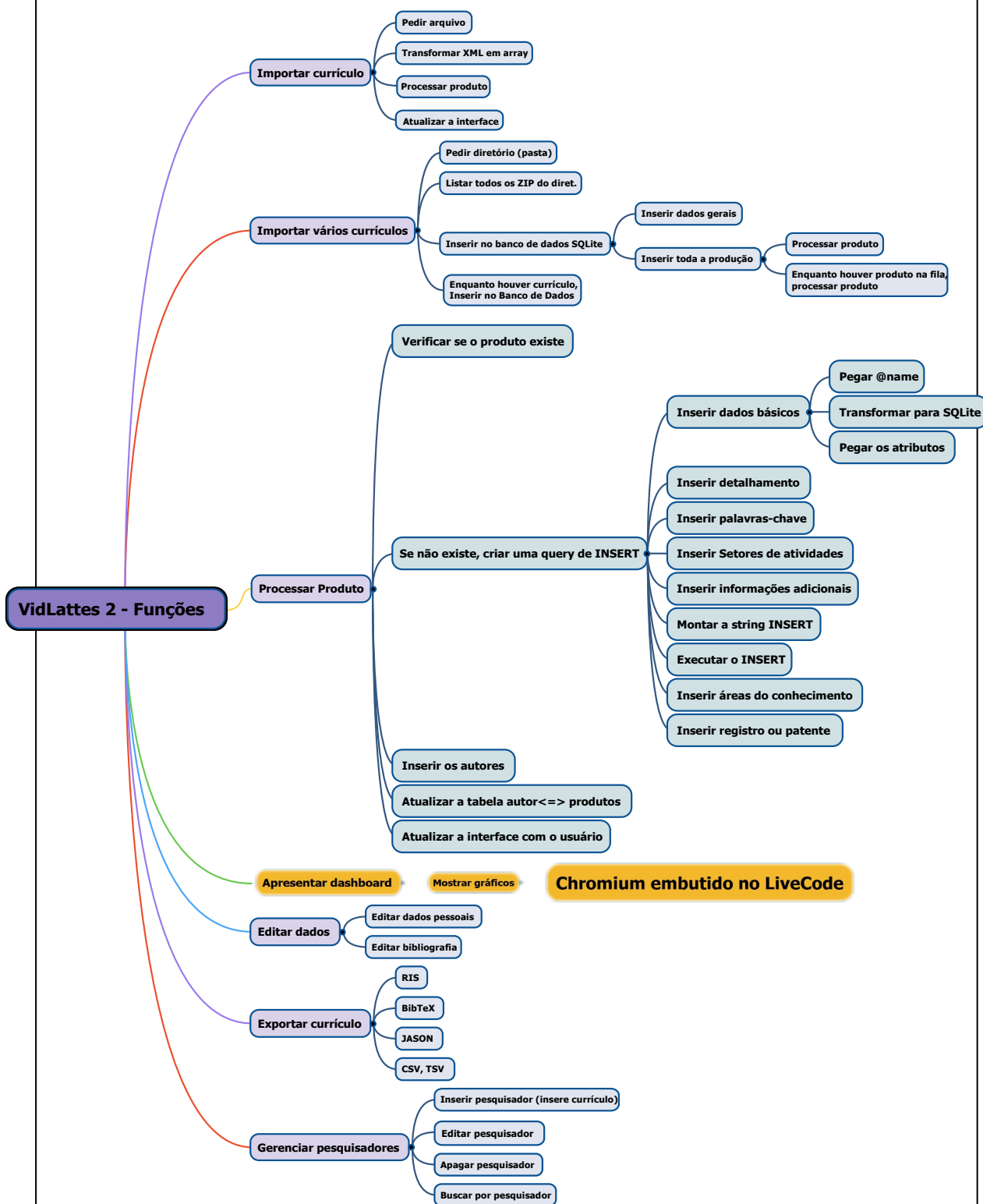
21 <https://plotly.com/javascript/>

Figura 51 – No *mockup*, visão completa do *dashboard*.



Fonte: este autor com imagens ilustrativas da internet.

Figura 52 – Estrutura do VidLattes. Em destaque, o módulo de visualização de dados.



Fonte: este autor.

4.5.5 Tipos de gráficos usados

A escolha dos tipos de gráficos foi feita a partir do tipo de dados a ser apresentado, conforme a Figura 27 — Fluxograma de escolha de gráfico conforme operação a ser apresentada. na página 72. A grande maioria dos dados da plataforma Lattes é sobre a evolução da produção acadêmica do pesquisador em comparação consigo mesmo e em comparação a outros pesquisadores, em geral ao longo do tempo, como no caso das avaliações quadrienais e a evolução ao longo da história da pós-graduação no Brasil.

Os gráficos escolhidos foram os gráficos de linhas, gráficos de áreas, gráficos de áreas empilhadas, gráfico de pizza (estes serão usados com parcimônia), gráficos de barras e colunas. Eventualmente, serão usados gráficos de radar.















4.5.6 Avaliação

Nesta subseção, procurou-se verificar se o artefato produzido foi adequado às teorias utilizadas. Avaliou-se em duas dimensões, a verificação e a validação, os itens propostos por Boehm (1979, 1984) para avaliação de software.

A verificação foi feita depois de completada a definição ou codificação de cada função do artefato, através de testes adequados ao escopo da função desenvolvida. Esta fase foi facilitada pela IDE LiveCode, pelo uso de rastreamento de erros e visualização de variáveis e funções em uso.









Quanto aos aspectos técnicos, a avaliação está descrita no Quadro 11, página 140, conforme alguns dos critérios apresentados no Quadro 6 - “Critérios de validação e verificação”, página 94.

Quadro 11 – Critérios de validação e verificação aplicados

Critério	Propriedade	Descrição	Avaliação
Completeness	Sem "ASD"	Não deve haver elemento "ASD" ("A Ser Definido") ou locais onde haja lacunas nas especificações.	
	Sem referências não-existentes	Não deve haver referência a elementos não definidos nas especificações.	
	Sem itens da especificação faltantes	Todos os itens especificados devem estar presentes.	
	Sem funções faltantes	Funções que deveriam fazer parte do sistema mas não estão nas especificações.	
Consistency	Interna	Quando não há conflito entre elementos internos do software.	
	Externa	Quando não há conflito entre elementos internos e externos com as especificações	
	Erros de digitação	Erros de escrita e de digitação (<i>typos</i>), em geral, levam a problemas de consistência.	
Viability	Engenharia humana	O sistema pode prover ao usuário uma maneira satisfatória de operação?	
		O sistema satisfaz as necessidades humanas em vários níveis?	
	Engenharia de recursos	O sistema pode ser desenvolvido de maneira a cumprir, a um aceitável custo, todas as especificações requeridas.	
		Em sua vida útil e a um custo-benefício favorável, o sistema poderá acomodar o crescimento esperado de requisitos operacionais? Ou seja, seria escalável ?	
	Engenharia de software	A manutenção do sistema terá custo-benefício vantajoso?	
		A portabilidade terá custo-benefício vantajoso?	
		O sistema terá suficiente precisão, confiança e disponibilidade ?	
<p>Fonte: este autor. Legenda:  Cumpre o item.  Não cumpre o item.</p>			

Quanto aos aspectos funcionais, avaliação foi conforme apresentado no Quadro 12, página 141:

Quadro 12 – Atendimento de requisitos funcionais do projeto

	Critério	Avaliação
1	Levantamento rápido da produção docente e discente do PPG registrada no banco de dados do Currículo Lattes.	
2	Geração de relatórios com a produção docente e discente.	
3	Análise quantitativa da produção docente e discente ao longo de vários anos, tendo como <i>default</i> o intervalo de anos entre as avaliações, determinado pela CAPES (atualmente, quadrienal).	
4	Capacidade de gerar gráficos a partir dos dados.	
5	Capacidade de fazer busca por todo o conteúdo	
6	Capacidade de imprimir os resultados.	
7	Capacidade de exportar os resultados para outro sistema de representação gráfica, caso necessário, em formato CSV.	
8	Capacidade multiplataforma. O sistema deve funcionar pelo menos nos ambientes mais usados por usuários de computador <i>desktop</i> : Windows , Linux e MacOS.	

Fonte: este autor. Legendas:  = Atende .  = Não atende.

Os itens 1, 2, 3, 7 e 8 já foram implementados. Os itens 5 e 6 ainda não foram desenvolvidos. O item 4, “Capacidade de gerar gráficos a partir dos dados”, por questões de dinâmica da pesquisa, foi implementado até o estágio de design da interface e desenvolvimento do algoritmo, necessitando a codificação ser desenvolvida em pesquisa posterior.

5 Explicitação do aprendizado

A penúltima fase da *Design Science Research* determina que se faça a explicitação do aprendizado, legando um conhecimento construído na elaboração da pesquisa. Isto é tratado nas seções a seguir, “5.1 Documentação do processo de projeto e desenvolvimento do artefato”, “5.2 Discussão dos resultados” e “5.3 Outras questões”.

5.1 Documentação do processo de projeto e desenvolvimento do artefato

Devido à dinâmica da pesquisa, não foi implementado o código funcional, apenas o estágio de projeto da interface com o usuário. A partir da Arquitetura da Informação, do Design da Informação e de preceitos da Visualização de Dados, foi criado o painel de dados mostrado na Figura 50, página 136, e Figura 51, página 137.

5.2 Discussão dos resultados

A elaboração de um protótipo, a partir de levantamento parcial de necessidades de um programa de pós-graduação (PPG), mostrou que o artefato final poderá ser bastante útil, no que toca à adequação às exigências da CAPES. O protótipo do artefato digital consolidou, em cerca de 3 minutos, a coleção de dados sobre um PPG dispersos em duas dezenas de currículos.

Entre os pontos a serem observados estão as inconsistências do próprio conteúdo do banco de dados Lattes, como se espera de um conjunto de dados inseridos por 7 milhões de pessoas. Sempre haverá inconsistências. Por exemplo, coautores de uma mesma obra podem grafar seu próprio nome ou o nome da publicação de forma diferente, em seus currículos. A plataforma Lattes ainda não dispõe de um banco de dados de obras acadêmicas que permita a diversos coautores fazerem

referências padronizadas. A entrada da citação é responsabilidade de cada pesquisador. Um único caractere do título que seja diferente, entre coautores, já fará com que a produção seja vista no sistema como duas obras diferentes. Portanto, o artefato digital aqui proposto, ao consolidar as obras dos pesquisadores, deverá fazer uma busca por palavras-chave dos títulos e testar mais alguns campos, como data, ISSN, ISBN, para certificar-se de que a referência aponta para a mesma obra. Esta operação é chamada **deduplicação**.

No âmbito da Arquitetura da Informação, a utilização de um objeto LiveCode da classe “*browser*” (navegador *Web* embutido), derivado do projeto de código aberto *Chromium*, pode oferecer grande flexibilidade e grandes possibilidades ao artefato digital proposto, pois poderá usar qualquer biblioteca gráfica escrita em JavaScript. Este arranjo também pode ser usado também em linguagens de programação como Free Pascal/IDE Lazarus.

Como fonte de dados, decidiu-se usar os arquivos XML do currículo Lattes, em função de supostamente terem maior precisão, pois são gerados diretamente a partir do banco de dados, sem a necessidade de extração do html.

Conhecendo-se o passado da CAPES e da plataforma Lattes, pode-se começar a pensar o futuro Lattes. Como o sistema brasileiro vai sobreviver em um mundo com redes sociais acadêmicas como [Academia.edu](https://www.academia.edu/) e [ResearchGate.Net](https://www.researchgate.net/), além de sistemas de indexação como ORCID e DOI? Neste contexto, deixar o Lattes sucumbir seria condenar o Brasil a ser mero usuário de comunidades científicas comerciais — pagando assinatura ou se submetendo a anúncios intrusivos (Righetto, 2021) — ou usuário de comunidades públicas, mas sem o protagonismo de ser o proprietário e coautor da base de dados.

Analisando-se a evolução na área dos dados acadêmicos, pode-se arriscar uma previsão: o ORCID se transformará no “Lattes mundial”, banco de dados internacional de pesquisadores, grupos de pesquisa e instituições, junto com o DOI, banco de dados de publicações digitais. Portanto, deve-se começar a pensar em integra-

ção de sistemas, ou como o Lattes vai interagir com o ORCID e o DOI. Por exemplo, o Lattes necessitará de uma API (*Application Programming Interface*), uma maneira padronizada de permitir, sem o limitador de CAPTCHAs, a consulta ao banco de dados através de programação de máquinas.

Outro ponto que salientamos é que o conhecimento obtido pela exploração de um pequeno *sub-set* dos dados sobre produção acadêmica — poucas dezenas de pesquisadores do banco de dados Lattes da CAPES — pode servir de base para exploração, como *big data*, de mais bancos de dados como os do CNPq e os do IBGE. Isto pode levar a um ótimo uso dos dados de sete milhões de usuários acadêmicos.

Por fim, pode-se perguntar: qual a utilidade de uma pesquisa que compara gerenciadores de currículos Lattes, se já existem vários concorrentes no mercado? Neste nosso caso, o artefato funcionando é importante, pois pretende ser um somatório das melhores características de todos os concorrentes. Mas não é mais importante que o processo todo de se aplicar a metodologia *Design Science Research* na elaboração do artefato. Esta escolha de ferramenta ajuda a difundir pesquisas que vêm resolver problemas concretos da sociedade que subsidia os estudos, embora nem sempre com o melhor artefato do mundo, mas o que foi possível planejar com os recursos que se dispunha.

5.3 Outras questões

O software livre, característica da maior parte das ferramentas usadas no design do artefato, não é uma panaceia universal. É necessário desenvolver um ecossistema com milhares de usuários para que um projeto FOSS se torne viável, tenha uma vida longa e uma boa base de conhecimentos para resolver problemas encontrados durante o desenvolvimento de *software*. O caso LiveCode é ilustrativo. O sistema usado para o planejamento do artefato nesta pesquisa, LiveCode, abriu o código em 2013 mas encerrou as licenças de código aberto em 2020 alegando que

o projeto, depois de 8 anos, não reuniu uma comunidade de desenvolvedores significativa e os custos de desenvolvimento de uma versão comercial e uma versão FOSS eram muito altos. O código já publicado não pode mudar a licença, e foi copiado pela comunidade e colocado no site de desenvolvimento de software GITHUB (Miller, 2021). No entanto, até 2023 não havia reunido voluntários que pudessem entender a complexidade do sistema. O LiveCode é um sistema multi-plataforma muito complexo: o mesmo código pode ser montado para ser instalado nos sistemas operacionais Windows, MacOS, Linux, iPhone e Android. É possível que o projeto LiveCode FOSS tenha continuidade, mas para futuro desenvolvimento do projeto VidLattes achamos mais seguro usar outra ferramenta com grande comunidade de desenvolvedores, como *Free Pascal* com *IDE Lazarus* ou Python com interface gráfica. Isto, além de garantir a continuidade do projeto VidLattes, poderá facilitar a participação de mais desenvolvedores, já que se usará uma linguagem com centenas de milhares de programadores.

Foi um aprendizado o aprofundamento dos conhecimentos sobre representação visual de dados numéricos. Desde o que *não fazer* em gráficos de dados até o processo de escolha do tipo de gráfico correto.

Também foi um grande aprendizado estudar e conhecer mais profundamente a CAPES e o Sistema Brasileiro de Pós-Graduação e as questões emergentes com a aplicação da autoavaliação nos programas de pós-graduação *stricto sensu*.

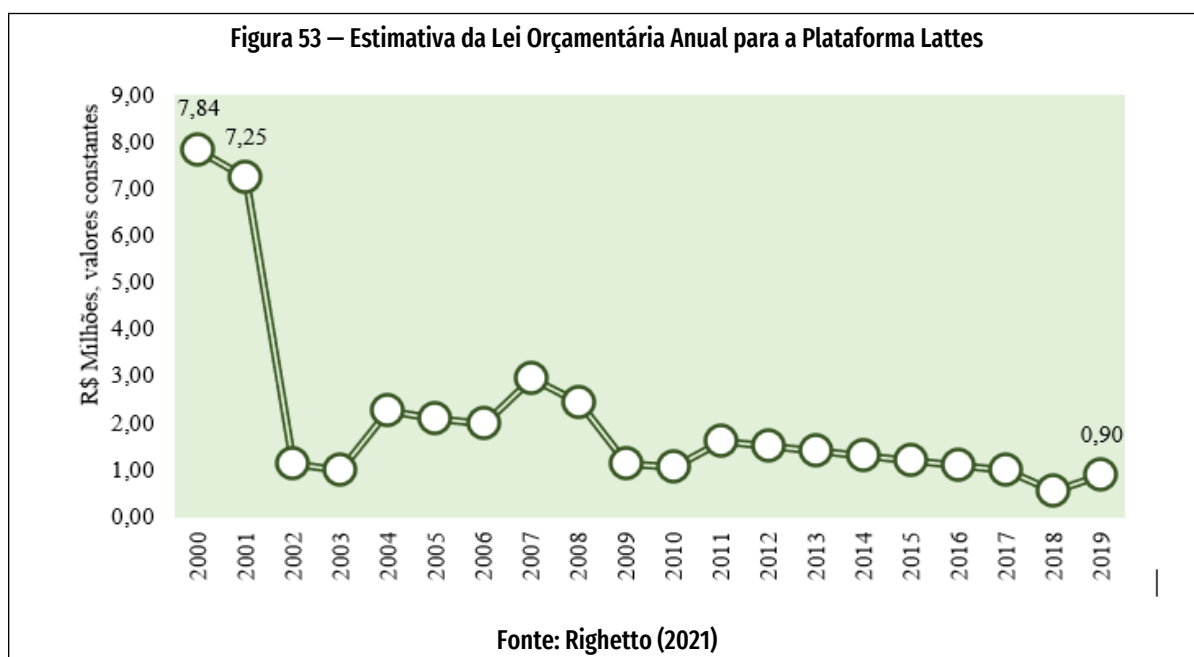
Construído pela comunidade de pesquisadores, o banco de dados Plataforma Lattes é uma iniciativa praticamente inédita no mundo, contando, em 2023, com mais de 7 milhões de pesquisadores registrados.

Ao chegar a este ponto, a plataforma Lattes tem alguns cenários pela frente. Aparentemente, o CNPq está tentando criar um ambiente diverso e ativo de artefatos que analise os dados das bases brasileiras de ciência e tecnologia. Neste sentido, organizou a EXPOLattes (CNPq, 2018), uma mostra de artefatos e sistemas que utilizam do banco de dados Lattes. O evento reuniu diversos desenvolvedores de

soluções para o problema da geração de conhecimento a partir dos dados extraídos do Lattes.

Se o Currículo Lattes se impôs com sucesso como referência científica brasileira, isto se deve em parte a seu ineditismo. Mas internacionalmente começam a aparecer iniciativas similares, tanto sem fins lucrativos.— como o DOI (*Digital Object Identifier*) e ORCID (*Open Research and Contributor ID*) — quanto redes sociais comerciais de pesquisadores, como Academia.Edu e ResearchGate.net.

Para fazer frente a concorrência, o Currículo Lattes precisa de muito mais apoio governamental, mas o orçamento anual na plataforma (Figura 53, página 146) aparentemente sugere baixo interesse pelo projeto.



6 Considerações finais

Este capítulo fecha a pesquisa, e está organizado em dois tópicos: “6.1 Conclusões”, apresentando a confirmação da hipótese apresentada na subseção “1.3 Hipótese”; e “6.2 Sugestões para trabalhos futuros”, que sugere caminhos futuros para o desenvolvimento do artefato.

6.1 Conclusões

Aplicada a teoria no objeto desta pesquisa, chegamos à conclusão de que o artefato cumpriu parte do que se propôs.

A autoavaliação é um dos itens que fazem parte do processo de avaliação dos programas de pós-graduação pela CAPES. Parte importante da autoavaliação é o levantamento da produção bibliográfica da produção docente e discente. Da forma em que está estruturado hoje, o banco de dados Plataforma Lattes se posiciona como uma fonte de dados inédita no mundo. No entanto, a consulta e extração de dados do Lattes em *batch* (em batelada) é complicada e demorada se for feita manualmente.

Para resolver problemas como estes, propusemos um artefato para fazer a contagem da produção acadêmica automaticamente e apresentar o resultado em diversos estilos de gráficos. Para os requisitos da autoavaliação, buscou-se resgatar a história da CAPES e da consolidação do ensino superior no Brasil. A partir deste ponto, definiu-se um objetivo principal e objetivos específicos.

A fim de se descobrir quais os tipos de dados e de gráficos mais adequados para a tarefa, foi feita pesquisa bibliográfica sobre Arquitetura da Informação, Design da Informação, software livre e sobre a política de dados públicos abertos adotada no Brasil.

A busca por artefatos similares e a entrevista com pesquisadores experientes em coordenação de PPG permitiu se ter uma visão clara de que as soluções para as necessidades de um programa de pós-graduação em Design podem ser aplicadas em programa de pós-graduação de qualquer área do conhecimento. O artefato tem mecanismos que permitem a personalização da produção acadêmica a ser analisada, conforme as exigências da CAPES e a necessidade do curso. Por exemplo, um PPG em Design pode considerar que um item da produção chamado “patente” deve fazer parte dos relatórios, enquanto um item de avaliação chamado “partitura” não faz sentido ao programa. No entanto, um PPG de Artes pode considerar “partitura” relevante, enquanto “patente” não faz sentido. Essa personalização vem ao encontro do que prega a *Design Science Research*: uma solução particular pode ser generalizada para resolução de outros problemas que tenham alguma semelhança entre si.

6.2 Sugestões para trabalhos futuros

A elaboração desta pesquisa sobre a implementação de gráficos sobre a produção acadêmica levou à identificação de novas questões que merecem estudo.

O primeiro ponto é a programação do módulo de construção de gráficos, que chegou a ser projetado mas não foi implementado em função da dinâmica da pesquisa. A busca *full text* também não foi implementada, e é uma das candidatas a integrar uma próxima versão do artefato.

Também se vê a necessidade do acesso e cruzamento de dados entre os bancos de dados Lattes com os dados das plataformas Sucupira, Qualis e CNPq. Principalmente se sentiu a necessidade de uma base de dados sobre a produção acadêmica que permita a referência padronizada a obras, sem diferenças de nomes de autores ou títulos. A ausência de uma base de dados de obras permite a duplicação, criando a necessidade de deduplicação.

Um passo adiante seria um módulo do artefato para qualificar as publicações do Qualis, aumentando a precisão do levantamento de produção do currículo Lattes, ao avaliar qualitativamente a produção acadêmica.

Outro ponto que pode ser estudado e implantado no VidLattes é o de rede de colaboradores, para mostrar visualmente, em um grafo, as relações entre pesquisadores através de coautorias.

Referências

- ABELA, Andrew. **Choosing a good chart**. [S. l.], 2009. Site de apresentação da metodologia. Disponível em: <https://extremepresentation.com/wp-content/uploads/choosing-a-good-chart-09-1.pdf>. Acesso em: 3 jun. 2020.
- ABREU, Lucas dos Santos. **Extractor Lattes CNPq**. [S. l.], 2020. Site de programação em grupo Github. Disponível em: <https://fabricadesoftwareuniville.github.io/ExtratorLattesCNPq/>. Acesso em: 6 out. 2018.
- BOEHM, Barry W. Guidelines for Verifying and Validating Software Requirements and Design Specifications. *In: EUROPEAN CONFERENCE ON APPLIED INFORMATION TECHNOLOGY OF THE INTERNATIONAL FEDERATION FOR INFORMATION PROCESSING, 1979, Londres. Anais [...]*. Londres: North Holland, 1979. Disponível em: <http://csse.usc.edu/TECHRPTS/1979/usccse79-501/usccse79-501.pdf>. Acesso em: 22 out. 2020.
- BOEHM, Barry W. Verifying and Validating Software Requirements and Design Specifications. **IEEE Software**, [s. l.], v. 1, n. 1, p. 75–88, 1984.
- BRASIL. Executivo Federal. Decreto nº 8.777. Institui a Política de Dados Abertos do Poder Executivo federal. **Código**, n. 8777, 11 maio 2016. Disponível em: http://www.planalto.gov.br/ccivil_03/_Ato2015-2018/2016/Decreto/D8777.htm. Acesso em: 9 set. 2018.
- BRASIL. **O que são dados abertos?**. [S. l.], [s. d.]. Portal de dados. Disponível em: <http://dados.gov.br/pagina/dados-abertos>. Acesso em: 9 set. 2018.
- CAPES, Coordenação de Aperfeiçoamento de Pessoal do Ensino Superior. **Autoavaliação de Programas de Pós-Graduação**. Brasília, DF: Coordenação de Aperfeiçoamento de Pessoal de Nível Superior, 2019. Relatório de Grupo de Trabalho. Disponível em: <https://www.gov.br/capes/pt-br/centrais-de-conteudo/10062019-autoavaliacao-de-programas-de-pos-graduacao-pdf>. Acesso em: 25 fev. 2022.

CAPES, Coordenação de Aperfeiçoamento de Pessoal do Ensino Superior. **GEOCAPES (1998-2021)**. [S. l.], 2021. Sítio de dados. Disponível em: <https://geocapes.capes.gov.br/geocapes/>. Acesso em: 18 out. 2018.

CAPES, Coordenação de Aperfeiçoamento de Pessoal do Ensino Superior. **Plano Diretor de Tecnologia da Informação 2017-2019**. [S. l.], 2017a. Portal de dados. Disponível em: <http://www.capes.gov.br/tecnologia-da-informacao/pdti-plano-diretor-de-tecnologia-da-informacao>. Acesso em: 8 set. 2018.

CAPES, Coordenação de Aperfeiçoamento de Pessoal do Ensino Superior. **Plano Nacional de Pós-Graduação (PNPG), 2011-2020, volume 1**. Brasília: CAPES, 2010. v. 1

CAPES, Coordenação de Aperfeiçoamento de Pessoal do Ensino Superior. **Proposta de Aprimoramento da Avaliação da Pós-Graduação Brasileira para o Quadriênio 2021-2024 - modelo multidimensional**. Brasília, DF: Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), 2020. Disponível em: <https://www.gov.br/capes/pt-br/centrais-de-conteudo/25052020-relatorio-final-2019-comissao-pnpg-pdf>. Acesso em: 22 jan. 2022.

CAPES, Coordenação de Aperfeiçoamento de Pessoal do Ensino Superior. **Relatório de avaliação 2013-2016 - quadriênio 2017 - Arquitetura, Urbanismo e Design**. Brasília, DF: CAPES, 2017b. Disponível em: <https://docs.google.com/viewer?a=v&pid=sites&srcid=Y2FwZXMuZ292LmJyfGF2YWxpYWVhby1xdWFkcmllbmFsfGd4OjQ4ZTEwNjgzN2VjYmE0ZTA>. Acesso em: 4 set. 2018.

CAPES, Coordenação de Aperfeiçoamento de Pessoal do Ensino Superior. **Relatório Final 2016 - Sumário Executivo**. Brasília, DF: CAPES, 2017c. Disponível em: <http://www.capes.gov.br/images/stories/download/relatorios/231117-Relatorio-PNPG-Final-2016-CS.pdf>. .

CAPES, Coordenação de Aperfeiçoamento de Pessoal do Ensino Superior. **Sobre a Avaliação**. Brasília, DF: [s. n.], 2018. Disponível em: <http://capes.gov.br/avaliacao/sobre-a-avaliacao>. Acesso em: 3 set. 2018.

CNPQ, Conselho Nacional de Desenvolvimento Científico e Tecnológico. ANAIS do I Seminário Avaliação de Políticas de CTI. *In:* , 1., 2018, Brasília/DF. **Seminário Avaliação de Políticas de CTI**. Brasília/DF: CNPq, 2018. Disponível em: <https://www.cgee.org.br/web/seminarioavaliacaociti/anais>. Acesso em: 4 fev. 2023.

COMUNIDADE CONSCIENTIAS-LMPL. [S. l.], [s. d.]. Disponível em: <http://lmpl.cnpq.br/lmpl/>. Acesso em: 6 out. 2018.

DB BROWSER FOR SQLITE. versão 3.10.99. [S. l.: s. n.], 2018. Multiplataforma: Windows, Macintosh, GNU/Linux. Disponível em: <http://sqlitebrowser.org/>.

DIGITAL OBJECT IDENTIFIER. *In*: WIKIPÉDIA, A ENCICLOPÉDIA LIVRE. [S. l.: s. n.], 2022. Disponível em: https://pt.wikipedia.org/w/index.php?title=Digital_object_identifier&oldid=63209667. Acesso em: 30 jan. 2023.

DILLON, Andrew; TURNBULL, Don. **Information architecture**. *In*: ENCYCLOPEDIA OF LIBRARY AND INFORMATION SCIENCE. [S. l.]: Taylor & Francis, 2009. p. 1–9. Disponível em: <https://donturn.com/publications/Dillon-Turnbull-2005-Information-Architecture.pdf>. Acesso em: 17 set. 2018.

DRESCH, Aline; LACERDA, Daniel Pacheco; ANTUNES JR., José Antônio Valle. **Design Science Research: método de pesquisa para avanço da ciência e tecnologia**. 2015. ed. Porto Alegre: Bookman, 2014.

DTD INTRODUCTION. *In*: XML FILES. [s. d.]. Disponível em: <https://www.xmlfiles.com/dtd/dtd-intro/>. Acesso em: 7 out. 2018.

ENGELHARDT, Yuri *et al.* **On information design**. Ljubljana: The Museum of Architecture and Design: Društvo Pekinpah, 2016. Disponível em: <http://www.dlib.si/stream/URN:NBN:SI:doc-2N1R3N8A/8475a930-3883-43a1-b3d0-2a97bca66fea/PDF>. Acesso em: 4 ago. 2018.

FENTON, Steve. Pie Charts are Bad. *In*: STEVE FENTON. 17 abr. 2009. Disponível em: <https://www.stevefenton.co.uk/2009/04/pie-charts-are-bad/>. Acesso em: 2 nov. 2019.

FEW, Stephen. **Information dashboard design: the effective visual communication of data**. 1. ed. Beijing; Cambridge [MA]: O'Reilly, 2006a.

FEW, Stephen. **Information dashboard design: the effective visual communication of data**. 1st eded. Beijing; Cambridge [MA]: O'Reilly, 2006b.

FREITAS, Carla Maria Dal Sasso *et al.* Introdução à visualização de informações. **Revista de informática teórica e aplicada (RITA)**, [s. l.], v. 8, n. 2, p. 143–158, 2001.

FUNG, Kaiser. **The unspoken rules of visualisation (and when to break them)**. [S. l.], 2020. Disponível em: <https://datajournalism.com/read/longreads/the-unspoken-rules-of-visualisation-and-when-to-break-them>. Acesso em: 19 jun. 2020.

GERSHON, Nahum; EICK, Stephen G.; CARD, Stuart. Information Visualization. **interactions**, [s. l.], v. 5, n. 2, p. 9–15, 1998.

HARDIN, Malia *et al.* **Qual gráfico é o correto para você? Conte histórias impactantes com dados**. [S. l.]: Tableau Software, 2012. Disponível em: https://www.tableau.com/sites/default/files/whitepapers/which_chart_v6_ptb.pdf
f. Acesso em: 15 mar. 2019.

HENNUM, Erik *et al.* Design patterns for information architecture. **IBM developerWorks**, [s. l.], p. 12, 2005.

HEVNER, Alan R.; CHATTERJEE, Samir. **Design research in information systems: theory and practice**. New York; London: Springer, 2010. (Integrated series in information systems, v. v. 22).

HORN, Robert E. Information design: emergence of a new profession. In: JACOBSON, R. E. (ed.). **Information design**. Cambridge, MA: MIT Press, 1999. Disponível em: <http://steinhardtapps.es.its.nyu.edu/create/courses/2015/reading/Horn.pdf>.

IBGE, Coordenação de Trabalho e Rendimento. **Pesquisa Nacional por Amostra de Domicílios: Síntese de Indicadores**. Rio de Janeiro: IBGE, 2016. Relatório. Disponível em: <https://biblioteca.ibge.gov.br/visualizacao/livros/liv98887.pdf>. Acesso em: 19 out. 2018.

JOHANNESSON, Paul; PERJONS, Erik. **An Introduction to Design Science**. 1. ed. New York: Springer, 2014.

LIMA, Diogo de Faveri Pereira. **Quanto vale uma pós-graduação stricto-sensu no Brasil? Os efeitos do mestrado e doutorado na remuneração de seus egressos**. 2016. 137 f. Tese de doutorado - Fundação Getúlio Vargas. Escola Brasileira de Administração Pública e de Empresas. Centro de Formação Acadêmica e Pesquisa, Rio de Janeiro, 2016.

LOPES, Erick Luiz Fonsêca; BORGES, Eduardo Nunes. Research.NET Web: uma aplicação para análise e exibição de redes de colaboração acadêmica utilizando grafos dinâmicos. In: XIV ESCOLA REGIONAL DE BANCO DE DADOS, 2018, Rio Grande, RS, Brasil. **Anais [...]**. Rio Grande, RS, Brasil: Universidade Federal do Rio Grande (FURG), Centro de Ciências Computacionais, 2018. p. 127-. Disponível em: http://erbd2018.c3.furg.br/downloads/180168_1.pdf. Acesso em: 29 nov. 2022.

MACCARI, Emerson Antonio *et al.* Proposta de um modelo de gestão de programas de pós-graduação na área de Administração a partir dos sistemas de avaliação do Brasil (CAPES) e dos Estados Unidos (AACSB). **Revista de Administração**, [s. l.], v. 49, n. 2, p. 369–383, 2014.

MATIAS, Mesailde Souza de Oliveira. **Base referencial para o povoamento de repositórios institucionais: coleta automatizada de metadados da Plataforma Lattes**. 2015. 86 f. Dissertação de Mestrado em gestão de organizações e Sistemas Públicos - Universidade Federal de São Carlos, São Carlos, SP, 2015. Disponível em: <https://repositorio.ufscar.br/handle/ufscar/6932>. Acesso em: 13 dez. 2022.

MEDIANA (ESTATÍSTICA). *In*: WIKIPÉDIA, A ENCICLOPÉDIA LIVRE. [S. l.: s. n.], 2019. Disponível em: [https://pt.wikipedia.org/w/index.php?title=Mediana_\(estat%C3%ADstica\)&oldid=56838883](https://pt.wikipedia.org/w/index.php?title=Mediana_(estat%C3%ADstica)&oldid=56838883). Acesso em: 24 jun. 2020.

MENA-CHALCO, Jesús Pascual; CESAR JR., Roberto Marcondes. ScriptLattes: an open-source knowledge extraction system from the Lattes platform. **Journal of the Brazilian Computer Society**, [s. l.], v. 15, n. 4, p. 31–39, 2009.

MILLER, Kevin. LiveCode. *In*: | THE HOME OF LIVECODE OPEN SOURCE. 31 ago. 2021. Disponível em: <https://livecode.org/>. Acesso em: 9 maio 2023.

MORVILLE, Peter; ROSENFELD, Louis. **Information architecture for the World Wide Web**. 3. ed. Sebastopol, CA: O'Reilly, 2007.

NÃO SOU UM ROBÔ. **Revista Galileu**, [s. l.], p. 8–9, 2016.

O QUE É SOFTWARE LIVRE? [S. l.], 2018. Disponível em: <https://www.gnu.org/philosophy/free-sw.pt-br.html#content>. Acesso em: 19 out. 2018.

OLIVEIRA, Eder; BERMEJO, Paulo de Souza; KERN, Vinícius. GeraLattes: extração de informação gerencial de currículos de pesquisadores usando XML. [s. l.], 2004. Disponível em: https://www.researchgate.net/publication/237394025_GeraLattes_extracao_de_informacao_gerencial_de_curriculos_de_pesquisadores_usando_XML. Acesso em: 28 jan. 2023.

ORCID. *In*: WIKIPÉDIA, A ENCICLOPÉDIA LIVRE. [S. l.: s. n.], 2022. Disponível em: <https://pt.wikipedia.org/w/index.php?title=ORCID&oldid=64868442>. Acesso em: 30 jan. 2023.

PETTERSSON, Rune. **Information design: an introduction**. Amsterdam; Philadelphia: John Benjamins Pub. Co, 2003. (Document design companion series, v. 3).

PINO, Ingrid. **DataViz Basics: 3 de 4 - Guia de gráficos básicos**. [S. l.], 2017. Site empresarial da agência DP6. Disponível em: <https://blog.dp6.com.br/guia-de-graficos-basicos-dataviz-basics-3-de-4-c12e39c8f027>. Acesso em: 1 jun. 2020.

PINO, Ingrid; GALINDO, Renata. **Como criar dashboards eficientes**. [S. l.], 2017a. Divulgação de agência de marketing digital DP6. Disponível em: <https://blog.dp6.com.br/como-criar-dashboards-eficientes-dataviz-intermediate-2-de-4-f073b96f0d89>. Acesso em: 1 out. 2019.

PINO, Ingrid; GALINDO, Renata. **DataViz Intermediate: 3 de 4 - Guia de Gráficos Avançados**. [S. l.], 2017b. Site empresarial da agência DP6. Disponível em: <https://blog.dp6.com.br/guia-de-graficos-avancados-dataviz-intermediate-3-de-4-2c29fd0e7bc6>. Acesso em: 17 jun. 2020.

PRAKASH, Abhishek. **Linux is Running on Almost All of the Top 500 Supercomputers - It's FOSS**. [S. l.], 2019. Site de notícias sobre tecnologia. Disponível em: <https://itsfoss.com/linux-supercomputers-2017/>. Acesso em: 7 mar. 2019.

QUARTIL. *In*: WIKIPÉDIA, A ENCICLOPÉDIA LIVRE. [S. l.: s. n.], 2019. Disponível em: <https://pt.wikipedia.org/w/index.php?title=Quartil&oldid=55160166>. Acesso em: 24 jun. 2020.

RIGHETTO, Flavio. **A Plataforma Lattes corre o risco de se tornar um tecnossauero?**. [S. l.], 2021. Disponível em: <http://www.ipea.gov.br/cts/pt/central-de-conteudo/artigos/artigos/277-a-plataforma-lattes-corre-o-risco-de-se-tornar-um-tecnossauero>. Acesso em: 13 fev. 2023.

RÍOS, Cristina. Quality Assurance in Higher Education in Spain: An Overview of the Accreditation System. **International Research and Review**, [s. l.], v. 5, n. 1, p. 25–45, 2015.

RYAN, Tricia. Quality assurance in higher education: A review of literature. **Higher Learning Research Communications**, [s. l.], v. 5, n. 4, 2015. Disponível em: <https://scholarworks.waldenu.edu/hlrc/vol5/iss4/1/>. Acesso em: 28 mar. 2022.

SUCUPIRA, Newton *et al.* **Parecer CFE 977/1965**. Brasília, DF: CAPES, 1965. Disponível em: <https://www.gov.br/capes/pt-br/centrais-de-conteudo/parecer-cesu-977-1965-pdf>. .

THE 8 PRINCIPLES OF OPEN GOVERNMENT DATA. [S. l.], [s. d.]. Disponível em: <https://opengovdata.org/>. Acesso em: 18 set. 2018.

TUFTE, Edward R. **The Visual Display of Quantitative Information**. 2nd. ed. [S. l.]: Graphics Pr, 2001.

WEBER, Luc. Quality Assurance in Higher Education: A Comparison of Eight Systems. **Knowledge Brief**, [s. l.], v. 35, 2010. Disponível em: <http://www.worldbank.org/eca>. Acesso em: 23 abr. 2019.

WIERINGA, Roel. **Design science methodology for information systems and software engineering**. New York, NY: Springer Berlin Heidelberg, 2014.

ZELAZNY, Gene. **Say It With Charts: The Executive's Guide to Visual Communication**. 4. ed. New York: McGraw-Hill Education, 2001.