Requisitos de Privacidade no Ciclo de Vida do Software: Uma Análise Bibliométrica e de Redes de Colaboração Global e Brasileira

André Gheventer[1], Jonice Oliveira[1], Rafael Maiani de Mello[1], Silas Lima Filho[1]

[1] Universidade Federal do Rio de Janeiro (UFRJ), Rio de Janeiro, RJ - Brasil

agheventer@ufrj.br, {jonice, rafaelmello, silaslfilho}@dcc.ufrj.br

Abstract. *Context: Digital transformation and current regulations (GDPR/LGPD) demand integrating privacy across the Software Product Life Cycle (SPLC). Objectives: To investigate and compare global and Brazilian scientific production on data privacy within the SPLC. Method: Bibliometric and social network analysis applied to 496 SCOPUS database documents (2002-2025). Results: The US and Europe lead globally with cohesive networks; Brazil surges post-LGPD (2018), but its research network is highly fragmented into isolated clusters. 70% of the analyzed publications are concentrated in conference proceedings. Topically, global literature features discussions on AI, whereas Brazilian studies focus on empirical studies. Conclusions: The literature primarily addresses early development phases, suggesting a research gap in SPLC's late stages and highlighting the need for more integrated collaborative hubs in the Brazilian context.*

Resumo. *Contexto: A transformação digital e os regulamentos vigentes (GDPR/LGPD) exigem a integração da privacidade ao longo de todo o Ciclo de Vida do Produto de Software (SPLC). Objetivos: Investigar e comparar a produção científica global e brasileira sobre privacidade de dados no SPLC. Método: Análise bibliométrica e de redes sociais aplicada a 496 documentos da base de dados SCOPUS (2002-2025). Resultados: EUA e Europa lideram globalmente com redes coesas; o Brasil cresce pós-LGPD (2018), mas sua rede de pesquisa é altamente fragmentada em clusters isolados. 70% das publicações analisadas concentram-se em anais de conferências. Tematicamente, a literatura global evidencia discussões sobre IA, enquanto a nacional concentra-se em estudos empíricos. Conclusões: A literatura prioriza as fases iniciais do desenvolvimento, o que sugere uma lacuna de pesquisa nas etapas finais do SPLC e evidencia a necessidade de hubs colaborativos mais integrados no cenário nacional.*

1. Introdução

A evolução tecnológica e a digitalização acelerada dos serviços impulsionaram um crescimento substancial na coleta e no processamento de dados pessoais por sistemas de software [Saraiva et al. 2023]. Como consequência direta, os incidentes de segurança cibernética e os vazamentos de dados atingiram níveis significativos. Apenas nos primeiros seis meses de 2024, cerca de 1 bilhão de cidadãos foram afetados globalmente por violações de dados mantidos por prestadoras de serviços [ITRC 2024], um cenário

de risco contínuo que vem se agravando e gerando impactos severos também no Brasil. Como evidência dessa escalada, apenas em 2025, a Autoridade Nacional de Proteção de Dados (ANPD) registrou 395 comunicações oficiais de incidentes de segurança envolvendo dados pessoais [ANPD 2025]. Dentre eles, um incidente recente no Brasil afetou os sistemas operados pelo Conselho Nacional de Justiça (CNJ) e resultou no acesso indevido a informações pessoais de mais de 11 milhões de cidadãos brasileiros [CNJ 2025].

Em resposta a essas vulnerabilidades, consolidou-se um rigoroso arcabouço regulatório global, com destaque para a General Data Protection Regulation [GDPR 2016] na União Europeia e a Lei Geral de Proteção de Dados Pessoais [LGPD 2018] no Brasil. Tais legislações influenciam diretamente o design e a manutenção de sistemas de software para garantir a conformidade normativa, exigindo que a proteção de dados não seja apenas uma camada reativa de segurança, mas um princípio integrado desde a concepção do produto, o chamado "*Privacy by Design"*. No entanto, essas regulamentações diferem dos padrões tradicionais da engenharia de software por não prescreverem requisitos ou funcionalidades de implementação específicos e obrigatórios, delineando apenas princípios abrangentes que devem ser seguidos [Campanile et al. 2022]. Esse tipo de abordagem oferece flexibilidade para as equipes de desenvolvimento, ao mesmo tempo que garante que as considerações de privacidade e proteção sejam incorporadas ao longo do ciclo de vida de produto de software (SPLC - Software Product Life Cycle). De acordo com o SWEBOK (2024), o SPLC inclui todas as atividades necessárias para definir, construir, operar, manter e desativar um produto de software.

Apesar do avanço normativo, a literatura evidencia desafios relevantes na tradução desses princípios em práticas concretas. Não obstante o conhecimento dos profissionais envolvidos acerca da importância e das obrigações das medidas de proteção e privacidade de dados pessoais no SPLC, a complexidade inerente à legislação, aliada à ausência de modelos consolidados, gera incertezas na operacionalização de requisitos de privacidade pelos profissionais de software, limitando o campo de atuação e expondo as organizações aos riscos de penalidades legais [Shapiro, S.S. 2010, Canedo, E.D. et al. 2020, Cerqueira et al. 2023]. Além disso, a frequente sobreposição conceitual entre os conceitos de "segurança" e "privacidade" contribui para que requisitos de privacidade sejam negligenciados ou tratados de forma superficial, resultando na desatenção quanto ao entendimento e posterior aplicação na prática [Peixoto M. et al. 2023, Andrade V.C. et al. 2023]. Kalloniatis et al. (2008) destacam que muitos modelos ainda tratam os requisitos de privacidade como uma extensão dos requisitos de segurança em sistemas de software. Assim, esses modelos não fornecem técnicas específicas para identificar os requisitos de privacidade ou traduzi-los em componentes de sistemas de software, nem sugerem alternativas relevantes de implementação.

Embora a produção científica sobre o tema esteja em expansão, as pesquisas tendem a abordar a privacidade de forma isolada, carecendo de uma visão integrada que perpasse todas as fases do SPLC. Somado a essa lacuna temática, o expressivo volume de publicações dispersas dificulta a identificação clara de quais redes de colaboração, atores influentes e tendências tecnológicas estão de fato liderando o avanço da área. De acordo com a Conferência das Nações Unidas sobre Comércio e Desenvolvimento

[UNCTAD 2026], 158 dos 195 países-membros já possuem alguma regulamentação própria sobre o tema. Mapear essa complexa estrutura de conhecimento, contrastando a maturidade do estado da arte global com a evolução do cenário nacional, torna-se, portanto, um passo fundamental para direcionar a indústria e a academia. Diante desse contexto, o presente estudo é guiado pela seguinte Questão de Pesquisa Principal (QP1): *Como a produção científica mundial e brasileira tem respondido às demandas por privacidade de dados pessoais ao longo das etapas do Ciclo de Vida do Produto de Software (SPLC)?*

Para responder a esta questão, conduzimos um estudo bibliométrico comparativo, apoiado por uma análise de redes sociais, com base em documentos indexados na base de dados SCOPUS. Este trabalho diferencia-se ao investigar, de forma sistemática, padrões de produção científica, colaboração e evolução temática, contrastando o contexto internacional com o contexto brasileiro.

As principais contribuições deste artigo incluem:

1. *Mapeamento de Influência na Engenharia de Privacidade*: Identificação dos autores com maior influência estrutural na rede e dos países com maior centralidade nas redes de coautoria, evidenciando como marcos regulatórios se correlacionam com a formação dos grandes hubs de pesquisa globais e nacionais.
2. *Identificação de Veículos de Disseminação*: Levantamento dos eventos científicos e periódicos com maior concentração de publicações na área, revelando que a comunidade concentra 70% de suas publicações em anais de conferências.
3. *Descoberta de Tendências e Lacunas no SPLC*: Evidência, por meio de redes de coocorrência de palavras-chave, de que a literatura global e nacional concentra seus esforços quase exclusivamente nas fases iniciais do ciclo de vida, como "*Privacy by Design*" e "*Privacy Requirement*". Esse padrão revela uma lacuna crítica na pesquisa sobre as fases avançadas do SPLC (manutenção, operação e desativação). Adicionalmente, observa-se uma assimetria temática entre os contextos brasileiro e internacional quanto à presença de discussões sobre Inteligência Artificial na proteção de dados.

O restante deste artigo está estruturado da seguinte forma: inicialmente, na Seção 2, analisamos trabalhos relacionados para posicionar nossa contribuição; em seguida, a Seção 3 detalha a metodologia e o protocolo de coleta de dados; posteriormente, a Seção 4 apresenta os resultados das análises bibliométricas e de redes sociais; a Seção 5 discute as descobertas, interpretando seu significado e situando-as no contexto mais amplo da pesquisa; finalmente, a Seção 6 apresenta as conclusões, limitações e direções para trabalhos futuros.

2. Trabalhos Relacionados

Estudos recentes têm empregado análises bibliométricas e de redes sociais para mapear o panorama da segurança e da privacidade de dados em ecossistemas computacionais. Em um escopo voltado à conscientização e à ciência da informação, Santanna et al. (2023) avaliaram publicações entre 1999 e 2021, identificando um aumento na produção científica sobre privacidade a partir de 2016, impulsionado pelas preocupações da sociedade com vazamentos e uso indevido de dados pessoais. Avaliando um período

semelhante, Muhammad et al. (2023) investigaram tendências em segurança cibernética de 2018 a 2023, apontando uma escassez de pesquisas focadas na privacidade em mídias sociais e na mitigação de danos aos usuários.

Ampliando a análise para sistemas computacionais em geral, Ali et al. (2024) conduziram uma revisão cobrindo o período de 1974 a 2020. Os autores observaram um crescimento anual médio de 34,1% nas publicações da área, destacando a liderança dos Estados Unidos (38,89% da produção) e a atuação do Brasil (1,36%) em pesquisas sobre privacidade de dados em sistemas computacionais. Sob uma ótica geopolítica e regulatória recente, Valero-Ancco et al. (2025) examinaram a literatura de 2014 a 2024, demonstrando que eventos como a implementação da GDPR geraram picos de produção científica. Direcionando o escopo especificamente ao desenvolvimento de sistemas, Conceição et al. (2026) analisaram publicações entre 2015 e 2026 com foco em segurança e gestão de riscos. Os resultados indicaram forte concentração nas áreas de Ciência da Computação e Engenharia de Software, com destaque para temas como modelos seguros de Ciclo de Vida de Desenvolvimento (SDLC) e decomposição de requisitos. Nesse mapeamento, os autores identificaram que os Estados Unidos mantêm a liderança global, enquanto o Brasil já se consolida como o principal contribuinte na América do Sul. Contudo, o estudo aponta que os requisitos de segurança ainda são definidos sem vínculos explícitos com riscos previamente identificados.

Embora a literatura supracitada ofereça um panorama valioso sobre privacidade e segurança em software, as revisões existentes tendem a adotar um escopo mais geral. O presente artigo preenche essa lacuna metodológica ao desenhar uma estratégia de busca abrangente, estruturada para capturar terminologias de todas as etapas do Ciclo de Vida do Produto de Software (SPLC), desde a concepção e elicitação, passando por testes e revisão de código, até a manutenção, envolvendo processos de software, profissionais de desenvolvimento e produtos de software. Por meio dessa captura sistemática de metadados, o estudo mapeia a distribuição da atenção da comunidade científica ao longo do SPLC. Adicionalmente, o estudo utiliza a visualização temporal do VOSviewer e a Análise de Redes Sociais para mapear a posição do Brasil no cenário científico global, destacando a evolução do ecossistema nacional impulsionada pela LGPD.

3. Metodologia

Para esse estudo aplicamos uma abordagem híbrida, utilizando métodos de análise bibliométrica e de redes sociais. O termo “contexto internacional” foi adotado para referenciar as publicações científicas de autores afiliados a instituições não brasileiras e o termo “contexto nacional” para referenciar as publicações científicas de autores afiliados a instituições brasileiras. Nossa pesquisa foi estruturada em cinco etapas distintas, baseadas, com adaptações, no protocolo proposto por Öztürk et al. (2024) para pesquisas bibliométricas. Para atender ao propósito do estudo, inserimos uma etapa de normalização e desambiguação dos dados, a fim de evitar problemas comuns em repositórios bibliográficos que podem comprometer a integridade e a validade dos dados analisados [Rodrigues et al. 2023].

Para guiar esse estudo e atender aos objetivos propostos, a Questão de Pesquisa Principal (QP1) foi desdobrada nas seguintes subquestões:

QP1.1. Quais países e autores apresentam maior centralidade e influência na rede, identificando os principais polos de produção e de referência científica na área?

QP1.2. Quais são os eventos e periódicos preferidos para a publicação de documentos de pesquisa?

QP1.3. Quais áreas de pesquisa estão ganhando destaque e atraindo a atenção da comunidade científica?

Para fins de reprodutibilidade, o repositório online do estudo preserva os dados brutos exportados da base de dados SCOPUS, a string de busca completa, o protocolo metodológico, os arquivos gerados pelo VOSviewer e as tabelas complementares com indicadores adicionais, disponíveis em https://zenodo.org/records/19340977.

3.1. Estratégia de busca e coleta de dados

A base de dados da SCOPUS foi escolhida como fonte de metadados para esta análise por adotar um processo rigoroso de curadoria, que inclui a seleção e a reavaliação contínua das publicações indexadas [Singh et al., 2021]. Adicionalmente, a SCOPUS indexa publicações de importantes editoras da área de computação, como IEEE e ACM, ampliando a cobertura de estudos relevantes [Almagribi et al. 2025], sendo utilizada por estudos na área de Engenharia de Software como base principal para a identificação de literatura e extração de dados bibliométricos (Cartaxo et al. 2018, Baldassarre et al. 2021, Silva et al. 2025).

A estratégia de busca foi elaborada com base no modelo PICo (Problema, Interesse e Contexto) [Hosseini, M. et al. 2024], a partir da definição de três conceitos principais: (P) privacidade / segurança de dados e conformidade legal; (I) etapas do SPLC; e (Co) processos, pessoas e produtos de software. Esses conceitos foram combinados por meio dos operadores booleanos “AND” e “OR”, o que permitiu a construção de uma string de busca abrangente e alinhada ao objetivo do estudo. A string de busca foi aplicada nos campos de título, resumo e palavras-chave. Como critérios de inclusão, foram consideradas publicações em inglês e em português. Como critério de exclusão, foram removidos registros classificados como “*conference review*”, por consistirem em resumos descritivos de eventos científicos, sem contribuição técnica direta. Não foram aplicadas restrições temporais, com o objetivo de capturar a evolução histórica da produção científica relacionada ao tema. Os metadados foram coletados em maio de 2025[1] por meio do Portal de Periódicos da CAPES, o que resultou em um total de 496 publicações: 34 do contexto nacional e 462 do contexto internacional. Os dados de replicação e a string de busca estão disponíveis no *repositório online do estudo*.

3.2. Processamento e visualização dos dados

Utilizamos o Microsoft Excel e os arquivos de “dicionário de sinônimos (*thesaurus files*)” do software VOSviewer (V1.6.20) para normalizar e desambiguar manualmente

[1]O fechamento da amostra bibliométrica ocorreu em maio de 2025. Referências posteriores a este período citadas ao longo do texto pertencem à literatura de fundamentação, não integrando o conjunto de 496 publicações. Todos os hiperlinks apresentados neste estudo foram verificados como ativos em 27 de março de 2026.

os dados de autores, palavras-chave, eventos e periódicos, além de gerar as tabelas e as ilustrações apresentadas neste estudo. Para desambiguar os nomes de autores, aplicamos o *Scopus Author ID* (SA ID) como critério de validação. Após o procedimento, foi possível identificar 41 dos 1.216 autores iniciais com nomes ambíguos ou até mesmo cadastrados em mais de um SA ID, e 42 palavras-chave entre 1.125. O software VOSviewer foi utilizado para construir e visualizar os mapas de colaboração bibliométrica apresentados nesse estudo.

4. Resultados

A Figura 1 ilustra a distribuição anual das publicações científicas coletadas. Embora o gráfico apresente uma redução no volume de publicações em 2025, esse dado reflete apenas o período de indexação incompleto até o momento da coleta, não configurando uma queda na produção científica. Os primeiros registros científicos identificados remontam a 2002 **[1]**. Nesse período, trabalhos seminais já destacavam a complexidade de traduzir textos legais em requisitos de software [Toval et al. 2002] e a necessidade de avaliar riscos de privacidade em componentes de código aberto [Hansen et al. 2002]. Desde essa época, autores como Iris et al. (2002) já sinalizavam que as proteções de privacidade deveriam ser especificadas nas fases iniciais de análise e de design do desenvolvimento.

A partir de 2016 **[2]**, houve um aumento significativo nas publicações internacionais sobre proteção de dados, passando de 18 em 2016 a um pico de 61 em 2021. Esse crescimento coincide com a promulgação do Regulamento Geral de Proteção de Dados da União Europeia (GDPR), aprovado em 2016 e implementado em 2018. A produção científica nacional sobre proteção de dados teve início em 2018 **[3]**, com as duas primeiras publicações identificadas. Esse marco coincide com a promulgação da Lei Geral de Proteção de Dados (LGPD – Lei nº 13.709/2018) no Brasil. Observa-se ainda um novo salto nas publicações em 2020, ano em que a LGPD entrou efetivamente em vigor, o que indica que a legislação também funcionou como um catalisador do interesse acadêmico no tema.

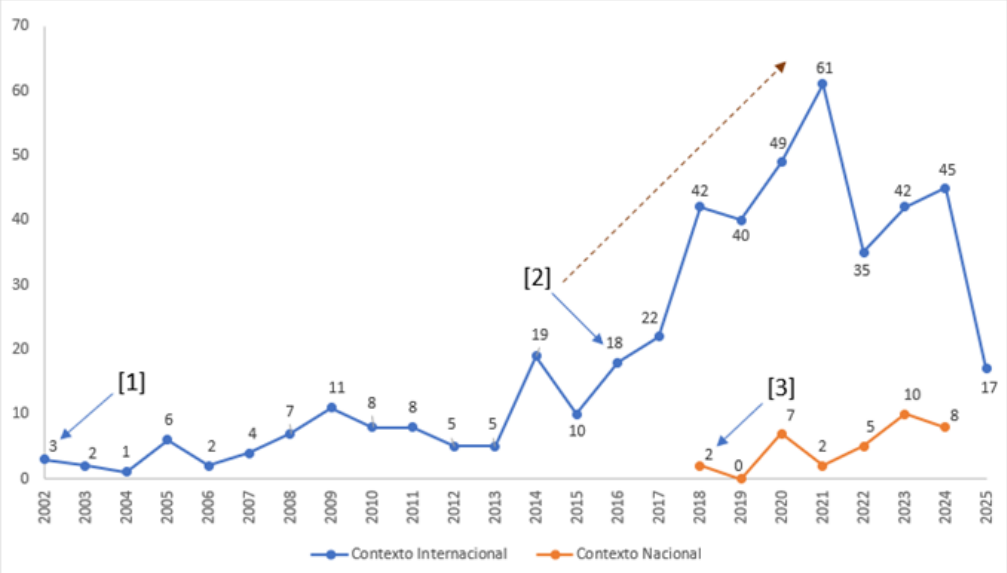

Figura 1. Produção científica do "contexto internacional" e do "contexto nacional".

Avançando para a análise de redes de colaboração e para a geração dos mapas, adotou-se, no software VOSviewer, o método de normalização *LinLog/modularity*, que enfatiza a estrutura dos *clusters* e a intensidade de suas relações na rede. Adicionalmente, utilizou-se o parâmetro *Weights: Total link strength* (TLS) para representar a força total das conexões. Ao todo, foram identificados 75 países que contribuíram com publicações científicas sobre a temática, totalizando 9.265 citações. Entretanto, apenas 51 países registraram colaborações internacionais em suas publicações no período analisado. A Figura 2 ilustra o mapa de coautoria dos países com as arestas (ou conexões) representando a relação de colaboração científica. Seis clusters foram identificados: vermelho (13 países), verde (12 países), azul (10 países), amarelo (8 países), lilás (6 países) e cinza (2 países).

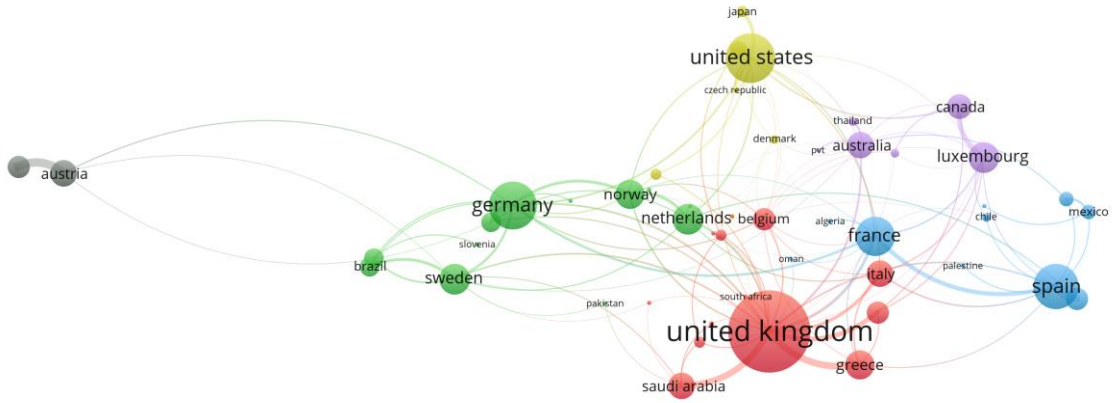

Figura 2. Mapa de coautoria de países

A Tabela 1 detalha os indicadores quantitativos de produção científica dos principais países contribuidores.

Tabela 1. Indicadores de produção científica por país

PAÍS	TLS
United kingdom	66
United states	31
Spain	31
France	26
Germany	23
Greece	21
Italy	20
Belgium	20
Sweden	20
Australia	18
TOTAL	276

PAÍS	QD
United states	64
Germany	57
United kingdom	56
Brazil	32
China	28
Italy	26
Greece	25
Australia	25
Spain	25
Canada	24
TOTAL	362

PAÍS	QC
United states	1.250
United kingdom	933
Germany	615
Canada	590
Greece	543
Netherlands	447
Australia	353
Spain	304
France	300
Italy	266
TOTAL	5.601

(TLS) Total Link Strength (QD) Quantidade de documentos (QC) Quantidade de citações

Para a rede de coocorrência de palavras-chave um mapa de visualização de sobreposição de tendência temporal foi gerado no VOSviewer adotando o parâmetro Score: Avg. pub. year para posicionar a palavra-chave dentro do tempo. Utilizando um gradiente de cores, onde tons azuis representam termos mais antigos e tons lilás indicam

termos mais recentes, é possível detectar mudanças no interesse de pesquisa e avaliar como determinados conceitos estão ganhando destaque.

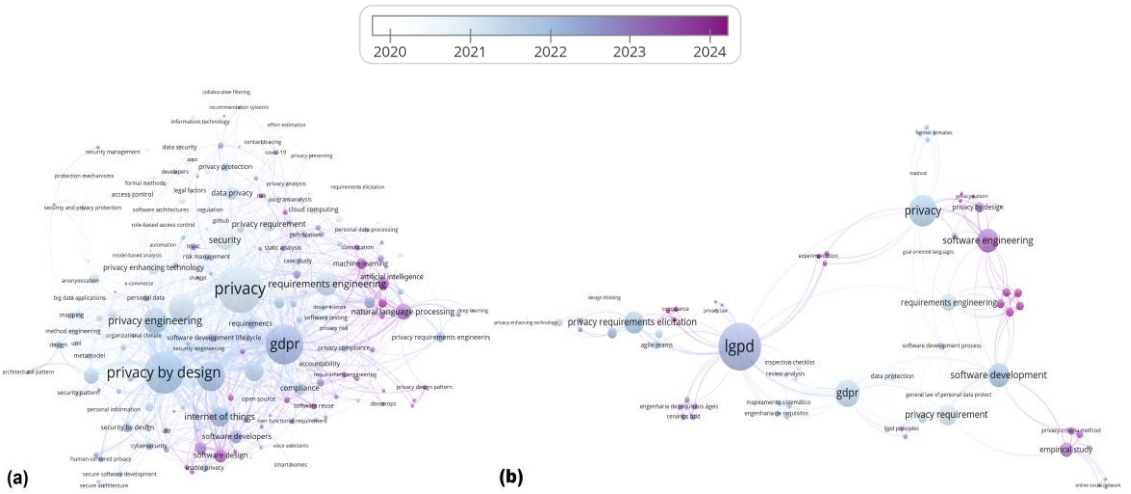

Figura 3. Mapa de coocorrência de palavras-chave.

A Tabela 2 apresenta as palavras-chave de maior força na rede e suas respectivas ocorrências em ambos os contextos.

Tabela 2. Palavras-chave com maior força na rede

Contexto Internacional					Contexto Nacional				
Palavras-chave	**TLS**	**%TLS**	**OC**	**%OC**	**Palavras-chave**	**TLS**	**% TLS**	**OC**	**%OC**
privacy	516	5,25	115	5,90	lgpd	32	8,04	8	6,84
privacy by design	389	3,96	83	4,26	gdpr	21	5,28	5	4,27
gdpr	359	3,65	76	3,90	privacy	20	5,03	7	5,98
data protection	208	2,12	39	2,00	privacy requirements elicitation	17	4,27	5	4,27
privacy engineering	178	1,81	40	2,05	privacy requirement	16	4,02	5	4,27
software engineering	176	1,79	33	1,69	empirical study	11	2,76	4	3,42
requirements engineering	140	1,42	33	1,69	software development	11	2,76	4	3,42
software development	122	1,24	24	1,23	software engineering	11	2,76	4	3,42
internet of things	118	1,20	22	1,13	data privacy	8	2,01	2	1,71
security	112	1,14	27	1,38	privacy by design	8	2,01	3	2,56
TOTAL	2.318	23,59	492	25,23	**TOTAL**	155	38,94%	47	40,17

(TLS) Total Link Strength (%TLS) Percentual de ocorrências dentre o total (OC) Quantidade de ocorrências (%OC) Percentual de ocorrências dentre o total

Para esta análise utilizamos apenas as palavras-chave dos autores. Também realizamos a desambiguação das palavras-chave para melhorar a qualidade e a precisão da análise, entretanto, neste caso focando apenas em variações como singular, plural, abreviações e formas com hífens, buscando assim evitar uma abordagem de entendimento com viés do pesquisador. No mapa do contexto internacional (Figura 3a) identificamos 982 palavras-chave, com 1.950 ocorrências e um TLS da rede de 9.828. Já no mapa do contexto nacional (Figura 3b) identificamos 70 palavras-chave, com 117 ocorrências e um TLS da rede de 398. As palavras-chave "*privacy*" e "*lgpd*" tiveram as maiores ocorrências, com TLS 516 e 32 respectivamente.

No contexto internacional, observa-se a incidência das palavras-chave "*natural language processing*", "*artificial intelligence*" e "*machine learning*", associadas a tonalidades mais recentes no mapa, indicando uma tendência de investigação voltada a tecnologias emergentes e à automação. Por outro lado, no contexto nacional, a palavra-chave "*empirical study*" também se destaca entre os termos mais recentes, sugerindo uma orientação da comunidade científica para pesquisas baseadas em evidências do mundo real.

Complementarmente à análise temática, investigou-se a estrutura de colaboração entre os autores por meio de redes de coautoria. A Figura 4a ilustra a rede de coautoria do contexto internacional, onde os autores estão afiliados em instituições não brasileiras. São 1.188 autores, 274 *clusters* e 2.341 arestas. Os maiores *clusters* são compostos por 51 autores (*cluster* 1 - vermelho), 29 autores (*cluster* verde) e 25 autores (*cluster* azul). Já a rede de colaboração do contexto nacional, entre os autores afiliados em instituições brasileiras é demonstrada na Figura 4b. A rede conta com um total de 69 autores, 14 clusters e 138 arestas. Os maiores *clusters* são compostos por 17 autores (*cluster* 1 - vermelho), 11 autores (*cluster* 2 - verde) e 7 autores (*cluster* 3 - azul).

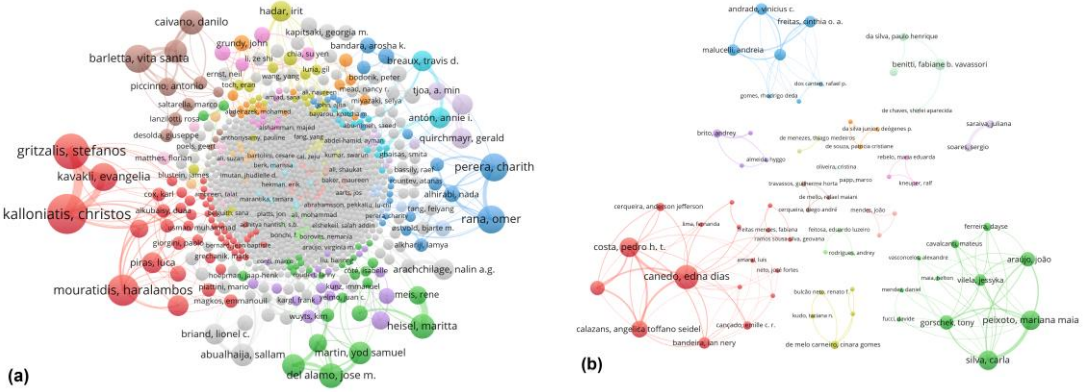

Figura 4. Redes de coautoria.

A Tabela 3 apresenta os autores com maior força na rede em ambos os contextos.

Tabela 3. Autores com maior força na rede

Contexto internacional	TLS	QD	QC
Mouratidis, Haralambos	58	12	158
Kalloniatis, Christos	41	17	449
Pavlidis, Michalis	41	6	80
Gritzalis, Stefanos	38	15	448
Piras, Luca	38	6	75
Perera, Charith	35	10	234
Barletta, Vita Santa	33	10	97
Rana, Omer	32	10	112
Al-obeidallah, Mohammed Ghazi	28	4	55
Magkos, Emmanouil	28	3	65
TOTAL	372	93	1.773

Contexto nacional	TLS	QD	QC
Canedo, Edna Dias	32	8	66
Costa, Pedro H. T.	27	6	65
Peixoto, Mariana Maia	25	7	48
Silva, Carla	25	6	45
Calazans, Angelica Toffano S.	21	5	64
Gorschek, Tony	21	4	25
Araújo, João	20	4	30
Vilela, Jessyka	17	3	25
Masson, Eloisa Toffano Seidel	16	4	54
Bandeira, Ian Nery	15	3	17
TOTAL	219	50	439

(TLS) Total Link Strength (QD) Quantidade de documentos publicados (QC) Quantidade de citações

Para aprofundar essa análise visual, extraíram-se métricas quantitativas de Análise de Redes Sociais, especificamente a densidade da rede (d) e o grau médio de

colaboração (k). Utilizando as fórmulas padrão para grafos não-direcionados $d = 2L / [N(N-1)]$ e $k = 2L / N$, baseadas no número de nós (***N***) e de arestas (***L***), identificou-se que a rede do contexto nacional apresenta uma densidade de 0,0588 (5,88%) e um grau médio de 4,0 coautores. Em contrapartida, a rede do contexto internacional registra uma densidade de 0,0033 (0,33%) e um grau médio de colaboração de 3,94.

A comparação direta dessas densidades, contudo, exige cautela metodológica: redes de tamanhos muito distintos — 69 autores no contexto nacional contra 1.188 no contexto internacional — tendem naturalmente a produzir densidades mais elevadas nas menores, uma vez que o número de conexões possíveis cresce proporcionalmente ao quadrado do número de autores [Wasserman e Faust, 1994]. Nesse sentido, o grau médio de colaboração oferece uma métrica mais comparável entre os dois cenários. Os valores praticamente idênticos (4,0 vs. 3,94) sugerem que os pesquisadores tendem a manter, em média, cerca de quatro conexões de coautoria na rede, independentemente da geografia. A grande diferença estrutural reside, portanto, não na intensidade da colaboração local, mas na sua distribuição: enquanto o contexto internacional organiza sua produção em torno de um núcleo principal coeso, a rede nacional fragmenta esse mesmo esforço em 14 clusters desconexos. Essa configuração estrutural sugere que a rede nacional se organiza em ilhas institucionais desconexas, com escassa presença de autores que conectem os diferentes clusters identificados.

Por fim, no que se refere aos veículos de publicação, os dados extraídos apontam para uma preferência por submissões a conferências, tanto na comunidade internacional quanto na nacional, com cerca de 70% das publicações no total. Das 296 publicações em conferências do cenário internacional, 43 (14%) concentraram-se no "*International Requirements Engineering Conference (RE)*" e no "*International Conference on Software Engineering (ICSE)*". No contexto nacional, das 23 publicações em conferências, o "*Brazilian Symposium on Software Engineering (SBES)*" obteve a preferência, com 6 publicações (26%).

5. Discussão

Países como a Alemanha, os Estados Unidos e o Reino Unido são identificados como hubs de distribuição de conhecimento, com alta centralidade e influência evidenciada pelo volume de citações recebidas **(QP1.1)**. Esses três países não apenas produziram uma quantidade substancial de publicações, mas também receberam 33% do total de todas as citações, sugerindo que as suas pesquisas são amplamente reconhecidas e valorizadas. Apesar de não figurar entre os 10 países com publicações mais citadas, o Brasil se destaca como o 4º maior produtor de publicações, demonstrando um significativo potencial de contribuição para o avanço científico, ainda que com desafios para ampliar sua visibilidade internacional. A Espanha e a França também aparecem entre os quatro países com mais TLS, o que demonstra que possuem uma rede de colaboração ampla e forte com outros países, bem como relevância no cenário científico internacional.

Na análise da rede de coautoria do contexto internacional, observa-se a formação de um grande componente principal denso. A estrutura revela uma topologia centro-periferia, onde um núcleo central atua na conectividade entre diferentes clusters. Essa configuração favorece a aproximação de hubs influentes, como Haralambos Mouratidis

(TLS 58), Christos Kalloniatis e Michalis Pavlidis (ambos com TLS 41). A centralidade desse grupo é impulsionada por trabalhos focados na integração de privacidade nas fases iniciais do SPLC, como o método PriS (Kalloniatis et al., 2008) e a plataforma DEFeND Architecture (Piras et al., 2019), desenhada para apoiar a conformidade técnica e organizacional exigida pela GDPR. Em contraste, a rede de coautoria do contexto nacional apresenta uma topologia fragmentada, caracterizada por múltiplos componentes desconexos. Essa configuração estrutural sugere que a pesquisa brasileira no tema se organiza em pequenos grupos, desprovida de um núcleo central aglutinador com autores atuando como nós-ponte para fomentar a colaboração interinstitucional. Além disso, embora a rede nacional apresente uma densidade matematicamente maior que a internacional (5,88% vs 0,33%), o grau médio (4,0) diluído em 14 clusters para apenas 69 autores sugere que a colaboração é intensa, porém restrita. Apesar dessa dispersão, a professora Edna Dias Canedo consolida-se como o principal nó de influência, com a maior centralidade (TLS 32). Seu trabalho de maior impacto (Canedo et al., 2020) atua na interseção entre a academia e a prática, revelando que os profissionais de TIC reconhecem o impacto da LGPD, mas relatam déficit prático para implementar o "*Privacy by Design*".

A forte concentração de estudos em anais de eventos ratifica as conferências como o principal veículo de disseminação científica sobre este tema (**QP1.2**). Destaca-se que RE e ICSE configuram-se como os principais eventos internacionais com foco em engenharia de requisitos e de software. Em paralelo, o SBES consolida-se como o principal evento de engenharia de software na América Latina, promovido pela Sociedade Brasileira de Computação (SBC) em conjunto com o Congresso Brasileiro de Software (CBSoft). As análises também apontam para direções temáticas divergentes entre as comunidades (**QP1.3**). Conforme demonstrado na evolução temporal das palavras-chave, enquanto o cenário internacional apresenta uma crescente coocorrência de termos associados a tecnologias emergentes (como Inteligência Artificial e Processamento de Linguagem Natural), o que indica um interesse exploratório nessas interseções, o contexto nacional concentra esforços em diagnósticos empíricos e na adaptação prática das organizações às regulamentações vigentes, como a LGPD.

6. Conclusão

Este estudo investigou como a produção científica mundial e brasileira tem respondido às demandas por privacidade de dados pessoais ao longo das etapas do Ciclo de Vida do Produto de Software (SPLC), por meio de uma abordagem bibliométrica e de análise de redes sociais. Os resultados apontam que o cenário internacional apresenta maior influência e atuação no tema, enquanto o Brasil se destaca como um relevante produtor científico latino-americano, embora ainda enfrente desafios relacionados à visibilidade e à integração em redes internacionais de colaboração.

No que se refere aos veículos de disseminação, observou-se que cerca de 70% das publicações concentram-se em anais de conferências. Além disso, tanto no contexto internacional quanto no nacional, as investigações concentram-se majoritariamente nas fases iniciais do desenvolvimento de software, com ênfase na elicitação e na especificação de requisitos de privacidade, frequentemente orientadas por demandas regulatórias, como a GDPR e a LGPD. Entretanto, os resultados também revelam lacunas importantes. A estratégia de busca foi deliberadamente estruturada para capturar

terminologias de todas as etapas do SPLC — da concepção e elicitação à manutenção e desativação. O fato da produção científica identificada concentrar-se nas fases iniciais, evidenciado tanto pela predominância de palavras-chave como "*Privacy by Design*" e "*Privacy Requirements*", quanto pela preferência por eventos especializados em requisitos e design, constitui um indício metodologicamente sustentado da limitada integração da privacidade ao longo de todo o SPLC, bem como da ausência de abordagens sistemáticas que conectem os requisitos de privacidade à gestão de riscos.

Adicionalmente, identificou-se uma assimetria entre os contextos analisados, na qual o cenário internacional já incorpora discussões emergentes envolvendo a inteligência artificial e sua interseção com a privacidade de dados, enquanto o contexto nacional ainda se concentra em estudos empíricos e na adaptação às exigências regulatórias. Esses achados reforçam a necessidade de abordagens mais integradas, que considerem a privacidade de forma transversal ao longo de todo o SPLC, bem como a ampliação da participação brasileira em redes globais de pesquisa.

Apesar das contribuições, este estudo apresenta limitações, como a utilização de uma única base de dados, possíveis vieses no processo manual de desambiguação e a restrição do VOSviewer quanto à completude das métricas topológicas de redes. Como trabalhos futuros, destaca-se a necessidade de investigações que explorem a integração da privacidade ao longo de todo o SPLC, bem como o uso de tecnologias emergentes, como a inteligência artificial, para apoiar a identificação e especificação de requisitos de privacidade. Recomenda-se, ainda, a ampliação das fontes de dados e a importação dos dados compartilhados em plataformas especializadas em análise de redes sociais, permitindo a extração, a análise e a comparação por meio de métricas complementares.

Agradecimentos

Os autores agradecem ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), à Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) — particularmente por meio do programa PROEXT-PG — e à Fundação Carlos Chagas Filho de Amparo à Pesquisa do Estado do Rio de Janeiro (FAPERJ), processo nº E-26/210.936/2024.

Referencias

Ali, A.S., Zaaba, Z.F., and Singh, M.M. (2024). The rise of “security and privacy”: bibliometric analysis of computer privacy research. International Journal of Information Security.

Almagribi, A. B., Ardianto, F., Taufan, A., and Kristomo, D. (2025). How is Software Engineering Linked to Business? A Scopus-Based Bibliometric and Visualization. The Indonesian Journal of Computer Science, 14(1).

Andrade V.C., Freitas, C.O.A, Reinehr S., and Malucelli A. (2023). Personal Data Privacy in Software Development Processes: A Practitioner’s Point of View. IEEE 22nd International Conference on Trust, Security and Privacy in Computing and Communications (TrustCom).

ANPD (2025). Autoridade Nacional de Proteção de Dados. Comunicado de Incidente de Segurança (CIS). https://www.gov.br/anpd/pt-br/canais_atendimento/agente-de-tratamento/comunicado-de-incidente-de-seguranca-cis.

Baldassarre, M.T., Caivano, D., Dimauro, G., Romano, S., and Scanniello, G. (2021). On internet of-things devices in ambient assisted living solutions. In: International Conference on Information Systems Development.

Canedo, E.D., Calazans, A.T.S., Masson, E.T.S., Costa, P.H.T., and Lima, F. (2020). Perceptions of ICT practitioners regarding software privacy. Entropy.

Cerqueira, D.A., Mello, R.M., and Travassos, G.H. (2023). Um checklist para inspeção de privacidade e proteção de dados pessoais em artefatos de software. CIbSE 2023 - XXVI Ibero-American Conference on Software Engineering.

Campanile, L., Iacono, M., and Mastroianni, M. (2022). Towards privacy-aware software design in small and medium enterprises. IEEE DASC/PiCom/CBDCom/CyberSciTech.

Cartaxo, B., Pinto, G., and Soares, S.(2018). The role of rapid reviews in supporting decision making in software engineering practice. In: Proceedings of the 22nd International Conference on Evaluation and Assessment in Software Engineering 2018 (EASE '18), pp. 24–34. Association for Computing Machinery (ACM).

CNJ. Conselho Nacional de Justiça. Comunicado de Incidente de Segurança. (2025). Disponível em: https://www.cnj.jus.br/comunicado-de-incidente-de-seguranca/.

Conceição, F., Dias Lousã, M. J., and Pereira de Morais, J. C. (2026). Security and Risk in Software Development Projects: A Bibliometric Review.

GDPR. (2016). General Data Protection Regulation. https://gdpr-info.eu/.

Hansen, M., Köhntopp, K., and Pfitzmann, A. (2002). The open source approach - Opportunities and limitations with respect to security and privacy. Computers and SecurityVolume 21, Issue 5, Pages 461 – 471.

Hosseini, M., Jahanshahlou F., Akbarzadeh M.A., Zarei M., and Vaez-Gharamaleki Y. (2024). Formulating research questions for evidence-based studies. Journal of Medicine, Surgery, and Public Health.

Iris, R., Dov, D., and Shmuel, K. (2002). OPM/Web - Object-process methodology for developing Web applications. Annals of Software EngineeringVolume 13, Issue 1-4, Pages 141 - 161.

ITRC. (2024). Identity Theft Resource Center. ITRC H1 Data Breach Analysis. https://www.idtheftcenter.org/publication/itrc-h1-data-breach-analysis/.

Kalloniatis, C., Kavakli, E., and Stefanos, G. (2008). Addressing privacy requirements in system design: The PriS method. Requirements Engineering.

LGPD (2018), Presidência da República. Lei Geral de Proteção de Dados (LGPD - Lei nº 13.709, de 14 de agosto de 2018). https://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/l13709.htm.

Muhammad, G., Pratama, A.R., Shaloom, C., and Cassandra, C. (2023). Cybersecurity Awareness Literature Review: A Bibliometric Analysis. 2023 International Conference on Informatics, Multimedia, Cyber and Informations System (ICIMCIS).

Öztürk, O., Kocaman, R., and Kanbach, D. (2024). How to design bibliometric research: an overview and a framework proposal. Review of Managerial Science. 10.1007/s11846-024-00738-0.

Peixoto M., Ferreira D., Cavalcanti M., Silva C., Vilela J., Araújo J., and Gorschek T (2023). The perspective of Brazilian software developers on data privacy. The Journal of Systems \& Software

Piras, L., Al-Obeidallah, M. G., Praitano, A., Tsohou, A., Mouratidis, H., Gallego-Nicasio Crespo, B., ... & Zorzino, G. G. (2019). DEFeND architecture: a privacy by design platform for GDPR compliance. In International conference on trust and privacy in digital business (pp. 78-93). Cham: Springer International Publishing.

Rodrigues, N.S, Mariano, A.M, and Ralha, C.G. (2023). Author name disambiguation literature review with consolidated meta-analytic approach. International Journal on Digital Libraries.

Santanna, J., Weber, C., Prado, J.M.K., and Ardigo, J.D. (2023). A questão da privacidade no regime de informação contemporâneo no contexto da Ciência da Informação. Revista Ibero-Americana de Ciência da Informação.

Saraiva, J., Soares, S. (2023). Privacy and Security documents for Agile Software Engineering: An experiment of LGPD Inventory adoption. 2023 ACM/IEEE International Symposium on Empirical Software Engineering and Measurement (ESEM).

Shapiro, S.S. (2010). Privacy by design: moving from art to practice. Communications of the ACM v. 53, n. 6, p. 27–29.

Silva, D. G., Countinho, C., and Costa, C. J. (2025). A Bibliometric Analysis of Free Open-Source Software Adoption (2001-2023). Procedia Computer Science, 263, 1-8.

Singh, V.K., Singh, P., Karmakar M., Leta J., and Mayr P. (2021). The journal coverage of Web of Science, Scopus and Dimensions: A comparative analysis. Scientometrics.

SWEBOK (2024). Guide to the Software Engineering Body of Knowledge v4.0. IEEE Computer Society.

Toval, A., Olmos, A., and Piattini, M. (2002). Legal Requirements Reuse: A Critical Success Factor for Requirements Quality and Personal Data Protection. Proceedings of the IEEE Joint International Conference on Requirements Engineering (RE'02).

UNCTAD (2026). United Nations Conference on Trade and Development. Data protection and privacy legislation worldwide. https://unctad.org/page/data-protection-and-privacy-legislation-worldwide.

Valero-Ancco, V. N., Lujano-Ortega, Y., Calderon-Quino, K. M., Gutierrez, F. S., Pari-Orihuela, M., and Bustinza-Choquehuanca, S. (2025). Personal Data Protection in the Era of Digital Surveillance: A Bibliometric Analysis of Scientific Production (2014–2024). Revista Electrónica de Ciencia Penal y Criminología, 27(1).

Wasserman, S. and Faust, K. (1994). Social Network Analysis: Methods and Applications. Cambridge University Press, Cambridge. ISBN 0521387078.