Universidade de Brasília
Instituto de Ciências Exatas
Departamento de Ciência da Computação

Gestão de Riscos no Direito Fundamental à Privacidade de Dados Pessoais no Processo Judicial Eletrônico / Diário de Justiça Eletrônico

Frank Ned Santa Cruz de Oliveira

Brasília
Março de 2020

Universidade de Brasília
Instituto de Ciências Exatas
Departamento de Ciência da Computação

Gestão de Riscos no Direito Fundamental à Privacidade de Dados Pessoais no Processo Judicial Eletrônico / Diário de Justiça Eletrônico

Frank Ned Santa Cruz de Oliveira

Dissertação apresentada como requisito parcial para conclusão do
Mestrado Profissional em Computação Aplicada

Orientador
Prof. Dr. João Mello da Silva

Coorientador
Prof. Dr. Edgard Costa Oliveira

Brasília
2020

REFERÊNCIA BIBLIOGRÁFICA E CATALOGAÇÃO

OLIVEIRA, Frank Ned Santa Cruz de. Gestão de Riscos no Direito Fundamental à Privacidade de Dados Pessoais no Processo Judicial Eletrônico / Diário de Justiça Eletrônico. Departamento de ciência da computação, Universidade de Brasília, 2020, 137f. Dissertação de mestrado em computação aplicada.

Ficha catalográfica elaborada automaticamente,
com os dados fornecidos pelo(a) autor(a)

SS231g

SANTA CRUZ DE OLIVEIRA, FRANK NED
SANTA CRUZ DE OLIVEIRA, FRANK NED Gestão de Riscos no Direito Fundamental à Privacidade de Dados Pessoais no Processo Judicial Eletrônico / Diário de Justiça Eletrônico / FRANK NED SANTA CRUZ DE OLIVEIRA;
Orientador João Mello da Silva Co-orientador Edgard Costa Oliveira. -- Brasília, 2020.
137 p.

Dissertação (Mestrado - Mestrado Profissional em Computação Aplicada) -- Universidade de Brasília, 2020.

1. mineração de dados. 2. CRISP-DM. 3. ISO 27701. 4. estrutura de privacidade. 5. TEMAC.
I. Mello da Silva, João, orient. II. Costa Oliveira, Edgard, co-orient. III. Título.

Universidade de Brasília
Instituto de Ciências Exatas
Departamento de Ciência da Computação

Gestão de Riscos no Direito Fundamental à Privacidade de Dados Pessoais no Processo Judicial Eletrônico / Diário de Justiça Eletrônico

Frank Ned Santa Cruz de Oliveira

Dissertação apresentada como requisito parcial para conclusão do
Mestrado Profissional em Computação Aplicada

Prof. Dr. João Mello da Silva (Orientador)
FT/UnB

Prof. Dr. Edgard Costa Oliveira (Coorientador)
FT/UnB

Prof. Dr. Sergio da Costa Côrtes
IESB

Prof. Dr. Ari Melo Mariano
FT/UNB

Prof.ª Dr.ª Aletéia Patrícia Favacho de Araújo
Coordenadora do Programa de Pós-graduação em Computação Aplicada

Brasília, 19 de março de 2020

DEDICATÓRIA

Aos meus pais Lenita e Joaquim. Meus exemplos de amor incondicional, suporte, lealdade, honestidade e caráter. Muitas vezes fizeram escolhas em detrimento de suas vontades, mas pensando nos estudos dos filhos, que somente vim a perceber e compreender depois de adulto. À vocês, meus queridos e amados pais, meu mais profundo amor e gratidão!!!

Aos meus sobrinhos Ana Flor e Benício Moreno (rãzinha) alegria dos meus olhos e felicidade do meu coração.

Aos colegas operadores do Direito, em sua luta incansável. Como ressalta Rudolf Von Ihering, "Enquanto o direito tiver de rechaçar o ataque causado pela injustiça – e isso durará enquanto o mundo estiver de pé – ele não será poupado […] O direito não é mero pensamento, mas sim força viva".

AGRADECIMENTOS

Gostaria de expressar minha gratidão para com muitas pessoas que forneceram inspiração, ideias, críticas e toda a sorte de ajuda para esta pesquisa, assim, quero agradecer em particular à:

Minha mãe, Lenita Santa Cruz de Oliveira, que desde sempre, pacientemente me incentivou, acolheu em seus braços e sempre será meu porto seguro;

Meu pai, Joaquim Rosa de Oliveira, também um porto seguro, silenciosamente cuidando de todos nós! Não conheço ninguém mais inteligente que ele;

Meu orientador, Prof. Dr. João Mello da Silva, um empreendedor, sempre cheio de ideias e paciente o suficiente para compreender e tolerar meus atrasos;

Meu coorientador, Prof. Dr. Edgard Costa Oliveira, sempre disponível e com senso prático. Há mais de 20 anos, tive o prazer de trabalhar com ele em projetos de segurança da informação, como as primeiras eleições eletrônicas do país;

Ao Prof. Dr. Marcelo Ladeira, Foco, Força e Fé! Agradeço aos vários ensinamentos como pessoa humana e professor. Em nosso primeiro dia de aula nos alertou sobre a carteira de estudantes: "Vocês irão receber a carteira de estudante, que dá desconto no cinema. Abra a gaveta e guarde-a, pois serão dois anos sem vida social". Ele não estava errado;

A Prof.ª Dr.ª Aletéia Patrícia Favacho de Araújo, pela dedicação e cuidados com o PPCA;
Meus alunos da Faculdade de Direito da Universidade de Brasília, jovens pesquisadores que tiveram a duvidosa alegria de verem algumas destas ideias em sala de aula e de contribuir com suas observações;

A Prof.a Dr.ª Maristela Terto de Holanda, amiga desde os tempos do projeto PROSSIGA/CNPq, em 1998. Um exemplo de força e dedicação;

Aos colegas do PPCA, pelo ambiente de aprendizado, companheirismo e por confiarem a mim a missão de representante da turma, junto ao conselho acadêmico;

A Prof.a Dr.a Claúdia Rosane Roesler e Prof. Dr. Isaac Costa Reis da Faculdade de Direito da UnB pela generosa acolhida e debates instigantes no Grupo de Pesquisa Retórica, Argumentação e Juridicidades (GPRAJ);

Aos meus alunos e amigos das mais diversas práticas esportivas, nos tempos livres sou instrutor de mergulho recreacional, mergulho técnico, mergulho de caverna, primeiros socorros e resgate aquático e de selva. Eles sempre confiaram nas minhas orientações e com quem aprendo muito a cada oportunidade de interação;

Não se equivoca quem acreditar perceber, aqui e ali, entre as linhas, uma brisa. Largas porções do texto foram elaboradas ao ar livre, durante corridas de trilha, no sublime Cerrado de Brasília. Assim vários comentários feitos pela luz, pelo vento, pelas nuvens, pelas flores ou pelo Anbe, meu cachorro, foram imediatamente incorporados;

Sejam especialmente contemplados aqueles que suportaram meu silêncio.

“Poucos são os que têm privacidade para ficar tristes.
Nesse mundo de vigília e patrulha constantes, é um luxo poder sofrer
sem ter ninguém nos observando "
(Martha Medeiros)

RESUMO

O sistema judiciário do Brasil instituiu o Processo Judicial Eletrônico (PJe) e o Diário de justiça Eletrônico (DJe) que vêm promovendo mudanças profundas em todo o ecossistema do judiciário. Por um lado, busca-se aumentar a celeridade processual, reduzir custos e facilitar o acesso a justiça. Por outro lado, o sistema tem o potencial de, ao alcance de um clique, expor a privacidade das partes e, ao mesmo tempo, os advogados e escritórios de advocacia, por falta de treinamento e adoção de boas práticas têm, eventualmente, colocado em risco os dados pessoais dos assistidos. Estes dois fatos contribuem para a violação latente e efetiva da privacidade. O presente estudo investiga a exposição de dados pessoais tanto no Processo Judicial Eletrônico como no Diário de Justiça Eletrônico, principalmente em processos que tramitam sob sigilo de justiça. Também investiga as práticas adotadas pela advocacia na proteção da privacidade dos assistidos em meio digital. Assim a construção de *big data analytics*, e explorados por meio de jurimetria apoiada em algoritmos de inteligência artificial potencializam a exposição da privacidade em uma economia movida a dados que alimenta o "capitalismos da vigilância[1]". Neste sentido, não só em virtude da Lei Geral de Proteção de Dados Pessoais – Lei 13.709/2018, mas também pautado em outros princípios constitucionais, é proposto um modelo de de-identificação por pseudonimização para resguardar a privacidade das partes no processo eletrônico, tendo por base o *Health Insurance Portability and Accountability Act* (HIPAA), também é proposto um ajuste no formato do Dje, para a efitivação do Direito ao Esquecimento, tendo em vista, principalmente, o grande potencial da WEB 3.0 e até mesmo a WEB 4.0, ao mesmo tempo em que organiza um conjunto de recomendações de boas práticas focado em pessoas, "processo" e tecnologia, para advogados e escritórios de advocacia, à luz da ABNT NBR ISO 31000:2018, com o objetivo minimizar riscos de exposição de dados pessoais dos assistidos.

Palavras-chave: mineração de dados, CRISP-DM, ISO 27701, estrutura de privacidade, TEMAC

[1] **Capitalismo de vigilância** é um termo cunhado e popularizado em 2016, pela Profa. Shoshana Zuboff, da Escola de Negócios de *Harvard*, que denota um novo gênero de capitalismo que monetiza dados adquiridos por vigilância.

ABSTRACT

The judicial system in Brazil instituted the Electronic Judicial Process (PJe) and the Electronic Justice Journal (DJe), which have been promoting profound changes in the entire judicial ecosystem. On the one hand, the aim is to increase procedural speed, reduce costs and facilitate access to justice. On the other hand, the system has the potential, within the reach of a click, to expose the privacy of the parties and, at the same time, lawyers and law firms, due to lack of training and adoption of good practices, have eventually put into risk the personal data of the beneficiaries. These two facts contribute to the latent and effective breach of privacy. The present study investigates the exposure of personal data both in the Electronic Judicial Process and in the Electronic Justice Journal, mainly in processes that are under confidentiality. It also investigates the practices adopted by the legal profession in protecting the privacy of those assisted in digital media. Thus, the construction of big data analytics, through jurimetry supported by artificial intelligence algorithms, enhance the exposure of privacy in a data-driven economy that feeds the "surveillance capitalisms[2]". In this sense, not only in virtue of the General Law on Protection of Personal Data - Law 13.709 / 2018, but also based on other constitutional principles, it is proposed a model of de-identification by pseudonymization to safeguard the privacy of the parties in the electronic process, based on the Health Insurance Portability and Accountability Act (HIPAA), while organizing a set of good practice recommendations focused on people and processes, for lawyers and law firms, in the light of ABNT NBR ISO 31000: 2018, with the objective minimize risks of exposure of the personal data of the beneficiaries.

Keywords: data mining, CRISP-DM, ISO 27701, privacy framework, TEMAC

[2] ***Surveillance capitalism*** *is a term coined and popularized in 2016 by Profa. Shoshana Zuboff, from the Harvard Business School, who denotes a new kind of capitalism that monetizes data acquired by surveillance*.

SUMÁRIO

LISTA DE FIGURAS

LISTA DE TABELAS

LISTA DE ABREVIATURAS E SIGLAS

ABNT	Associação Brasileira de Normas Técnicas.
AVE	Valores de Variância Média Extraída.
CGI	Comitê Gestor da Internet no Brasil.
CNJ	Conselho Nacional de Justiça.
CRFB	Constituição da República Federativa do Brasil.
CRISP-DM	CRoss Industry Standard Process for Data Mining.
DJe	Diário de justiça Eletrônico.
FI	Fator de Impacto.
GDPR	General Data Protectiom Regulation.
IA	Inteligência Artificial.
IBGE	Instituro Brasileiro de Geografia e Estatística.
ISO	International Organization for Standardization.
KDD	Knowledge Discovery in Databases.
LGPD	Lei Geral de Proteção de Dados Pessoais.
MEE	Modelo de Equações Estruturais.
MP	Ministério Público.
NBR	Normas Brasileiras.
NCPC	Novo Código de Processo Civil.
NIST	National Institute of Standards and Technology.
TCU	Tribunal de Contas da União.
TEMAC	Teoria do Enfoque Meta Analítico Consolidado.
TTAT	Technology Threat Avoidance Theory.
PEC	Proposta de Emenda à Constituição.
PII	Personal Identificable Informations.
PJe	Processo Judicial Eletrônico.
RIPDP	Relatório de Impacto à Proteção de Dados Pessoais.
SEMMA	Sample, Explore, Modify, Model e Assess
SGPI	Sistema de Gestão de Privacidade da Informação.
SGSI	Sistema de Gestão de Segurança da Informação.
STF	Supremo Tribunal Federal.
VIF	fator de variância.
OAB	Ordem dos Advogados do Brasil.
WOS	Web of Science.

Capítulo 1
Introdução

1.1 Contextualização

As sucessivas inovações da tecnologia da informação e comunicação (TIC), desde o final do século XX, conjugadas com boas práticas de gestão que promovem maior eficiência e eficácia no fornecimento de serviços têm permeado todos os setores da economia. As empresas, desde muito, passaram a adotar sistemas computacionais para a condução da gestão de cadeias produtivas e tomada de decisões a partir de dados, levando, desta forma, ao que hoje conhecemos como *data-driven* economy, ou seja, economia movida a dados, na qual percebe-se a busca, extração e processamento massivo de dados pessoais, que têm gerado vários questionamentos no tocante ao uso ético, respeito à dignidade dos titulares de dados e proteção da privacidade desses usuários (CIURIAK, 2018, on-line).

Gradualmente o sistema judiciário do Brasil, como os demais setores da economia, aderiu ao movimento de atualização não ficando imune às boas práticas de gestão e adoção de solução de TIC e, desta forma iniciou o processo de informatização da justiça, por meio da Lei nº 11.419, de 19 de dezembro de 2006 (Brasil, 2006), instituiu-se o Processo Judicial Eletrônico (PJe) e o Diário de justiça Eletrônico (DJe) que vêm promovendo mudanças profundas em todo o ecossistema judiciário brasileiro, compreendendo advogados, magistrados, serventuários, membros do Ministério Público, partes e peritos, com profundos reflexos na sociedade. Ainda neste caminhar, processos em meio físico (até então guardados em depósitos) estão passando pelo processo de digitalização, sendo convertidos em dados e meta dados, tudo acessível à distância de alguns cliques. Em Brasília, no Tribunal de Justiça do Distrito Federal e dos Territórios (TJDFT) 100% das Unidades Judiciárias de 1° e 2º grau, Turmas Recursais, Unidade de apoio direto às Unidades Judiciárias e o Cartório Judicial Único já operam no PJe[3].

Neste contexto, percebe-se o avanço do "capitalismo de vigilância" sob o judiciário brasileiro, ao constatar o crescimento da Associação Brasileira de Jurimetria[4]. Com acelerado desenvolvimento de sistemas baseados em algoritmos de Inteligência Artificial, como o

[3] https://www.tjdft.jus.br/pje/aqui-tem-pje. Acesso em: 15 mar. 2020

[4] Associação Brasileira de Jurimetria (ABJ) - https://abj.org.br/. Acesso em: 15 abr. 2020

Victor do Supremo Tribunal de Justiça (STJ)[5], a celeridade processual é apontada como um dos principais benefícios. No entanto, pouco se discute a questão da proteção de dados pessoais, criação de perfis, predição e vigilância, especialmente em relação aos processos que tramitam sob sigilo de justiça, nos quais a de-identificação substituindo o nome das partes pelas suas iniciais é, atualmente, insuficiente e ineficiente para resguardar a privacidade.

Por outro lado, cada vez mais, plataformas tecnológicas oferecem soluções "gratuitas" para advogados e escritórios de advocacia que estão baseadas na coleta maciça de dados de comportamento de usuários, além de dados de peças processuais.

1.2 Descrição do problema

Sendo o processo judicial regido pelo princípio da publicidade, tem-se observado a indexação e construção de bancos de dados (*Big Data Analytics*) com dados processuais e também com dados pessoais das partes, e no caso de processos que tramitam sob sigilo de justiça, onde há limitações quanto aos dados tornados públicos, nota-se que nem sempre são respeitadas as limitações e, mesmo quando adotadas o modelo atual, elas têm se mostrado insuficientes para resguardar a privacidade e proteção dos dados pessoais das partes, gerando sua exposição com profundas repercussões negativas na vida desses indivíduos.

Assim, o problema central da presente pesquisa diz respeito aos riscos à privacidade das partes processuais, tendo como vetor de impacto duas vertentes:

- Os efeitos decorrentes da relação máquina-pessoa natural, ou seja, as exposições de dados pessoais geradas pelo PJe e DJe, mesmo nos processos que tramitam sob sigilo de justiça;
- E os efeitos da relação pessoa natural-máquina[6], no qual atores do ecossistema judicial, no caso, advogados e escritórios de advocacia não possuem em geral formação em proteção de dados e, muito menos "processos" para proteger os dados pessoais, que recebem dos assistidos e passam a estar sob sua guarda e responsabilidade.

[5] http://www.stf.jus.br/portal/cms/verNoticiaDetalhe.asp?idConteudo=380038

[6] No presente trabalho a relação homem-máquina / máquina-homem é nomeada de relação homem-pessoa natural / pessoa natural-homem, em aderência ao Art. 5, V da Lei Geral de Proteção de Dados Pessoais (Lei 13.709/2019), que tem a seguinte redação para o Titular de dados: pessoa natural a quem se referem os dados pessoais que são objeto de tratamento.

Desta forma, diante dos fatos descritos anteriormente e evidenciada a situação, este estudo foi elaborado a partir de duas óticas distintas, mas complementares, com os seguintes questionamentos:

1. O atual modelo de sigilo de justiça adotado pelo Processo Judicial Eletrônico (PJe) e Diário de Justiça Eletrônico (DJe) é adequado e suficiente para assegurar que as partes não se tornem identificadas ou identificáveis?

2. Os advogados e escritórios de advocacia adotam medidas mínimas e adequadas de gestão de riscos no que diz respeito à proteção dos dados pessoais e dados pessoais sensíveis dos assistidos?

1.3 Justificativa

Os tribunais sempre atuaram, para além dos encaminhamentos processuais, como fonte de informação.

A lei 11.419/2006, que determinou o uso de meio eletrônico na tramitação de processos judicias, comunicação de atos e transmissão de peças processuais, veio com o objetivo de melhorar a qualidade dos serviços judiciais entregue aos cidadãos. Entre as melhorias almejadas pode-se citar:

- Celeridade na tramitação processual;
- Razoável duração do processo;
- Facilitar acesso à justiça;
- Promover transparência da evolução processual para o cidadão em litígio;
- Aprimorar a gestão do arquivo legado de processos já finalizados que eram realizados todos em papel, gerando diversas despesas como a manutenção de espaço físico para armazenamento de processos, rotinas de combate a umidade e fungos, controles de risco de incêndio e dificuldade de localização, entre outros.

Entretanto, é de comum conhecimento que todo sistema computacional está sujeito a ameaças que podem afetar a disponibilidade, integridade, confidencialidade, a autenticidade dos serviços e o não repúdio, gerando no cidadão uma frustação e descrédito na qualidade dos serviços prestados pelo Estado e perda de credibilidade. E atualmente, um risco adicional diz

respeito ao tratamento de dados pessoais e dados pessoais sensíveis com finalidade diversa afetando, desta forma, a privacidade do cidadão.

Neste sentido, um risco não considerado no momento da imposição da adoção do PJe e DJe é o de exposição das partes na rede mundial de computadores, uma vez que seus nomes e iniciais são indexadas por mecanismos de buscas e, também, são passíveis de se tornarem identificadas, mesmo em processos sob sigilo de justiça, uma vez que os sistemas de *Big Data Analytics* torna possível tal identificação.

Nas palavras de Boaventura de Sousa Santos (SANTOS, 2005, p. 82):

> [...] à distinção entre informação relevante e informação irrelevante. As novas tecnologias e interesses informacionais e comunicacionais têm critérios de relevância que não coincidem com os dos tribunais e, como tal, tentam impor os seus, sujeitando os operadores a um *stress* específico que designo por *stress* comunicacional.

Estes critérios de relevância não coincidentes com os dos tribunais acabam expondo as partes na rede mundial de computadores, gerando desta forma o *stress* comunicacional, impondo às partes, para além dos inerentes ao processo, questões relacionadas a extrema exposição.

Não pode a justiça permanecer com esta visão, de um mundo analógico, de quando o processo tramitava em meio físico e tinha certa limitação de exposição das partes, apesar do importante princípio da publicidade.

Nas observações de Rodrigues Cunha (1999, p. 51, *apud* SANTOS, 2005, p. 82):

> [...] Cunha Rodrigues (1999, p. 51) identificou melhor que ninguém esses perigos:
> a) o de, pelo "excesso de informação", se transmitir uma dimensão totalizante dos factos, susceptível de estigmatizar grupos ou classes sociais, gerando sentimentos de indignação, por um lado, e de indignidade, por outro;
> b) a "sofisticação do escândalo", pela amplificação desproporcionada dos factos, provocando fracturas entre a opinião pública e a realidade;
> c) a sobrepenalização dos arguidos, pelas formas de mediatização utilizadas, sobretudo quando não se chama a atenção para a garantia constitucional de que os arguidos devem considerar-se inocentes até ao trânsito em julgado da decisão;
> [...]

Certo é, que o despertar da comunidade para o fenômeno judiciário, o chamamento ao exercício da cidadania ativa exigida pela democracia dos atuais Estados de Direito e a inevitável tensão entre Justiça e meios de comunicação social trouxeram para a Ágora a discussão em torno da necessidade de zonas de opacidade processual. Sobretudo, a

importância do princípio da publicidade dos atos processuais, em tempo de *Big Data Analytics* faz-se necessário e urgente a necessidade de aprimoramento do mecanismo de sigilo de justiça de forma a se ter uma compatibilização entre os diversos interesses conflitantes, como por exemplo, a eficácia das investigações, a presunção de inocência, a privacidade das partes e a liberdade de imprensa.

A publicidade é um princípio que salvaguarda a possibilidade, constitucional, de se ter meios para o conhecimento dos atos processuais, algo bem diferente de os tornar amplamente públicos na rede mundial de computadores.

Por outra via, na relação pessoa natural-máquina, também os operadores do direito, advogados, que recebem e tratam um conjunto de dados pessoais e dados pessoais sensíveis estão vulneráveis no sentido de não terem o treinamento necessário e nem usarem ferramentas adequadas para a proteção destes dados.

De acordo com Frank Pasquale (2015, p. 9, *apud* FRAZÃO, 2019, p. 27):

> [...] os dados pessoais têm sido utilizados por governos e grandes *players* econômicos para a criação do que chama de *one-way mirror*, possibilitando que tais agentes saibam tudo dos cidadãos, enquanto estes nada sabem dos primeiros. E tudo isso acontece por meio de um monitoramento e vigília constantes sobre cada passo da vida das pessoas, o que leva a um verdadeiro capitalismo de vigilância, cuja principal consequência é a constituição de uma sociedade também de vigilância.

Assim sendo, ter um mecanismo efetivo de de-identificação do nome das partes nos processos que tramitam sob sigilo de justiça e mesmo nos processos públicos é fundamental para garantir a efetivação da qualidade do serviço prestado pelo advogado ao cidadão, na garantia de seu direito fundamental da privacidade. Ao mesmo tempo, um processo maduro de Gerenciamento de Riscos, para além da disponibilidade, integridade, confidencialidade, autenticidade e não repúdio, contribui para que os escritórios de advocacia tenham melhores mecanismos de proteção dos dados pessoais e sensíveis.

A importância do presente estudo não é somente para a OAB ou CNJ, mas para toda a sociedade, uma vez que quase 100% do judiciário já está utilizando o PJe e desta forma, atualmente, a única forma de gerar demandas e acompanhamentos processuais é via sistema. E o mesmo princípio, em breve, irá aplicar-se ao legado processual, qual seja, os processos existem em papel que estão sendo, gradualmente, digitalizados e inseridos no sistema.

Assim, a relevância da pesquisa para a OAB, no sentido de ser um dos principais atores, de "fiscalização", no acompanhamento da evolução, manutenção, estabilidade e disponibilidade do PJe e Dje que se trata, atualmente, do projeto de maior envergadura do judiciário brasileiro é propor um mecanismos efetivo que garanta a privacidade das partes nos

processos que tramitam sob sigilo de justiça e também, disponibilizar para a advocacia, um modelo de processo para Gestão de Risco da Privacidade em relação aos dados pessoais e dados pessoais sensíveis.

Vislumbra-se que tal estudo, tendo como focos principais, o PJe e DJe sob o prisma da privacidade, irá contribuir de forma positiva no aprimoramento e amadurecimento do ambiente do PJe e DJe, sobre os seguintes aspectos:

- Maior maturidade no enfrentamento de riscos associados a ameaças de exposição das partes e violação de direitos fundamentais relacionadas à privacidade, em virtude da exposição e impacto na vida pessoal das partes;
- Melhoria na qualidade do serviço prestado e entregue ao cidadão.

Por outra via, no que diz respeito ao advogado e escritórios de advocacia, a proposta de um modelo de gestão de riscos de dados pessoais e dados pessoais sensíveis, contribui na:

- Salvaguarda dos dados pessoais e dados pessoais sensíveis dos assistidos;
- Redução dos riscos de exposição da imagem e violações de privacidade.

1.4 Objetivos

1.4.1 Objetivo Geral

Propor o aprimoramento do PJe e DJe de modo a resguardar o cidadão contra o uso abusivo e indiscriminado dos seus dados pessoais, nas situações de decretação do sigilo de justiça e elaborar processos mínimos a serem adotados por advogados e escritórios de advocacia no sentido de proteger os dados pessoais e dados pessoais sensíveis aos quais tem acesso.

1.4.2 Objetivos Específicos

- Identificar riscos de exposição da privacidade das partes no PJe e DJe;
- Propor um mecanismo eficiente de de-identificação do nome das partes para os processos que tramitam sob sigilo de justiça;
- Mapear e analisar cenários no escritório de advocacia na relação pessoa natural-máquina que possa oferecer risco de exposição da privacidade dos assistidos;

- Desenhar as boas práticas no tratamento de dados pessoais e dados pessoais sensíveis pelos advogados e escritórios de advocacia.

1.5 Organização desta Dissertação

Este trabalho está estruturado em cinco capítulos.

- **Capítulo 1** – Desenvolve a introdução a partir da contextualização do tema, além de apresentar a problemática, justificativa e objetivos geral e específicos da pesquisa;
- **Capítulo 2** – Estrutura do judiciário nacional, as comarcas e municípios, os foros de competência, o Processo Judicial Eletrônico no Brasil e na Europa, enquadramento legal da proteção de dados pessoas no Brasil e Europa, bem como uma análise geral das implicações da Lei Geral de Proteção de Dados Pessoais;
- **Capítulo 3** – Apresenta a revisão de literatura e conceitos que tratam dos aspectos relacionados à gestão de riscos, relação máquina-pessoa natural e pessoa natural-máquina, publicização de dados, anonimização, de-identificação, pseudonimização, bem como o estado da arte;
- **Capítulo 4** – Apresenta as metodologias de pesquisa utilizadas;
- **Capítulo 5** – Pesquisa de campo por meio do levantamento e análise de dados;
- **Capítulo 6** – Apresenta a conclusão da pesquisa e sugestões de trabalhos futuros.

1.6 Conclusão do capítulo

Observa-se que a informatização da justiça apresenta aspectos positivos, entretanto como efeito colateral surge o risco de extrema exposição das partes processuais, atingindo desta forma os direitos da personalidade, imagem, honra, reputação, privacidade entre outros, que são direitos fundamentais garantidos constitucionalmente.

Por outra via, os advogados e escritórios de advocacia possuem um volume imenso de dados pessoais e dados pessoais sensíveis, ao mesmo tempo que não possuem os processos e formação adequada para adotar medidas de proteção dos dados em tela.

Também se apresentou o objetivo geral e objetivos específicos da pesquisa, a delimitação e áreas de pesquisas relacionadas ao tema e por fim a estrutura dos capítulos. Partindo dessa análise, esta é a importância e relevância da pesquisa para a OAB, CNJ e

sociedade como um todo. No próximo capítulo apresenta-se uma revisão geral do funcionamento do sistema judiciário.

Capítulo 2
O sistema judiciário e outras questões

2.1 Municípios e comarcas

De acordo com os dados do Instituto Brasileiro de Geografia e Estatística (IBGE), o Brasil tem 5.570 (cinco mil quinhentos e setenta) municípios para as 27 (vinte e sete) unidades da federação, sendo que a maior parte dos municípios (68,2%, ou 3.670) com até 20 mil habitantes (IBGE, 2019).

Estes 5.570 municípios são assistidos por 2.702 comarcas (48,5% dos municípios brasileiros), segundo dados do Relatório Justiça em Números, ano base 2019 (CNJ, 2019). A criação de comarcas segue alguns critérios como: extensão territorial, número de habitantes, número de eleitores, receita tributária e movimento forense.

2.2 Estrutura do judiciário

O sistema de justiça no Brasil é "dividido" em instância:

- **Primeira instância** – onde, em geral, iniciam as ações judiciais;
- **Segunda instância** – onde são julgados recursos contra decisões de primeira instância e alguns processos originários (os réus são autoridades com foro privilegiado), é formada pelos tribunais de Justiça e pelos tribunais regionais federais, eleitorais e do trabalho;
- **Instância superior** - é constituída pelos tribunais superiores (Supremo Tribunal Federal, Superior Tribunal de Justiça, Tribunal Superior do Trabalho, Tribunal Superior Eleitoral, Superior Tribunal Militar).

Já no que diz respeito à organização, o poder judiciário brasileiro é agrupado em justiça especial e justiça comum. A Justiça Estadual, integrante da justiça comum (junto com a Justiça Federal), é responsável por julgar matérias que não sejam da competência dos demais

segmentos do Judiciário (justiça especial) - Federal, do Trabalho, Eleitoral e Militar, ou seja, sua competência é residual.

A grande porta de entrada das demandas judiciais está a cargo da justiça comum estadual, 1º grau, que inclusive possui maior estrutura frente às demais justiças, conforme a Figura 2.1:

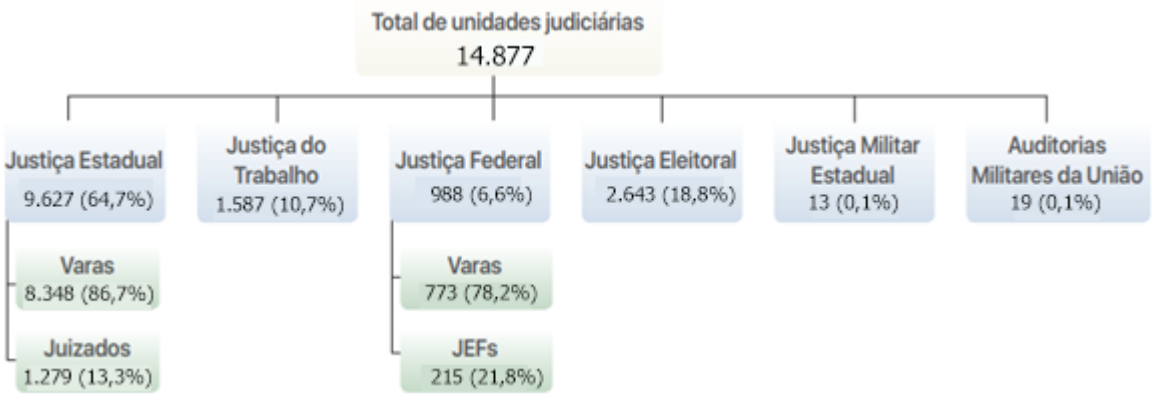

Figura 2.1: Número de unidades judiciárias de 1º grau, por ramo de justiça
Fonte: Conselho Nacional de Justiça

2.3 O foro de competência

A definição do foro competente é feita a partir de uma série de critérios que considera questões como limite territorial da comarca, funcionalidade, valor da causa, questão material (civil, penal, trabalhista, eleitoral, etc.).

Para este trabalho, o critério de maior relevância é a regra geral, ou seja, não incidindo nenhuma das inúmeras normas especiais sobre competência de foro, será aplicada a regra geral da competência territorial do domicílio do réu, conforme estabelece o artigo 46, caput, do Código de Processo Civil (CPC): "A ação fundada em direito pessoal ou em direito real sobre bens móveis será proposta, em regra, no foro de domicílio do réu".

Ainda, o CPC, em determinadas circunstâncias, inverte a regra processual do foro comum, determinando como foro competente, por exemplo, o domicílio ou residência do alimentando, para a ação em que se pedem alimentos (artigo 53, inciso II), bem como, também, a residência da pessoa idosa (artigo 53, inciso III, alínea e), para a causa que verse sobre direito previsto no respectivo estatuto. Cita-se também o Código de Defesa do Consumidor, um microssistema legislativo que assegura ao consumidor, pessoa classificada como hipossuficiente e vulnerável, no seu artigo 101, inciso I, o foro de seu domicílio, para

fins de proteção e a facilitação da defesa de seus direitos (artigo 6º, inciso VIII). Em especial no caso de pessoas com deficiência o foro será seu domicílio.

Notar que muitos são os casos em que o foro é o de domicílio. No âmbito da privacidade dos dados pessoais este é um ponto relevante, pois, mesmo nos casos de processos que tramitam sob sigilo de justiça, com duas peças de informação já é possível tornar a pessoa identificável, por exemplo: iniciais e cidade. Mais fácil ainda se tiver, por exemplo, três peças de informação: iniciais, cidade e pessoa com deficiência.

2.4 O Processo Judicial Eletrônico no Brasil

A Constituição da República Federativa do Brasil (CRFB) de 1988 disciplinou o acesso ao Judiciário como direito fundamental do cidadão, sendo as instituições nacionais da época consideradas demasiado morosas e ineficientes para conferir provimentos legais adequados e ágeis. O Judiciário chegou à era digital com a urgência de atender os objetivos constitucionais, como prover um mais amplo acesso dos cidadãos aos tribunais, uma razoável duração do processo judicial, bem como respeito aos princípios da publicidade processual.

No espírito de efetivar a disposição constitucional que exigiu um Judiciário mais acessível e ágil criou-se uma estrutura legal/informacional para o Processo Judicial Eletrônico (PJe) e o Diário de Justiça Eletrônico (DJe). A lei determinou a transição dos procedimentos dos tribunais brasileiros de uma arcaica massa de documentos de papel para bancos de dados virtuais.

Assim, tanto o PJe, como o DJe foram instituídos pela lei 11.419/2006, que determinou o uso de meio eletrônico na tramitação de processos judicias, que englobariam todas as decisões judiciais, transmissão de peças processuais das partes e comunicação de atos dos tribunais. Tendo como abrangência os processos civis, penais e trabalhistas, em qualquer grau de jurisdição, inclusive os juizados especiais, ou seja, toda a tramitação processual no sistema judiciário à luz da referida Lei nº 11.419, de 19 de dezembro de 2006 (Brasil, 2006), deve dar-se por meio de sistema computacional:

> Art. 1º O uso de meio eletrônico na tramitação de processos judiciais, comunicação de atos e transmissão de peças processuais será admitido nos termos desta Lei.
> § 1º Aplica-se o disposto nesta Lei, indistintamente, aos processos civil, penal e trabalhista, bem como aos juizados especiais, em qualquer grau de jurisdição.
> [...]
> Art. 4º Os tribunais poderão criar Diário de justiça eletrônico, disponibilizado em sítio da rede mundial de computadores, para publicação de atos judiciais e

> administrativos próprios e dos órgãos a eles subordinados, bem como comunicações em geral.
> § 1º O sítio e o conteúdo das publicações de que trata este artigo deverão ser assinados digitalmente com base em certificado emitido por Autoridade Certificadora credenciada na forma da lei específica.
> § 2º A publicação eletrônica na forma deste artigo substitui qualquer outro meio e publicação oficial, para quaisquer efeitos legais, à exceção dos casos que, por lei, exigem intimação ou vista pessoal.

Se por um lado, com tal medida, busca-se a adoção de mecanismos tecnológicos que dê vazão aos milhões de processos existentes no judiciário brasileiro, que segundo dados do Conselho Nacional de Justiça (CNJ), disponível no relatório analítico "Justiça em números 2019" (CNJ, 2019)[7], ano-base 2018, informa que o Poder Judiciário finalizou o ano de 2018 com 78,7 milhões de processos em tramitação, aguardando alguma solução definitiva.

Por outro, procura-se atender preceitos fundamentais presentes na Constituição da República Federativa do Brasil de 1988, (Brasil, 1988, on-line), em seu artigo 5°, incisos LX e LXXVIII, estabelecem o princípio da publicidade do processo e da razoável duração do mesmo além de meios para garantir a celeridade:

> LX - a lei só poderá restringir a publicidade dos atos processuais quando a defesa da intimidade ou o interesse social o exigirem;
> [...]
> LXXVIII - a todos, no âmbito judicial e administrativo, são assegurados a razoável duração do processo e os meios que garantam a celeridade de sua tramitação.

Também o inciso IX do artigo 93 da CRFB trata da publicidade:

> IX - todos os julgamentos dos órgãos do Poder Judiciário serão públicos, e fundamentadas todas as decisões, sob pena de nulidade, podendo a lei limitar a presença, em determinados atos, às próprias partes e a seus advogados, ou somente a estes, em casos nos quais a preservação do direito à intimidade do interessado no sigilo não prejudique o interesse público à informação; (Redação dada pela Emenda Constitucional nº 45, de 2004).

Após a promulgação da Lei 11.419/2006, o Conselho Nacional de Justiça (CNJ), lançou em junho/2011 a primeira versão do PJe. Em junho/2019 o sistema completou 8 (oito) anos de funcionamento e segundo relatório de auditoria do Tribunal de Contas da União (TCU, 2019), de julho de 2019, com o objetivo de avaliar a implementação e o funcionamento da informatização dos processos judiciais, em especial o Processo Judicial Eletrônico (PJe) o

[7] Disponível em: https://www.cnj.jus.br/wp-content/uploads/conteudo/arquivo/2019/08/justica_em_numeros20190919.pdf. Acesso em: 23 Jan. 2020

sistema único ainda não é uma realidade. "Os 155 sistemas que foram passíveis de identificação apresentaram um dispêndio decorrente de falhas de interoperabilidade e duplicidade de esforços da ordem de, pelo menos, R$ 374 milhões, de 2013 a 2017, apenas na esfera da União" (TCU, 2019, p. 42).

O Tribunal de Contas da União (TCU) no Acórdão 1.534/2019, de 08/07/2019 (TCU, 2019), do Plenário, de relatoria do Ministro do TCU Raimundo Carreiro, determina o prazo de 180 (cento e oitenta dias) para o Conselho Nacional de Justiça (CNJ) apresentar um plano de ação para resolver o problema:

> [..]
> 9.1 determinar ao Conselho Nacional de Justiça, com fundamento no art. 250, inciso II, do Regimento Interno deste Tribunal, que:
> 9.1.1 apresente a este Tribunal, no prazo de 180 (cento e oitenta) dias, a contar da ciência, plano de ação voltado ao aprimoramento da eficiência, da efetividade e da transparência das ações de implementação e operação do Processo Judicial Eletrônico (PJe), indicando as medidas a serem adotadas, os responsáveis pelas ações e os respectivos prazos de implementação, abrangendo, no que couber, os objetivos específicos descritos no subitem 252 do Relatório que integra este Acórdão.
> 9.1.2 nos termos do art. 46 da Resolução-CNJ 185/2013, interrompa, no prazo de 15 (quinze) dias, a contar da ciência, as transferências voluntárias de ativos a tribunais que não tenham implantado efetivamente o PJe;
> [...]

Em um primeiro momento as leis federais que instituíram esta estrutura legal não estipularam um banco de dados ou plataforma centralizada para o Judiciário nacional ou ao menos uniformizada por meio de um padrão único. Desta forma, a profusão de várias plataformas com diversos graus de implementação, uso e eficiência são prejudiciais ao acesso dos cidadãos e partes ao Judiciário e, portanto, violam os princípios constitucionais norteadores do processo eletrônico. O Conselho Nacional de Justiça estabeleceu, em 18 de dezembro de 2013, por intermédio da Resolução 185/2013[8], a obrigação de todos os tribunais brasileiros usarem uma só base de dados centralizada (PJe). Em 2015, foi promulgado um Novo Código de Processo Civil (NCPC) nacional, trazendo consigo novas regras que regulamentam o processo eletrônico e as plataformas processuais eletrônicas.

Assim, que pese a importância e a relevância de modernização da justiça, diversos são os problemas que os advogados, os cidadãos e mesmo os jurisdicionados estão enfrentando.

[8] Disponível em: https://atos.cnj.jus.br/atos/detalhar/1933. Acesso em: 23 Jan. 2020

Como exemplos específicos é possível citar que em virtude da diversidade de sistemas com diferentes requisitos de versão de *software*, não é incomum escritórios de advocacia e mesmo advogados terem diferentes máquinas ou sistemas virtuais para poderem acessar os vários sistemas espalhados pelos estados.

De forma a atender a determinação do TCU, o CNJ como órgão nacional de controle, supervisão e coordenação do Poder Judiciário, em seus diversos segmentos, à exceção do Supremo Tribunal Federal, conclamou todos os tribunais, órgãos e instituições que exercem funções essenciais à Justiça (Ministério Público, advocacias públicas, Defensorias Públicas, OAB etc.) a colaborarem no sentido de adotar, nacionalmente, um único sistema, o PJe, que na verdade já havia sido definido na resolução 185/2013 do CNJ[9].

Outro problema que começa a transparecer em relação à privacidade de dados pessoais à medida que mais e mais processos são ajuizados em formato digital, é a exposição das partes uma vez que as ferramentas de indexação permitem, ao pesquisar o nome de um cidadão, chegar a processos, que a depender do teor expõem a imagem, a honra, a intimidade, a privacidade e o nome, mesmo que decretado sigilo de justiça. O armazenamento de dados em sistemas de *Big Data Analytics* com a simples omissão do nome completo e a troca pelas suas iniciais não é suficiente para garantir o sigilo de justiça.

Segundo dados do relatório Justiça em números (CNJ, 2019) durante o ano de 2018, apenas 16,2% do total de processos novos ingressaram fisicamente. Em apenas um ano, entraram 20,6 milhões de casos novos eletrônicos.

Também os escritórios de advocacia e advogados, frente à nova realidade no uso de sistemas de informação, para além da relação pessoa natural-máquina, encontram-se com seus sistemas em situação de risco, especialmente com a sanção da Lei Geral de Proteção de Dados Pessoais (LGPD), Lei 13.709/2018, que passa a produzir efeitos a partir de agosto de 2020, que entre outras obrigações exige políticas de Governança de dados, em relação aos dados pessoais e dados pessoais sensíveis que tratam.

Estes fatos demonstram claramente a ausência de um plano de gestão de riscos, tanto no que diz respeito à violação dos Direitos da Personalidade do cidadão, quanto aos sistemas internos nos escritórios de advocacia, que podem também expor dados pessoais.

A proteção de dados pessoais e privacidade é um dos principais debates so que diz respeito aos direitos humanos no século XXI. Este crescimento na importância da proteção de

[9] Disponível em: https://atos.cnj.jus.br/atos/detalhar/1933 Acessado em: 19 fev. 2020

dados pessoais e, no Brasil a exemplo de outros países como os membros da Comunidade Europeia, a possível conversão da proteção de dados pessoais como Direito Fundamental, por meio da Proposta de Emenda à Constituição (PEC 17/2019)[10], reflete a necessidade do uso de meios adequados para resguardar tanto os dados pessoais como os direitos fundamentais.

2.5 O Processo Judicial Eletrônico na Europa e o caso França

Tem-se observado mudanças legislativas nos Estados-Membros da União Europeia com o objetivo de permitir a utilização de Tecnologias de Comunicação e Informação (TIC) nos processos judiciais. De acordo com informações constantes no site da *European E-justic e*[11], existe, atualmente, uma diversidade de modelos de operação (EUROPEAN E-JUSTICE, 2020):

> Em alguns casos, os processos podem ser abertos mediante uma petição inicial eletrônica enviada pelo requerente ao tribunal; noutros casos, os tribunais têm poder para notificar documentos às partes de forma eletrônica; e noutros casos ainda, todas as comunicações são feitas eletronicamente. Em alguns Estados-Membros, a utilização da Internet para efeitos de comunicação é aceite em quase todos os tipos de processos civis, mas noutros esta utilização é limitada a determinados tipos especiais de processos.

Um dos princípios orientadores na informatização do judiciário na União Europeia é que as legislações nacionais devem garantir que a comunicação por meio da Internet não ameace a adequada proteção de dados sensíveis durante a tramitação do processo.

No site da *European E-justice*, existem respostas a 17 questões sobre o tema para cada um dos Estados-Membros. Na Tabela 2.1 estão destacados os pontos de maior relevância para esta pesquisa em relação a alguns dos países da União Europeia:

Tabela 2.1: Situação da informatização do judiciário em alguns países da União Europeia

PAÍS	É possível iniciar um processo via internet?	Podem as partes ou seu representante legal consultar a tramitação on-line?
Alemanha	Em princípio a legislação permite o	Em geral não é possível.

[10] Acrescenta o inciso LXXIX, ao art. 5º, e o inciso XXX, ao art. 22, da Constituição Federal para incluir a proteção de dados pessoais entre os direitos fundamentais do cidadão e fixar a competência privativa da União para legislar sobre a matéria.

[11] Disponível em: https://beta.e-justice.europa.eu/home?action=home. Acesso em: 24 Fev. 2020.

	processo eletrônico. Entretanto, na prática, não é possível em todas as localidades e em todos os tipos de processos.	
Bélgica	Desde 2012 existem iniciativas no sentido de informatizar o processo judicial. Entretanto lentas.	Não
Espanha	Sim, em implantação desde Jan. 2017	Não. Nem os profissionais da justiça nem as partes podem consultar os processos on-line, uma vez que essa funcionalidade não está disponível a nível nacional para os processos cíveis. A funcionalidade está sendo introduzida em algumas áreas do país.
Finlândia	Sim	Sim
França	Sim	Não*
Grécia	Sim, para primeira instância	Sim, para primeira instância.
Irlanda	Sim, para ações de pequenos valores.	Sim, as partes podem acompanhar on-line.
Itália	Sim	Sim
Portugal	Sim	À exceção dos processos de execução em que a consulta pode ser feita pelas próprias partes, nos restantes processos judiciais a consulta on-line só é disponibilizada aos mandatários (advogados, solicitadores), aos agentes de execução e, em processos de insolvência, aos administradores de insolvência.
Suécia	Não	Não

Fonte: *European E-justice*

A França é uma situação especial, pois, em março de 2019, a Assembleia Nacional Francesa sancionou uma lei que diz respeito à organização judiciária do país até 2022, Lei 2019-222[12]. Entre outras questões, a referida legislação regulamentou o acesso aos dados do judiciário, no Art. 33, parte integrante da seção 3 que tem como título "Conciliar as decisões

[12] Disponível em: https://www.legifrance.gouv.fr/eli/loi/2019/3/23/JUST1806695L/jo/texte. Acesso em: 24 Fev. 2020.

da justiça com o direito à privacidade" (*Section 3: Concilier la publicité des décisions de justice et le droit au respect de la vie privée*).

No Art. L111-14, a reforma sobre o acesso aos dados processuais em meio eletrônico fortalece a ideia da importância em se ter mecanismos de disponibilização on-line e chama a atenção para a inconveniência de processos físicos ou em papel, além de enfrentar questões relativas às formas e meios de disponibilização dos dados, custos de acesso, de-identificação de dados e coleta de dados na construção de modelos preditivo.

No tocante à conciliação entre o acesso a dados públicos e, de outro, o direito à privacidade das pessoas envolvidas, o mesmo artigo 33 modifica o artigo L10-1 do Código de Administração da Justiça e estabelece restrições e diretrizes para a de-identificação de parte das informações incluídas nos acórdãos.

Segundo a redação do referido artigo, os sobrenomes e nomes próprios das pessoas físicas mencionadas nos acórdãos, quando são partes ou terceiros envolvidos, devem ser ocultados antes da disponibilização ao público. Ainda afirma que, nos casos onde a divulgação for suscetível de prejudicar a segurança ou violar a privacidade das pessoas, deverão ser ocultados os elementos que permitam identificar as partes, os terceiros, os magistrados e os serventuários.

O dispositivo determina ainda que os dados relativos à identidade de juízes e membros do judiciário não podem ser reutilizados com o objetivo de avaliar, analisar, comparar ou prever suas práticas profissionais, ou seja, não podem ser utilizados em modelos preditivos, em especial jurimetria.

Caso estas proibições sejam violadas, o infrator está sujeito às penalidades previstas nos artigos 226-18, 226-24 e 226-31 do Código Penal Francês, sem prejuízo das medidas e sanções previstas pela Lei n° 78-17, de 6 de janeiro de 1978, relativas ao tratamento de dados, arquivos e liberdades, que pode chegar à pena máxima de até 5 anos de reclusão.

2.6 Dados pessoais

Na Lei Geral de Proteção de Dados Pessoais – Lei 13.709/2018 (LGPD), dados pessoais são definidos como todas as informações relativas a uma pessoa natural identificada ou identificável, abrangendo aspetos pertinentes à vida privada, profissional ou pública, conceito este copiado da lei Europeia de proteção de dados, o *General Data Protection Regulation* (GDPR). Já nos Estados Unidos da América (EUA) os dados pessoais são usualmente designados por *Personal Identifiable Information* (PII).

O conceito de "identificável" abrange todas as situações em que a pessoa, embora não explicitamente identificada, possa vir a sê-lo, caso sejam realizadas pesquisas cruzadas com outras fontes de dados.

É importante realçar que, mesmo encriptados, de-identificados ou pseudonimizados, os dados continuam a ser considerados dados pessoais.

A LGPD preconiza também um regime especial para a salvaguarda de dados pessoais sensíveis, incluindo dados de saúde, biométricos, genéticos, vida sexual, filiação sindical, origem étnica ou racial, assim como opiniões e convicções políticas e religiosas.

O PJe e DJe estão repletos de dados pessoais, tais como nome das partes, atividade, profissão, perfiz em plataformas digitais, dados relacionados a questões de saúde, cidade, estado, números IPs, placas de veículos, endereços de imóveis entre outros. Estes dados estão dispersos, publicamente, a depender do estilo de redação nas decisões, sentenças, votos e acórdãos.

Importante observar que mesmo processos que tramitam sob sigilo de justiça, onde o nome completo é substituído pelas iniciais das partes é possível tornar alguém identificável ao se realizar correlacionamentos das iniciais com base de dados dos assistidos em determinada comarca, conjugado com cidade, profissão entre outras peças de informações. Considerando essas afirmações, parece natural a necessidade do poder judiciário voltar seus olhos para a definição de regras de acesso a seus bancos de dados.

Coincidentemente durante o desenvolvimento desta pesquisa, em abril de 2019 o presidente do Conselho Nacional de Justiça (CNJ) publicou portaria constituindo uma comissão para estudar o tema e propor soluções, a portaria 63/2019 do CNJ[13] que "Institui Grupo de Trabalho destinado à elaboração de estudos e propostas voltadas à política de acesso às bases de dados processuais dos tribunais e dá outras providências". E mais recentemente, em 13 de fevereiro de 2020, nova portaria sobre o tema, 24/2020[14] prorroga o prazo para julho de 2020.

2.7 A publicidade do processo e a privacidade

Na sociedade atual muitas relações são realizadas de forma remota. Nas palavras de Doneda (2018):

[13] Disponível em: https://atos.cnj.jus.br/atos/detalhar/2890. Acesso em: 20 Nov. 2019.
[14] Disponível em: https://atos.cnj.jus.br/atos/detalhar/3191. Acesso em: 13 Fev. 2020.

> [...] na sociedade atual, caracterizada pelas relações remotas, os dados pessoais acabam por se constituir na única forma de representação das pessoas perante as mais diversas organizações estatais e privadas, sendo determinantes para abrir ou fechar as portas de oportunidades e acessos.

Tanto a doutrina como a jurisprudência reconhecem que o direito à privacidade abrange, hoje, não apenas a proteção à vida íntima e dos espaços privado do indivíduo, mas também a proteção de seus dados pessoais, pois esses constituem uma relevante projeção da personalidade do indivíduo perante o Estado e sistemas comerciais privados.

Poderia o Estado, por meio de plataformas tecnológicas, disponibilizar, dados e informações dos cidadãos, para cumprir mandamentos constitucionais e desta forma, estar, indiretamente expondo certas dimensões da personalidade do cidadão?

O problema central da presente pesquisa dá-se em relação aos riscos à privacidade, tendo como vetor de impacto duas vertentes dos efeitos decorrentes da relação máquina-pessoa natural e pessoa natural-máquina, sendo que se percebe cada vez uma maior dependência tecnológica para solucionar problemas do ser humano e, neste sentido, de forma ilusória, achar que a tecnologia é simplesmente um meio para alcançar o fim.

Diante das mudanças em velocidade exponencial na adoção de algoritmos de inteligência artificial (I.A.) e rapidez com que as informações circulam mundialmente pela internet, ganha relevo o tensionamento entre o direito à publicidade do processo e à privacidade das partes.

Para Aline Marinho Bailão Iglesias (IGLESIAS, 2015, p. 12), "ambos têm status constitucional, inexistindo hierarquia entre eles, ou seja, nenhum tem primazia absoluta sobre o outro, de modo que cada um constitui limite constitucional do outro e vice-versa".

Ainda nas palavras de Iglesias[15], esta aponta que:

> O direito à privacidade surgiu com a ascensão da burguesia, no século XVIII. As atividades que eram feitas de forma comunitária passaram a se desenvolver na órbita privada. No âmbito do direito à privacidade, estão as relações pessoais do indivíduo com seus familiares e amigos; tutela de tudo o que o indivíduo não deseja que seja de conhecimento público, isto é, o direito de excluir do conhecimento de terceiros aquilo que somente a ele se refira, que diga respeito ao seu modo de ser na sua vida privada; ao passo que a intimidade constitui o núcleo mais essencial da privacidade, atinente ao que diz respeito somente à pessoa, à sua esfera mais reservada. Portanto, tem-se que a privacidade é mais ampla e engloba a intimidade.

Todavia, é importante destacar que o direito à privacidade garantida na Constituição Federal em seu Art. 5°, inciso X não é absoluta. Ademais o homem público nada pode

[15] Id., 2015, p. 12.

esconder enquanto no exercício de suas funções daqueles que representam ou servem. Neste sentido as palavras de Ives Gandra da Silva Martins (MARTINS, 2001, p. 6):

> Entendemos que a privacidade a que se refere o art. 5º inciso X diz respeito àquele cidadão comum, ao homem corrente, e não aos que devem representar a cidadania. Estes devem ter sua vida como um livro aberto. Os "homens públicos" – neles incluo as mulheres, para diferenciá-las daquelas em que o adjetivo indica profissão diversa – como a própria qualificação está a indicar são "públicos", e não "privados", nada podendo esconder dos que representam ou servem.

Tratam-se de conceitos subjetivos a privacidade e intimidade, pois, cabe a cada indivíduo determinar quais são as dimensões de sua vida que pretende impedir que chegue ao conhecimento público.

A privacidade, nos últimos anos, vem ganhando cada vez mais relevo em decorrência, principalmente: da ampliação da capilaridade da internet e disseminação massiva de dispositivos computacionais, fezendo com que todos passassem a manter seus dados na forma digital, facilitando sua coleta.

Segundo a pesquisa TIC Domicílios 2018, do Comitê Gestor da Internet no Brasil (CGI), 70% da população brasileira está na rede, o que equivale a 126,9 milhões de pessoas (CGI, 2018). A estruturação dos *Big Data Analytics* a qual permite o cruzamento e correlacionamento das informações fornecidas pelo usuário e capturas na rede, inclusive sem o seu conhecimento, nos mais diversos tipos de aplicações e aplicativos e a construção de perfilização dos usuários.

A disponibilização dos dados processuais na internet possibilita que qualquer um tenha acesso a eles, e vasculhem, na rede, a vida privada de quem interessar. Segundo Aline Marinho Bailão Iglesias (IGLESIAS, 2015, p. 14), “se por um lado o processo moderno deve se adequar às novas tecnologias, estas, por sua vez, não podem olvidar princípios consagrados, como o direito de personalidade, que abrange a privacidade”.

A publicidade dos atos processuais, por sua vez, constitui garantia inerente ao Estado Democrático de Direito, destinada a combater arbitrariedades, a proporcionar a transparência processual, a fiscalização popular dos atos praticados por juízes, promotores e advogados no processo, além de constituir uma forma de evitar ingerências externas nas manifestações processuais dos agentes já mencionados.

A publicidade do processo é a regra e está prevista nos artigos 5º, LX, e 93º, IX, da Constituição Federal. Ambos os dispositivos ressalvam que a lei poderá restringir a publicidade dos atos processuais, a fim de defender a intimidade das partes envolvidas.

O art. 189 do Novo Código de Processo Civil (NCPC)[16], Lei nº 13.105, de 16 de março de 2015 (Brasil, 2015), trata dos processos protegidos pelo segredo de justiça, ou seja, enumera as hipóteses em que é possível restringir a publicidade dos atos processuais, conforme autorizado pela Constituição Federal:

> Art. 189. Os atos processuais são públicos, todavia tramitam em segredo de justiça os processos:
> I - em que o exija o interesse público ou social;
> II - que versem sobre casamento, separação de corpos, divórcio, separação, união estável, filiação, alimentos e guarda de crianças e adolescentes;
> III - em que constem dados protegidos pelo direito constitucional à intimidade;
> IV - que versem sobre arbitragem, inclusive sobre cumprimento de carta arbitral, desde que a confidencialidade estipulada na arbitragem seja comprovada perante o juízo.
> § 1º O direito de consultar os autos de processo que tramite em segredo de justiça e de pedir certidões de seus atos é restrito às partes e aos seus procuradores.
> § 2º O terceiro que demonstrar interesse jurídico pode requerer ao juiz certidão do dispositivo da sentença, bem como de inventário e de partilha resultantes de divórcio ou separação.

Importante fazer a distinção entre os atos classificados como segredo de justiça, conforme Art. 189/NCPC, já citado e os atos aos quais são atribuídos o sigilo de justiça, ou seja, sobre estes atos recai uma restrição sobre as próprias partes envolvidas – inclusive seus advogados – na relação processual. Desta forma, não terão acesso a alguns conteúdos do processo, serão esses os documentos sigilosos com acesso apenas ao juiz, ao Ministério Público (MP) e aos serventuários autorizados. Geralmente, o segredo ocorre na fase investigativa sobre aqueles elementos ainda não documentados, sendo assegurado o direito de vista aos advogados pelo art. 7º, XIII, do Estatuto da OAB e pela Súmula Vinculante 14 do Supremo Tribunal Federal (STF) em respeito ao exercício do direito de defesa.

Assim, excetuadas as hipóteses legais, os atos processuais são públicos. A fim de regulamentar a divulgação dos dados processuais eletrônicos na rede mundial de computadores, o Conselho Nacional de Justiça expediu a Resolução nº 121, de 5 de outubro de 2010[17]. Merecem destaque os artigos 1º, 2º e 4º:

> Art. 1.º A consulta aos dados básicos dos processos judiciais será disponibilizada na rede mundial de computadores (internet), assegurado o direito de acesso a informações processuais **a toda e qualquer pessoa, independentemente de prévio cadastramento ou de demonstração de interesse**.
> Parágrafo único. No caso de processo em sigilo ou segredo de justiça não se aplica o disposto neste artigo.
> Art. 2.º Os dados básicos do processo de livre acesso são:
> I – número, classe e assuntos do processo;
> II – **nome das partes** e de seus advogados;

[16] Disponível em: http://www.planalto.gov.br/ccivil_03/_Ato2015-2018/2015/Lei/L13105.htm. Acesso em: 24 Jan. 2020
[17] Disponível em: https://atos.cnj.jus.br/atos/detalhar/atos-normativos?documento=92. Acesso em: 24. Jan. 2020

III – movimentação processual;
IV – inteiro teor das decisões, sentenças, votos e acórdãos.

A Figura 2.2 apresenta a categorização dos atos processuais quanto à privacidade:

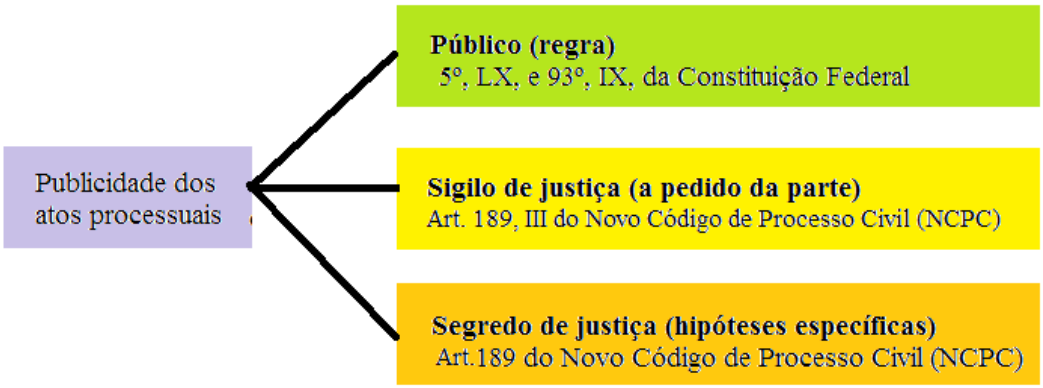

Figura 2.2: Publicidade dos atos processuais
Fonte: autoria própria

Ponto relevante para o presente trabalho de pesquisa é que nos casos de decretação do sigilo de justiça a implementação do mesmo dá-se por meio da substituição do nome das partes pelas iniciais, permanecendo todas as demais informações acessíveis a qualquer um, principalmente quando disponibilizadas no Diário de Justiça Eletrônico (DJe), ou seja, os atos processuais após tramitação no PJe são disponibilizados no DJe e nos processos sob sigilo o nome das partes é substituído pelas iniciais da mesma, Figura 2.3:

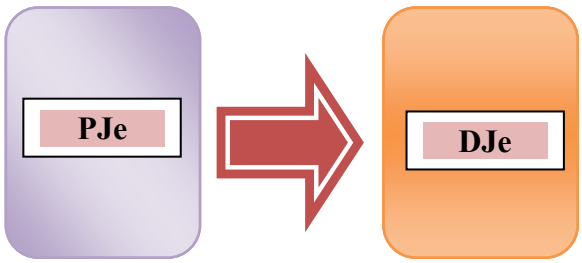

Figura 2.3: Publicização dos atos processuais no DJe
Fonte: autoria própria

Neste sentido, considerando o contexto anteriormente descrito, especialmente na vertente do relacionamento máquina-pessoa natural, no que diz respeito à exposição do nome ou iniciais das partes pelo PJe e DJe em cenários que podem caracterizar violação de direitos fundamentais da personalidade, questiona-se a efetividade do atual modelo de sigilo de justiça pautado pela substituição do nome das partes pelas, simples, iniciais das mesmas frente a mecanismos de *Big Data Analytics*, correlacionamento massivo de dados e aprendizado de máquina (*machine learning*).

A outra vertente, na qual foi identificado outro problema, desta vez na via pessoa natural-máquina são dos diversos dados pessoais e sensíveis de posse dos escritórios de advocacia que nem sempre são adequadamente protegidos, podendo, também, levar à exposição das partes.

Em termo esquemático a Figura 2.4 representa os pontos de investigação do presente projeto de pesquisa:

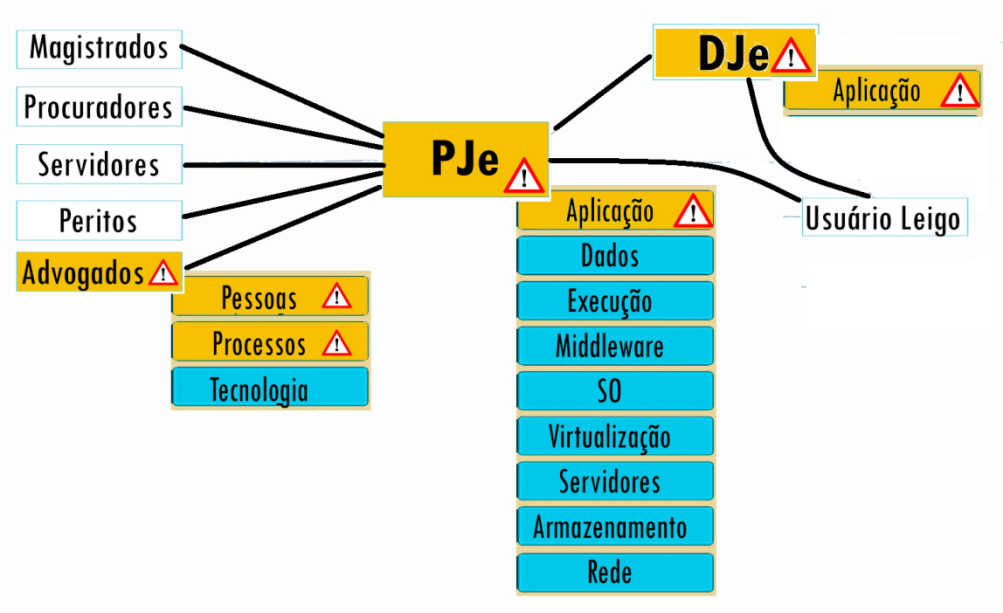

Figura 2.4: Pontos de investigação da pesquisa
Fonte: autoria própria

Dentro do PJe e DJe, como consequência reflexa, será investigada a relação máquina-pessoa natural, na camada de aplicação, referente à violação da privacidade das partes processuais nos processos que tramitam sob sigilo de justiça.

Já na relação pessoa natural-máquina, será objeto de investigação, dentro do escritório de advocacia as pessoas e processos no que diz respeito à gestão dos riscos de exposição de dados pessoais e sensíveis.

Sendo assim, diante dos fatos descritos anteriormente, evidenciada a situação, os problemas de pesquisa que norteiam este estudo foram elaborado a partir de duas óticas distintas, com os seguintes questionamentos:

1. O atual modelo de sigilo de justiça adotado pelo Processo Judicial Eletrônico (PJe) e Diário de Justiça Eletrônico (DJe) é adequado e suficiente para garantir que as partes não sejam identificadas ou identificáveis?

2. Os escritórios de advocacia adotam medidas mínimas e adequadas, de gestão de riscos, no tocante à proteção dos dados pessoais e dados pessoas sensíveis dos assistidos?

2.8 A Lei Geral de Proteção de Dados Pessoais

A Lei Geral de Proteção de Dados Pessoais (LGPD) foi sancionada em 14 de agosto de 2018 e tem previsão para início de produção dos efeitos plenos em agosto de 2020. A lei nacional foi fortemente inspirada na *General Data Protection Regulation* (GDPR), lei europeia que começou a produzir efeitos em maio de 2018.

A LGPD aplicada aos dados pessoais tanto em meio digital como em meio físico, e logo no seu artigo primeiro já deixa claro que o titular de dados é a parte vulnerável, assim todos os esforços devem ser direcionados a sua proteção:

> Art. 1º Esta Lei dispõe sobre o tratamento de dados pessoais, inclusive nos meios digitais, por pessoa natural ou por pessoa jurídica de direito público ou privado, com o objetivo de **proteger** os direitos fundamentais de liberdade e de privacidade e o livre desenvolvimento da personalidade da **pessoa natural**. (grifo nosso)

Desta forma a LGPD estabelece uma série de medidas como o estabelecimento de uma política de governança de dados, adoção de boas práticas de segurança e privacidade dos dados, tais como: consentimento informado, livre, com finalidade determinada, inequívoco, expresso e específico, ainda a coleta e registro do consentimento, pseudonimização, encriptação ou anonimização de dados, notificação de violação de privacidade, nomeação de encarregado de dados pessoais nas instituições, adoção do conceito de desenvolvimento de sistemas tendo privacidade como concepção e por padrão (*privacy by design* e *privacy by default*).

No Art. 5° da referida lei é apresentada uma série de conceitos onde se destacam os seguintes:

Tabela 2.2: Conceitos presentes na Lei Geral de Proteção de Dados Pessoais

INCISO	DEFINIÇÃO
I - dado pessoal	informação relacionada a pessoa natural identificada ou identificável;

II - dado pessoal sensível	dado pessoal sobre origem racial ou étnica, convicção religiosa, opinião política, filiação a sindicato ou a organização de caráter religioso, filosófico ou político, dado referente à saúde ou à vida sexual, dado genético ou biométrico, quando vinculado a uma pessoa natural;
III - dado anonimizado	dado relativo a titular que não possa ser identificado, considerando a utilização de meios técnicos razoáveis e disponíveis na ocasião de seu tratamento
V – titular	pessoa natural a quem se referem os dados pessoais que são objeto de tratamento;
VI – controlador	pessoa natural ou jurídica, de direito público ou privado, a quem competem as decisões referentes ao tratamento de dados pessoais;
VII – operador	pessoa natural ou jurídica, de direito público ou privado, que realiza o tratamento de dados pessoais em nome do controlador;
VIII – encarregado	pessoa indicada pelo controlador e operador para atuar como canal de comunicação entre o controlador, os titulares dos dados e a Autoridade Nacional de Proteção de Dados (ANPD)
X – tratamento	toda operação realizada com dados pessoais, como as que se referem a coleta, produção, recepção, classificação, utilização, acesso, reprodução, transmissão, distribuição, processamento, arquivamento, armazenamento, eliminação, avaliação ou controle da informação, modificação, comunicação, transferência, difusão ou extração;
XI - anonimização	utilização de meios técnicos razoáveis e disponíveis no momento do tratamento, por meio dos quais um dado perde a possibilidade de associação, direta ou indireta, a um indivíduo;
XII – consentimento	manifestação livre, informada e inequívoca pela qual o titular concorda com o tratamento de seus dados pessoais para uma

	finalidade determinada;
XVII - relatório de impacto à proteção de dados pessoais	documentação do controlador que contém a descrição dos processos de tratamento de dados pessoais que podem gerar riscos às liberdades civis e aos direitos fundamentais, bem como medidas, salvaguardas e mecanismos de mitigação de risco;
XIX - autoridade nacional	órgão da administração pública responsável por zelar, implementar e fiscalizar o cumprimento desta Lei em todo o território nacional.

Fonte: Lei Geral de Proteção de Dados Pessoais

A Lei aplica-se a todos os setores, seja privado ou público. O Art. 4° apresenta algumas exceções de não aplicabilidade da LGPD e o judiciário não está entre elas. Desta forma, faz-se necessário a adequação plena aos ditames legais.

Para que se possa tratar dados pessoais há necessidade de atendimento ao menos de uma das 10 hipóteses estabelecidas no Art. 7º:

> Art. 7º O tratamento de dados pessoais somente poderá ser realizado nas seguintes hipóteses:
> I - mediante o fornecimento de consentimento pelo titular;
> II - para o cumprimento de obrigação legal ou regulatória pelo controlador;
> III - pela administração pública, para o tratamento e uso compartilhado de dados necessários à execução de políticas públicas previstas em leis e regulamentos ou respaldadas em contratos, convênios ou instrumentos congêneres, observadas as disposições do Capítulo IV desta Lei;
> IV - para a realização de estudos por órgão de pesquisa, garantida, sempre que possível, a anonimização dos dados pessoais;
> V - quando necessário para a execução de contrato ou de procedimentos preliminares relacionados a contrato do qual seja parte o titular, a pedido do titular dos dados;
> **VI - para o exercício regular de direitos em processo judicial, administrativo ou arbitral, esse último nos termos da Lei nº 9.307, de 23 de setembro de 1996 (Lei de Arbitragem);**
> VII - para a proteção da vida ou da incolumidade física do titular ou de terceiro;
> VIII - para a tutela da saúde, em procedimento realizado por profissionais da área da saúde ou por entidades sanitárias;
> VIII - para a tutela da saúde, exclusivamente, em procedimento realizado por profissionais de saúde, serviços de saúde ou autoridade sanitária; (Redação dada pela Lei nº 13.853, de 2019) Vigência
> IX - quando necessário para atender aos interesses legítimos do controlador ou de terceiro, exceto no caso de prevalecerem direitos e liberdades fundamentais do titular que exijam a proteção dos dados pessoais; ou
> X - para a proteção do crédito, inclusive quanto ao disposto na legislação pertinente.
> (grifo nosso)

Neste sentido, o inciso VI apresenta a hipótese de uso de dados pessoais para o exercício regular do direito em processos em geral. Diferente não poderia ser a CRFB que possui

expressa previsão referente à inafastabilidade da apreciação pelo Poder Judiciário (Art. 5°, XXXV/CRFB) e também à ampla defesa e contraditório (Art. 5°, LV/ CRFB).

Por outra via, a hipótese que autoriza o tratamento de dados pessoais em processos em geral, mesmo que para esta única e exclusiva finalidade, não isenta o Poder Judiciário de adotar as medidas necessárias e adequadas para a proteção dos dados e salvaguarda da privacidade dos titulares de dados ou, aqui, dos jurisdicionados.

Os princípios norteadores da proteção de dados estão presentes no Art. 6°, da LGPD

> Art. 6º As atividades de tratamento de dados pessoais deverão observar a boa-fé e os seguintes princípios:
> I - finalidade: realização do tratamento para propósitos legítimos, específicos, explícitos e informados ao titular, sem possibilidade de tratamento posterior de forma incompatível com essas finalidades;
> II - adequação: compatibilidade do tratamento com as finalidades informadas ao titular, de acordo com o contexto do tratamento;
> III - necessidade: limitação do tratamento ao mínimo necessário para a realização de suas finalidades, com abrangência dos dados pertinentes, proporcionais e não excessivos em relação às finalidades do tratamento de dados;
> IV - livre acesso: garantia, aos titulares, de consulta facilitada e gratuita sobre a forma e a duração do tratamento, bem como sobre a integralidade de seus dados pessoais;
> V - qualidade dos dados: garantia, aos titulares, de exatidão, clareza, relevância e atualização dos dados, de acordo com a necessidade e para o cumprimento da finalidade de seu tratamento;
> VI - transparência: garantia, aos titulares, de informações claras, precisas e facilmente acessíveis sobre a realização do tratamento e os respectivos agentes de tratamento, observados os segredos comercial e industrial;
> **VII - segurança: utilização de medidas técnicas e administrativas aptas a proteger os dados pessoais de acessos não autorizados e de situações acidentais ou ilícitas de destruição, perda, alteração, comunicação ou difusão;**
> **VIII - prevenção: adoção de medidas para prevenir a ocorrência de danos em virtude do tratamento de dados pessoais;**
> IX - não discriminação: impossibilidade de realização do tratamento para fins discriminatórios ilícitos ou abusivos;
> X - responsabilização e prestação de contas: demonstração, pelo agente, da adoção de medidas eficazes e capazes de comprovar a observância e o cumprimento das normas de proteção de dados pessoais e, inclusive, da eficácia dessas medidas.
> (Grifo nosso)

Merece destaque os incisos VII e VIII, que estão "interligados" e tratam, respectivamente, da segurança e prevenção, ou seja, deve o Poder Judiciário, de forma preventiva, inciso VIII do Art. 6° da LGPD, adotar as medidas necessárias para garantir a segurança dos dados. Tanto em sua dimensão referente à segurança da informação como à privacidade. Ou seja, adotar medidas para proteger os dados pessoais e dados pessoais sensíveis dos jurisdicionados de acessos não autorizados, de eventos acidentais ou ilícitos de destruição, alteração, perda, comunicação ou difusão.

Aqui, novamente surge o conflito entre dois direitos jusfundamental: Publicidade X Privacidade. Como já indicado anteriormente a Constituição Federal tem como princípio a publicidade do processo, entretanto há que haver mecanismos mínimos que resguardem a privacidade dos jurisdicionados, principalmente nos processos que tramitam sob sigilo de justiça.

Importante observar que aos princípios da segurança e prevenção, aproximam-se, também os princípios da responsabilidade e prestação de contas, podendo-se, inclusive, buscar o exercício do direito de defesa perante o Estado.

2.9 Conclusão do capítulo

Observa-se que a informatização da justiça apresenta aspectos positivos, entretanto como efeito colateral surge o risco de extrema exposição das partes processuais, atingindo desta forma os direitos da personalidade, imagem, honra, reputação, privacidade e entre outros, que são direitos fundamentais garantidos constitucionalmente.

Por outra via, os advogados e escritórios de advocacia possuem um volume imenso de dados pessoais e dados pessoais sensíveis, ao mesmo tempo que não possuem os processos e formação adequada para adotar medidas de proteção dos dados em tela.

No próximo capítulo, apresenta-se uma revisão sistemática da literatura com os principais conceitos empregados na presente pesquisa.

Capítulo 3

Revisão de literatura

Este capítulo apresenta os principais conceitos referentes às técnicas de privacidade de dados pessoais, gestão de riscos e boas práticas de mercado referentes à segurança e privacidade. Com o objetivo de identificar as melhores contribuições para o estado da arte, foi utilizada a teoria do Enfoque Meta-analítico Consolidado (TEMAC). O *Health Insurance Portability and Accountability Act* (HIPAA) foi utilizado como referência em parte deste estudo, como modelo de de-identificação, com base na área de saúde.

3.1 Teoria do Enfoque Meta Analítico Consolidado

A Teoria do Enfoque Meta Analítico Consolidada (TEMAC) é uma metodologia de pesquisa bibliográfica, quantitativa, de caráter exploratório. Segundo Mariano e Rocha (2017), o TEMAC consegue integrar as exigências atuais da literatura a respeito de trabalhos científicos como precisão, robustez, validade, funcionalidade, tempo e custos, possibilitando, desta forma obter os melhores autores, artigos e revistas.

3.1.1 Preparação da pesquisa

A preparação da pesquisa consiste em estabelecer os parâmetros da busca (palavra chave). Foram utilizadas duas bases de dados: *Web of Science* (WOS) e *Scopus*, por representarem bases sólidas e internacionais da ciência.

A pesquisa na base de dados *Web of Science* foi realizada com o filtro de resultados com a palavra-chave *pseudonymization* entre os anos 2010 e 2020, no qual obteve-se 77 resultados, dos quais 11 resultados com mais citações foram exportados em formato de texto sem formatação para a utilização posterior.

Já a pesquisa na base de dados *Scopus*, realizada através da base de dados Publish or Perish, utilizando o mesmo termo, qual seja, *pseudonymization* no campo *All of the Words* entre os anos 2011 e 2020, obtendo 171 resultados, que foram filtrados mais uma vez pela característica Publisher, excluindo aqueles que não possuíam a característica, obtendo 34 resultados que foram exportados no formato RIS para utilização em etapas seguintes.

Pesquisas semelhantes foram realizadas nas duas bases com o termo *de-identification*, e os resultados são apresentados na Tabela 3.1:

Tabela 3.1: Quantidade de artigos publicados por termo período 2010 a 2020

TERMO	*Web of Science*	*Scopus*
	Quantidade de artigos	Quantidade de artigos
pseudonymization	77	171
de-identification	430	539

Fonte: *Web of Science* e *Scopus*

3.1.2 Seleção dos artigos de maior relevância

Nas Figuras 3.1 e 3.2 são apresentadas as áreas de conhecimento mais ativas na produção de artigos para o tema *pseudonymization* e ***de-identification***:

Figura 3.1: Áreas mais ativas na produção de artigos – Tema: *pseudonymization*
Fonte: *Web of Science*

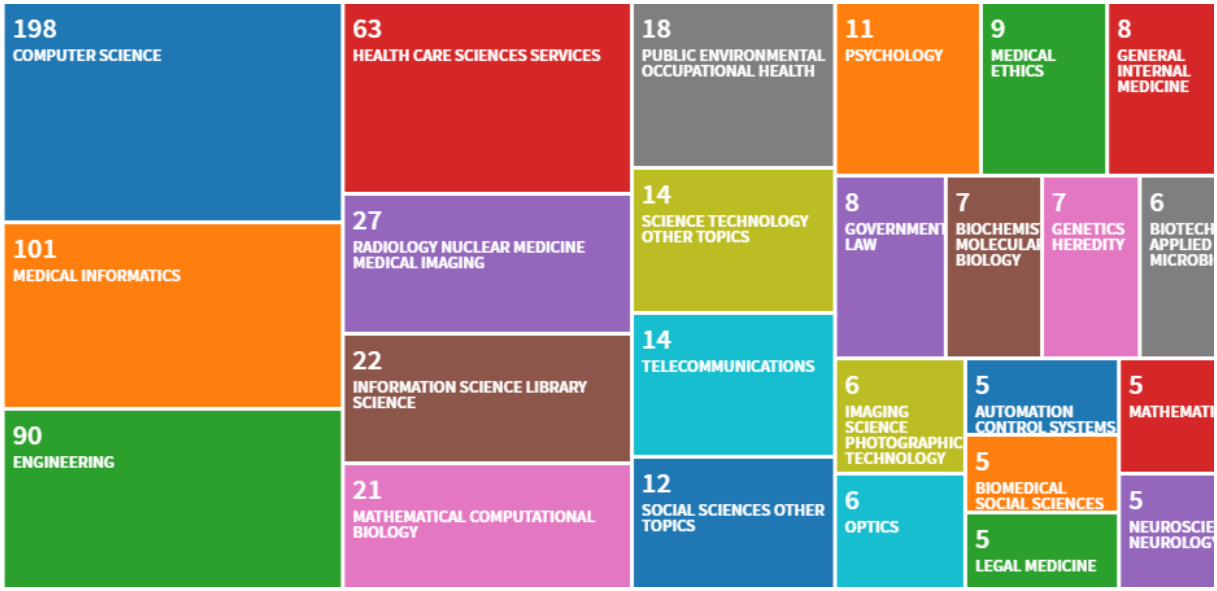

Figura 3.2: Áreas mais ativas na produção de artigos – Tema: *de-identification*
Fonte: *Web of Science*

Após a identificação das áreas de conhecimento com maior atividade na produção de artigos sobre o tema, foi realizada a classificação dos artigos de acordo com o número de citações, tendo como critérios de agrupamento os artigos com 20 (vinte) ou mais citações.

Tabela 3.2: Quantidade de artigos com 20 ou mais citações no período 2010 a 2020

TERMO	*Web of Science*		*Scopus*	
	Quantidade de artigos	Quantidade de artigos com 20 ou mais citações	Quantidade de artigos	Quantidade de artigos com 20 ou mais citações
pseudonymization	77	04	171	20
de-identification	430	36	539	58

Fonte: *Web of Science* e *Scopus*

3.1.3 Análise dos artigos e autores

Segundo Calazans (CALAZANS, 2015) o objetivo dessa análise consiste em avaliar a importância do tema ao longo dos anos. Essa análise permite visualizar a relação dos artigos selecionados em relação à pesquisa no período estudado. Após uma análise dos artigos em relação a proposta do estudo, foram selecionados os trabalhos dos autores abaixo, ordenados pelo maior impacto, levando-se em consideração a quantidade de citações, para investigação inicial.

AUTOR	TÍTULO	CITAÇÕES	ANO	QUALIS	PUBLICADO
Bayardo, R.J., Agrawal, R.	Data privacy through optimal k-anonymization	753	2005		Proceedings - International Conference on Data Engineering pp. 217-228
Neubauer, Thomas; Heurix, Johannes	A methodology for the pseudonymization of medical data	43	2011	A1	INTERNATIONAL JOURNAL OF MEDICAL INFORMATICS
Ganslandt, T.; Mate, S.; Helbing, K.; et al.	Unlocking Data for Clinical Research - The German i2b2 Experience	26	2011	B1	APPLIED CLINICAL INFORMATICS
Sahi, Muneeb Ahmed; Abbas, Haider; Saleem, Kashif; et al.	Privacy Preservation in e-Healthcare Environments: State of the Art and Future Directions	20	2018	A2	IEEE ACCESS
Jawurek, Marek; Johns, Martin; Rieck, Konrad	Smart Metering De-Pseudonymization	33	2011		27TH ANNUAL COMPUTER SECURITY APPLICATIONS CONFERENCE
Bolognini, Luca; Bistolfi, Camilla	Pseudonymization and impacts of Big (personal/anonymous) Data processing in the transition from the Directive 95/46/EC to the new EU General Data Protection Regulation	20	2017		COMPUTER LAW & SECURITY REVIEW

Figura 3.3: Artigos mais relevantes – Tema: *pseudonymization*
Fonte: *Web of Science*

AUTOR	TÍTULO	CITAÇÕES	ANO	QUALIS	PUBLICADO
Meystre, Stephane M.; Friedlin, F. Jeffrey; South, Brett R.; et al.	Automatic de-identification of textual documents in the electronic health record: a review of recent research	97	2010	A1	BMC MEDICAL RESEARCH METHODOLOGY
El Emam, Khaled; Jonker, Elizabeth; Arbuckle, Luk; et al.	A Systematic Review of Re-Identification Attacks on Health Data	96	2011	B1	PLOS ONE
Fernandes, Andrea C.; Cloete, Danielle; Broadbent, Matthew T. M.; et al.	Development and evaluation of a de-identification procedure for a case register sourced from mental health electronic records	67	2013	A2	BMC MEDICAL INFORMATICS AND DECISION MAKING
Ford, Elizabeth; Carroll, John A.; Smith, Helen E.; et al.	Extracting information from the text of electronic medical records to improve case detection: a systematic review	64	2016	A1	JOURNAL OF THE AMERICAN MEDICAL INFORMATICS ASSOCIATION
Aberdeen, John; Bayer, Samuel; Yeniterzi, Reyyan; et al.	The MITRE Identification Scrubber Toolkit: Design, training, and assessment	49	2010	A1	INTERNATIONAL JOURNAL OF MEDICAL INFORMATICS
Dernoncourt, Franck; Lee, Ji Young; Uzuner, Ozlem; et al.	De-identification of patient notes with recurrent neural networks	31	2017	A1	JOURNAL OF THE AMERICAN MEDICAL INFORMATICS ASSOCIATION
El Emam, Khaled	Risk-Based De-Identification of Health Data	20	2010	A2	IEEE SECURITY & PRIVACY
Friedman, Carol; Rindflesch, Thomas C.; Corn, Milton	Natural language processing: State of the art and prospects for significant progress, a workshop sponsored by the National Library of Medicine	48	2013		JOURNAL OF BIOMEDICAL INFORMATICS

Figura 3.4: Artigos mais relevantes – Tema: *de-identification*
Fonte: *Web of Science*

3.2 Categorias de de-identificação

Será utilizado, neste estudo, de forma genérica, o termo de-identificação como referência a um processo ou procedimento de eliminação ou modificação da informação pessoal ou parte desta, existente em um banco de dados ou sistema, com o objetivo de dificultar ou impedir a identificação unívoca dos titulares de dados. Na prática, esse processo poderá traduzir-se em uma anonimização, ou pseudonimização, a depender das técnicas

empregadas. A definição e desambiguação destes termos encontram-se em (SIMSON, 2015), (NELSON, 2015) e (ARTICLE 29, 2014):

- **De-identificação**: como gênero, consiste na remoção ou ofuscação de toda a informação pessoal armazenada em um banco de dados, com o objetivo de impedir a identificação dos titulares de dados. A de-identificação não é necessariamente um processo irreversível, podendo prever-se a existência de uma tabela de mapeamento que permita reverter o processo (associando os registos originais aos registos de-identificados).

  Além da supressão de todos os atributos identificadores, a de-identificação usualmente implica a modificação dos *quasi-identifiers*, por meio de processos de generalização (Ex.: modificando a escala de um atributo) ou por introdução de fatores de incerteza, com variação no grau de entropia, tendo por base os valores originais.

  - **Anonimização**: é considerado um caso "forte" de de-identificação, através do qual se pretende tornar inviável ou mesmo impossível (utilizando todos os meios considerados razoáveis) a re-identificação (inclusive pelo próprio técnico que realizou a operação inicial). Ou seja, por princípio deverá ser um processo irreversível, análogo à destruição. Cabe observar que a definição é adaptável conforme o contexto tecnológico empregado: "todos os meios considerados razoáveis", permitindo assim ponderar os recursos necessários, o custo e conhecimento necessários para realizar uma re-identificação.

  - **Pseudonimização**: como caso "fraco" da de-identificação, é um processo que visa substituir todos os identificadores pessoais (exemplo: nomes, endereço, CPF, etc.) por pseudônimos: palavras ou códigos gerados artificialmente, os quais poderão funcionar como representações mascaradas dos dados originais. Uma pseudonimização "forte" tem adicionalmente a preocupação de incidir sobre os atributos *quasi-identifiers* (Ex.: data de nascimento), e que a atribuição de códigos seja realizada de forma aleatória e independente dos valores originais (embora possam eventualmente continuar relacionados entre si). Por norma, a pseudonimização mantém todos os atributos de um banco de dados relacional, permitindo salvaguardar a respetiva estrutura e sintaxe dos dados.

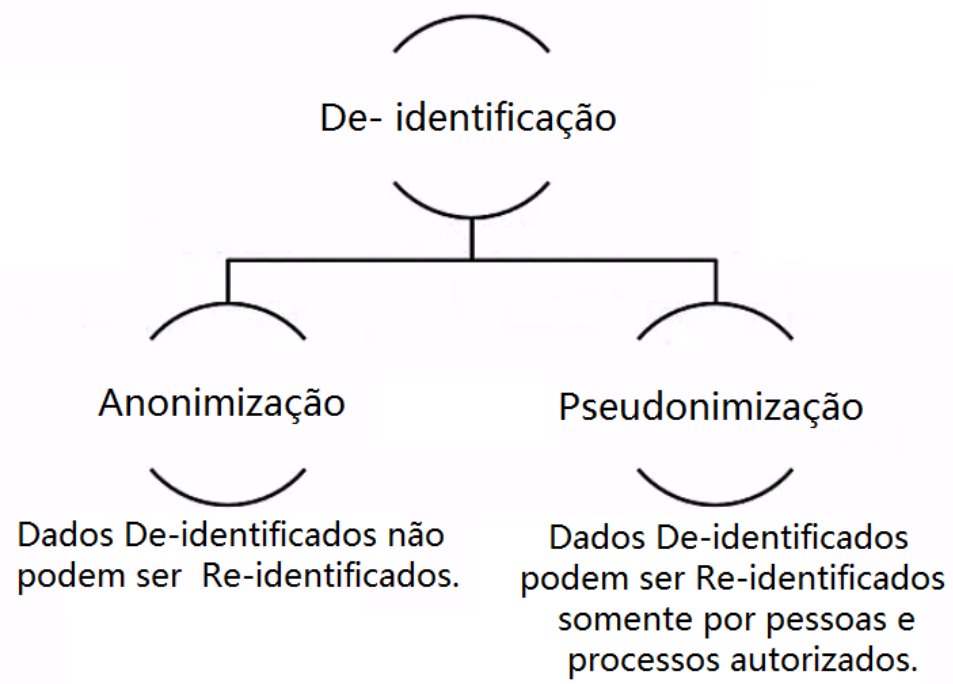

Figura 3.5: Espécies de De-identificação
Fonte: autoria própria

3.3 Modelo HIPAA

No *Health Insurance Portability and Accountability Act* (HIPAA), que diz respeito ao modelo Estadunisense de tratamento de dados pessoais na área da saúde é comum utilizar-se o termo de-identificação, estando este inclusivamente explicitado na metodologia HIPAA *Safe Harbor*, que define como padrão a remoção de 18 atributos sensíveis, como uma das condições necessárias para que uma base de dados possa ser considerada de-identificada (HIPAA, 2012):

Tabela 3.3: Atributos sensíveis do HIPAA

1. Nomes	7. SSN ou NIF	13. Identificador de dispositivos
2. Códigos postais (até 3 dig.)	8. N°s de registro médico	14. URLs
3. Datas (exceto o ano)	9. Nºs de apólice de saúde	15. IPs
4. Nºs de telefone	10. Nºs de conta	16. Identificadores biométricos
5. Nºs de fax	11. Nºs de certificado ou licença	17. Fotos com face completa
6. Endereços de email	12. Identificador de veículos	18. Outros atrib. Identificadores

Fonte: HIPAA *Safe Harbour*

Por motivos técnicos ou funcionais o modelo do HIPAA permite a preservação dos IDs únicos das tabelas iniciais, garantindo a ligação entre os registos e os titulares de dados e dessa forma possibilitando uma posterior re-identificação (usualmente sujeita a um processo de autorização, já que implica o acesso a uma tabela de mapeamento ou mecanismo similar).

3.4 Eficácia de um processo de de-identificação

O grupo de trabalho *Article* 29, no relatório *Data Protection Working Party, Opinion 05/2014 on Anonymisation Techniques* (ARTICLE 29, 2014), indicou que a eficácia de um processo de de-identificação pode ser aferida tendo como referência a dificuldade em:

1. **Re-identificação dos indivíduos** (também designado por *single-out*): A aplicação de uma "máscara" simples sobre o nome ou ID dos titulares de dados, ou mesmo o uso de criptografia simétrica, permite sempre que a entidade que realizou a operação volte a identificar os indivíduos, isto é, são processos reversíveis;
2. **Ligação entre registos, dentro ou entre bancos de dados**: A capacidade de associar ou correlacionar dois ou mais registos de um mesmo titular de dados (violando também a 1ª regra), ou de um grupo (neste caso violando apenas esta 2ª regra);

3. **Inferência de qualquer informação sobre um titular de dados**: deduzir o valor de um atributo de um titular de dados, com uma significativa probabilidade de sucesso, a partir de outros atributos.

Os processos de anonimização e pseudonimização são, geralmente, computacionalmente caros, ou seja, demandam muitos recursos e via de regra implicam em diversos ajustes até se obter a desejada utilidade dos dados finais. A situação ideal implicaria maximizar, simultaneamente, a privacidade e a utilidade dos dados, o que, via de regra, é impossível de se alcançar, pois as dificuldades advêm não só das limitações dos modelos, mas também da necessidade de interligação dos registos entre bases de dados de sistemas distintos (LUK e EMAM, 2013):

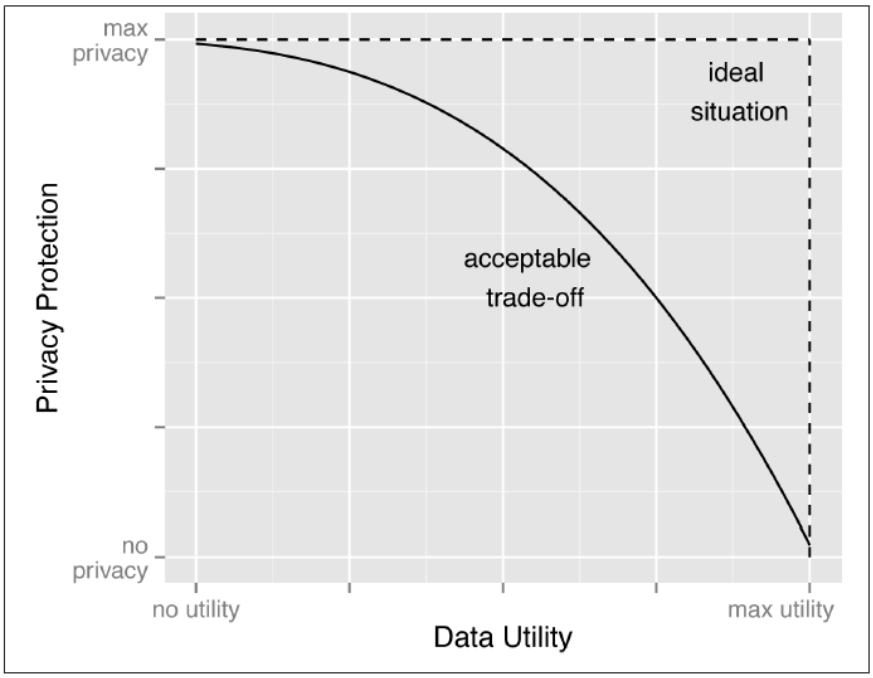

Figura 3.6*: trade-off* entre privacidade e utilidade dos dados
Fonte: *Anonymizing Health Data* (LUK e EMAM, 2013, p. 20)

Desta forma, no processo de de-identificação deve-se buscar a melhor relação possível entre:

- **Utilidade dos dados**: para permitir que os utilizadores consigam trabalhar com os dados, extraindo informação e estatísticas;
- **Privacidade dos dados**: para garantir que parte da informação mantém-se ocultada.

De uma forma geral as técnicas de de-identificação usualmente utilizadas podem ser categorizadas como Randomização, Generalização e Pseudonimização:

- Randomização: *Noise Addition*, *Shuffling* e *Differential Privacy*;
- Generalização: *K-Anonymity* e *L-Diversity / T-Closeness*;
- Pseudonimização: Substituição (ou codificação), Encriptação (usualmente com chave simétrica), *Hash* e Máscara de caracteres.

Estas são somente algumas das técnicas possíveis em cada uma das classificações apresentadas, não sendo objeto da presente pesquisa o aprofundamento das vantagens e força de casa uma delas.

3.5 Conceito de risco

Risco é contrariar objetivos, ou seja, impactar no atingimento dos objetivos da organização. Uma definição formal é a apresentada na Norma ABNT ISO Guia 73 na qual risco é "o efeito da incerteza nos objetivos" (ABNT, 2009a, p. 4). Todavia, a incerteza pode gerar resultados tanto negativos quanto positivos. Neste diapasão, um olhar pelo prisma positivo deve ser entendido como uma oportunidade a ser trabalhada em favor da organização para alcance de seus objetivos.

Na obra Desafio aos Deuses (BERNSTEIN, 1997, p. 29), é apresentado que desde os primórdios a humanidade convive com o risco, porém há um marco que muda esta relação, no sentido de que "*a ideia revolucionária que define a fronteira entre os tempos modernos e o passado é o domínio do risco: a noção de que o futuro é mais do que um capricho dos deuses e de que homens e mulheres não são passivos ante a natureza*".

Ainda de acordo com o autor citado, o estudo do risco teve sua origem no Renascimento, quando as pessoas, ao desafiarem crenças consagradas, libertaram-se de restrições de seu passado e abriram caminho para descobertas. Desde então, o risco tem sido objeto de estudo por nomes como o matemático Blaise Pascal (para decifrar o enigma proposto por Paccioli), Pierre de Fermat (teoria das probabilidades), Gottfriedvon Leibniz (retorno dos eventos), Daniel Bernoulli (lei dos grandes números e amostragem estatística) e Thomas Bayes (teorema de Bayes). Contudo, a gestão de riscos moderna passou a ser estudada e aperfeiçoada após a Segunda Guerra Mundial por pesquisadores como Markowitz, Lintner, Treynor, Sharpe e Mossin (DIONNE, 2013).

Diversos são os tipos de riscos existentes conforme o modelo adotado. Assim, não existe uma classificação única na literatura, sendo necessário avaliar as especificidades de cada organização. Contudo, os riscos, em termos amplos podem ser classificados tendo por base a origem dos eventos, se internos ou externos (GESPUBLICA, 2013, on-line):

- **Riscos externos**: são aqueles associados ao ambiente onde a organização opera. Em geral, a organização não tem controle direto sobre estes eventos, entretanto, mesmo assim, ações podem ser tomadas (exemplo: raios e instalação de para-raios);
- **Riscos internos**: são os riscos associados à própria estrutura da organização, seus processos, governança, quadro de pessoal, recursos ou ambiente de tecnologia. Aqui a organização pode e deve agir diretamente de forma proativa.

Os riscos externos podem, ainda, ser subdivididos em outras categorias, tais como: Políticos (Nacional e Internacional), Econômico/Financeiros (Nacional/internacional), socioculturais, tecnológicos, legal/regulatório e ambiental. Já os riscos internos podem ser subdivididos em recursos financeiros, recursos humanos, processos internos, sistemas de informação, parceiros, fornecedores e outros riscos específicos da organização e seu contexto. (GESPUBLICA, 2013, on-line).

Importante ressaltar a diferença entre os riscos inerentes e os riscos residuais. Aquele refere-se à exposição proveniente de um risco específico antes que qualquer ação seja tomada para gerenciá-lo, já o risco residual é a exposição remanescente de um risco específico após uma ação ser tomada para gerenciá-lo (GESPUBLICA, 2013, on-line). Neste sentido, a análise de riscos deve considerar o risco inerente e não o residual. Pois assim pode-se obter uma real percepção do risco ao qual a organização está exposta. O risco residual deve ser aceitável e justificável, isto é, deve estar dentro do apetite de risco da organização.

Além de interpretações jurídicas, há de se considerar a partir dos conceitos de erro latente e erro efetivo, existentes na Engenharia da Confiabilidade, o comportamento do sistema quanto à falha na preservação da privacidade nos processos que tramitam sob sigilo de justiça.

Laprie (1995) define falha como um problema que ocorre externamente a um componente do sistema computacional, mas que pode ser percebido por esse, tal como falhas físicas ou de especificação do componente. Uma vez percebida, essa falha pode resultar um erro latente ou efetivo na execução do componente.

A partir desta visão sugere-se a construção dos termos "violação latente da privacidade" e "violação efetiva da privacidade", sendo:

- **Violação latente da privacidade**: fragmentos de dados pessoais ou não que geram situações de risco que, ao serem combinarem com outros fragmentos de dados, criam a oportunidade para a exposição da privacidade;
- **Violação efetiva da privacidade**: é a resultante da violação latente da privacidade, ou seja, a exposição da privacidade em sí, que causa real prejuízo à parte.

3.6 Gestão de riscos - Norma ABNT/NBR ISO 31000:2018

A adoção de boas práticas de mercado além de facilitar a implementação de bons modelos é, segundo FONSECA (2010, p. 73) "um sistema de padronização e desta forma um dos pilares, mais importantes, para empresas que desejam implementar sistemas de gestão baseados na qualidade". Por meio dos sistemas de padronização a organização garante que os processos e atividades sejam executados segundo regras definidas de forma clara, minimizando, desta forma, os riscos e a ocorrência de falhas operacionais. Um destes modelos, amplamente reconhecido e adotado pelo mercado é a norma ABNT NBR ISO 31000:2018 – Gestão de Riscos, basilar na elaboração do presente estudo.

A ABNT NBR ISO 31000, lançada incialmente em 2009 pela ISO –*International Organization for Standardization*, passou por um processo de revisão e aprimoramento e teve uma nova versão publicada em março de 2018. A norma trata dos aspectos positivos e negativos da manifestação de um risco e apresenta princípios, guias e terminologias comuns para a gestão de riscos de forma a se indicar uma padronização e uniformização dos processos já existentes.

O modelo proposto pela norma viabiliza sua utilização em organizações de qualquer tamanho ou segmento de atividade. Possuindo elevado grau de flexibilidade permiti, desta forma, que dentro de uma mesma empresa, áreas distintas tratem a incerteza de acordo com as regras específicas de cada uma, utilizando-se de um processo único e integrado.

Segundo a ABNT NBR ISO 31000:2018, gerenciar riscos, de forma eficaz, baseia-se em princípios, estrutura e processos. Este conjunto de ações deve ser documentado e seguido pela organização em seus diferentes níveis, permeando todos os setores que busquem melhorias e aprimoramento de seus processos de gestão de riscos. A estrutura para a implementação da 31000:2018 é apresentada abaixo:

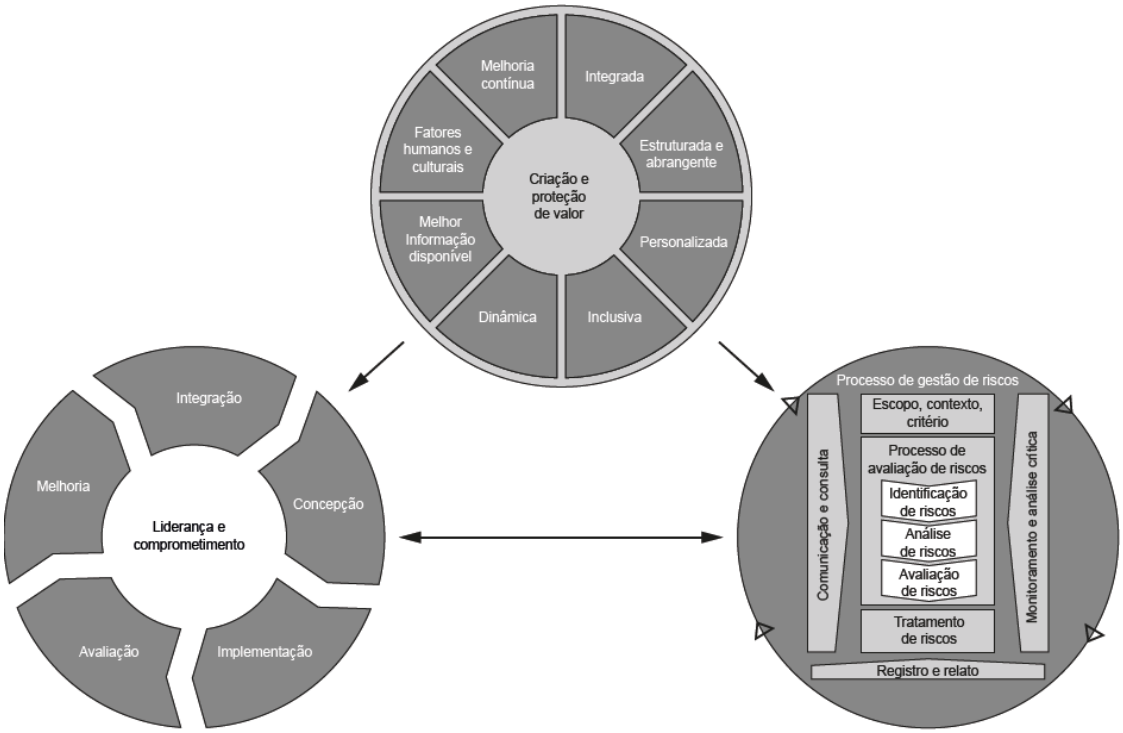

Figura 3.7*:* Princípios, estrutura e processo (ISO 31000:2018)
Fonte: ABNT NBR ISO 31000:2018

Ainda de acordo com a ABNT NBR ISO 31000:2018, o processo de gestão de riscos envolve a aplicação sistemática de políticas, procedimentos e práticas para as atividades de comunicação e consulta, estabelecimento do contexto e avaliação, tratamento, monitoramento, análise crítica, registro e relato de riscos", ou seja, trata-se de um processo que tem por objetivo controlar as incertezas de uma organização, contexto, atividade, processo ou projeto, etc.

Convém que o processo de gestão de riscos seja parte integrante da gestão e da tomada de decisão, e seja integrado à estrutura, operações e processos da organização.

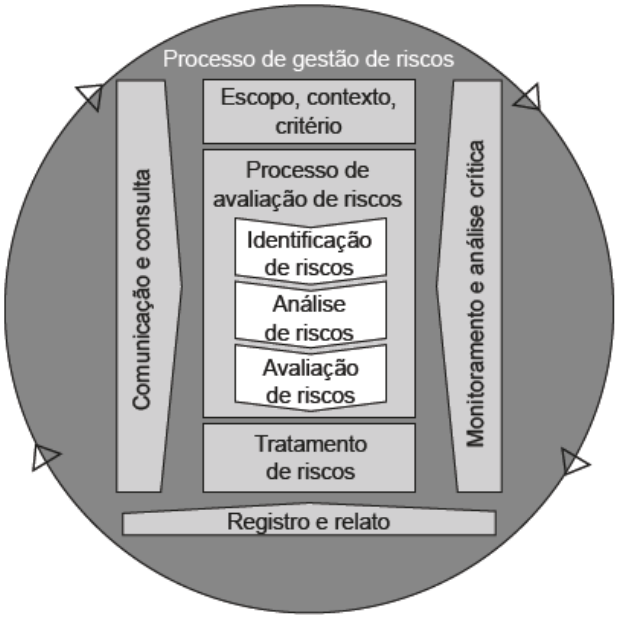

Figura 3.8: O processo de gestão de riscos
Fonte: ABNT NBR ISO 31000:2018

O processo proposto pela ABNT NBR ISO 31000:2018 é composto por sub-processos e atividades principaiss: comunicação e consulta; estabelecimento do contexto; avaliação dos riscos; tratamento, registro e relatório, e por último, processo de monitoramento e controle. Desta forma são estes os processos propostos para a gestão dos riscos inerentes à relação máquina-pessoa natural (PJe e DJe) como também na relação pessoa natural-máquina (advogados e escritórios) e abordados nas próximas seções do presente trabalho.

3.7 Gestão da privacidade da informação - Norma ABNT/NBR ISO 27701:2019

Lançada, no Brasil, em dezembro de 2019 a ABNT/NBR ISO 27701:2019 é uma Extensão da ABNT NBR ISO/IEC 27001 e ABNT NBR ISO/IEC 27002 para gestão da privacidade da informação.

A norma especifica os requisitos e fornece as diretrizes para o estabelecimento, implementação, manutenção e melhoria contínua de um Sistema de Gestão de Privacidade da Informação (SGPI) dentro do contexto da organização.

Interessante destacar os seguintes anexos da ISO 27701:

- Anexo C: Princípios da 29100 x controles da 27701;

- Anexo D: Mapeamento com a *General Data Protection Regulation* (GDPR);
- Anexo E: Mapeamento da ISO 27701 com a ISO 27018 e ISO 29151;
- Anexo G: Mapeamento com a Lei Geral de Proteção de Dados Pessoais (LGPDP).

Como extensão, a ISO 27701 depende das normas ISO 27001, ISO 27002 e ISO 29100 para sua implementação. De fato, a nova norma amplia os requisitos da ISO 27001, levando em consideração a proteção da privacidade dos titulares de dados pessoais.

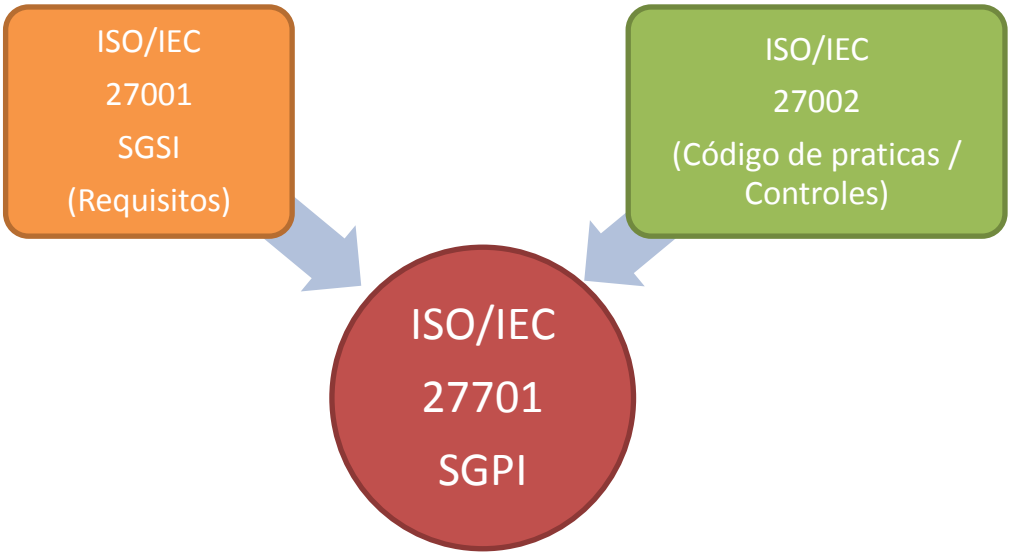

Figura 3.9*:* Relação da ISO 27701 com as ISOs 27001 e 27002
Fonte: autoria própria

3.7.1 ISO 27001 e ISO 27002

A norma 27.001 trata dos Requisitos de Sistemas de Gestão de Segurança da Informação (SGSI), tendo por objetivo a adoção de um conjunto de requisitos e controles para mitigar e gerir adequadamente os riscos de uma organização.

O SGSI busca aumentar o nível de segurança das informações e dos sistemas, tendo como pilares tradicionais a Disponibilidade, Integridade, Confidencialidade, Autenticidade e Não repúdio. Atualmente dois outros princípios são referenciados por alguns segmentos de mercado, a luz da governança: Autenticidade e Legalidade.

O modelo utiliza como ferramenta de suporte o tradicional cilo P-D-C-A, também conhecido como gestão de quatro passos: Planejar (*Plan*), Fazer (*Do*), Verficar (*Check*) e

Atuar (*Act*). Busca identificar, de forma continuada, oportunidades de melhorias, transformando em um processo contínuo.

A norma recomenda um modelo de 4 (quatro) camadas:

Figura 3.10*:* Camadas do modelo ISO 27001
Fonte: autoria própria

Assim, conforme o modelo:

- A **Política de Segurança da Informação** (PSI) apresenta as diretrizes, o que deve ser feito, de uma instituição no tocante à segurança da informação e está associada ao nível estratégico da mesma;
- Já o conjunto de **normas** regulamenta quais as regras, o tratamento dados aos riscos no sentido de gerenciar os mesmos estando associado ao nível tático;
- Os **procedimentos**, como fazer, é o registro do passo a passo para a execução das tarefas e está associado ao nível operacional. Este documento permite que, em caso de afastamento de algum colaborar por motivos, por exemplo de férias, doença ou mesmo desligamento, a instituição possa manter a continuidade operacional sendo executada por outra pessoa;
- Por fim, as **evidências**, referem-se à geração de métricas e registro de *Accountability*.

Por sua vez, a 27.002 apresenta recomendações, código de práticas e controles, de um conjunto de normas a serem desenvolvidas, tais como: controle de acesso, criptografia,

política de mesa limpa, uso da Internet, cópia de segurança, descarte de dados e informações entre outros. Orienta, também, quanto à estrutura desejável das mesmas:

1. Objetivo do documento;
2. Campo de aplicação;
3. Terminologia;
4. Objetivo da Norma;
5. Exceções;
6. Responsabilidades;
7. Punições;
8. Metodologia de revisão e atualização;
9. Referência;
10. Anexos.

3.8 NIST – *Privacy Framework*

Em janeiro de 2020 o *National Institute of Standards and Technology* (NIST) lança o NIST *Privacy Framework: a tool for improving privacy through enterprise risk management*.

Já na abertura do documento é apresentado o argumento central do seu objetivo a partir da construção das consequências da economia baseada em dados na vida dos titulares de dados (NIST, 2020, p. 2. tradução nossa)[18]:

> Por mais de duas décadas, a Internet e as tecnologias de informação associadas impulsionaram inovações, crescimentos econômicos e melhorias nos serviços sociais, sem precedentes. Muitos desses benefícios são alimentados por dados sobre indivíduos que fluem através de um ecossistema complexo. Como resultado, os indivíduos podem não ser capazes de entender as possíveis consequências para sua privacidade à medida que interagem com sistemas, produtos e serviços. Ao mesmo tempo, as organizações podem não perceber a extensão completa dessas consequências para os indivíduos, para a sociedade ou suas empresas, o que pode afetar suas marcas, seus resultados financeiros e suas perspectivas futuras de crescimento.

[18] For more than two decades, the Internet and associated information technologies have driven unprecedented innovation, economic value, and access to social services. Many of these benefits arefueled by data about individuals that flow through a complex ecosystem. As a result, individuals may not be able to understand the potential consequences for their privacy as they interact with systems, products, and services. Organizations may not fully realize the consequences either. Failure to manage privacy risks can have direct adverse consequences at both the individual and societal levels, with follow-on effects on organizations' brands, bottom lines, and *future prospects for growth. Finding ways to continue to derive benefits from data processing while simultaneously protecting individuals' privacy is challenging, and not well-suited to one-size-fits-all solutions*.

O *Privacy Framework* é um conjunto de medidas que tem por objetivo melhorar a privacidade por meio do gerenciamento de riscos corporativos, para permitir melhores práticas de engenharia de privacidade e ajuda as organizações a proteger a privacidade dos indivíduos. O *Framework* informa que foi elaborado de forma a ser utilizado por organizações de todos os portes e independente de qualquer tecnologia. Usa uma abordagem simplificada, adaptável às funções de qualquer organização no ecossistema de processamento de dados.

É estruturado em três partes principais:

- ***Core*** – trata-se de uma série de atividades de proteção da privacidade que viabiliza a comunicação e priorização das ações de proteção da privacidade em toda a instituição, desde as áreas executivas até as áreas operacionais.
- ***Profile*** – representas as atividades de privacidade atuais de uma organização ou os resultados desejados. Para desenvolver um perfil (*profile)*, a organização pode revisar todos os resultados e atividades no *core* a fim de determinar quais são os mais importantes com base nos objetivos de negócios ou missão, funções do ecossistema de processamento de dados, tipos de processamento de dados e necessidades de privacidade.
- ***Implementation tier*** – fornece um ponto de referência sobre como uma organização vê o risco à privacidade e verifica se possui processos e recursos suficientes para gerenciar esse risco. Os níveis (*tiers*) servem para refletir o nível de maturidade da organização no tocante à privacidade.

Por sua vez o *core* possui uma série de funções que permitem a gestão dos riscos à privacidade:

- ***Identify-P*** – desenvolve o entendimento da organização para o gerenciamento dos riscos à privacidade do titular de dados em função do tratamento dos mesmos;
- ***Govern-P*** – desenvolve e implementa a estrutura de governança organizacional de forma a permitir um entendimento contínuo das prioridades de gerenciamento de riscos da privacidade;
- ***Control-P*** – desenvolve e implementa atividades adequadas de forma a permitir que a organização ou o titular de dados realize o gerenciamento dos dados pessoais com bom nível de granularidade;
- ***Communicate-P*** – desenvolve e implementa atividades para garantir que tanto a organização como os titulares de dados tenham o real entendimento de como o tratamento de dados está associado aos riscos de privacidade;

- ***Protect***-P - desenvolve e implementa medidas adequadas de segurança. Trabalha a dimensão dos riscos cibernéticos é há uma sobreposição com os riscos à privacidade.

A Figura 3.11 apresenta a sobreposição entre às funções relacionadas aos riscos cibernéticos e as funções dos riscos à privacidade:

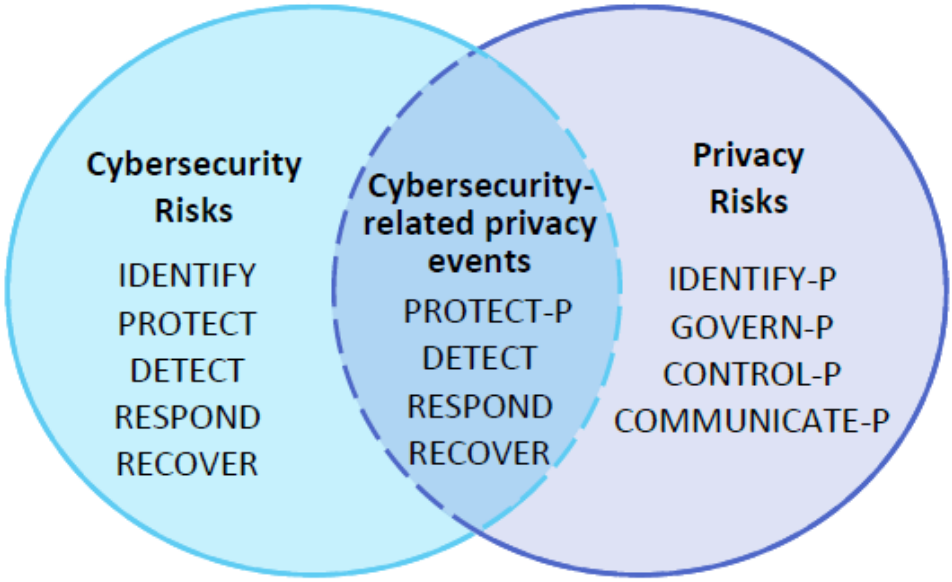

Figura 3.11: Funções para gerenciamento dos riscos cibernéticos e riscos à privacidade
Fonte: NIST, 2020

3.9 Conceitos gerais de *Big Data Analytics*

Na presente pesquisa o conceito de *Big data analytics* é utilizado a partir de duas perspectivas: *Big Data* (BD) e *Business Analytics* (BA). BD originalmente refere-se aos 3 Vs, ou seja, grande Volume, Velocidade e Variedade de dados dinâmicos que se diferencia dos modelos tradicionais de gerenciamento de dados (RUSSOM, 2011, p. 6). Atualmente outros dois Vs entraram na equação, sendo eles Veracidade e Valor. Dados verídicos são de grande relevância ao propósito da análise, sendo um ponto chave na agregação de valor ao processamento de dados. Já o Valor tem relação com os custos envolvidos no processo de coleta, análise e processamento dos dados, mas também de forma fundamental possui relação com a riqueza dos dados e o quanto agrega aos negócios.

Neste ponto não restam dúvidas que os dados do sistema judiciário são de extremo valor, uma vez que, em virtude do devido processo legal, são Verídicos e Valorosos, ao mesmo tempo que, com a maior capilarização do PJe / Dje (CNJ, 2019), estão em conformidade com Volume, Velocidade e Variedade.

Já o BA refere-se às tecnologias, habilidade e práticas utilizadas pelos setores corporativos em termos de estratégias e operações continuadas para a obtenção de *insights* que orientam os planos de negócios corporativos, envolvendo desde estratégias para o desenvolvimento de produtos, oferta de serviços, análise estatística, modelos de predição, além de técnicas de otimização dos processos (RUSSOM, 2011, p. 26).

BDBA oferecem novas oportunidades para empresas com a geração de valor a partir da análise massiva de dados. BDBA efetivamente auxiliam empresas a tomarem melhores decisões, aprimorar suas estratégias, operar de forma mais eficiente e a ter melhor performance financeira (WANG et al, 2016, on-line).

Pela quantidade de pessoas jurídicas, pessoas físicas, *lawtechs* e *legaltechs* vinculadas à Associação Brasileira de Jurimetria (ABJ)[19], não restam dúvidas que o mercado brasileiro já descobriu e vem atuando intensamente na extração e exploração do dados disponibilizados pelo PJe / DJe.

Entretanto, no presente trabalho, o foco concentra-se na exposição da privacidade em processos que tramitam sob sigilo de justiça e o risco da violação latente e efetiva da privacidade.

3.10 Conceitos gerais de estatística

Considerando que o presente trabalho possa despertar o interesse de operadores do direito, que nem sempre possuem familiaridade com a estatística, de forma a facilitar o entendimento de alguns trechos, da presente pesquisa, são apresentados neste tópico alguns conceitos fundamentais e básicos de estatística.

O poder de um teste estatístico corresponde à probabilidade de rejeitar corretamente a hipótese nula (COHEN, 1992a, p. 98) e depende de três aspetos, incluindo o tamanho do efeito/TDE, o nível de significância e o tamanho da amostra (COHEN, 1988, p.4; COHEN, 1992a, p. 98; COHEN, 1992b, pp. 99-100).

[19] Associação Brasileira de Jurimetria (ABJ) - https://abj.org.br/

No campo da probabilidade, um conceito fundamental está associado à curva normal ou curva em sino ou curva de erros ou, ainda, curva de Gauss, Figura 3.12. É a lei de probabilidade contínua mais importante na Análise Estatística. Essa curva descreve tanto fenômenos físicos como financeiros e tem uma propriedade que é enunciada como Teorema Central do Limite, dizendo que se pode aproximar outras distribuições sob determinadas hipóteses gerais pela normal quando o número de observações fica grande. (CAIRE, 2013, p. 15)

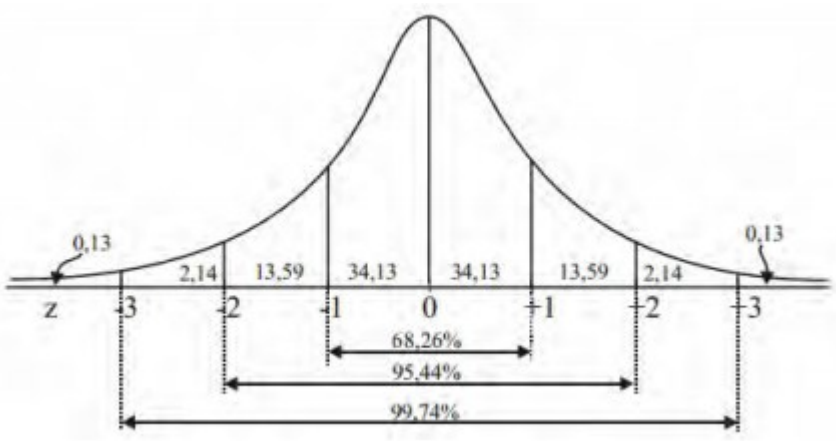

Figura 3.12*:* Curva Normal com $\mu = 0$ *e* $\sigma = 1$
Fonte: CAIRE, 2013, p. 11

A média (μ) posiciona o centro, enquanto o desvio-padrão(σ) fornece o grau de dispersão.

A curva normal possui 6 propriedades, a saber:

1. A curva é uma função de *x*, e o seu domínio estende-se de -∞ a +∞;
2. A curva é assintótica; nunca toca o eixo horizontal, e, portanto a função de *x* jamais se anula;
3. A área compreendida pela curva nesse intervalo é exatamente igual a 1;
4. A função tem um máximo, que corresponde ao seu ponto médio, ou seja, à média da distribuição;
5. A distribuição é simétrica em torno da média;
6. A curva tem dois pontos de inflexão, simétricos em relação à média . Esses pontos de inflexão são desvio padrão da distribuição normal.

Pela propriedade 3, a área delimitada pela distribuição normal padrão e o eixo horizontal é igual à probabilidade e pode ser calculada pela função densidade de probabilidade na integral $F(x) = P(X \leq x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{(y-\mu)^2}{2\sigma^2}} dy$, que no exemplo gera a tabela de

distribuição normal padrão, ou seja, tabela *z*. Sendo assim a probabilidade é algum valor entre 0 e 1.

Por sua vez a propriedade da área no tocante ao formato da distribuição permite a análise da normal em relação ao coeficiente de assimetria e curtose.

Assimetria indica o grau e o sentido do afastamento da simetria, de uma distribuição. Denotado por a_3, é obtida utilizando o segundo e o terceiro momentos centrados na média pela expressão $a_3 = \frac{m_3}{m_2\sqrt{m_2}}$, sendo que (BALIEIRO, sem ano, p. 47):

- **Assimétrica à direita ou negativa**: Se $a_3 < 0$, indicando que a maioria dos valores são maiores ou se localizam à direita da média aritmética, ou seja, a "cauda" à esquerda é alongada;
- **Simétrica**: Se $a_3 = 0$, indica que os valores estão uniformemente distribuídos em torno da média aritmética;
- **Assimétrica à esquerda ou positiva**: Se $a_3 > 0$, indicando que a maioria dos valores são menores ou se localizam à esquerda da média aritmética, ou seja, a "cauda" à direita é alongada.

A curtose é o grau de achatamento da curva de uma distribuição em relação a uma distribuição padrão, denominada curva normal. O coeficiente de curtose é denotador por a_4, sendo calculado a partir do segundo e do quarto momentos centrados na média pela expressão $a_4 = \frac{m_4}{m_2^2}$ (BALIEIRO, sem ano, p. 53):

- **Curva mesocúrtica ou normal**: Se $a_4 = 3$, é considerada como curva padrão, pois tem o mesmo achatamento que a distribuição normal ou Gaussiana;
- **Curva leptocúrtica**: Se $a_4 > 3$, é a curva mais alta do que o normal, apresenta o topo relativamente alto, o que significa que os valores se acham mais agrupados em torno da moda. A curva leptocúrtica também possui caudas grossas, devido a presença de valores de grande amplitude, com sinal negativo e/ou positivo;
- **Curva platicúrtica**: Se $a_4 < 3$, quando o valor da curtose é menor que zero, é uma curva mais baixa do que a normal, apresenta o topo achatado, significando que várias classes apresentam frequências quase iguais.

O nível de significância corresponde à evidência de que o fenômeno existe ou o risco de rejeitar erradamente a hipótese nula (COHEN,1988; 1992b).

Em 1925, Fisher defendeu o cálculo do valor da probabilidade (*p-valor*) da estatística e Neyman e Pearson em 1933 apresentaram o conceito de nível de significância, ou alfa ($\propto$). Para Fisher (1925; 1959), o valor de *p* de um teste estatístico mede a força da Evidência contra a Hipótese nula (*p*(E|H). O autor defendeu um valor de *p-valor* inferior a 0,05 como

um “valor suficientemente pequeno”, mas não como uma regra de ouro (SANTO, DANIEL, 2015, on-line).

No método de Neyman e Pearson, o valor de alfa associa-se ao nível de confiança com que o teste estatístico permite rejeitar a hipótese nula (H_0 = não há um determinado resultado), tendo sido acordados o valor fixos de 0,05 (i. e., os resultados são “significativos”, havendo 5% de probabilidade de obter os dados se a H_0 fosse verdadeira), 0,01 (i. e., os resultados são “altamente significativos”) ou 0,001 (i. e., os resultados são “muito altamente significativos”). Neste método, há que distinguir o erro do Tipo I (rejeição falsa de H_0) e o erro do Tipo II (manutenção falsa de H_0). A probabilidade do erro do Tipo I é $\propto$, e a probabilidade do erro do Tipo II é β. O poder é o complemento de beta (1 - β), definindo-se como a probabilidade de rejeitar corretamente a H_0 quando a hipótese alternativa é verdadeira (H_1 = há um determinado resultado), algumas literaturas nominam a hipótese alternativa como H_a(COHEN, 1992a; KLINE, 2004). Estes conceitos estão consolidados na Tabela 3.4:

Tabela 3.4: Teste de hipótese

	H_0 Verdadeiro	H_0 Falso
Aceitar H_0	Decisão correta $1-\propto$	Erro tipo II β
Rejeitar H_0	Erro tipo I $\propto = 0{,}05$	Poder do teste $1-\beta$

Fonte: autoria própria

Assim, a maioria dos pesquisadores normalmente espera obter um *p-valor* suficientemente pequeno (i. e., *p-valor* $< \propto$), para que possa rejeitar a hipótese nula, aceitando a hipótese alternativa.

Todos os testes estatísticos têm o seu próprio índice de Tamanho do Efeito (TDE) (COHEN, 1992b). Estes índices do TDE, para além de terem a vantagem de não dependerem do tamanho da amostra, informam sobre o significado dos resultados e consistem numa métrica comum para comparar resultados de estudos diferentes (SANTO, DANIEL, 2015, on-line).

3.11 Conclusão do capítulo

Verifica-se que a disciplina de gestão de riscos já possui vasto material e conceitos bem desenvolvidos, sendo "novidade" os modelos de gestão de risco da privacidade como a ABNT/NBR ISO 27701:2019 e o *Privacy Framework* do NIST, de 2020.

Também o conceito de de-identificação com suas espécies pseudonimização e anonimização ganha relevância no tocante à proteção da privacidade.

Foram apresentados os conceitos gerais de *Big Data Analytics* e estatísticos utilizados na pesquisa.

Por fim, áreas como da saúde que tratam dados pessoais sensíveis, possuem modelos próprios como o HIPAA que define 18 atributos sensíveis, a serem tratados como uma das condições necessárias para que uma base de dados possa ser considerada de-identificada.

No próximo capítulo, serão apresentadas as metodologias de pesquisa empregadas no presente estudos.

Capítulo 4

Metodologia de pesquisa

A proposta desta dissertação, em certa medida é interdisciplinar, uma vez que o suporte da computação realiza análise quanti-quali de dados tanto do judiciário, como de escritório de advocacia e dos próprios advogados, para constatar o risco de violação de certos direitos fundamentais da privacidade dos cidadãos. Nos parágrafos a seguir estão descritas informações referentes à metodologia da pesquisa.

4.1 Método da pesquisa

O método de produção de conhecimento ou método científico trata de como se produz o conhecimento. Envolve planejamento, organização e registro para pesquisas futuras.

Desse modo, produção de conhecimento e o conceito de ciência estão ligados, de forma direta, ao conceito de método científico (RICHARDSON, 1999), sendo o objetivo primordial da ciência alcançar a veracidade dos fatos (GIL, 2008). O conhecimento caracteriza-se pela investigação do porquê de um fenômeno, pela necessidade de explicar a ocorrência do fenômeno (RICHARDSON, 1999). Desta forma, o método é o conjunto das atividades, sistemáticas e racionais que, com maior segurança e economia, permite alcançar o objetivo (MARCONI, 2009). Caso os passos definidos no método forem executados, os resultados obtidos deverão ser convincentes (WAZLAWICK, 2014). Já método científico como o conjunto de procedimentos intelectuais e técnicos adaptados para se atingir o conhecimento trata-se da metodologia. A metodologia são as regras estabelecidas para o método científico, por exemplo: a necessidade de observar os fatos, a necessidade de formular hipóteses para buscar melhor compreensão dos fatos observados, a elaboração e uso de instrumentos, etc. (RICHARDSON, 1999). Para tanto, a pesquisa caracteriza-se como o processo formal e sistemático de desenvolvimento do método científico, cujo objetivo fundamental é descobrir respostas para problemas mediante o emprego de procedimentos científicos (RICHARDSON, 1999). Ela requer um tratamento científico e se constitui no caminho para conhecer a realidade ou para descobrir verdades parciais (MARCONI, 2009).

No contexto científico, a pesquisa pode ser classificada de acordo com diferentes critérios. Entre eles, é possível diferenciar tipos de pesquisa de acordo com sua natureza, objetivos ou procedimentos técnicos (WAZLAWICK, 2014). Para o presente estudo, a natureza

da pesquisa é aplicada pois tem como principal característica a aplicação da solução em problemas que ocorrem na realidade (MARCONI, 2009), tendo como motivação a necessidade de resolver problemas concretos (VERGARA, 2000). Neste sentido, a natureza de pesquisa aplicada se justifica pela aplicação da gestão de riscos, a partir da identificação dos pontos que podem expor, de forma extrema, a privacidade do cidadão, propondo, assim, aprimoramento tanto do PJe, como do DJe, além de procedimentos para a advocacia e escritórios advocatícios.

Em relação aos objetivos, a pesquisa é classificada como exploratória, devido à necessidade de estudo para definição dos principais riscos de exposição da privacidade e as principais abordagens de gestão de riscos para mitigar a exposição. Durante a pesquisa exploratória foi examinado um conjunto de fenômenos, buscando anomalias desconhecidas que possam ser a base para uma pesquisa mais elaborada (WAZLAWICK, 2014). São desenvolvidas com o objetivo de proporcionar visão geral, de tipo aproximativo, acerca de determinado fato (GIL, 2008).

Já em relação à estratégia, a classificação da pesquisa é bibliográfica, documental e também um estudo de caso. Na pesquisa bibliográfica estuda-se artigos, dissertações, teses, livros e outras publicações usualmente disponibilizadas por editoras indexadas (WAZLAWICK, 2014), além de leis e jurisprudência. Já a pesquisa documental foi realizada por meio de materiais que ainda não receberam tratamento analítico, ou seja, as fontes primárias como relatórios, arquivos obtidos em órgãos públicos, o questionário que será aplicado aos advogados etc. Por fim, estudo de caso, pois, o trabalho será desenvolvido a partir da análise do Processo Judicial Eletrônico (PJe), do Diário Judicial Eletrônico (DJe) e do escritório de advocacia de uma forma geral. O estudo de caso busca, dentre outros objetivos, explorar situações da vida real cujos limites não estão claramente definidos e descreve situações do contexto em que está sendo feita determinada investigação (GIL, 2008). Com isso, foi realizado um estudo de caso minucioso e exaustivo para propor a gestão de riscos como instrumento que irá auxiliar os advogados e escritórios de advocacia no aprimoramento da guarda e proteção de dados pessoais e dados pessoais sensíveis, bem como ajuste no PJe e DJe.

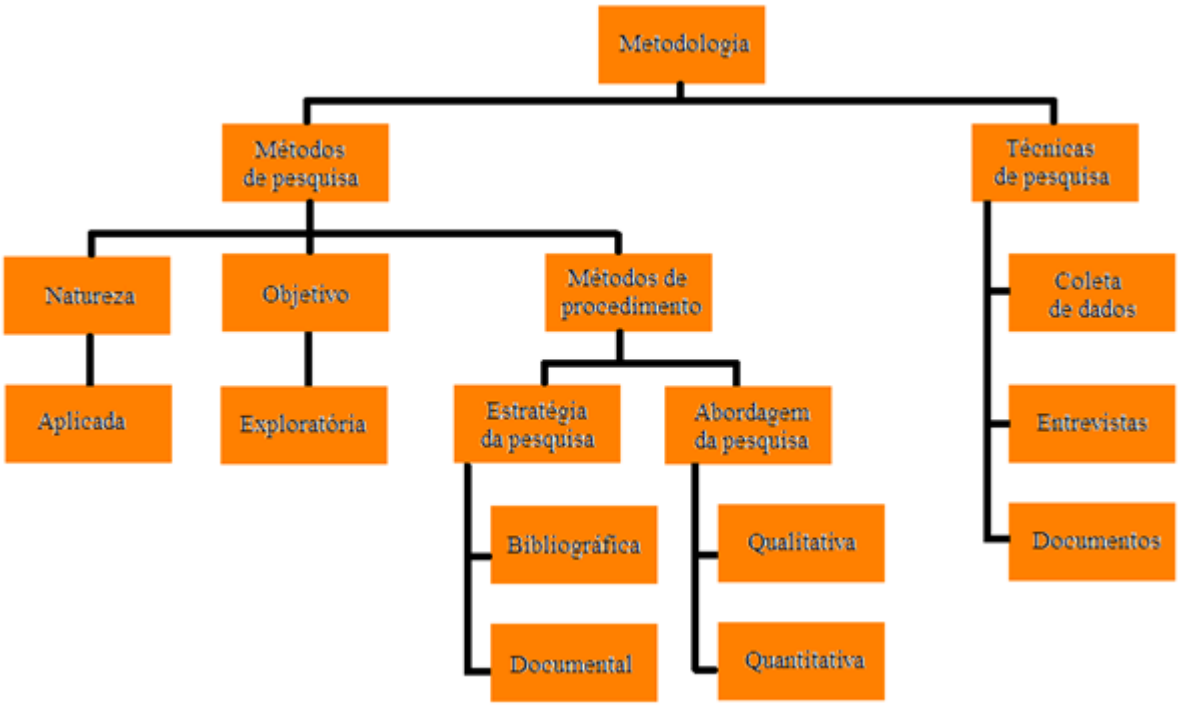

Figura 4.1: Método de pesquisa
Fonte: autoria própria

Além de pesquisa de campo foram empregadas duas abordagens distintas e complementares para alcançar os objetivos deste trabalho.

Inicialmente utilizou-se uma abordagem de pesquisa bibliográfica quantitativa, baseada na TEMAC descrita no tópico 3.1, com o objetivo de selecionar a literatura com maior Fator de Impacto (FI) para os temas aqui tratados.

Depois foi elaborado um questionário para coleta de dados, sendo que a acurácia do mesmo foi validada por meio do Modelo de Equações Estruturais (MEE).

Quanto a delimitação durante as pesquisas foram identificados diversas situações na quais há exposição das partes, pessoas e inclusive menores, mesmo mediante a decretação do sigilo de justiça. Entre os casos mais notórios podemos observar:

- Publicação nos Diários de Justiça Eletrônico de despachos onde consta o nome completo das partes, mesmo estando o processo sob sigilo de justiça. Em alguns casos, publica-se inclusive o perfil dos envolvidos nas plataformas digitais. No caso de crianças e adolescentes, considerados vulneráveis, há diversas normas que garantem o direito a proteção à privacidade, como os artigos 70, 100 e 45 do Estatuto da Criança e do Adolescente (Lei 8.069/90). Adicionalmente a Lei 13.431/2017 estabelece o sistema de garantia de direitos da criança e do adolescente vítima ou testemunha de violência;

- Estilo de redação do magistrado que prefere utilizar no corpo das sentenças, votos, decisões etc. o nome das partes ao invés de requerente/requerido, autor/réu, embargado/embargante etc. Nestes casos, de estilo, quando decretado o sigilo de justiça, o mesmo aplica-se à qualificação das partes no PJe/DJe e não aos nomes registrados e existentes no corpo do processo;
- Processos já arquivados definitivamente em formato físico que estão passando por digitalização e, desta forma, expondo de forma absoluta e ilimitada os nomes das partes, a natureza da demanda, as decisões proferidas no decorrer do processo em sua integralidade, o valor da causa e demais informações.

No que concerne as áreas relacionadas, a presente pesquisa contempla em maior ou menor profundidade os seguintes campos:

- Estruturação e uniformização de sistemas do judiciário empregados pelos tribunais para dar publicidade aos atos processuais, em especial os Diários de Justiça Eletrônica, que atualmente cada estado/tribunal adota o modelo de sua preferência;
- Mecanismos de *data scrap* [20]e coleta massiva de dados;
- Construção de *Big Data Analytics* a partir de dados do judiciário;
- Processamento de linguagem natural com o emprego de algoritmos de inteligência artificial (I.A.);
- Emprego de *machine-learning* e *deep-learning* para o processamento e perfilização de titulares de dados.

As técnicas dos campos de pesquisa citados, quando empregados em processos nos quais são determinados decretação do sigilo de justiça, porém não apresentam eficiência, sendo potenciais geradores de exposição e impacto na vida das partes.

[20] *Data scraping* - são técnicas computacionais, automatizadas, na qual um programa extrai dados de saída legível somente para humanos, geralmente de um serviço ou aplicativo. Os dados extraídos são, com frequência, utilizados na construção dos *Big Data Analytics* com o objetivo de serem minerados.

4.2 Estrutura da pesquisa

De acordo com Wazlawick, o método consiste na sequência de passos necessários para demonstrar que o objetivo proposto foi atingido. Trata-se de uma fase fundamental a ser executada logo após a definição do objetivo (WAZLAWICK, 2014).

Assim, o primeiro objetivo foi fundamentado por meio de pesquisa documental e bibliográfica. Para a pesquisa documental, foram realizadas buscas nos sites dos Tribunais de Justiça a fim identificar processos em sigilo de justiça. Para a pesquisa bibliográfica foi utilizada a metodologia baseada no enfoque meta-analítico. A partir do resultado dessa pesquisa, identifica-se os principais critérios a serem utilizados na gestão de riscos no tocante aos cuidados a serem adotados pelos advogados e escritórios de advocacia.

O segundo objetivo foi realizado por meio da análise dos processos sob sigilo de justiça para identificação do nível de exposição das partes. Ainda nesta etapa foi elaborado um questionário a ser aplicado aos operadores do direito, advogados, para identificar a cultura em proteção de dados. O resultado das duas pesquisas foi a base para construção de uma proposta de aprimoramento do PJe e do DJe, bem como de recomendações para a advocacia no tocante a proteção de dados pessoais e dados pessoais sensíveis.

Já em relação ao terceiro objetivos foi a proposição de mecanismos para reduzir a exposição das partes no tocante à privacidade em virtude da publicidade do processo, com foco nos que tramitam sob sigilo de justiça.

Por fim, o quarto objetivo constitui-se de construção de recomendações, gerais, de forma a elevar o nível de proteção de dados pessoais e dados pessoais sensíveis nos escritórios de advocacia e melhora da cultura de proteção de dados entre os advogados, estabelecendo desta forma um maior nível de prontidão em relação ao tema.

4.3 Teoria do Enfoque Meta Analítico Consolidado

Conforme o tópico 3.1 a TEMAC é uma metodologia de pesquisa bibliográfica, quantitativa, de caráter exploratório, que consegue integrar as exigências atuais da literatura a respeito de trabalhos científicos como precisão, robustez, validade, funcionalidade, tempo e custos, possibilitando, desta forma obter os melhores autores, artigos e revistas.

De acordo com as Figuras 3.1 e 3.2, pode-se observar que os principais trabalhos selecionados são de periódicos com conceitos QUALIS variando entre B1 a A1, em sua grande

maioria. Diante disso, a partir do estudo foi possível identificar as contribuições científicas mais relevantes para a presente pesquisa.

4.4 Metodologia[21] *Cross Industry Standard Process for Data Mining* (CRISP-DM)

Incialmente realizou-se uma análise de três possíveis metodologias para a mineração de dados do PJe e DJe:

- ***CRoss Industry Standard Process for Data Mining*** **(CRISP-DM)** – É uma metodologia que inicia o fluxo com o entendimento do negócio, avançando para as técnicas de mineração de dados até a fase de implementação. (MD SELAMAT, 2018, on-line)
- ***Sample, Explore, Modify, Model e Assess*** **(SEMMA)** – A principal característica desta metodologia é focar principalmente nas tarefas de criação do modelo, deixando as questões de negócio de fora. Por outro lado, possui várias semelhanças com a CRISP-DM; (MD SELAMAT, 2018, on-line)
- ***Knowledge Discovery in Databases*** **(KDD)** – Essa metodologia, diferente das anteriores, não foca em questões de negócio ou geração de modelos, mas sim na descoberta de conhecimentos a partir dos dados. (FAYYAD, 1996ª, on-line)

Para a mineração de dados do PJe e DJe optou-se por utilizar a metodologia *CRoss Industry Standard Process for Data Mining* (CHAPMAN; etc, 2000), em virtude de ser um padrão de fato em mineração de dados, por não ser proprietário, poder ser aplicada a qualquer tipo de negócio e não ter dependência de ferramentas para ser executada. Conforme apresentado na Figura 4.2, o modelo CRISP-DM é composto de 6 fases:

[21] Em relação ao CRISP-DM alguns autores preferem usar o termo modelo, já outros metodologia. No presente trabalho optou-se pela última.

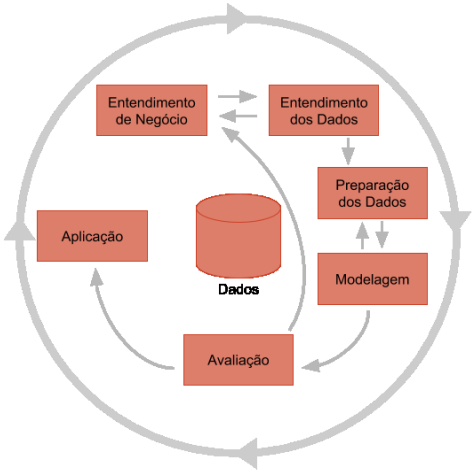

Figura 4.2: Metodologia CRISP-DM
Fonte: CRISP-DM

1. **Entendimento do negócio**: esta fase parte do entendimento do negócio e dos objetivos do projeto de *Data Mining*. É uma perspectiva negocial, ou seja, é orientado para resolver um problema de negócio;
2. **Entendimento dos dados:** esta etapa consiste na captura, entendimento até a identificação de problemas relacionados à qualidade dos dados, tais como: *missing values*, *outliers* entre outros, ou seja, busca-se ter noção de como o conjunto de dados é composto. É também nesta fase que se formula a hipótese a partir da análise dos dados;
3. **Preparação dos dados:** é a fase de preparação da modelagem dos dados. Envolve as atividades para construir o conjunto e/ou ajustar os dados a partir dos dados iniciais coletados, com foco em seleção de atributos, limpeza, calibragem, balanceamento, construção, integração e formatação dos dados de entrada;
4. **Modelagem:** esta etapa envolve a aplicação de diferentes técnicas de modelagem. Esta etapa pode ser repetida diversas vezes, inclusive revisitando etapas anteriores. Seus parâmetros são calibrados para obtenção de valores ótimos, tais como: acurácia, sensibilidade, especificidade etc.;
5. **Avaliação**: os modelos obtidos são avaliados para validar se atendem as necessidades do negócio;
6. **Aplicação:** o conhecimento obtido com os modelos gerados deve ser aplicado na organização, isso pode envolver desde a apresentação dos resultados da modelagem para a tomada de decisão até a aplicação do modelo em um outro conjunto de dados.

4.5 Modelo de Equações Estruturais – Validação do questionário

A coleta de dados reais e precisos é de fundamental importância para a condução de um projeto de pesquisa. Neste sentido investigar a relação pessoa natural-máquina no que diz respeito aos advogados e escritórios de advocacia, permitiu medir a relevância e pertinência do tema para a advocacia. A coleta dividir-se-á em duas fases:

- **Fase 1**: Aplicação do questionário a um grupo de advogados com o objetivo de coletar respostas para os cálculos de validação do questionário por meio do modelo de equações estruturais;
- **Fase 2**: Aplicação do questionário a um grupo de advogados afim de coletar dados para análise.

Na fase 1 foi solicitado que advogados e estudantes do curso de direito respondessem as questões elaboradas sobre o tema, no sentido de coletar um número de respostas suficientes para a validação de acurácia do questionário, por meio do modelo de equações estruturais (MEE), sendo que tal abordagem proporciona ao pesquisador a flexibilidade de poder desenvolver (CHIN. 1998, p297):

- Modelos de relações múltiplas entre predições e variáveis;
- Construtos não observáveis de variáveis latentes;
- Erros no modelo na medida das variáveis observadas;
- Testar estatisticamente a priori a teoria e comparar as conjunturas e hipóteses com os dados empíricos.

Já na fase 2, o questionário foi disponibilizado, on-line, para os estudantes e operadores do direito para ser respondido.

Como já dito, o questionário foi de natureza quali-quanti. Nas palavras de Sônia Vieira, "A pesquisa qualitativa não é generalizável, mas exploratória, no sentido de buscar conhecimento para uma questão sobre a qual as informações disponíveis são, ainda, insuficientes". (VIEIRA, 2009, p. 6)

A elaboração do questionário pautou-se em metodologia apresentada por Vieira (VIEIRA, 2009), onde se destaca os seguintes passos: Elaboração das questões e definição das escalas de mensuração.

4.5.1 Elaboração das questões

O questionário foi elaborado considerando-se uma das premissas que norteia a presente pesquisa, que diz respeito aos pilares "centrais" para alcançar um adequado nível de prontidão em proteção da privacidade, que são: Pessoas, Tecnologia e Processos.

Em uma primeira aproximação do tema percebe-se algumas vulnerabilidades em relação aos comportamentos dos operadores do direito e sua relação com tecnologias. A grande maioria dos operadores do direito, por óbvio, não são profissionais de tecnologia, apesar de serem 100% dependentes, principalmente em função do Processo Judicial Eletrônico. Neste sentido, é comum não possuírem ferramentas de segurança além de adotarem comportamento de risco nos hábitos de navegação na Internet, realizando o *downloading* de programas suspeitos, compartilhando senhas e até mesmo do *token* com seus certificados pessoais, descuidos na interação com e-mails recebidos. Estas são apenas algumas das práticas que oferecem extremo perigo em um mundo conectado em rede.

Ao mesmo tempo, a linha divisória entre o ambiente doméstico e o ambiente corporativo, em termos de acesso, mostra-se cada vez mais difusa, assim brechas de segurança em que equipamentos domésticos, *gadgets*, entre outros acabam por vulnerabilizar o ambiente corporativo.

Diante deste cenário, o modelo de pesquisa adotado, com base na *Technology Threat Avoidance Theory* (TTAT) (LIANG e XUE, 2009), para examinar a relação entre tecnologias, pessoas, processos e prontidão é apresentado na Figura 4.3:

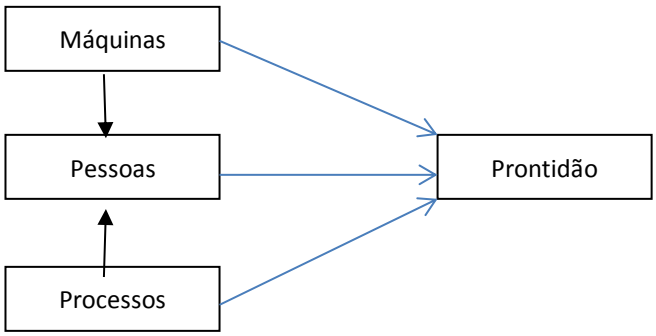

Figura 4.3: Relação entre tecnologias, pessoas, processos e prontidão
Fonte: Autoria própria

A partir do modelo apresentado foi elaborado um questionário com uma série de questões agrupadas pelos *constructos* apresentados na Tabela 4.1:

Tabela 4.1: Questionário para coleta de dados

Tecnologia	Tec1	Eu consigo, com sucesso, identificar vulnerabilidades tecnológicas em meus dispositivos
	Tec2	Eu consigo, com sucesso, proteger meu certificado digital
	Tec3	Eu consigo, com sucesso, identificar uma mensagem de SPAM
	Tec4	Eu consigo, com sucesso, identificar uma rede Wi-Fi não legítima
	Tec5	Eu consigo, com sucesso, usar recursos de criptografia
Pessoas		
	Pess1	Eu consigo, com sucesso, aprender sobre segurança cibernética, mesmo sem ter lido nada a respeito antes
	Pess2	Eu consigo, com sucesso, aprender sobre segurança cibernética, somente tendo referência de informações
	Pess3	Eu consigo, com sucesso, aprender sobre segurança cibernética, mesmo sem ajuda de ninguém
	Pess4	Eu consigo, com sucesso, apender sobre segurança cibernética, se tiver muito tempo disponível
	Pess5	Eu consigo, com sucesso, aprender sobre segurança cibernética, mesmo sem ninguém me ensinar
Processos		
	CultOrg1	Na sua instituição existe uma pessoa dedicada a cuidar das questões de segurança da informação
	CultOrg2	Existe uma verificação de vida pregressa em relação às pessoas que acessam e manipulam dados da instituição
	CultOrg3	A instituição possui uma política de segurança da informação
	CultOrg4	A instituição possui uma política que detalha as regras para formação se senhas
	CultOrg5	Os colaboradores recebem treinamento sobre cuidados com segurança da informação
	CultOrg6	Você acredita que o nível de segurança, da informação, em seu ambiente de trabalho/estágio/estudo é elevado
	CultOrg7	Existe compartilhamento de usuário/senha para acesso a informações
	CultOrg8	Existe controle de acesso aos sistemas que armazenam dados sensíveis
	CultOrg9	Existe controle de uso das portas USB dos equipamentos
	CultOrg10	Existe controle de acesso a sites da Internet
	CultOrg11	Para acessar a rede Wi-Fi, existe uma senha única para todos
Prontidão		
	Pront1	Todos possuem responsabilidade de proteger seu computador de ataques e garantir que as informações estão armazenadas de forma segura
	Pront2	Se um atacante quer invadir um sistema, eu posso fazer muito pouco para impedi-lo
	Pront3	Segurança da informação é algo que deve ser foco somente das equipes técnicas
	Pront4	Eu consigo, com segurança, reconhecer uma mensagem de e-mail suspeita

Fonte: Autoria própria

Os domínios de respostas são apresentados na Tabela 4.2:

Tabela 4.2: Domínio de resposta do questionário

Questão	Domínio
Tec1, Tec2, Tec3, Tec4 e Tec5 CultOrg1, CultOrg2, CultOrg3, CultOrg4, CultOrg5, CultOrg7, CultOrg8, CultOrg9, CultOrg10 e CultOrg11	Sim ou Não
Pront1, Pront2, Pront3 e Pront4	Concordo completamente Concordo Nem discordo, nem concordo Discordo Discordo completamente
CultOrg6	Sim Não Não sei responder Outra resposta

Fonte: Autoria própria

4.5.2 Validação do questionário

O Modelo de Equações Estruturais (MEE) possui duas subdivisões, quais sejam: mensuração e estrutural. Sendo que o modelo de mensuração indicada como os construtos são mensurados pelas variáveis observadas, pelos prismas da validade e da confiabilidade. Por outro lado, o modelo estrutural indica as relações de causa e efeito entre os construtos, dando relevo aos efeitos causais e o total de variância não explicada (PEREIRA, 2014, p. 10). Nesta seção, apresentam-se os resultados obtidos nos dois modelos, por meio de análise realizada com auxílio do *software SmartPLS* 3. O número de respondentes para esta fase foram 30 (trinta) pessoas divididas entre advogados e estudantes de direito. As respostas para a validação encontram-se no apêndice.

A. Resultados do Modelo de Medida

Nesta etapa, verificou-se o nível de confiabilidade dos indicadores em termos de capacidade de explicação da variável. Primeiramente, o conjunto de dados foi avaliado levando-se em consideração os parâmetros de curtose e de assimetria. Os indicadores demonstraram-se consistentes, visto que as distribuições das variáveis observadas não foram grandes (HAIR, 2014).

Para análise das cargas fatoriais foi estabelecido que as cargas exteriores das construções reflexivas deveriam estar acima do limite de valor de 0,70 de carregamento. Dessa forma, foram excluídas do modelo as variáveis: PESS1, TEC1, TEC3, TEC4, PRONT1, PRONT4, PRONT5, CUTORG2, CUTORG7, CUTORG9, CUTORG10 E CUTORG11.

O modelo resultante, com as outras variáveis restantes, apresentou os valores de Variância Média Extraída (AVE) e Confiabilidade, medidas por meio do Alfa de *Cronbach*, Confiabilidade Composta e Rho-A, apresentados na Tabela 4.3.

Tabela 4.3: Descrição das variáveis da utilidade percebida

Variáveis / Critérios	AVE	Alfa de Cronbach	Confiabilidade Composta	Rho-A
Pessoas	0.683	0.843	0.896	0.860
Processos	0.638	0.888	0.913	0.898
Prontidão	0.701	0.608	0.822	0.798
Tecnologia	0.764	0.693	0.866	0.701
Limite	0.5	0.7	0.6	0.7

Fonte: Autoria própria

Conforme literatura (HAIR, 2014), considerando a natureza exploratória do estudo, a confiabilidade composta foi escolhida como parâmetro para avaliação. Os valores da confiabilidade composta de 0,60 a 0,70 são aceitáveis, sendo que em estágios mais avançados de estudo, como no presente trabalho, valores entre 0,70 e 0,90 são considerados satisfatórios (HAIR, 2014).

A AVE é a porção dos dados explicada por cada um dos construtos, respectivos aos seus conjuntos de variáveis ou, em média, quanto as variáveis correlacionam-se positivamente com os seus respectivos construtos. Para AVEs maiores que 0,50 admite-se que o modelo converge para um resultado satisfatório (HAIR, 2014), (FORNELL e LARCKER, 1981).

A Tabela 4.4 apresenta os resultados da validade discriminante observando os critérios de Fornell-Larcker (HAIR, 2014), (FORNELL e LARCKER, 1981), considerada uma abordagem mais conservadora para avalição da validade discriminante dos itens.

Tabela 4.4: Validade discriminante

	Pessoas	Processos	Prontidão	Tecnologia
Pessoas	0.827			
Processos	0.303	0.799		
Prontidão	0.488	0.563	0.837	
Tecnologia	0.640	0.256	0.435	0.874

Fonte: Autoria própria

A Tabela 4.5 apresenta a validade discriminante dos indicadores em termos de cargas cruzadas.

Tabela 4.5: Validade discriminante em termos de cargas cruzadas

	Pessoas	Processos	Prontidão	Tecnologia
CutOrg1	0.148	0.893	0.464	0.019
CutOrg3	0.409	0.845	0.431	0.299
CutOrg4	0.234	0.870	0.456	0.112
CutOrg5	0.067	0.725	0.492	0.297
CutOrg6	0.009	0.719	0.172	0.104
CutOrg8	0.364	0.721	0.500	0.299
Pess2	0.775	0.203	0.197	0.604
Pess3	0.904	0.236	0.427	0.661
Pess4	0.740	0.272	0.527	0.326
Pess5	0.875	0.294	0.456	0.502
Pront2	0.043	0.403	0.724	0.270
Pront3	0.615	0.529	0.936	0.430
Tec2	0.464	0.333	0.437	0.856
Tec5	0.644	0.129	0.332	0.892

Fonte: Autoria própria

Verifica-se o respeito aos critérios delimitadores para todas as variáveis analisadas. O mesmo ocorre para a análise das cargas cruzadas dos indicadores, os quais se mantiveram superiores às cargas cruzadas dos outros construtos.

B. Resultados do Modelo Estrutural

O modelo estrutural foi avaliado considerando a colinearidade, significância e relevância dos relacionamentos do modelo estrutural, avaliação da qualidade e capacidade preditiva. Nesse contexto, um valor de tolerância de inferior a 0,20 ou superior a 5, indicam um potencial problema de colinearidade (HAIR, 2009). Como apresentado na Tabela 4.6, a análise da colinearidade por meio do fator de variância (VIF) obteve valores considerados satisfatórios indicando que todas as variáveis podem permanecer no modelo (HAIR, 2009).

Tabela 4.6: Resultados de VIF obtidos

Critério	Prontidão
Pessoas	1.754
Processos	1.109
Prontidão	-
Tecnologia	1.705
Limite	5

Fonte: Autoria própria

A avaliação do modelo interno considera a avaliação dos coeficientes de determinação de Pearson (R^2). Os valores de R^2 variam de 0 a 1, com níveis maiores indicando maior exatidão preditiva, sendo difícil apresentar regra prática para valores aceitáveis de R^2, já que eles dependem da complexidade do modelo e da área de pesquisa (HAIR, 2014).

Hair, Ringle e Sarstedt (HAIR, 2011) exemplificam que $R^2 = 0,20$ é considerado alto em disciplinas tais como o comportamento do consumidor e que em estudos de pesquisa de marketing, R^2 de 0,75, 0,50 ou 0,25 por variável latente endógena do modelo estrutural valores estes que, podem configurar a força da variável como substancial, moderada ou fraca, respectivamente.

No presente trabalho, a variável endógena do modelo Prontidão (PRONT) apresentou o valor de $R^2 = 0,443$ e o valor de R^2 ajustado de 0,379, considerados adequados e com efeito moderado dada a natureza do estudo.

Na sequência, verificou-se a significância dos coeficientes de caminho. Conforme Hair et al. (HAIR, 2014), os coeficientes de caminho têm valores padronizados entre -1 e 1, e os valores estimados próximos a 1 representam relação positiva forte (e vice-versa para valores negativos),

que são quase sempre estatisticamente significativos (isto é, diferente de zero na população) e quanto mais próximos de zero, mais fraca é a relação, conforme a Tabela 4.8.

Outro teste importante apresentado na Tabela 4.7, é o de Tamanho do Efeito (f^2) ou indicador de Cohen (RINGLE, 2014), cujo valor é obtido pela inclusão e exclusão de construtos do modelo (um a um), de modo a avaliar quanto cada construto é "útil" para o ajuste do modelo. Valores próximos a 0,02, 0,15 e 0,35 são considerados pequenos, médios e grandes, respectivamente (RINGLE, 2014).

Tabela 4.7: tamanho do efeito (f²) e coeficientes de caminho (ß)

Caminho	f^2	β
Pessoas ➔Prontidão	0,064	0,297
Processos ➔Prontidão	0,322	0,417
Tecnologia ➔Prontidão	0,027	0.206

Fonte: Autoria própria

Mediante a realização do procedimento de *bootstrapping*, foi possível encontrar os resultados para os coeficientes de caminho e valores da estatística t e *p-value*, conforme Tabela 4.8.

Tabela 4.8: Significância estatística dos coeficientes de caminho

	Beta do caminho	Valor Médio	Desvio Padrão	Estatística *t*	*P-Value*
Pessoas ➔ Prontidão	0.251	0.246	0.258	0.973	0.331
Processos ➔ Pessoas	0.149	0.170	0.174	0.855	0.393
Processos ➔ Prontidão	0.446	0.456	0.153	2.907	0.004
Tecnologia ➔ Pessoas	0.602	0.600	0.138	4.359	0.000
Tecnologia ➔ Prontidão	0.160	0.152	0.202	0.795	0.427

Fonte: Autoria própria

A literatura indica que os valores ideais para estatística *t* devem ser ≥ 1,96. e valores de *p-value* ≤ 0,05 (HAIR, 2014). Neste contexto, o indicador Tecnologia (TEC) apresentou estatística $t = 4,359$ e *p*-value = 0,000 respectivamente, em aderência com os valores preconizados.

C. Modelo

O modelo final, com os indicadores, as variáveis latentes e os valores de β para as relações causais entre elas para determinação da satisfação são apresentados na Fig. 4.8. Por meio da análise da Tabela 4.10, combinada com o modelo apresentado na Fig. 4.8, observa-se que o módulo do valor padronizado do coeficiente β para Tecnologia (4,359) é o maior entre os

preditores, ou seja, a relação pessoa natural-máquina. Em seguida, o coeficiente *β* para Processo (2,907) indicando ser essa a variável mais importante na previsão da prontidão.

Os resultados também mostram que Processos é significativo (p-value = 0,002) ao nível de significância de 0,05. Isso indica que existe relação significativa entre Processos e Prontidão.

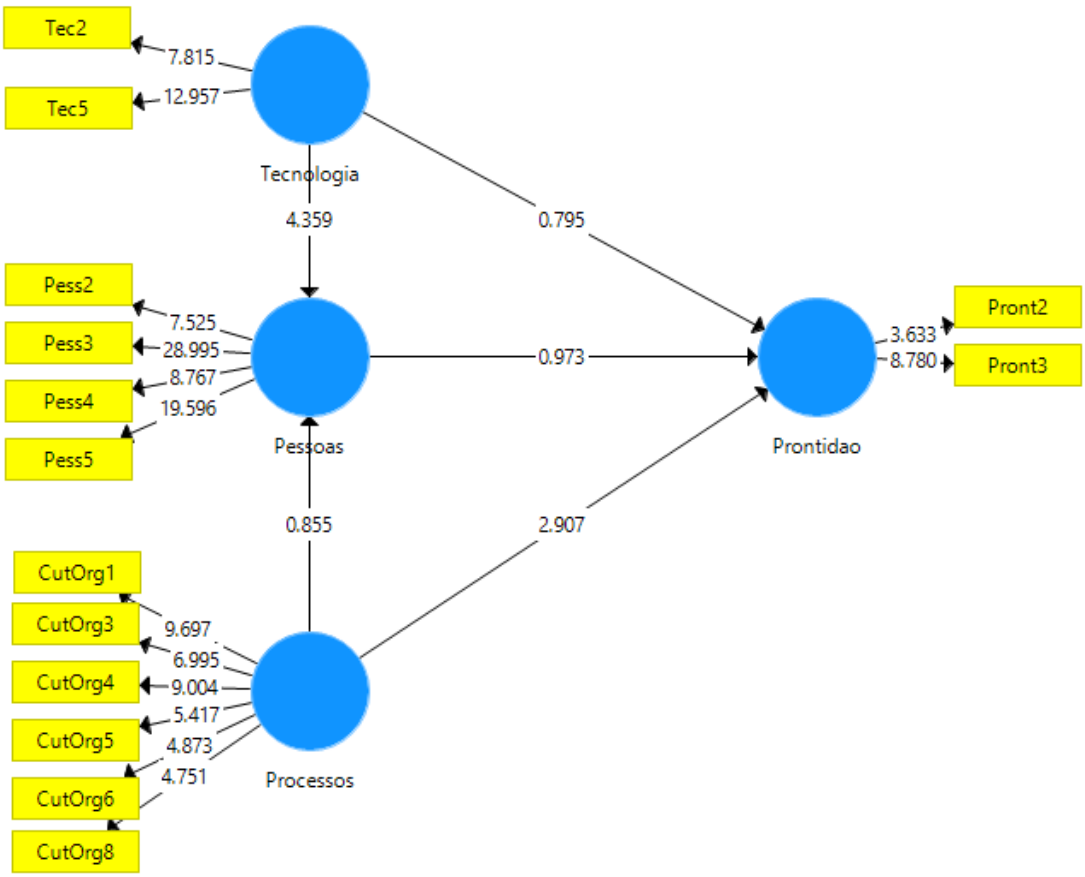

Figura 4.4: Modelo Final com cargas fatoriais e valor beta
Fonte: Autoria própria

A partir do ponto de vista estatístico, o estudo atendeu todos os critérios para validação do modelo, além de apresentar um bom nível de ajuste, mostrando-se estatisticamente adequado. Do ponto de vista de confrontação das hipóteses, entende-se que estas foram confirmadas.

Assim, fica demonstrado, no presente modelo, a força da relação pessoa natural-máquina e a importância de processos para a prontidão.

4.6 Conclusão do capítulo

Neste capítulo foi apresentada a metodologia de pesquisa que envolve planejamento, organização e registro para pesquisas futuras. Trata-se de um estudo interdisciplinar, que envolve principalmente a área da computação em proximidade com o direito.

Foram apresentadas as principais metodologias empregadas na pesquisa. Em primeiro momento utilizou-se a TEMAC para a identificação, levantamento e análise da bibliografia com maior quantidade de citações. Também foram utilizados o metodologia CRISP-DM, para mineração de dados, além de Equações Estruturais (MEE) para validar o questionário de levantamento de dados junto aos advogados e escritórios de advocacia.

O próximo capítulo apresenta o resultado das pesquisas de campo tanto na relação máquina-pessoa natural, como na interação pessoa natural-máquina.

Capítulo 5

Pesquisas de Campo

Ao propor alternativas para proteção de dados pessoais foram realizadas duas vertentes de análise na presente pesquisa.

Na primeira vertente da pesquisa de campo foi analisada a relação máquina-pessoa natural, ou seja, em que medida a publicidade de dados pessoais presentes nos processos judiciais podem expor a privacidade do titular de dados, uma vez que eles são passíveis de serem tratados de forma automatizada por algoritmos de predição. Como elemento de pesquisa foi realizado, de forma automática a análise de todas as publicações do DJe do TJDFT no ano de 2019, ou seja, 248 documentos, empregou-se a metodologia CRISP-DM.

Já na segunda vertente de pesquisa de campo analisou-se a relação pessoa natural-máquina, tratando dos riscos existentes nos processos e pessoas no que concerne a salvaguarda de dados pessoais em posse de advogados e escritórios de advocacia. Para tanto, foi realizada uma pesquisa exploratória qualitativa disponibilizada, on-line, por meio de questionário, respondido por 82 advogados. Na sequência os dados foram analisados e a partir das vulnerabilidades identificadas elaborou-se um conjunto de recomendações de boas práticas, tanto na dimensão de processos como comportamental, que atua no pilar pessoas.

Por razões de preservação da privacidade não serão divulgados nomes ou outras informações relevantes dos titulares de dados. Nos casos de nomes ou fragmentos de dados que possam tornar o titular de dados identificado ou identificável os mesmos serão substituídos por uma máscara de asteriscos, Ex.: (******).

5.1 Relação máquina-pessoa natural (DJe)

Na primeira vertente da pesquisa de campo foi analisada a relação máquina-pessoa natural, ou seja, em que medida a publicidade de dados pessoais presentes nos processos judiciais podem expor a privacidade do titular de dados uma vez que os mesmos são passíveis de serem tratados de forma automatizada por algoritmos de predição. Como elemento de pesquisa foi realizada, de forma automática, a análise de todas as publicações do DJe do TJDFT do ano de 2019, ou seja, 248 documentos, empreganso-se a metodologia CRISP-DM. A Figura 5.1 apresenta o objeto de estudo nesta primeira vertente:

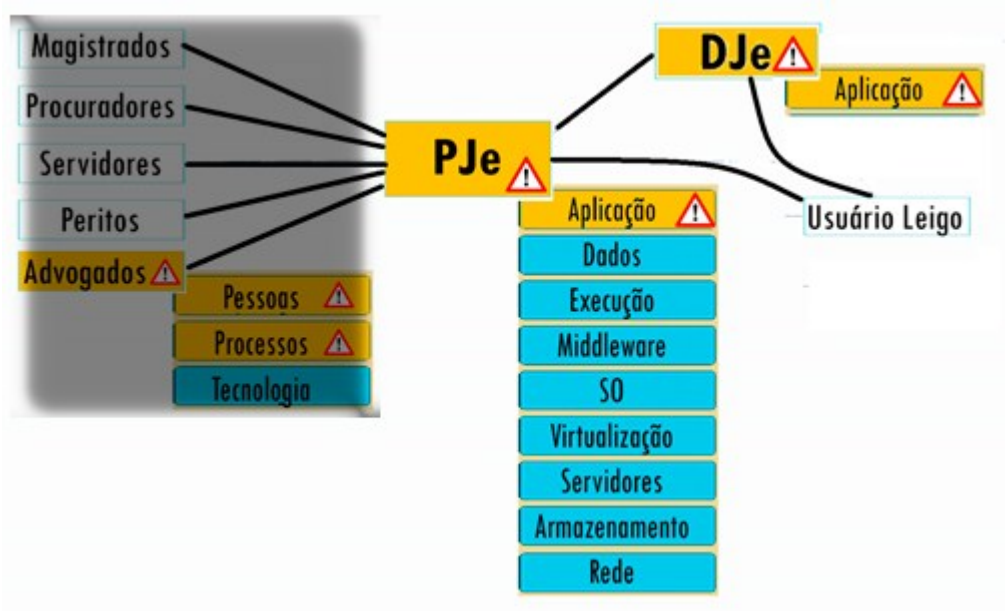

Figura 5.1: Pontos de investigação da pesquisa
Fonte: autoria própria

5.1.1 Entendimento do negócio e Coleta de dados no DJe

Na fase de Entendimento do Negócio, da metodologia CRISP-DM, traçou-se o objetivo do problema a ser resolvido, à luz da Mineração de Dados, em alinhamento com os problemas do presente trabalho, qual seja, identificação de situações de exposição da privacidade, mesmo estando o processo sob sigilo de justiça. Foi identificado que os dados estavam disponíveis no portal do TJDFT no formato *Portable Document Format* (PDF).

A coleta de dados no DJe foi realizada com um programa desenvolvido na linguagem R, sendo que os recursos computacionais empregados estão representados e descritos na Figura 5.2:

Figura 5.2: Estrutura utilizada para *scrap* do DJe do TJDFT
Fonte: autoria própria

A Tabela 5.1 apresenta o tempo gasto para a realização do *download* dos arquivos em dois diferentes cenários, sendo que foi realizado o *download* de todos os arquivos publicados

no ano de 2019. Os arquivos são disponibilizados no DJe no formato .PDF e possuem tamanho médio de 16.1 Mbytes:

Tabela 5.1: Tempo de processamento na coleta de dados

Período	QTD arquivos	Conexão 4G 10MB download/ 10MB upload Duração	Conexão ADSL 10MB download/ 1MB upload Duração
01 dia (1° publicação 2019)	01	18”11ms	1”38ms
01 semana (1° semana Jan/2019)	05	3’25”98ms	10”81ms
01 mês (Jan./2019)	22	8’37”69ms	3’16”47ms
01 ano (2019)	248	3h34’22”	1h07’44”

Fonte: autoria própria

Código fonte do *script* utilizado para a realização do *download* dos dados:

Tabela 5.2: Código fonte do programa de *scrap*

```
###
# Programa: scrap_DJe_TJDFT.r
# Versão: 1.0
# Autor: Frank Ned – info at santacruzadv.com
# Descrição: Realiza o download de TODOS os arquivos do DJe do TJDFT
#


#
# Total de publicações em 2019
#
TotalPub_2019 <- 248

#
# Contador de controle do loop e usado no nome dos arquivos
#
Cont <- 1



#
# Loop para realizar o download, automatizado, dos arquivos
#
while ( Cont <= TotalPub_2019)
        {
#
# Composição do nome do arquivo - Concatenação de Cont com .pdf (n.pdf)
#
    NFile <- paste(as.character(Cont), "pdf", sep=".")
    print(NFile)

#
# Composição da URL para download
#
        url <- paste("https://pesquisadje-api.tjdft.jus.br/v1/diarios/pdf/2019/", NFile, sep="")
```

```
#
# Composição do ´caminho do diretório onde será salvo o arquivo
#
            dst <- paste("/home/santacruz/UnB/ppca/DJe/", NFile, sep="")

#browseURL(url, encodeIfNeeded = T, browser = "firefox.exe")

#
# Download do arquivo
#
            download.file(url, dst, mode = "wb")

#
# Incremento do contador (Cont) de controle do loop
#
     Cont <- Cont + 1
           }
```

Fonte: autoria própria

5.1.2 Entendimento dos dados

Na fase de Entendimento dos dados houve, inicialmente, a necessidade de tratamento de todas as 248 publicações do DJe do TJDFT, do ano de 2019, cada uma com uma média de 2.000 páginas, onde as mesmas foram convertidas do formato .pdf para o formato .txt, com o objetivo de facilitar o processamento nas fases seguinte da análise dos dados. Para tanto utilizou-se o programa desenvolvido na linguagem R, sendo os tempos de processamento apresentados na Tabela 5.3 e o programa na Tabela 5.4:

Tabela 5.3: Tempo de processamento na conversão de pdf para txt

Corpus	**Processador: Intel Core I3-6006U CPU @ 2.00GHz Memória: 4GB**	**Processador: Intel Core I7-7700K CPU @ 4.20GHz Memória: 8GB**
248 arquivos pdf; média de 2.000 páginas por arquivo.	3h47’11”	35’23”50ms

Fonte: autoria própria

Tabela 5.4: Código fonte do programa de conversão do formato .pdf para .txt

```
###
# Programa: converte PDF2TXT.r
# Versão: 1.0
# Autor: Frank Ned – info at santacruzadv.com
# Descrição: Realiza a conversão de arquivos no formato .pdf para .txt (para mineração de dados)
#


#
# Diretório com os arquivos PDF
```

```
#
dest  <-  "/home/santacruz/UnB/ppca/DJe/"

#
# Criação de um vetor com o nome dos arquivos
#
arquivos  <-  list.files(path = dest, pattern = "pdf", full.names = TRUE)



#
# Converte cada arquivo pdf em txt
#
lapply (arquivos, function(i) system(paste(' "/usr/bin/pdftotext.exe" ', paste0(' " ', I, ' ' " ')), wait =
FALSE) )
```

Fonte: autoria própria

Os dados encontram-se semiestruturados, ou seja, possuem algum nível de organização imposto ao conteúdo. Uma melhoria considerável na organização dos dados seria a disponibilização dos mesmos, pelo TJDFT, no formato *eXtensible Markup Language* (XML), como exemplo.

Na análise de conteúdo não foram encontrados *missing values* – valores faltantes ou *outliers* – valores atípicos, sendo assim, o conjunto de dados, DJe, classificado para os propósitos da presente pesquisa como de boa qualidade.

5.1.3 Preparação e modelagem dos dados

A preparação dos dados consistiu em extrair, somente, as decisões, sentenças, votos e acórdãos do conjunto original de dados coletados do DJe no TJDFT. O novo conjunto de dados não incorporou informações como: nome das partes, advogados, magistrado, número do processo etc. Dois foram os motivos para a geração de um novo conjunto de dados (*data set*).

O primeiro e mais importante motivo diz respeito à preocupação em resguardar, formalmente, a privacidade das partes. Entretanto conforme demonstrados nos casos estudados a frente, tal medida não foi suficiente, uma vez que o modelo atual de sigilo processual não é mais suficiente para a forma como os dados são disponibilizados. O segundo motivo deu-se em virtude dos demais atributos não serem de interesse para o presente estudo.

Em termos formais da construção do *data set* utilizado na presente pesquisa tm-se o conjunto de dados original, X, com *n* exemplares representado como $X = \{\vec{x}_1, \vec{x}_2, \dots, \vec{x}_i, \dots, \vec{x}_n\}$, sendo os atributos dos dados representados por x_{ij}. Assim, o *i*-

ésimo exemplar do conjunto de dados original deve ser denotado por: $\vec{x}_i = \{x_{i1}, x_{i2}, \dots, x_{ij}\}$. Por fim, o novo conjunto de dados é denotado como $Y = \{ x_{11}, x_{12}, \dots, x_{1j}\}$.

A partir de Fayyad et al. (1996); Han, Kamber e Pei (2011); e Rokach e Maimon (2008) foi elaborada a Figura 5.3 com as taxonomias, tradicionais, das tarefas de mineração de dados:

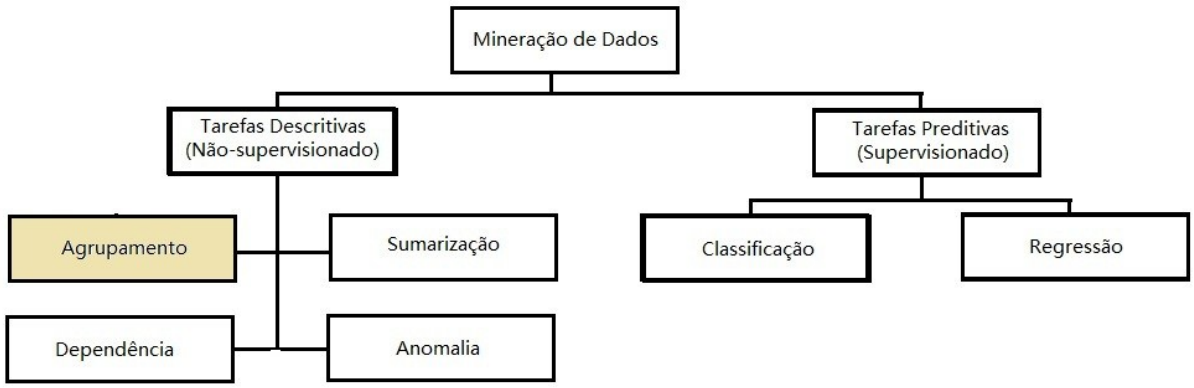

Figura 5.3: Tarefas de mineração de dados
Fonte: autoria própria

Na presente pesquisa a tarefa de mineração realizada envolve agrupamento ou *clustering*. É um tipo de tarefa que permite descobrir relações existentes entre atributos de um conjunto de dados, diferentemente das tarefas preditivas, não há rótulos associados (atributo de classe). As relações serão identificadas por meio do estudo, apenas, das descrições sobre os eventos presente no domínio de dados, por exemplo, sigilo de justiça, violência contra a mulher, violência contra crianças, termos específicos como crimes sexuais, etc.

Segundo Silva, Peres e Boscarioli (2016) as seguintes condições devem ser tradicionalmente respeitadas na tarefa de agrupamento, em que *c* é o número de grupos no modelo de grupos resultante:

$$C_k \neq \emptyset, k = 1, \dots, c;$$
$$\bigcup_{k=1}^{c}(C_k = X);$$
$$C_k \cap C_l = \emptyset, k, l = 1, \dots, c \; e \; k \neq l.$$

Após a conversão dos arquivos para um formato que permite melhor exploração dos dados, foi desenvolvido um programa na linguagem R, Tabela 5.5, com o objetivo de extrair dados de interesse para a presente pesquisa:

Tabela 5.5: Código fonte do programa de extração de dados

```
###
# Programa: converte minera_dados.r
# Versão: 1.0
# Autor: Frank Ned – info at santacruzadv.com
# Descrição: Realiza a mineração e extração de dados
#



#
# Realiza a extração apenas dos processos sob sigilo de justiça
# Duas opções de REGEXPR:
#  1) (\b[A-Z]\.)  => Extrai somente as iniciais das partes;
#  2) (?<=Processo).*?(?=Relator) => Extrai todas as informações referente aos processos em geral
#

 dest  <-  "/home/santacruz/UnB/ppca/DJe/"

 meuTXTarq <- list.files(path = dest, pattern = "txt",  full.names = TRUE)
 processos <- lapply(meuTXTarq, function(i)
                            {
                              j <- paste0(scan(i, what = character()), collapse = " ")
                              regmatches(j, gregexpr("(\b[A-Z]\.)", j, perl=TRUE))
                              }
                            )

#
# Grava os processos em um arquivo .txt a parte
#

 lapply(1:length(processos),  function(i) write.table(processos[i], file=paste(meuTXTarq[i],
"processos", "txt", sep="."), quote = FALSE, row.names = FALSE, col.names = FALSE, eol = " " )
```

Fonte: autoria própria

O TJDFT é o tribunal mais demandado entre os da justiça estadual. São 12.154 casos novos por 100.000 habitantes (CNJ, 2019). Só em 2019 no PJe foram 666.025 casos novos[22]. O primeiro termo procurado foi "pornografia infantil" em virtude da natureza da ofensa e também das partes envolvidas que deveriam ter seus dados em tramitação sob sigilo de justiça.

Durante todo o ano de 2019 foram encontradas 12 processos que tratam do tema. Destes, foram selecionados 3 casos para serem analisados e validados na pesquisa e descritos no item "5.1.4 Cenário 1 – Pornografia infantil".

A segunda extração de dados concentrou-se em todos os processos nos quais as partes, na qualificação estavam de-identificadas com as iniciais. Foram encontrados 3.942 processos distribuídos entre violência doméstica, violência contra o idoso, divórcio com menores,

[22] Disponível em: https://www.tjdft.jus.br/institucional/imprensa/noticias/2020/janeiro/mais-de-666-mil-processos-eletronicos-foram-distribuidos-em-2019-no-tjdft. Acesso: 18 Mar. 2020

violência contra a criança entre outros. A análise de um caso geral está descrita no item "5.1.5 Cenário 2 – Sigilo de justiça".

Já a quinta fase da metodologia CRISP-DM, avaliação, deu-se por meio dos cenários apresentados e discutidos a seguir.

5.1.4 Cenário 1 – Pornografia infantil

5.1.4.1 – Caso 1

Ação promovida pelo requerente, em juizado de pequenas causas, reivindicando o reestabelecimento de seus acessos (e-mail, nuvem, jogos, planilhas, documentos, fotos, vídeos, plataforma de *stream,* etc.) a determinada plataforma, uma vez que após várias tentativas administrativas não se obteve retorno.

Na plataforma, após audiência de conciliação sem sucesso, contesta a petição inicial do requerente informando que os diversos acessos à plataforma foram cancelados em virtude do seu algoritmo ter identificado um arquivo de vídeo, armazenado na nuvem associada ao requerente, com **aparente** conteúdo de pornografia infantil. Nesta mesma contestação apresenta o e-mail do requerente:

> [...] no sentido de que a conta ***** da parte autora (**********@*****.com) foi detectada em seu sistema de segurança interno em razão de aparente atividade relacionada à pornografia infantil [...]

A juíza de primeira instância não acata os pedidos do requerente e em sentença, no mérito, expõe: nome completo, e-mail e cidade do requerente, julgando improcedentes os pedidos da parte autora.

O requerente contrata escritório de advocacia especializado em Direito Digital para recorrer. O escritório peticiona em segunda instância, solicitando sigilo de justiça e fundamentando o processo no sentido de reverter a decisão de primeira instância. Três desembargadores acatam o pedido. Decreta-se o sigilo de justiça e reverte-se a decisão de primeira instância, pela PROCEDÊNCIA dos pedidos. O processo retorna para primeira instância para cumprimento de sentença e, em ato contínuo, o escritório solicita sigilo de justiça. O pedido de sigilo que é ignorado.

Dados expostos

A sentença de primeira instância publicada no DJe expôs: nome completo, e-mail e cidade do requerente.

Violação da privacidade

Ao se pesquisar dados do requerente em mecanismos de indexação, encontra-se o nome do mesmo associado à pornografia infantil. Além de seu e-mail pessoal e cidade de residência.

Sendo o requerente um jovem adulto, não é difícil constatar os danos que esta exposição promovida pelo sistema do judiciário pode causar em sua vida profissional, acadêmica e até mesmo pessoal. O agravante é que o processo não tem relação com pornografia infantil e sim com o reestabelecimento de suas credenciais de acesso às plataformas já citadas.

5.1.4.2 – Caso 2

Ação penal com três integrantes no polo passivo (Réu) envolvendo entre outras questões eventual delito de pornografia infantil. O processo tramita sob sigilo de justiça, desta forma o nome dos réus é substituído pelas iniciais nos atos publicados.

Na sentença, ao qualificar as partes, o magistrado expõe o nome completo dos pais e cidade:

> [...] Cuida-se de denúncia oferecida pelo M.P., *** [...] ***, contra *.*.*., brasileiro, natural de **********/**, filho de **** *** **** ***** e ***** ********* *****, contra *.*.*.*., brasileiro, natural de ********/**, filho de ****** ******* ***** e ****** ******** *****, e contra *.*.*., brasileiro, natural de *******/**, filho de ******** ** ***** ****** e ***** ***** ** *********, devidamente qualificados nos autos, como incursos nas sanções [...]

Ainda na mesma sentença na parte de individualização das penas, o magistrado registra o nome completo de um dos réus, mesmo o processo estando sob sigilo de justiça:

> [...] Deixo de efetuar o cálculo para a detração prevista no § 2º, do art. 387 do Código de Processo Penal, vez que o regime, já fixado no grau mais brando, não será modificado, não obstante o período de prisão preventiva do Sentenciado. ********** ***** ** ***** [...]

Dados expostos

A sentença publicada no DJe expôs: nome completo dos pais dos condenados, nome completo de um dos réus e cidade.

Violação da privacidade

Ao pesquisar o nome dos pais dos condenados em mecanismos de indexação encontra-se o nome deles vinculado à condenação dos filhos em crime de pornografia infantil e outras modalidades delituosas. Expõe, também, o nome completo de um dos réus.

Sobre o nome do réu, estando o processo sob sigilo de justiça não pode o mesmo ser exposto. Mais grave ainda é a exposição dos nomes completos dos pais, uma vez que não possuem nenhuma relação com os atos praticados pelos filhos. A punição e julgamento social acabam por transcender a figura dos apenados, atingindo diretamente o nome dos pais e consequentemente sua reputação perante a sociedade.

5.1.4.3 – Caso 3

Caso ocorrido em 2008, em plataforma de rede social já extinta, em que o requerente invoca o direito ao esquecimento em virtude de matéria jornalística publicada a época dos fatos onde consta o nome do requerente envolvido no armazenamento de imagens de pornografia infantil.

No acórdão o magistrado cita as datas da reportagem bem como o seu título:

> [...] o requerente (ora recorrido) teve o nome mencionado em matéria disponibilizada em **.**.**** (atualizada em **.**.****), intitulada ?*** ** ********* **** ******* ***** ***** ***** ****** ** *****, em razão de acusação de armazenar em computador da casa ****** ******* imagens de pornografia infantil [...]

Observa-se que se trata de caso de 2008, que volta à luz com a publicação do acórdão.

Em sua decisão o magistrado reconhece que o requerente faz jus ao direito de esquecimento, para proteger sua privacidade em contraposição à liberdade de imprensa:

> [...] No particular, sopesados a tutela da dignidade do recorrido e o direito da coletividade à informação, mostra-se justificada a intervenção do Poder Judiciário com o objetivo de se resguardar o denominado direito ao esquecimento e de se determinar a exclusão da reportagem, consoante pretendido na petição inicial [...]

Dados expostos

O acórdão publicado no DJe expôs: nome completo do requerente, indicação da matéria e datas de publicação.

Violação da privacidade

Ao se pesquisar o nome do requerido em mecanismos de indexação encontra-se o nome associado à pornografia infantil. Ademais, ao se pesquisar pelo nome da

matéria, é possível localizá-la em várias URLs diferentes, além dos sistemas de *cache* da Internet.

É sabido que no direito penal após cumprida a pena o apenado não pode novamente ser condenado sob o risco de *bis in idem*[23].

5.1.5 Cenário 2 – Sigilo de justiça

Quando da decretação do sigilo de justiça o nome completo das partes é substituído pelas suas iniciais e certas informações processuais não são publicizadas. Entretanto, as decisões, sentenças, votos e acórdãos mesmo estando sob sigilo de justiça, da forma como hoje são divulgadas no DJe e com o uso de *Big Data Analytics* acaba por expor as partes. Ou seja, a de-identificação por meio da troca do nome das partes pelas iniciais não é efetiva, muito menos eficiente para preservar a privacidade das partes mediante os atuais mecanismos de mineração de dados.

Dados expostos

Com o uso de mineração de dados e técnicas de inteligência artificial é possível o correlacionamento de fragmentos de informações e por meio deles construir o perfil da parte.

Violação da privacidade

Nos casos analisados, considerando tão somente a aplicação do sigilo de justiça por meio de pseudonimização com a técnica de substituição do nome completo das partes pelas iniciais e conjugado com outros fragmentos de dados, há de se realizar a análise por dois distintos prismas:

Matemático

Analisando-se de forma isolada o risco de re-identificação, das partes, tomando **unicamente** como fragmento de dados as iniciais dos nomes, a pseudonimização, por substituição, mostra-se eficiente.

Pois em termos de probabilidade, considerando-se a regra dos eventos independentes, a ocorrência de um não afeta a probabilidade de ocorrência do outro, ou seja, a simples análise das iniciais nos diz tão somente a frequência de ocorrência das mesmas, em nada contribuindo, de forma isolada, para a re-identificação das partes.

[23] o princípio *non bis in idem* (não repetir sobre o mesmo) estabelece que ninguém pode ser julgado duas vezes pelo mesmo fato delituoso. O *bis in idem* no direito penal seria a não observância desse princípio, apenando um indivíduo pelo mesmo crime mais de uma vez.

Entretanto o risco surge quando as iniciais são correlacionadas com outros fragmentos de informações, *quasi-identifiers*, como comarca, cidade, estado, idade ou atividade/profissão, ponto analisado no próximo tópico.

Teoria dos mosaicos

Fulgencio Madrid Conesa (CONESA, 1984) elaborou a Teoria do Mosaico, pois entendeu que a Teoria dos círculos concêntricos já não era suficiente para enfrentar as novas e sofisticadas formas de ataque a privacidade:

> [...] existem, a priori, dados irrelevantes sobre o ponto de vista do direito, que se correlacionados com outros dados, também irrelevantes, podem servir para tornar totalmente visível a personalidade de um cidadão, assim como ocorre com as pequenas pedras que formam um mosaico, que sozinhas nada revelam, mas unidas podem formar conjuntos pelos de significados. [...] (CONESA, 1984, p.45, tradução nossa)[24]

Ocorre que privacidade, intimidade e segredo são vivências do ser humano que não existe fora de um momento, de uma cultura e de uma sociedade. Entretanto, a vida no mundo atual é de tal sorte mutável que não permite traçar contornos tão cartesianos. A teoria do mosaico baseia-se na potencialidade que um fragmento de dado tem para desnudar o indivíduo de qualquer proteção a sua privacidade, por meio da construção de perfis, a perfilização, tão cara ao "capitalismos da vigilância".

Assim, não é relevante se o fragmento de informação diz respeito à privacidade, intimidade ou segredo, mas sim o uso que é feito dela. No caso concreto, dos bancos de dados do judiciário, as iniciais das partes, conjugado com *quasi-identifiers* tais como a cidade e/ou comarca, estado, profissão; e outros fragmentos de dados existentes nas publicações uma vez correlacionados com *Big Data Analytics*, torna a parte plenamente identificada e mais, permite construções e deduções que tornam o indivíduo vulnerável, ferindo a intimidade e privacidade do mesmo.

A partir da mineração dos dados coletados no DJe construiu-se a Tabela 5.6 com a quantidade de casos por número de iniciais:

[24] *Existen datos a priori irrelevantes desde el punto de vista del derecho a la intimidad y que, sin embargo, en conexión con otros, quizá también irrelevantes, pueden servir para hacer totalmente transparente la personalidad del ciudadano, al igual que ocurre con las pequeñas piedras que forman los mosaicos, que en sí no dicen nada, pero que unidas pueden formar conjuntos plenos de significados.*

Tabela 5.6: Quantidade de ocorrência de processos sob sigilo por número de iniciais do nome

QTD INICIAIS	NUM. OCORRÊNCIA	PERCENTUAL
01[25]	11	0,27
02	56	1,42
03	827	20,97
04	1397	35,43
05	763	19,35
06	253	6,41
07	166	4,21
08	239	6,06
09	136	3,45
10	94	2,38
TOTAL	3942	100%

Fonte: autoria própria

O processo, quando sob sigilo de justiça não omite o município, comarca/fórum no qual tramita o feito. Neste sentido é possível aplicar um funil de filtragem para identificar a parte, conforme Figura 5.4:

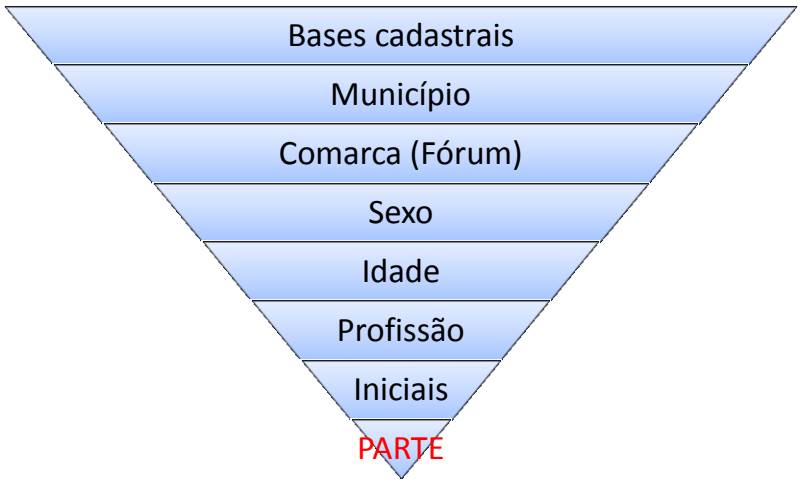

Figura 5.4: Funil de filtragem
Fonte: autoria própria

O caso concreto não será apresentado justamente para não haver exposição das partes, sendo apenas descrito para visualização da aplicação do funil, correlacionamento, em funcionamento.

O processo tramita em fórum do DF, que trata de obrigação de ********** ******, na qual a parte ré tem como profissão ******, ** anos e nome com 03 iniciais.

Partíndo-se do correlacionamento de bases cadastrais, tem-se os seguintes passos:

[25] Alguns magistrados quando o sigilo envolve menor informa somente a primeira inicial do nome.

1. Aplica-se um filtro na base selecionando todos os moradores na localidade do fórum (mais de 150 mil);
2. A localidade possui um único Fórum;
3. Seleciona-se todos moradores do sexo masculino (quase 80 mil);
4. Moradores do sexo masculino na faixa de ** anos (quase 800);
5. Todos com profissão de ****** (Base cadastral profissional mostra mais de 8.000 ******* do sexo másculino ativos);
6. Todos que tem o nome reduzido a 3 letras (6);
7. Os que coincidem com o padrão de letras das iniciais (1).

De fato, dada as características das iniciais, já no passo 05 foi possível identificar de forma única a parte.

5.1.6 Cenário 3 – Magistrados(as) (Casos gerais)

Como no cenário 2 - sigilo de justiça e apesar de não ser objeto de investigação direta do presente trabalho é possível, na verdade já existe[26], a construção de *Big Data Analytics* com dados dos magistrados(as) em relação a decisões, sentenças, votos e acórdãos, mesmo não existindo um padrão entre os DJes. Ainda que sejam implementados mecanismos para dificultar a extração de dados, como por exemplo CAPTCH *(Completely Automated Public Turing test to tell Computers and Humans Apart)* ou o bloqueio de IPs após determinado número de consultas, estas questões podem ser contornadas com uso de *proxy* randômico e solucionadores de CAPTCH.

Uma vez alimentado o *Big Data Analytics* e por meio da aplicação de técnicas de jurimetria (estatística aplicada ao direito), apoiada em mineração baseadas, por exemplo, em *deep learning*, é possível analisar o passado para prever o futuro. Neste sentido, é relativamente fácil identificar questões como:

- Viés nas decisões dos julgadores;
- Valor médio de condenações em casos semelhantes;
- Jurisprudências, doutrinadores e artigos mais citados nas fundamentações;
- Indicar os tipos de provas mais aceitas pelo julgador;

[26] Associação Brasileira de Jurimetria. Disponível em: https://abj.org.br/. Acesso em: 03 Mar. 2020.

- Identificar a probabilidade de determinado processo ter êxito, a partir da análise do posicionamento do magistrado em casos semelhantes.

Nestes tipos de aplicações, por ser o processo público e o(a) magistrado(a) uma pessoa também público, parece plenamente razoável tal aplicação, diferentemente do entendimento existente na França.

O uso, predominante, de estatística (probabilidade) no Processamento de Linguagem Natural (PLN) ocorre para compensar a dificuldade de criar uma gramática que cubra adequadamente todo o idioma. Assim, o uso de probabilidades ajuda um sistema a melhorar sua acurácia. Probabilidades associadas a termos textuais com calculados sobre grandes quantidades de dados geralmente podem fornecer muito resultados úteis (RINDFLESCH, 2013), como os já descritos acima.

O mesmo princípio aplicado a mineração de dados dos magistrados pode ser aplicado aos dados dos advogados, de forma simples como a identificação da natureza dos processos que tratam a quantidade de clientes, o nome dos clientes, instância que possui maior número de processos. Minerações mais sofisticadas de dados podem identificar estratégias utilizadas com maior frequência, índice de sucesso e fracasso por tipo de demanda entre outros.

5.1.7 Proposta de aprimoramento

Claro está que o método de de-identificação baseado, simplesmente, na pseudonimização de troca dos nomes pelas iniciais não é suficiente para garantir a não identificação das partes quando associado a *quasi-identifiers* e correlacionados com *Big Data Analytics*. Neste sentido a proposta de aprimoramento do método de pseudonimização visa justamente fortalecer a dimensão da privacidade, objetivo maior do sigilo de justiça.

Deve-se observar que, mesmo havendo a publicidade dos dados, por força de lei à luz da Lei Geral de Proteção de Dados Pessoais (LGPD) os mesmos não podem, ao menos em tese, ser utilizados para finalidades distintas das motivadoras da publicização:

> Art. 7º O tratamento de dados pessoais somente poderá ser realizado nas seguintes hipóteses:
> I - mediante o fornecimento de consentimento pelo titular;
> II - para o cumprimento de obrigação legal ou regulatória pelo controlador;
> [...]
> V - quando necessário para a execução de contrato ou de procedimentos preliminares relacionados a contrato do qual seja parte o titular, a pedido do titular dos dados;

> VI - para o exercício regular de direitos em processo judicial, administrativo ou arbitral, esse último nos termos da Lei nº 9.307, de 23 de setembro de 1996 (Lei de Arbitragem) ;
> [...]
> § 3º O tratamento de dados pessoais cujo acesso é público **deve considerar a finalidade**, a boa-fé e o interesse público que justificaram sua disponibilização.
> § 4º É dispensada a exigência do consentimento previsto no caput deste artigo para os dados tornados manifestamente públicos pelo titular, resguardados os direitos do titular e os princípios previstos nesta Lei.
> [...]
> § 7º O tratamento posterior dos dados pessoais a que se referem os §§ 3º e 4º deste artigo poderá ser realizado para **novas finalidades**, desde que observados os propósitos legítimos e específicos para o novo tratamento e a preservação dos direitos do titular, assim como os fundamentos e os princípios previstos nesta Lei. (Incluído pela Lei nº 13.853, de 2019). (grifos nosso)

Ou seja, nos casos observados, processos que tramitam sob sigilo de justiça, o tratamento deve considerar a finalidade, a boa-fé e o interesse público. Assim, surge uma questão relacionada à legalidade de empresas de mineração de dados tratarem os dados para **finalidades distintas** da motivação que levou à publicidade (Art. 7°, § 7º), mesmo que respeitando os princípios e fundamentos da LGPD, não afastando os direitos do titular de dados de solicitar, por exemplo, a exclusão dos seus dados quando tratados por instituições terceiras com finalidade econômica.

É uma questão extremamente delicada, que gera interpretações jurídicas distintas, sendo necessária análise do caso concreto para um melhor entendimento.

Há de se avaliar e gerar entendimento se: O dado tornado público por força de lei por entes públicos, como no caso do judiciário, pode ser explorado comercialmente?; valendo-se as empresas privadas do princípio da liberdade econômica e da bandeira do interesse público?. É uma discussão para além do escopo do presente trabalho a ser enfrentada no âmbito jurídico.

Caso os dados pessoais sejam de-identificados, afasta a aplicabilidade da LGPD.

Vale reforçar que, além de interpretações jurídicas, há de se considerar os conceitos de erro latente e erro efetivo, existentes na Engenharia da Confiabilidade na adaptação destes conceitos no âmbito da privacidade:

- **Violação latente da privacidade**: fragmentos de dados pessoais ou não que geram situações de risco que, ao serem combinarem com outros fragmentos de dados, criam a oportunidade para a exposição da privacidade;
- **Violação efetiva da privacidade**: é a resultante da violação latente da privacidade, ou seja, a exposição da privacidade em si, que causa real prejuízo à parte.

Logo, a confiabilidade de um sistema, nesta pesquisa em relação à exposição da privacidade, diz respeito à probabilidade de um dispositivo desempenhar seu propósito adequadamente durante o tempo desejado, sob as condições operativas encontradas.

Já ficou demonstrado que o dispositivo de implementação do sigilo de justiça, na forma de substituição do nome das partes pelas iniciais, não atendem mais, de forma adequada, o propósito original.

Dai a importância em implementar mecanismos efetivo de de-identificação, principalmente em relação aos processos que tramitam sob sigilo de justiça.

5.1.7.1 – Tabela de atributos sensíveis no processo

Como proposta de aprimoramento do PJe e DJe tem-se inicialmente, a partir da experiência do *Health Insurance Portability and Accountability Act* (HIPAA), a sugestão da criação de uma tabela de atributos sensíveis para o sistema do judiciário, sendo que estes atributos, uma vez presentes no processo e, principalmente, nos processos que tramitam sob sigilo de justiça deveriam ser observados e aplicados neles a de-identificação (pseudonimização) no momento da publicização dos atos processuais no PJe e DJe:

Tabela 5.7: Atributos sensíveis no processo

1.	Nomes	7. N° de identificação (CPF, ID etc)	13. Identificador de dispositivos
2.	Códigos postais (até 3 dig.)	8. Atividade, profissão ou ocupação	14. URLs
3.	Datas (exceto o ano)	9. Nºs de apólice de saúde	15. IPs
4.	Nºs de telefone	10. Nºs de conta	16. Identificadores biométricos
5.	Perfil em plataforma digitais	11. Nºs de certificado ou licença	17. Fotos com face completa
6.	Endereços de email	12. Identificador de veículos	18. Outros atrib. identificadores

Fonte: autoria própria a partir do HIPAA

A aplicação do mecanismo de de-identificação deve obedecer alguns princípios gerais:

1. Aplicável a todos os processos que tramitam sob sigilo de justiça;
2. Incidência sob a identificação das partes nos cabeçalhos, mas também no corpo das decisões, sentenças, votos e acórdãos;
3. As partes e representantes legais terem acesso aos dados sem as máscaras de de-identificação;

4. Conforme assinalado no item "2.7 A publicidade do processo e a privacidade", em princípio a de-identificação não se aplica a pessoa público nos atos que representam a cidadania.

O mecanismo pode ser implementado com a substituição dos atributos, pseudonimização, por meio de IDs únicos e gerados a partir de processo randômico com elevado grau de entropia, garantindo a ligação entre os registos e os titulares de dados e dessa forma possibilitando uma posterior re-identificação.

Uma possibilidade de mitigação da violação efetiva a implementação de um modelo como o MITRE *Identification Scrubber Toolkit* (MIST), uma ferramenta de código aberto de de-identificação de textos (ABERDEEN, 2010. p. 858). Trata-se de uma implementação de *Conditional Random Fields* (CRFs) voltada especialmente para processamento de textos, exatamente como o caso tanto do PJe como do DJe. Segundo John Aberdeen (ABERDEEN, 2010, p. 854) o "CRF foi aplicado com sucesso em vários problemas de processamento de textos, tais como: marcação de trechos de discursos, análise de dependência, resumos, documentos biomédicos entre outros". A função de de-identificação é uma instância especial que identifica tipos especiais de nomes, como o nome de pessoas.

Ainda de acordo com John Aberdeen (ABERDEEN, 2010):

> [...] *CRFs are a type of classifier that labelswords with their type of phrase, such as PERSON, LOCATION, or HOSPITAL. Rather than assume the labels for each word are independent, CRFs capture dependencies between the labels. Typically, a firstorder Markov assumption is made, such that for a word at position T whose value is Xt, the corresponding label Yt is dependente on the labels for Yt−1 and Yt+1 and independent of all other labels Y1, . . ., Yt−2 and Yt+2, . . ., Yn given Yt−1 and Yt+1. With this assumption, CRFs then define the conditional probability of a label sequence, _Y = Y1, . . . , Yn, given an observed sequence of tokens, _X = X1, . . . , Xn, as follows*:

$$P(\vec{y} = y_1, \ldots, y_n | \vec{x} = x_1, \ldots, x_n) = \frac{\exp\left(\sum_t \sum_k \lambda_k f_k(y_t, y_{t-1}, \vec{x}, t)\right)}{\sum_{\hat{y}} \exp\left(\sum_t \sum_k \lambda_k f_k(\hat{y}_t, \hat{y}_{t-1}, \vec{x}, t)\right)}$$

A função f_k captura algumas propriedades da palavra _X = X1,..., Xn na posição T, como o identificador da palavra corrente, por exemplo, se inicia com letra maiúscula ou se a palavra anterior é "Dr.". A função correlaciona as propriedades das palavras com marcadores particulares.

5.1.7.2 – Formação dos magistrados

No curso de formação dos magistrados incluir ou aprimorar o módulo específico no tocante à técnica de elaboração de decisões, sentenças, votos e acórdãos, principalmente no tocante ao estilo de redação de forma a reduzir o risco de exposição da privacidade das partes tendo como balizador a tabela de atributos sensíveis. É desejável, nos processos sob sigilo de justiça, evitar a exposição dos atributos sugeridos, tendo como objetivo principal atuar no pilar/fator pessoa e também "processo" buscando-se mitigar a exposição da privacidade das partes. Contribui, também, o aprofundamento do tema privacidade, direito ao esquecimento e noções de como funciona a rede (Internet), tanto a atual WEB 3.0 com os avanços para a WEB 4.0.

5.1.7.3 – Criação no DJe de seção específica referente ao direito ao esquecimento

Uma vez que a grande maioria dos atos são públicos, em respeito ao princípio constitucional da privacidade e sabendo-se da dificuldade técnica em efetivar o direito ao esquecimento na Internet quando invocado pelas partes, esta seção teria como objetivo principal dar publicidade às decisões favoráveis à desindexação de certos conteúdos da internet e com isso, as plataformas de indexação podem automatizar a remoção das referências de seus índices. Uma vez que a WEB 3.0, Web semântica, esta em sua fase "inicial" de implementação de todos o potencial da comunicação máquina-máquina e a WEB 4.0 também conhecida como "A Web Simbiótica" ou "A Web Inteligente" tem como principal objetivo a interação pessoa natural-máquina de maneira simbiótica. Uma seção desta natureza pode potencializar a desindexação, automática, pelos mecanismos de busca, quando determinado judicialmente.

5.2 Relação pessoa natural-máquina (advocacia)

Já na segunda vertente de pesquisa de campo analisou-se a relação pessoa natural-máquina, tratando dos riscos existentes nos processos e pessoas no tocante a salvaguarda de dados pessoais em posse dos advogados e escritórios de advocacia. Para tanto, foi realizada uma pesquisa exploratória qualitativa disponibilizada, on-line, por meio de questionário, respondido por 82 advogados. Na sequência os dados foram analisados e, a partir deles,

elaborou-se um conjunto de recomendações de boas práticas, tanto na dimensão de processos como comportamental, que atua no pilar 'pessoas'.

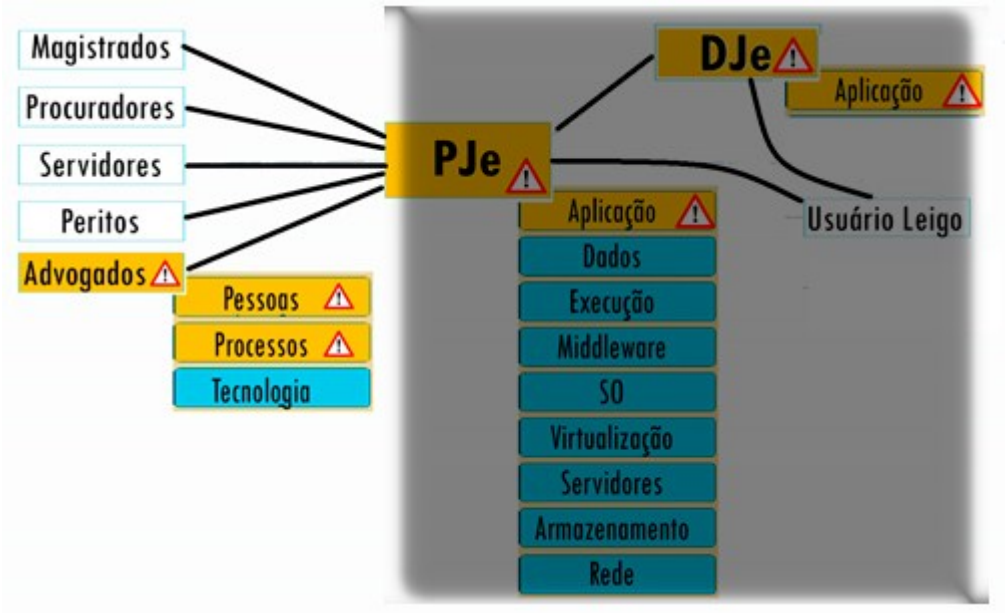

Figura 5.5: Pontos de investigação da pesquisa
Fonte: autoria própria

5.2.1 Coleta de dados por meio de questionário on-line

O questionário disponibilizado na Internet, anexo 2, composto de 28 questões foi previamente validado pelo cálculo de modelo de equações estruturais (MEE), conforme descrito em "4.5 Modelo de Equações Estruturais – Validação do questionário"

5.2.2 Tratamento e análise dos dados

Não houve necessidade de tratar os dados uma vez que o preenchimento do questionário pelos respondentes já gerou dados padronizados sem *missing values*, *outlier* ou outras inconsistências. As respostas estão no anexo 3.

Assim partiu-se para a análise e interpretação das respostas.

5.2.3 Cenário de riscos

Os advogados e escritórios de advocacia possuem um conjunto de informações extremamente sensíveis e confidenciais de seus clientes, tais como minutas de contratos,

dados de fusões e aquisições de empresas, fotos e registros diversos de situações delicadas, exames médicos, registros financeiros, questões a respeito de propriedade intelectual, detalhes de partilhas de bens, negociações de indenizações, dados de histórico criminal, enfim, um universo de dados e informações.

O Código de ética e Disciplina da OAB nos artigos 35 e 36 trata do sigilo[27]:

> Art. 35. O advogado tem o dever de guardar sigilo dos fatos de que tome conhecimento no exercício da profissão. Parágrafo único. O sigilo profissional abrange os fatos de que o advogado tenha tido conhecimento em virtude de funções desempenhadas na Ordem dos Advogados do Brasil.
>
> Art. 36. O sigilo profissional é de ordem pública, independendo de solicitação de reserva que lhe seja feita pelo cliente.
> § 1º Presumem-se confidenciais as comunicações de qualquer natureza entre advogado e cliente.

Boa parte dos fatos que chegam ao conhecimento do advogado e escritórios dá-se, e, atualmente, em sua grande maioria, em formato digital. São e-mails trocados com dados, informações e anexos. Relatórios, petições, arquivos diversos que são disponibilizados em nuvens públicas para facilitar e agilizar o acesso a dados e informações. Documentos diversos que são transportados em pendrive. Em outras situações, para adequar os documentos ao formato e tamanho exigido pelo PJe são utilizadas plataformas/sistemas públicos disponibilizados "gratuitamente" na Internet. Ou seja, quando há compartilhamento de documentos confidenciais com clientes ou terceiros autorizados, os advogados e escritórios de advocacia deparam-se com uma escolha que envolve dois interesses concorrentes: usabilidade ou segurança e privacidade.

De acordo com a análise dos dados coletados, claramente a usabilidade tem sido a regra. Assim, na presente pesquisa são apresentados alguns dos principais cenários de riscos, bem como propostas de mitigação dos mesmos, sendo eles descritos abaixo:

5.2.3.1 – e-mail

Primeiramente, não há dúvidas de que o e-mail se tornou uma das principais ferramentas de troca de informações da atualidade.

[27] Disponível em: https://www.oab.org.br/publicacoes/AbrirPDF?LivroId=0000004085. Acesso em 02 Mar. 2020

Segundo, os dados coletados no questionário 78,95% dos respondentes afirmaram utilizar e-mail "gratuito" para fins profissionais, conforme Figura 5.6 (Você utiliza algum e-mail gratuito para uso profissional? Ex.: gmail, uol, yahoo):

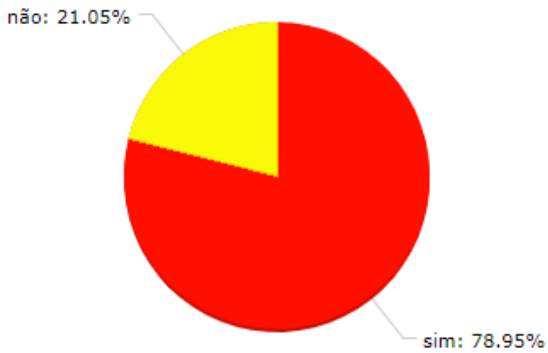

Figura 5.6: Uso de e-mail gratuito para fins profissional
Fonte: autoria própria

Entretanto ao utilizar este importante recurso, principalmente, os ofertados "gratuitamente" é importante ter em mente que todo conteúdo, bem com os anexos são analisados automaticamente e aplicados mecanismos de *data mine*. Neste sentido, o Termo de Privacidade do gmail não deixa dúvidas[28]:

> [...]Quando você faz upload, submete, armazena, envia ou recebe conteúdo a nossos Serviços ou por meio deles, **você concede ao Google (e àqueles com quem trabalhamos)** uma licença mundial para usar, hospedar, armazenar, reproduzir, modificar, criar obras derivadas (como aquelas resultantes de traduções, adaptações ou outras alterações que fazemos para que seu conteúdo funcione melhor com nossos Serviços), comunicar, publicar, executar e exibir publicamente e distribuir tal conteúdo. Os direitos que você concede nesta licença são para os fins restritos de operação, promoção e melhoria de nossos Serviços e de desenvolver novos Serviços. Essa licença perdura mesmo que você deixe de usar nossos Serviços (por exemplo, uma listagem de empresa que você adicionou ao Google Maps). Alguns Serviços podem oferecer-lhe modos de acessar e remover conteúdos que foram fornecidos para aquele Serviço. Além disso, em alguns de nossos Serviços, existem termos ou configurações que restringem o escopo de nosso uso do conteúdo enviado nesses Serviços. Certifique-se de que você tem os direitos necessários para nos conceder a licença de qualquer conteúdo que você enviar a nossos Serviços.
> Nossos sistemas automatizados analisam o seu conteúdo (**incluindo e-mails**) para fornecer recursos de produtos pessoalmente relevantes para você, como resultados de pesquisa customizados, propagandas personalizadas e detecção de spam e malware. Essa análise ocorre à medida que o conteúdo é enviado e recebido, e quando ele é armazenado. [...] (grifo nosso)

Ou seja, todos os e-mail trocados via plataforma gmail e outras plataformas "gratuitas" analisa e extrai dados e informações tanto do corpo da mensagem como de seus anexos. O

[28] Disponível em: https://policies.google.com/terms?hl=pt-BR#toc-content. Acesso em: 02 Mar. 2020

advogado, ao utilizar estas plataformas "gratuitas" para troca de mensagens e anexos que contenham dados confidenciais está, na verdade, permitindo que elas alimentem seu *Big Data Analytics* com os conteúdos minerados nos e-mail e anexos.

Outro ponto, levantado na pesquisa, diz respeito ao envio de e-mail, por engano, para o destinatário errado. Segundo os respondentes 45,61% dos mesmos já se encontraram nesta situação conforme a Figura 5.7 (Já aconteceu de você enviar um e-mail, com anexo, para destinatário incorreto, por engano?):

Figura 5.7: Envio de e-mail profissional para destinatário errado
Fonte: autoria própria

Neste quesito, depara-se novamente com o risco de exposição de dados confidenciais, que estão sob a guarda do operador do direito e/ou escritório, ao cometer o erro operacional de enviar o conteúdo para o destinatário incorreto.

5.2.3.2 – Uso de nuvem pública

Outro recurso tecnológico utilizado com frequência pelos operadores de direito são os serviços de nuvem para compartilhamento e armazenamento de documentos como petições, documentos de clientes, planilhas, vídeos, fotos e outra infinidade de dados e informações digitais.

Em relação ao serviço de nuvens, 94,74% dos respondentes afirmaram que utilizam algum tipo de serviço "gratuito", conforme Figura 5.8 (Você utiliza algum serviço, gratuito, de nuvem? Exemplo *dropbox*, *google drive*, *azure*):

Figura 5.8: Uso de nuvem pública
Fonte: autoria própria

Alguns dos serviços "gratuitos" mais populares são o *google drive*, *dropbox*, *Azure da Microsoft* entre outras.

A Política de Privacidade e Termo de Uso do *google drive* apresenta no trecho relacionado a coleta de dados as seguintes informações[29]:

> [...] Ao criar uma Conta do Google, você nos fornece informações pessoais que incluem seu nome e uma senha. Você também pode optar por adicionar um número de telefone ou informações de pagamento à sua conta. Mesmo se não estiver conectado a uma Conta do Google, você poderá optar por nos fornecer informações, como um endereço de e-mail para receber atualizações sobre nossos serviços.
> Também coletamos o conteúdo que você cria, de que faz upload ou que recebe de outras pessoas ao usar nossos serviços. Isso inclui e-mails enviados e recebidos, fotos e vídeos salvos, documentos e planilhas criados e comentários feitos em vídeos do YouTube. [...] (grifo nosso)

Novamente, em relação ao serviço de nuvem "gratuita" manifesta os riscos relacionados à violação da confidencialidade e privacidade, uma vez que todos os arquivos armazenados são inspecionados, minerados, por algoritmos.

[29] Disponível em: https://policies.google.com/privacy#infocollect. Acesso em: 02 Mar. 2020

5.2.3.3 – *Pendrive*

O *pendrive*, juntamente com o uso de nuvens, é um recurso de fácil armazenamento e transporte de dados e informações. Permite a flexibilidade de trabalhar em certos documentos com um simples acesso a computador, mesmo que este não esteja conectado à Internet.

Um dos maiores riscos associados ao *pendrive* são as questões relacionada ao esquecimento em computadores de uso público ou perda. De acordo com a pesquisa 38,60% já esqueceram ou perderam *pendrives* com conteúdo profissional, conforme aponta o gráfico da Figura 5.9 (Você já esqueceu em algum ambiente ou perdeu um *pendrive* com documentos de trabalho?):

Figura 5.9: Esquecimento ou extravio de pendrive
Fonte: autoria própria

Outro ponto de atenção é o risco de contaminação do mesmo com vírus ou *malwares*.

5.2.3.4 – Uso de computador público

O uso de computadores públicos, para fins profissionais foi reportado por 80,70% das pessoas que responderam o questionário, conforme exemplificado na Figura 5.10 (Você já teve que utilizar um computador público para editar um documento de trabalho? Ex.: LAN House, Hotel, Sala de apoio ao advogado(a) da OAB) :

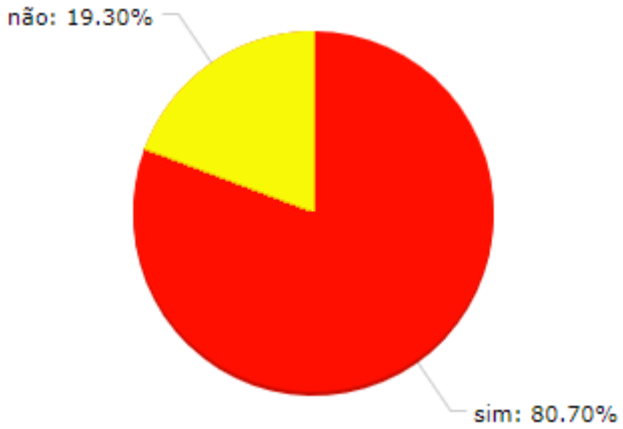

Figura 5.10: Uso de computador público para fins profissionais
Fonte: autoria própria

Os três grandes riscos associados à utilização de computadores públicos são: risco de esquecer documentos com dados pessoais e dados pessoais sensíveis em diretórios do computador; possibilidade de ter algum monitor de teclado (*keylog*) que captura tudo que é digitado pelo usuário e/ou monitor de tela que realiza *print* da tela do computador de tempos em tempos.

Outro risco diz espeito ao uso de pendrive nestes dispositivos e ter o meio de armazenamento infectado com vírus e/ou *malware*.

5.2.3.5 – Sistemas "gratuitos" de conversão para pdf/a e redução de tamanho

Em relação a plataformas de conversão de documentos para .pdf/A, padrão aceito pelo PJe ou redimensionamento do tamanho para até 10 Mb, tamanho máximo para arquivos pdf, 56,66% dos respondentes informaram que utilizam algum tipo de plataforma "gratuita" para tal finalidade.

Os principais riscos identificados no uso de plataformas e/ou sistemas "gratuitos" são:

- A coleta e processamento de dados, inclusive pessoais, do utilizados;
- O processamento de dados constantes no documento;
- O armazenamento do documento, mesmo que por curto período de tempo, na nuvem externa sobre a qual não se tem conhecimento sobre a utilização e eficiência dos mecanismos de segurança para proteger o acesso não autorizado aos documentos carregados na plataforma ou sistema;

- Vazamento de dados pessoais e dados pessoais sensíveis existente nos documentos carregados.

A título de exemplo a política de privacidade do aplicativo *CamScanner* informa[30]:

> [...] Ao usar nossos produtos e serviços, você consente e concede autorização a nossa coleta e processamento de suas informações, incluindo as informações do usuário do dispositivo e software, **informações pessoais e informações de conteúdo enviadas pelo usuário** (principalmente as informações de imagem que você enviou; sobre o conteúdo enviado, você garante que possui a propriedade intelectual envolvida ou obteve a licença relevante e não há violação do direito de terceiros). As informações são usadas apenas para a otimização da interação da interface do software. Nenhuma informação específica será usada ou divulgada [...].[31] (tradução nossa)

De fato, quase que 100% dos sistemas e plataformas "gratuitas" coletam e processam dados pessoais para finalidades diversas, fazendo parte da cadeia do "capitalismo da vigilância". As novas legislações como a GDPR da Europa, a LGPD no Brasil, buscam estabelecer princípios mínimos para o tratamento de dados pessoais, bem como bases legais de consentimento.

5.2.3.6 – Outras questões do pilar Pessoas

De forma complementar aos pontos até então apresentados as questões abaixo relacionadas, na Tabela 5.7, também dizem respeito ao comportamento das pessoas frente à tecnologia.

Tabela 5.8: Questões do pilar pessoas

#	QUESTÃO	% SIM
01	Eu consigo, com sucesso, aprender sobre segurança cibernética, mesmo sem ter lido nada a respeito antes?	6,8%
02	Eu consigo, com sucesso, aprender sobre segurança cibernética, mesmo sem ninguém me ensinar?	10,2%
03	Eu consigo, com sucesso, aprender sobre segurança cibernética, mesmo sem ajuda de ninguém?	5,1%
04	Eu consigo, com sucesso, apender sobre segurança cibernética, se tiver muito tempo	27,1%

[30] Disponível em: https://www.camscanner.com/app/privacy. Acesso em: 02 Mar. 2020.

[31] By using Our Products and services, you consent to and grant authorization to our collection and processing your information, including the device and software user information, personal information and content information uploaded by user (mainly the image information you uploaded; regarding the uploaded content, you warrant that you have the intellectual property involved therein or have obtained the relevant license and there is no infringement of third party's right). The information is only used for the optimization of software interface interaction. No specific information will be used or disclosed.

	disponível?	
05	Eu consigo, com sucesso, aprender sobre segurança cibernética, somente tendo referência de informações?	50,8%

Fonte: autoria própria

Este grupo de respostas gerou certo estranhamento, pois se forem consideradas as respostas que demostram maior autonomia de aprendizado, 01, 02, 03 e 04, o somatório é de 49,2%.

Confrontando esta elevada capacidade de aprendizado dos leigos no tema sobre segurança cibernética com as respostas anteriores, onde 78.95% utilizam e-mail "gratuito", para troca de informações e arquivos profissionais; 94,74% utilizam serviços de nuvem "gratuitos"; 80,70% já utilizaram computadores públicos para questões profissionais; 56,66% utilizam plataformas e/ou sistemas "gratuitos" para ajustar as petições e documentos diversos ao formato do PJe; pode-se chegar a algumas possíveis conclusões:

1. As perguntas não foram completamente entendidas;
2. Há certa inocência em relação ao real conceito de aprendizado sobre segurança cibernética;
3. Há dificuldade do advogado em admitir que não sabe algo.

Empiricamente, a hipótese três parece ser a mais plausível, uma vez que o operador do direito durante sua formação é treinado para responder às questões do cliente. Treinado para não se calar perante o oponente no momento da argumentação, não é incomum professores de direito reforçarem a questão de que "O advogado nunca se cala".

5.2.3.7 – As questões do pilar processos

Os processos são ferramentas que auxiliam na tomada de decisão e consequentemente na execução de certos procedimentos de forma "padronizada" e mais harmônica.

A tomada de decisão, de forma simplificada, é um processo de escolha entre várias opções que, entre outras questões, envolve a análise de riscos. Segundo Paul C. Nutt (Nutt, 2010), quatro são os possíveis processos de tomada de decisão:

- **Decisões baseada no Procedimento** - Utiliza um conjunto de instruções explícitas que atuam como guia na escolha de uma maneira de proceder;

- **Decisões baseada na Experiência** - Utiliza o aprendizado adquirido, pelo indivíduo, durante o processo de solucionar problemas e tomar decisões no decurso de sua vida e sua carreira;
- **Decisões Analíticas** - Envolve o diagnóstico de um problema, o levantamento e prova da evidência que se necessita para o enfrentar, e depois a comparação e seleção de uma opção;
- **Decisões Racionais** - É a escolha da opção mais eficiente, aquela que maximiza o resultado para dado insumo, ou que minimiza o insumo para dado resultado.

Para Christopher. L. Culp, o nível de processos de uma instituição reflete a maturidade da mesma em seus processos decisórios e de gestão de riscos. (CULP, 2002)

A existência, identificação e documentação de processos ajudam as instituições na gestão de riscos. Para identificar o nível de maturidade dos escritórios de advocacia, as seguintes questões, Tabela 5.9, foram realizadas:

Tabela 5.9: Questões do pilar processos

#	QUESTÃO	% SIM
01	Na sua instituição existe uma pessoa dedicada a cuidar das questões de segurança da informação? Ex.: escritório, local de trabalho, estágio	71,1%
02	Existe uma verificação de vida pregressa em relação às pessoas que acessam e manipulam dados da instituição?	37,3%
03	A instituição possui uma política de segurança da informação?	72,0%
04	A instituição possui uma política que detalha as regras para formação de senhas?	58,7%
05	Os colaboradores recebem treinamento sobre cuidados com segurança da informação?	45,3%
06	Você acredita que o nível de segurança, da informação, em seu ambiente de trabalho/estágio/estudo é elevado?	34,6%
07	Já houve ou Existe compartilhamento de usuário/senha para acesso a informações?	57,1%
08	Existe controle de acesso aos sistemas que armazenam dados sensíveis?	66,2%
09	No seu ambiente de trabalho ou estágio existe controle de uso das portas USB dos equipamentos?	31,6%
10	No seu ambiente de trabalho ou estágio existe controle de acesso a sites da Internet?	70,1%
11	No seu ambiente de trabalho ou estágio, para acessar a rede Wi-Fi, existe uma senha única para todos?	44,2%

Fonte: autoria própria

A partir da análise das respostas é possível identificar alguns pontos que merecem especial atenção e são indicadores de que os escritórios de advocacia apesar de em sua maioria terem Política de Segurança da Informação. Por exemplo apesar de 72% dos escritórios responder possuírem política de segurança da informação, tais políticas não estão adequadas à realidade, pois 41,3% responderam que o escritório não possui uma política detalhada de formação de senha, que é algo básico e uma das principais vulnerabilidades entre usuários finais.

Outras práticas identificadas, não alinhadas com as boas práticas de mercado, dizem respeito aos compartilhamentos de usuários e senhas, que representam 57,1% dos casos; e também a utilização de senha única para acesso a rede Wi-Fi, 55,8%.

Isso só constata que, apesar da importância de existirem processos, eles devem ser mantidos atualizados e revisitados com regularidade conforme recomendação da ISO 27.001, pois, a existência de processos desatualizados ou, sobre os quais os usuários não tenham conhecimento, não contribui para a gestão dos riscos cibernéticos.

5.2.3.8 – Visita a uma sala da advocacia

No Distrito Federal, em todos os fóruns existe uma sala reservada ao uso dos operadores do direito, equipada com computadores e recursos de escritório, tais como copiadora, espaço de reunião.

Foi realizada visita a um dos fóruns do Distrito Federal e analisada a área de *download* dos computadores para validar a existência ou não de documentos com informações pessoais. Na Figura 5.11 é apresentada listagem parcial dos arquivos encontrados no diretório *download*, sendo destacados alguns com dados pessoais, inclusive dados pessoais sensíveis:

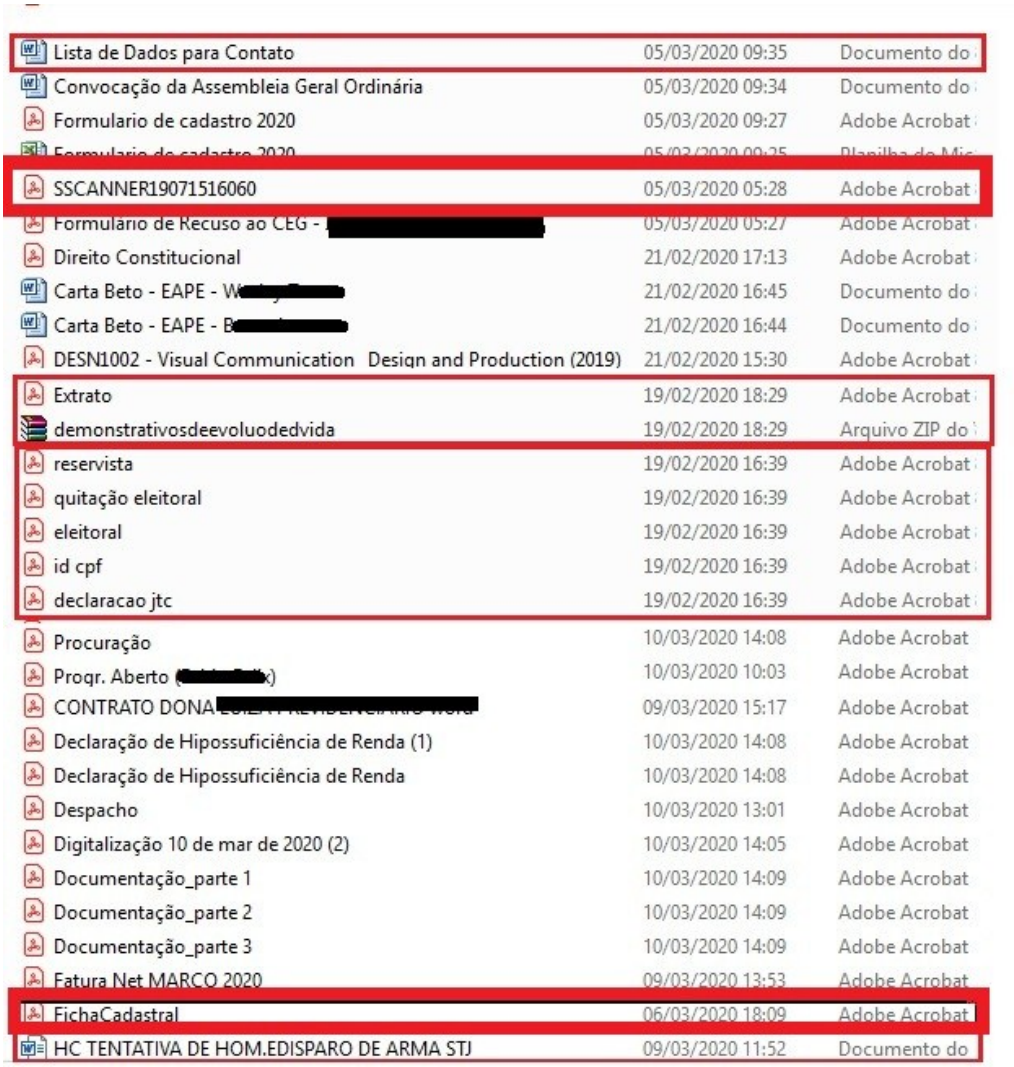

Figura 5.11: Lista parcial de arquivos encontrados na pasta *download*
Fonte: autoria própria

Entre os documentos encontrados merece destaque o arquivo **SSCANNER19071516060**, que trata do resultado de uma bateria de exames médicos com exposição de dados pessoais sensíveis, conforme demostrado nos apêndice: Exame de ressonância e Exame clínico – 22 exames.

Estas evidências confirmam o risco de esquecimento de arquivos com dados pessoais e dados pessoais sensíveis em computadores de uso comum. Conforme abordado no item "5.2.3.4 – Uso de computador público", onde 80,70% dos respondentes ao questionário afirmaram utilizar computador público para fins profissionais.

5.2.4 Recomendações

Ao analisar o resultado da pesquisa observa-se que as ameaças de cibersegurança e desafios que os escritórios e advogados enfrentam são tão diversos que se faz necessária a elaboração de algumas recomendações mínimas, no sentido de criar cultura e um ambiente minimamente protegido.

É importante ter em vista que medidas de segurança, sejam elas técnicas, de processos e de educação de pessoas não são projetos temporário. É uma jornada permanente de planejamento, implementação, acompanhamento, monitoramento e aprimoramento.

As recomendações aqui apresentadas estão baseadas principalmente no pilares pessoas e processos e tratam-se de atitudes simples que podem contribuir de forma significativa para a preservação da proteção de documentos com dados pessoais das partes e também para a privacidade das mesmas.

Tabela 5.10: Tabela de recomendações pessoas/processos

01	PILAR**:**	**Pessoa**	**RECURSO:**	**e-mail**
	Descr:	Envio de e-mail "a partir de" e/ou "para" domínios "gratuitos", corre-se o risco de análise automatizada do conteúdo (corpo) pelos algoritmos da plataforma e consequente coleta de dados pessoais.		
	RISCO:	Violação da privacidade	**RECOMENDAÇÃO**:	• Usos de criptografia. Ex.: GNuPG; • Registro de domínio próprio.
02	**PILAR:**	Pessoa	**RECURSO:**	e-mail
	Descr:	Envio de e-mail, com anexo, "a partir de" e/ou "para" domínios "gratuitos", corre-se o risco de análise automatizada dos anexos pelos algoritmos da plataforma e consequente coleta de dados pessoais.		
	RISCO:	Violação de privacidade	**RECOMENDAÇÃO:**	• Uso de senha nos anexos; • Usos de criptografia. Ex.: GNuPG; • Registro de domínio próprio.
03	**PILAR:**	Pessoa	**RECURSO:**	e-mail
	Descr:	Envio de e-mail para o destinatário errado (corpo).		
	RISCO:	Violação da privacidade	**RECOMENDAÇÃO:**	• Na assinatura do e-mail mensagem com orientações para o destinatário incorreto; • Uso de criptografia. Ex.: GNuPG;
04	**PILAR:**	Pessoa	**RECURSO:**	e-mail
	Descr:	Envio de e-mail para o destinatário errado (anexo).		
	RISCO:	Violação da privacidade	**RECOMENDAÇÃO:**	• Na assinatura do e-mail mensagem com orientações

				para o destinatário incorreto; • Uso de senha nos anexos; • Uso de criptografia. Ex.: GNuPG;
05	**PILAR:**	Pessoas / Processo	**RECURSO**:	Nuvem pública
	Descr:	Ao utilizar serviços de nuvem pública para armazenamento e troca de documentos com clientes. Ex.: comprovantes de residência, documentos de identificação, extratos diversos, exames médicos, minuta de contratos etc. Corre-se o risco de algoritmos das plataformas realizar análise dos documentos armazenados nas mesmas.		
	RISCO:	Violação da privacidade	**RECOMENDAÇÃO**:	• Uso de senha nos documentos; • Uso de criptografia. Ex.: GNuPG; • Contratação de plano comercial de nuvem.
06	**PILAR:**	Pessoa / Processo	**RECURSO**:	*pendrive*
	Descr:	Esquecer o dispositivo (*pendrive*), com informações pessoais do assistido, em computadores públicos. Terceiros podem ter acesso aos dados pessoais.		
	RISCO:	Violação da privacidade	**RECOMENDAÇÃO**:	• Uso de senha nos arquivos; • Criptografia do sistema de arquivos do *pendrive*. Ex.: Veracrypt
07	**PILAR:**	Pessoa / Processo	**RECURSO**:	*pendrive*
	Descr:	Perde/extraviar o dispositivo (*pendrive*), com informações pessoais do assistido. Terceiros podem ter acesso aos dados pessoais.		
	RISCO:	Violação da privacidade	**RECOMENDAÇÃO**:	• Uso de senha nos arquivos; • Criptografia do sistema de arquivos do *pendrive*. Ex.: Veracrypt
08	**PILAR:**	Pessoa / Processo	**RECURSO**:	*pendrive*
	Descr:	Uso de dispositivo (*pendrive*) em computador público.		
	RISCO:	Infecção por vírus / *trojan*	**RECOMENDAÇÃO**:	• Proteção da "raiz" quanto à escrita. Permitindo a realização de escrita somente em diretórios secundários. **Obs**: A maioria dos vírus oportunistas infectam a partir da raiz.
09	**PILAR:**	Pessoas	**RECURSO**:	Uso de computador público
	Descr:	Realizar a edição de peças com dados pessoas dos assistidos em computadores públicos. Os mesmo podem estar infectados com ferramentas de monitoramento e coleta de dados.		

	RISCO:	Violação da privacidade	**RECOMENDAÇÃO**:	• Evitar o uso de computadores públicos para trabalhar em documentos com dados pessoais.
10	**PILAR:**	Pessoas	**RECURSO**:	Uso de computador público
	Descr:	Realizar o *donwload* de documentos com dados pessoais do assistido em computador público. Pode-se, ao final do trabalho, esquecer os arquivos no diretório de *download*.		
	RISCO:	Violação da privacidade	**RECOMENDAÇÃO**:	• Mover os arquivos para um *pendrive* e trabalhar nos mesmo a partir deste dispositivo.
11	**PILAR:**	Processos	**RECURSO**:	Computador de uso público
	Descr:	Computadores existentes, por exemplo, nas salas da advocacia nos fóruns.		
	RISCO:	Violação da privacidade	**RECOMENDAÇÃO**:	• Habilitar *script* no processo de inicialização da máquina de forma a, minimamente, remover os arquivos do dia anterior da área de *download.*
12	**PILAR:**	Pessoas	**RECURSO**:	Uso sistemas on-line para converter arquivos no formato pdf/A.
	Descr:	Ao submeter arquivos com dados pessoais a plataformas on-line, "gratuita", de conversão de formato os algoritmos da mesma, de forma automatizada, podem realizar mineração de dados pessoais.		
	RISCO:	Violação da privacidade	**RECOMENDAÇÃO**:	• Conhecer os termos de privacidade da plataforma; • Uso de plataforma/aplicativos confiáveis. Ex.: Otimizador de PDF do TRT 14.
13	**PILAR:**	Pessoas	**RECURSO**:	Uso sistemas on-line para converter arquivos no formato pdf/A.
	Descr:	Ao submeter arquivos com dados pessoais a plataformas on-line, "gratuita", de conversão de formato a plataforma pode ficar com uma cópia do arquivo. E o remetente não ter informações a respeito do que é feito com a cópia.		
	RISCO:	Violação da privacidade	**RECOMENDAÇÃO**:	• Conhecer os termos de privacidade da plataforma; • Uso de plataforma/aplicativos confiáveis. Ex.: Otimizador de PDF do TRT 14.
14	**PILAR:**	Pessoas	**RECURSO**:	Uso sistemas on-line para ajustar o tamanho do arquivo ao padrão do PJe.

	Descr:	Ao submeter arquivos com dados pessoais a plataformas on-line, "gratuita", para ajustar o tamanho do arquivos ao limite do PJe, os algoritmos da mesma, de forma automatizada, podem realizar mineração de dados pessoais.		
	RISCO:	Violação da privacidade	**RECOMENDAÇÃO**:	• Conhecer os termos de privacidade da plataforma; • Uso de plataforma/aplicativos confiáveis. Ex.: Otimizador de PDF do TRT 14.
15	**PILAR:**	Pessoas	**RECURSO**:	Uso sistemas on-line para ajustar o tamanho do arquivo ao padrão do PJe.
	Descr:	Ao submeter arquivos com dados pessoais a plataformas on-line, "gratuita", para ajustar o tamanho do arquivos ao limite do PJe, a plataforma pode ficar com uma cópia do arquivo. E o remetente não ter informações a respeito do que é feito com a cópia.		
	RISCO:	Violação da privacidade	**RECOMENDAÇÃO**:	• Conhecer os termos de privacidade da plataforma; • Uso de plataforma/aplicativos confiáveis. Ex.: Otimizador de PDF do TRT 14.

Fonte: autoria própria

Estas são apenas algumas recomendações frente aos riscos mais emergentes identificados nos levantamentos realizados. Importante observar que a Lei Geral de Proteção de Dados Pessoais, em seu Art. 3° que trata da aplicabilidade da LGPD informa que ela se aplica a qualquer operação de tratamento realizada por pessoa natural ou por pessoa jurídica de direito público ou privado, independentemente do meio, do país de sua sede ou do país onde estejam localizados os dados. Assim, tanto o operador do direito como os escritórios de advocacia devem estar aderentes a nova legislação e implementar um programa de governança de dados pessoais que trata em profundidade da proteção dos dados pessoais, a partir do prisma da segurança da informação e também do tratamento dos dados pessoais e dados pessoais sensíveis.

5.3 Conclusão do capítulo

Foram apresentados neste capítulo estudos de casos concretos, na vertente da relação máquina-pessoa natural, onde por meio da coleta de dados nos sistemas DJe, foi demonstrada a exposição da privacidade das partes nos processos que tramitam sob sigilo de justiça,

comprovando que o atual mecanismo não é suficiente para garantir o direito fundamental à privacidade.

Em contrapartida a esta problemática foi apresentada a sugestão de criação de uma tabela de de-identificação, a ser aplicada aos dados do judiciário, a partir do modelo HIPAA e também um modelo de implementação, MITRE *Identification Scrubber Toolkit* (MIST), como referência inicial para estudos futuros.

Mesmo não sendo objeto central da presente pesquisa, traçaram-se breves considerações a respeito da análise de dados das sentenças para a criação de perfilização de magistrados(as) no que diz respeito, por exemplo à forma de decisão, bases legais mais utilizadas, provas aceitas com maior frequência, etc.

Já na outra vertente, relação pessoa natural-máquina, foram abordadas as fragilidades dos advogados e escritórios de advocacia nas dimensões pessoas e processos referentes às boas práticas para a proteção dos dados pessoais dos assistidos.

Para tanto, foi realizada uma pesquisa exploratória qualitativa disponibilizada, on-line, por meio de questionário, respondida por 82 operadores do direito. Foi também realizada visita a uma sala do advogado, existente em todos os fóruns do DF, onde existem recursos como espaço para reunião, impressão de documentos e computadores disponibilizados aos advogados. Nos computadores foi analisada o diretório de *downloads* e *cookies* que permite salvar a senha de e-mail "gratuito".

A partir da análise das respostas ao questionário e da visita à sala do advogado e, identificados os principais comportamentos de risco, elaborou-se um conjunto de recomendações de boas práticas de fácil adoção, que contribui para um cenário de melhor gestão dos riscos à privacidade.

No próximo capítulo, é apresentada a conclusão da presente pesquisa bem como sugestões de trabalhos futuros.

Capítulo 6

Conclusão da pesquisa e sugestões de trabalhos futuros

Inicialmente vale ressaltar o papel fundamental da metodologia TEMAC (detalhe no item 3.1), CRISP-DM (Item 4.4 e Cap. 5) e a aplicação do questionário (Item 4.5) para o desenvolvimento deste trabalho.

No decorrer desta dissertação mostrou-se os riscos existentes à privacidade tanto na relação máquina-pessoa natural, ou seja, nos sistemas PJe e DJe, como na relação pessoa natural-máquina, no tocante às práticas dos advogados e escritórios de advocacia.

Nos últimos anos o Poder Judiciário tem, sistematicamente, envidado esforços no sentido de promover cada vez mais a adoção do PJe e consequentemente DJe por todos os tribunais e instância da justiça. Ação importante no sentido de promover a celeridade processual, reduzir custos e facilitar o acesso a justiça.

Tal medida tem o potencial de, ao alcance de um clique, expor a privacidade das partes, ao mesmo tempo em que advogados e escritórios de advocacia, por falta de treinamento e adoção de boas práticas têm, eventualmente, colocado em risco dados pessoais dos assistidos.

O atual modelo de sigilo de justiça mostra-se ineficiente frente à mineração de dados. Logo a adoção de mecanismos efetivos para resguardar a privacidade das partes como o mecanismo, proposto, de pseudonimização com a substituição dos atributos, constante na tabela proposta, por IDs únicos e gerados a partir de processo randômico com elevado grau de entropia, garantindo a ligação entre os registos e os titulares de dados e dessa forma possibilitando uma posterior re-identificação. Também o aprimoramento da formação e atualização dos magistrados quanto aos riscos à privacidade frente à mineração de dados a partir de *quasi-identifiers* constantes no corpo das sentenças, votos, decisões etc.

Por outro lado, tanto os advogados como os escritórios de advocacia precisam incorporar a gestão de risco à privacidade em seus procedimentos e desenvolver a cultura de proteção de dados pessoais. Como demonstrando ao utilizar máquinas públicas, serviços de nuvem e e-mails “gratuitos”, *pendrive* sem criptografia corre-se imenso risco de exposição de

dados pessoais dos assistidos. Neste sentido o presente trabalho apresenta um conjunto de 15 sugestões que podem ser implementadas com baixo custo e que permitem elevar o grau de proteção dos dados pessoais dos assistidos, tais como adoção de mecanismos de criptografia em *pendrive*, utilização de plataformas confiáveis e atenção ao utilizar computadores públicos.

No Brasil, em virtude, principalmente, da LGPD e da PEC 17/2018, que eleva a proteção de dados ao status de Direito jusfundamental, é grande a preocupação com os dados expostos no sistema judicial.

Conforme mostrado no capítulo 3, já existem diversos modelos de Gestão de Riscos que podem ser aplicados tanto ao PJe/DJe como à práticas dos advogados e escritórios de advocacia. Os modelos de de-identificação de dados, com a tabela HIPAA, é uma importante referência a ser considerada ao analisar o sistema de processo judicial eletrônico. No mesmo sentido, a Norma ANT/NBR ISO 27701:2019, assim como o NIST *Privacy Framework: a tool for improving privacy through enterprise risk management*, são modelos de boas práticas de mercado a serem considerados pelos advogados e escritórios de advocacia. Também no capitulo 4, foi detalhada a metodologia de pesquisa que envolve busca biliométrica em bases como *Web of Science* (WOS) e *Scopus*, de forma a selecionar artigos que representem o estado da arte na proteção da privacidade de dados pessoais. É correto afirmar que em virtude do Brasil ser um dos países com maior adoção de informatização do judiciário e elevada quantidade de processos, comparando-se com os países Europeus, há escassez de artigos com recorte no judiciário brasileiro. Entretanto, existe vasta literatura sobre proteção de dados de saúde que serviram como referência na presente pesquisa, pois, apesar de ser outro setor, trata dados pessoais e dados pessoais sensíveis.

No capítulo 4 foi apresentado o modelo CRISP-DM utilizado como referência a condução das fases de mineração de dados e o Modelo de Equações Estruturais (MEE) para a validação do questionário que foi aplicado aos advogados e escritórios de advocacia.

No capítulo 5, após a coleta de todas as publicações do Diário de justiça Eletrônico do Tribunal de Justiça do Distrito Federal e Territórios do ano de 2019, foi realizada a extração e mineração de dados com foco em cenários que envolvem a exposição de dados pessoais e dados pessoais sensíveis das partes, principalmente em processos que tramitam sob sigilo de justiça.

Foi apresentada uma proposta de de-identificação, tendo como referência a tabela do HIPAA, bem como o modelo MITRE *Identification Scrubber Toolkit* (MIST), uma ferramenta de código aberto de de-identificação de textos, além de outras duas medidas que

passam pela questão cultural, sendo elas, a formação dos magistrados(as) e criação de uma seção referente ao direito ao esquecimento a ser investigada em trabalhos futuro.

No capítulo 5, por meio de coleta de dados por questionário, respondido por operadores do direito foram identificadas práticas adotadas pelos advogados e escritórios de advocacia referentes à gestão de risco da privacidade dos assistidos. Também foi realizada análise de uma sala da advocacia existe em um dos fóruns do Distrito Federal.

Após a análise dos dados coletados foi apresentada uma proposta, composta de um conjunto de boas práticas nos pilares Pessoas e Processos, a serem adotadas pelos advogados e escritórios de advocacia.

Assim, o mecanismo que materializa o princípio da publicidade no PJe e DJe deve ser sopesado com razoabilidade e proporcionalidade ao direito à privacidade e intimidade, devendo-se dar preferência à privacidade, uma vez que integra um dos direitos fundamentais da pessoa humana assegurado constitucionalmente e principalmente em momento no qual a proteção de dados pessoais está em vias de ser elevado a status constitucional por meio da PEC 17/2018.

O PJe e Dje, dessa forma, devem assegurar a restrição de acesso a certos dados e informações, de modo que possa resguardar o direito à privacidade das partes, e, por outro lado, permitir o acesso do público em geral às decisões proferidas pelo Poder Judiciário para que, eventualmente, possam exercer o poder de controle que a constituição os confere e a democracia exige.

Da mesma forma, os operadores do direito e escritórios de advocacia devem adotar boas práticas para garantir a proteção dos dados pessoais e dados pessoais sensíveis de seus assistidos.

Não há dúvidas de que, pior que a venda ou entrega da privacidade de alguém ao público em geral e empresas que vivem do "capitalismo de vigilância", é fazê-lo contra a vontade e desconhecimento do titular dos dados e do direito, muitas vezes um anônimo que pretende assim permanecer. Isto é, pior do que jogar ao povo dados e informações sobre a vida privada de uma pessoa é fazê-lo contra a sua vontade, desconhecimento e de forma anônima. A mais importante faceta do direito à privacidade é quando as informações expostas acabam por fazer perecer a confiança e a boa imagem que o indivíduo (ou sua família) levou para conquistar.

Assim foram alcançados os objetivos propostos de analisar e propor soluções quanto aos riscos à privacidade na relação máquina-pessoa natural e pessoa natural-máquina. Vale

ressaltar que esta preocupação já faz presente na medida em que todo o sistema judiciário passa a dotar o modelo do PJe.

Como trabalho futuro vislumbra-se as seguintes possibilidades:

- Aplicação do trabalho nos DJes dos demais Estados da federação;
- Construção de *Big Data Analytics* com perfilização de magistrados(as), sendo estes estudos de grande relevância não só para o judiciário mas para a sociedade como um todo;
- Perfilização de advogados e escritórios de advocacia;
- Anpalise de decisões, sentenças, votos, acórdão, etc. Com aplicação de Processamento de Linguagem Natural (PLN) para análise de retórica e argumentação;
- Analise da efetividade de normas diversas e setoriais como por exemplo Lei Maria da penha;
- Acompanhamento do volume e natureza das demandas por área do direito;
- Analise das jurisprudências mais citadas em decisões;
- Busca e identificação de vieses em decisões, sentenças, votos e acórdão;
- Uso da mineração de dados em outras vertentes a partir dos dados do PJe / DJe;
- Construção de modelos de predição a partir da análise de sentenças.

REFERÊNCIAS

ABERDEEN, John & Bayer, Samuel & Yeniterzi, Reyyan & Wellner, Ben & Clark, Cheryl & Hanauer, David & Malin, Bradley & Hirschman, Lynette. (2010). The MITRE Identification Scrubber Toolkit: Design, training, and assessment. International journal of medical informatics. 79. 849-59. 10.1016/j.ijmedinf.2010.09.007.

ABNT. ABNT ISO GUIA 73 - Gestão de riscos - Vocabulário. Associação Brasileira de Normas Técnicas, Rio de Janeiro, nov. 2009.

ABNT. ABNT ISO GUIA 73 - Gestão de riscos - Vocabulário. Associação Brasileira deNormas Técnicas, Rio de Janeiro, nov. 2009.

ABNT. ABNT NBR ISO/IEC 31000 – Gestão de riscos – princípios e diretrizes. Associação Brasileira de Normas Técnicas, Rio de Janeiro, nov. 2018.

ABNT. ABNT NBR ISO/IEC 31010 – Técnicas de avaliação de riscos. Associação Brasileira de Normas Técnicas, Rio de Janeiro, nov. 2009.

ABNT. ABNT NBR ISO/IEC 27701 – Técnicas de segurança — Extensão da ABNT NBR ISO/IEC 27001 e ABNT NBR ISO/IEC 27002 para gestão da privacidade da informação. Associação Brasileira de Normas Técnicas, Rio de Janeiro, dez. 2019.

ARTICLE 29. Data Protection Working Party, "Opinion 05/2014 on Anonymisation Techniques," 2014. Dsiponível em: https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2014/wp216_en.pdf. Acesso em: 02 Fev. 2020.

BALIEIRO, J. C. Introdução à estatística. USP. Disponível em: http://www.usp.br/gmab/discip/zab5711/aula1_slides.pdf. Acesso em: 15 abr. 2020.

BAYARDO, Roberto & Agrawal, Rakesh. (2005). *Data Privacy through Optimal k-Anonymization. Proceedings - International Conference on Data Engineering*. 2010. 217-228. 10.1109/ICDE.2005.42.

BERNSTEIN, Peter L. DESAFIO AOS DEUSES: A FASCINANTE HISTÓRIA DO RISCO. Rio de Janeiro: Campus, 1997.

BRASIL. Constituição da república federativa do Brasil de 1988. Disponível em: http://www.planalto.gov.br/ccivil_03/constituicao/constituicao.htm. Acesso em: 12 Jan. 2020.

BRASIL. Lei nº 11.419, de 19 de dezembro de 2006. Dispõe sobre a informatização do processo judicial. Disponível em: http://www.planalto.gov.br/ccivil_03/_ato2004-2006/2006/lei/l11419.htm. Acesso em: 12 Jan. 2020.

BRASIL. Lei nº 13.105, de 16 de março de 2015. Código de Processo Civil. Disponível em: http://www.planalto.gov.br/ccivil_03/_ato2015-2018/2015/lei/l13105.htm. Acesso em: 24 Jan. 2020.

BRASIL. Lei nº 13.709 Lei n° 13.709, de 14 de agosto de 2018.Lei Geral de Proteção de Dados Pessoais. Disponível em: http://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/L13709.htm. Acesso em: 24 Jan. 2020.

CAIRE, E. A HISTÓRIA DA ORIGEM DA CURVA NORMAL. Dissertação de mestrado. UNESP. Disponível em: https://repositorio.unesp.br/handle/11449/91024. Acesso em: 15 abr. 2020. 2013.

CALAZANS, A. T. S., ARI M. M. e ROBERTO A. P.: Uma revisão sistemática da bibliografia sobre m?tricas funcionais de tamanho de software utilizando o enfoque meta-analítico. Universitas: Gestão e TI, 5(2), dezembro 2015, ISSN 2179-832X, 2179-8311. Disponível em: http://www.publicacoes.uniceub.br/index.php/gti/article/view/3532. Acesso em: 27 Jan. 2020

CIURIAK, Dan, The Economics of Data: Implications for the Data-Driven Economy (February 4, 2018). Chapter 2 in "Data Governance in the Digital Age," Centre for International Governance Innovation, 5 March 2018. Disponível em: http://dx.doi.org/10.2139/ssrn.3118022. Acesso em: 04 abr. 2020

CGI. Comitê Gestor da Internet no Brasil. Pesquisa sobre o uso das tecnologias de informação e comunicação nos domicílios brasileiros : TIC domicílios [livro eletrônico] / Núcleo de Informação e Coordenação do Ponto BR, [editor]. -- São Paulo : Comitê Gestor da Internet no Brasil, 2019. 3.800 Kb; Disponível em: https://www.cetic.br/media/docs/publicacoes/2/12225320191028-tic_dom_2018_livro_eletronico.pdf. Acesso em: 23. Jan. 2020

CHAPMAN, Pete (NCR), CLINTON, Julian (SPSS), KERBER, Randy (NCR), KHABAZA, Thomas (SPSS), REINARTZ, Thomas (DaimlerChrysler), SHEARER, Colin (SPSS) e WIRTH, Rüdiger (DaimlerChrysler). Crisp 1.0. process and user guide, Disponível em: https://www.the-modeling-agency.com/crisp-dm.pdf. Acesso em: 12 Jan. 2020.

CHIN, W. W. (1998). The partial least squares approach for structural equation modeling. in G. A. Marcoulides (Ed.), Modern methods for business research (pp. 295–236). London: Lawrence Erlbaum Associates.

COHEN, J. (1988). *Statistical power analysis for the behavioral sciences* (2.ª ed.). Hillsdale: Lawrence Erlbaum Associates.

COHEN, J. (1992a). *A power primer. Psychological Bulletin*, 112(1), 155.

COHEN, J. (1992b). *Statistical power analysis. Current Directions in Psychological Science*, 1(3), 98–101.

RINGLE, C.; SILVA, D.; BIDO, D. Modelagem de Equações Estruturais com Utilização do Smartpls. REMark: Revista Brasileira de Marketing. 13. 10.5585/bjm.v13i2.2717, 2014.

FORNELL, C.; LARCKER, D. *Structural Equation Models with Unobservable Variables and Measurement Error: Algebra and Statistics. Journal of Marketing Research*, 18(3), 382-388. doi:10.2307/3150980, 1981.

CNJ. Conselho Nacional de Justiça. Justiça em números 2019: Ano-base 2018. 15º ed. Brasília: CNJ, 2019. Disponível em: https://www.cnj.jus.br/wp-content/uploads/conteudo/arquivo/2019/08/justica_em_numeros20190919.pdf. Acesso em: 23 Jan. 2020.

CONESA, Fugencio M. *Derechoa la intimidad, informática y Estado de Derecho*, Universidad de Valencia, Valencia, 2007.

CULP, Christopher. L. *The revolutionin corporateriskmanagement*: *A decade of innovations in process and products*. Journal of Applied Corporate Finance. V. 14, 2002.

DERNONCOURT, Franck & Lee, Ji & Uzuner, Ozlem & Szolovits, Peter. (2016). *De-identification of Patient Notes with Recurrent Neural Networks. Journal of the American Medical Informatics Association* : JAMIA. 24. 10.1093/jamia/ocw156.

DIONNE, Georges. RISK MANAGEMENT: HISTORY, DEFINITION AND CRITIQUE.2013.Disponível em: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2231635. Acesso em: 27 Jan.2020.

DONEDA, D. C. M; MENDES, L. S; SOUZA, C. A. P.; ANDRE, N. N. G., Considerações sobre inteligência artificial, ética e autonomia pessoal. Pensar – Revista de ciências jurídicas (Fortaleza), v. 23, n. 4, out./dez. 2018. P. 4.

EMAM, Khaled. (2010). *Risk-Based De-Identification of Health Data. Security & Privacy*, IEEE. 8. 64 - 67. 10.1109/MSP.2010.103.

EMAM, K.; JONKER, E.; ARBUCKLE, L.; MALIN, B. *A Systematic Review of Re-Identification Attacks on Health Data*. PloS one. 6. e28071. 10.1371/journal.pone.0028071, 2011.

EUROPAN E-JUSTICE. *Online processing of cases and e-communication with courts*. Disponível em: https://beta.e-

justice.europa.eu/280/EN/online_processing_of_cases_and_ecommunication_with_courts?init=true. Acesso em: 24 Fev. 2020.

FAYYAD et al. *Advances in Knowlwdge Discovery and Data Mining. Menlo Park*: AAAI Press, 1996. 560 p, 2011.

FAYYAD, U.M.; PIATETSKY-SHAPIRO, G.; SMYTH, P. *Advances in knowledge disc overy and data mining American Association for Artificial Intelligence, Menlo Park*, CA, USA (1996a), pp. 1-34 Disponível em: http://dl-acm-org.ez54.periodicos.capes.gov.br/citation.cfm?id=257938.257942. Acesso em: 27 mar. 2020.

FERNANDES, Andrea & Cloete, Danielle & Broadbent, Matthew & Hayes, Richard & Chang, Chin-Kuo & Jackson, Richard & Roberts, Angus & Tsang, Jason & Soncul, Murat & Liebscher, Jennifer & Stewart, Robert & Callard, Felicity. (2013). *Development and evaluation of a de-identification procedure for a case register sourced from mental health electronic records. BMC medical informatics and decision making*. 13. 71. 10.1186/1472-6947-13-71.

FONSECA, W.R. BENEFÍCIOS GERADOS COM A IMPLANTAÇÃO DO SISTEMA DE GESTÃO DA QUALIDADE: um estudo de caso de uma organização no segmento de fundição em Lagoa da Prata-MG, 2010.

FORD, Elizabeth & Carroll, John & Smith, Helen & Scott, Donia & Cassell, Jackie. (2016). *Extracting information from the text of electronic medical records to improve case detection: A systematic review. Journal of the American Medical Informatics Association*. 23. ocv180. 10.1093/jamia/ocv180.

FRAZÃO. A.; TEPEDINO, G.; OLIVA, M., Lei Geral de Proteção de Dados Pessoais e suas repercurssões no Direito Brasileiro. São Paulo: Thomson Reuters Brasil, 2019.

PEREIRA, G. A. Estimadores Ridge generalizados adaptados em modelos de equações estruturais: estudo de simulação e aplicação no perfil de consumidores de café. 2014. 80 p. PhD thesis, Tese (Doutorado em Estatística e Experimentação Agropecuária)– Universidade Federal de Lavras, Lavras, 2014. Disponível em: http://repositorio. ufla. br/handle/1/4461. Acesso em: 04 jan, 2020.

GANSLANDT, Thomas & Mate, Sebastian & Helbing, Krister & Sax, Ulrich & Prokosch, Hans-Ulrich. (2011). *Unlocking Data for Clinical Research – The German i2b2 Experience. Applied Clinical Informatics*. 2. 116-127.

GESPUBLICA/MPOG. GUIA DE ORIENTAÇÃO PARA O GERENCIAMENTO DE RISCOS. Disponível em: http://www.gespublica.gov.br/sites/default/files/documentos/p_vii_risco_oportunidade.pdf. Acessado em: 02 Jan. 2020. 2013.

GIL, A. C.: Métodos e técnicas de pesquisa social. Atlas, São Paulo, 2008, ISBN 978-85-224-5142-5. OCLC: 298931695.

HAN, J.; Kamber, M.; Pei, J. Data Mining: Concepts and Techniques. 3 ed Walthan, MA: Morgan Kaufmann Series in Data Management Systems, 2011, 744 p.

HIPAA. Guidance Regarding Methods for De-identification of Protected Health Information in Accordance with the Health Insurance Portability and Accountability Act (HIPAA) Privacy Rule," HHS - Office for Civil Rights (OCR), 2012. Disponível em: https://www.hhs.gov/sites/default/files/ocr/privacy/hipaa/understanding/coveredentities/De-identification/hhs_deid_guidance.pdf. Acesso em: 27 Jan. 2020.

IBGE. Instituro Brasileiro de Geografia e Estatística. Estimativas da população dos municípios para 2019. Disponível em: https://agenciadenoticias.ibge.gov.br/agencia-sala-de-imprensa/2013-agencia-de-noticias/releases/25278-ibge-divulga-as-estimativas-da-populacao-dos-municipios-para-2019. Acesso em: 22 Fev. 2020.

IGLESIAS, A. M. B.; OLIVEIRA, J. O. DA S. Z. DE; MARQUES, J. F. ASPECTOS CONTROVERSOS DO PROCESSO ELETRÔNICO. REVISTA ESMAT, v. 6, n. 7, p. 11-42, 12 jun. 2015.

JAWUREK, Marek & Johns, Martin & Rieck, Konrad. (2011). *Smart metering de-pseudonymization. ACM International Conference Proceeding Series*. 227-236. 10.1145/2076732.2076764.

HAIR, J. F.; BLACK, William C.; BABIN, Barry J.; ANDERSON, Rolph E.; TATHAM, Ronald L. Análise Multivariada de Dados. Tradução Adonai Schlup Sant'Anna. 6. ed., Porto Alegre, Bookman, 2009.

HAIR, J. F.; HULT, G. Tomas M.; RINGLE, Christian; SARSTEDT, Marko. *A Primer on Partial Least Squares Structural Equation Modeling*, 2014.

HAIR, J. F.; RINGLE, C. M.; SARSTEDT, M. *PLS-SEM: Indeed a Silver Bullet. Journal of Marketing Theory and Practice,* 19(2), 139-151. doi: 10.2753/MTP1069-6679190202, 2011.

KLINE, R. B. (2004). *Beyond significance testing: Reforming data analysis methods in behavioral research* (2.ª ed.). Washington, DC: American Psychological Association.

LAPRIE, J.C. *Dependable Computing and Fault Tolerance: Concepts and Terminology. Twenty-Fifth International Symposium on Fault-Tolerant Computing*, 1995.

LIANG, H., & XUE, Y. UNDERSTANDING SECURITY BEHAVIOURS IN PERSONAL COMPUTER USAGE: A THREAT AVOIDANCE PERSPECTIVE. Association for Information Systems, 11(7), 394–413. 2010.

LIANG, H., & XUE, Y. AVOIDANCE OF INFORMATION TECHNOLOGY THREATS: A THEORICAL PERSPECTIVE. MIS Quarterley (33), 2009, p. 71-90.

LUK Arbuckle, EMAM Khaled El. *Anonymizing Health Data: Case Studies and Methods to Get You Started*, 1st ed.: O'Reilly Media, 2013.

MARCONI, M. A.; EVA M. L.: Metodologia científica. Atlas, São Paulo, 2009, ISBN 978-85-224-4762-6. OCLC: 422876997.

MARIANO, A. M., & ROCHA, M. S. Revisão da literatura: apresentação de uma abordagem integradora. In AEDM International Conference–Economy, Business and Uncertainty: Ideas for a European and Mediterranean industrial policy. Reggio Calabria (Italia). Disponível em https://www.researchgate.net/publication/319547360_Revisao_da_Literatura_Apresentacao_de_uma_Abordagem_Integradora. Acesso em: 15 Fev. 2020.

MARTINS, Ives Gandra da Silva; MARTINS, Rogério Vidal Gandra da Silva. Privacidade na comunicação eletrônica. Ci. Inf., Brasília, v. 30, n. 1, p. 13-18, Apr. 2001. Disponível em: http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0100-19652001000100003&lng=en&nrm=iso. Acesso em: 23 Jan. 2020. http://dx.doi.org/10.1590/S0100-19652001000100003.

MD SELAMAT, S. A.; PRAKOONWIT, S.; SAHANDI, R.; Khan, W.; RAMACHANDRAN, M. (2018). *Big data analytics-A review of data-mining models for small and medium enterprises in the transportation sector*. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery. 8. e1238. 10.1002/widm.1238. Disponível em: https://onlinelibrary-wiley.ez54.periodicos.capes.gov.br/doi/full/10.1002/widm.1238. Acesso em: 27 mar. 2020

MEYSTRE, Stephane & Friedlin, F & South, Brett & Shen, Shuying & Samore, Matthew. (2010). *Automatic de-identification of textual documents in the electronic health record: A review of recent research. BMC medical research methodology*. 10. 70. 10.1186/1471-2288-10-70.

NELSON, Gregory. (2015). *Practical Implications of Sharing Data: A Primer on Data Privacy, Anonymization, and De-Identification*.

NEUBAUER, Thomas & Heurix, Johannes. (2010). *A methodology for the pseudonymization of medical data. International journal of medical informatics*. 80. 190-204. 10.1016/j.ijmedinf.2010.10.016.

NIST. National Institute of Standards and Technology (2020) Privacy Framework, Version 1.0. (National Institute of Standards and Technology, Gaithersburg, MD). Disponível em: https://www.nist.gov/system/files/documents/2020/01/16/NIST%20Privacy%20Framework_V1.0.pdf. Acessado em: 15 Fev. 2020.

NUTT, Paul C; Wilson, David C. *Crucial Trends and Issues in Strategic Decision Making*. *In: Handbook of Decision Making*. Edited by Paul C. Nutt and David C. W ilson. John Wiley & Sons, Ltd. 2010.

RICHARDSON, R. j., JOSÉ, A. S.: Pesquisa social: métodos e técnicas. Atlas, São Paulo, 3. ed. rev e ampl edição, 1999, ISBN 978-85-224-2111-4. OCLC: 46755954.

RINDFLESCH, Thomas & Corn, Milton. (2013). *Natural Language Processing: State of the Art and Prospects for Significant Progress, A workshop sponsored by the National Library of Medicine*. Journal of biomedical informatics. 46. 10.1016/j.jbi.2013.06.004.

ROKACH, L.; Maimon, O. Data Mining with Decision Trees: Theory and Applications. Series in Machine Perception and Artificial Intelligence. V. 69, Cingapura: World Scientific, 2008, 264 p.

RUSSOM, P., 2011. *Big Data Analytics*. TDWI *Best Practices Report, Fourth Quarter*. Disponível em: https://vivomente.com/wp-content/uploads/2016/04/big-data-analytics-white-paper.pdf. Acesso em: 28 mar. 2020.

SANTOS, Boaventura de Sousa. Os tribunais e as novas tecnologias de comunicação e de informação. Sociologias, Porto Alegre , n. 13, p. 82-109, June 2005 . Disponível em: http://www.scielo.br/scielo.php?script=sci_arttext&pid=S1517-45222005000100004&lng=en&nrm=iso. Acessado em: 14 Fev. 2020. https://doi.org/10.1590/S1517-45222005000100004.

SANTO, H. E.; DANIEL, F. Calcular e apresentar tamanhos do efeito em trabalhos científicos (1): As limitações do p < 0,05 na análise de diferenças de médias de dois grupos. Revista Portuguesa de Investigação Comportamental e Social. Disponível em: https://www.researchgate.net/profile/Helena_Espirito_Santo/publication/273143169_Calcular_e_apresentar_tamanhos_do_efeito_em_trabalhos_cientificos_1_As_limitacoes_do_p_005_na_analise_de_diferencas_de_medias_de_dois_grupos_Calculating_and_reporting_effect_sizes_on_scientific_papers/links/54fa10290cf2040df21b1b1c/Calcular-e-apresentar-tamanhos-do-efeito-em-trabalhos-cientificos-1-As-limitacoes-do-p-0-05-na-analise-de-diferencas-de-medias-de-dois-grupos-Calculating-and-reporting-effect-sizes-on-scientific-pap.pdf. Acesso em: 15 abr. 2020. 2015

SILVA, L. A., Peres, S. M. & Boscarioli, C. Introdução à Mineração de Dados: Com aplicações em R. Editora Campus – Série SBC, ISBN: 9788535284461, 296p., 2016.

SIMSON, L. Garfinkel, "NIST.IR.8053 - De-Identification of Personal Information," 2015. Disponível em: https://nvlpubs.nist.gov/nistpubs/ir/2015/NIST.IR.8053.pdf. Acesso em: 11 Fev. 2020.

SINGH, Ashish & Chatterjee, Kakali. (2019). *Security and Privacy Issues of Electronic Healthcare System: A Survey*.

TCU. Tribunal de Contas da União. Processo: TC 008.903/2018-2, 2019. Dispinível em: https://portal.tcu.gov.br/imprensa/noticias/tcu-aponta-atrasos-na-implementacao-do-processo-judicial-eletronico.htm. Acesso em: 27 Set. 2019.

TCU. Tribunal de Contas da União. Processo: Acórdão 1534/2019 – Plenário. Relator: Raimundo Carreiro. Dispinível em: https://pesquisa.apps.tcu.gov.br/#/documento/acordao-completo/1.534%252F2019/%20/DTRELEVANCIA%20desc,%20NUMACORDAOINT%20desc/0/%20?uuid=290af640-a1a9-11e9-8050-7d5646f645d1. Acesso em: 19 Fev. 2020.

VERGARA, S. C.: Projetos e relatórios de pesquisa em administração. São Paulo: Atlas, 2000. Métodos de pesquisa em administração, 3, 2009.

VIEIRA, S. Como elaborar questionários. São Paulo: Atlas, 2009.

WANG, G.; GUNASEKARAN, A.; NGAI, E.W.T.; PAPADOPOULOS, T. (2016) *Big data analytics in logistics and supply chain management: Certain investigations for research and applications. International Journal of Production Economics*, 176 . pp. 98-110. ISSN 0925-5273. (doi:10.1016/j.ijpe.2016.03.014) (KAR id:57194). Disponível em: https://www-sciencedirect.ez54.periodicos.capes.gov.br/science/article/pii/S0925527316300056?via%3Dihub#bib120. Acesso em: 27 mar. 2020.

WAZLAWICK, R. S.: Metodologia de pesquisa para ciência da computação. 2014, ISBN 978-85-352-7783-8. Disponível em: http://www.sciencedirect.com/science/book/9788535277821. Acesso em: 20 Dez. 2019 OCLC: 902734376.

ZUPIC, I; CATER, T.. Bibliometric Methods in Management and Organization. Organizational Research Methods, p.1-44, 2014.

APÊNDICE

EXAME DE RESSONÂNCIA

[redacted]

Nome: [redacted] **Idade:** [redacted]

Data Nascimento: [redacted] **Código:** [redacted]

Médico Solicitante: Dr. [redacted] **Data Exame:** [redacted]

RESSONÂNCIA MAGNÉTICA DO CRÂNIO

TÉCNICA

Exame realizado em equipamento de 1,5 T, com sequências em vários planos e ponderações, antes e após a infusão venosa do meio de contraste.
Exame tecnicamente prejudicado por artefatos de movimentação do paciente, dregadando algumas imagens

ACHADOS

Formação cística, extra-axial, na linha média da fossa posterior, retrocerebelar, sem realce pelo meio de contraste, com cerca de 4,5 x 2,7 x 2,1 cm.
Sulcos, cisternas e fissuras cerebrais de aspecto anatômico.
Parênquima cerebral com intensidade de sinal normal, com boa definição do córtex e substância branca.
Estruturas do tronco cerebral sem alterações.
Ventrículos laterais, terceiro e quarto ventrículo com dimensões dentro dos limites da normalidade.
Hemisférios cerebelares simétricos, com lóbulos e folias de morfologia preservada.
Não se observa captação anômala de contraste.
A sequência difusão não evidenciou lesões isquêmicas agudas.

IMPRESSÃO DIAGNÓSTICA

Provável cisto aracnoideo na fossa craniana posterior.

OBS.: Sinusopatia inflamatória maxiloetmoidal bilateral.

EXAME CLÍNICO – 22 EXAMES

[redacted]

[redacted]

Nome : [redacted]
RG : [redacted]
DN : [redacted]
Médico : [redacted]
Convênio: [redacted]
Unidade : [redacted]

Código da OS : [redacted]

Atendimento : [redacted]
Qnt de exames: 22
Página: 1/22

Responsável Técnico: [redacted]
Endereço da Unidade: [redacted]
Laboratório registrado no CRF/DF sob o número [redacted]

HEMOGRAMA COMPLETO

Método : Análise realizada por Citometria de fluxo fluorescente e impedância "XN10-Sysmex"
Material: SANGUE TOTAL COM EDTA

Eritrograma	Resultados		Valores de Referência Homens	Mulheres	Crianças	Acima de 70 anos
HEMACIAS	4,65	milhões/mm3	4,50 a 6,10	4,00 a 5,40	4,07 a 5,37	3,90 a 5,36
HEMOGLOBINA	13,7	g/dL	13,0 a 16,5	12,0 a 15,8	10,5 a 14,0	11,5 a 15,1
HEMATOCRITO	38,9	%	36,0 a 54,0	33,0 a 47,8	30,0 a 44,5	33,0 a 46,0
VCM	83,7	fl	80,0 a 98,0	80,0 a 98,0	70,0 a 86,0	80,0 a 98,0
HCM	29,5	pg	26,8 a 32,9	26,2 a 32,6	23,2 a 31,7	27,0 a 31,0
CHCM	35,2	g/dl	30,0 a 36,5	30,0 a 36,5	30,0 a 36,5	30,0 a 36,5
RDW	12,9	%	11,0 a 16,0	11,0 a 16,0	11,0 a 16,0	11,0 a 16,0

Leucograma	Resultado %	Resultado /mm3	Valores de Referência Adultos %	Adultos /mm3	Crianças (Menores de 8 anos) %	Crianças (Menores de 8 anos) /mm3
LEUCOCITOS		6.300	100	3.600 a 11.000	100	4.000 a 14.000
BASTONETES	0	0	0-5	0 a 550	0-6	0 a 450
SEGMENTADOS	35	2.205	40-70	1.480 a 7.700	30-64	1.200 a 9.600
EOSINOFILOS	9	567	0-7	0 a 550	0-5	0 a 550
BASOFILOS	2	126	0-2	0 a 220	0-2	0 a 300
LINFOCITOS	43	2.709	20-50	740 a 5.500	38-70	1.520 a 10.500
MONOCITOS	11	693	3-14	37 a 1.500	3-14	40 a 1.700

OBSERVAÇÃO: EOSINOFILIA CONFIRMADA

Série Plaquetária	Resultados		Valores de Referência Adultos	Crianças
PLAQUETAS	342	x 10^3/mm3	130 a 450 x 10^3/mm3	140 a 500 x 10^3/mm3
VMP	10,7 fl		6,8 a 12,6	6,8 a 12,6

Coleta: [redacted] - 08:46:00
Liberação: [redacted] - 12:01:03

RESPOSTAS AO QUESTIONÁRIO - (PARA VALIDAÇÃO SEM)

	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y
1	Tec2	Tec3	Tec4	Tec5	Pess1	Pess2	Pess3	Pess4	Pess5	CutOrg1	CutOrg2	CutOrg3	CutOrg4	CutOrg5	CutOrg6	CutOrg7	CutOrg8	CutOrg9	CutOrg10	CutOrg11	Pront1	Pront2	Pront3	Pront4
2	2	1	2	2	1	1	1	1	1	1	1	1	1	2	2	2	2	2	2	2	5	4	1	5
3	2	1	2	2	2	2	2	2	2	1	1	1	2	2	2	1	1	2	1	2	4	2	2	4
4	2	1	2	2	2	2	2	1	2	2	2	2	2	2	2	1	2	2	1	2	4	3	2	4
5	2	2	2	2	2	2	2	1	1	1	2	1	2	2	2	1	1	2	2	1	5	4	2	3
6	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	4	4	4	4
7	2	1	2	2	1	2	1	1	1	1	2	1	1	1	1	1	1	1	1	2	5	2	1	4
8	2	1	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	4	4	4	4
9	2	1	2	2	1	1	2	1	2	2	2	1	2	2	1	2	1	1	1	1	4	4	4	4
10	2	2	2	2	2	2	2	1	1	2	2	2	2	2	2	1	1	2	1	1	4	2	1	2
11	2	1	1	2	2	2	2	2	2	1	2	1	1	1	1	1	2	2	1	2	5	1	2	4
12	2	1	2	2	1	2	2	2	2	2	2	2	2	2	2	1	2	2	2	2	4	4	4	4
13	1	1	2	1	2	1	1	1	1	1	2	1	1	1	1	2	1	2	1	2	4	3	1	4
14	1	1	2	2	1	2	2	1	2	1	1	1	1	1	2	2	1	1	1	2	5	1	1	4
15	1	1	2	2	1	1	1	1	1	1	1	1	1	2	2	1	1	2	1	1	4	2	1	5
16	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	1	2	4	4	5	1
17	1	1	1	1	2	1	1	1	1	2	2	1	2	2	2	1	2	2	1	2	5	4	2	5
18	2	1	1	2	1	2	2	1	2	2	2	2	2	2	2	1	1	2	1	2	4	4	2	5
19	2	1	2	2	2	1	2	1	2	2	2	2	2	2	2	2	2	2	2	1	5	5	2	4
20	2	1	1	1	1	1	1	1	1	2	2	1	2	2	2	1	1	2	2	1	5	5	1	4
21	2	1	2	2	2	2	2	1	1	1	2	1	1	1	1	2	1	1	1	2	4	4	2	4
22	2	2	2	2	1	1	1	1	1	1	2	1	1	2	2	2	1	1	1	1	5	4	2	3
23	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	4	4	4	4
24	1	1	2	2	2	2	1	1	1	2	2	2	2	2	2	1	2	2	2	1	4	4	2	4
25	1	1	1	1	1	1	1	1	1	1	2	1	1	1	1	2	1	2	1	2	5	2	1	5
26	1	1	1	2	2	1	2	1	2	1	1	1	1	2	1	2	1	2	1	2	4	4	2	4
27	1	1	2	1	1	1	1	1	1	2	2	1	2	2	2	1	1	2	1	1	5	2	1	4
28	2	2	2	2	1	2	2	1	2	1	2	1	1	1	1	2	1	1	1	2	5	4	1	4
29	2	1	1	2	2	2	2	1	2	2	2	2	2	2	2	1	2	2	2	1	4	3	1	2
30	1	1	2	2	2	2	2	2	2	1	2	1	1	1	1	2	1	1	1	2	4	3	2	4
31	1	1	1	1	1	1	1	1	1	2	1	2	2	1	2	2	1	2	2	1	5	2	1	5

QUESTIONÁRIO

Levantamento da cultura de segurança digital entre operadores e estudantes de direito

0 %

Página 1

Bem-vindo(a),

Este questionário tem por objetivo coletar algumas percepções dos operadores e estudantes de direito em relação à segurança cibernética.

Servirá como um dos elementos para a elaboração de um artigo com recomendações de boas práticas de segurança digital no dia a dia do advogado(a) e escritórios.

É composto de algumas questões de respostas simples, a maioria sim ou não.

É possível responde-lo em menos de 2 minutos.

Conto com sua participação e ao final se informar seu e-mail receberá em primeira mão o artigo em elaboração.

= = =
= = = Desloque a página para baixo e para cima para responder as perguntas. Ao final clique em concluir = = =
= = =

Obrigado!

Frank Ned Santa Cruz de Oliveira

Você utiliza algum e-mail gratuito para uso profissional? Ex.: gmail, uol, yahoo *

- ◯ sim
- ◯ não

Já aconteceu de você enviar um e-mail, com anexo, para destinatário incorreto, por engano? *

- ◯ sim
- ◯ não

Você já esqueceu em algum ambiente ou perdeu um Pendrive com documentos de trabalho? *

- ◯ sim
- ◯ não

Você utiliza algum serviço, gratuito, de nuvem? Ex: dropbox, google drive, azure *

- ◯ sim
- ◯ não

Você já teve que utilizar um computador público para editar um documento de trabalho? Ex: LAN House, Hotel, Sala de apoio ao advogado(a) da OAB *

- ◯ sim
- ◯ não

Você utiliza algum sistema on-line para redução do tamanho seu pdf ou para converte-lo no formato pdf/A? Se sim, por gentileza, indique qual. Ex.: "I love pdf", "pdf resize" etc.

Eu consigo, com sucesso, identificar vulnerabilidades tecnológicas em meus dispositivos? Ex.: computador, smartphone, tablet etc

- ◯ sim
- ◯ não

Eu consigo, com sucesso, proteger meu certificado digital?

- ◯ sim
- ◯ não

Eu consigo, com sucesso, identificar uma mensagem de SPAM?

- ◯ sim
- ◯ não

Eu consigo, com sucesso, identificar uma rede Wi-Fi não legítima?

- ◯ sim
- ◯ não

Eu consigo, com sucesso, usar recursos de criptografia?

- ◯ sim
- ◯ não

Em relação ao aprendizado sobre segurança digital. Qual resposta melhor representa sua realidade?

- ( ) Eu consigo, com sucesso, aprender sobre segurança cibernética, mesmo sem ter lido nada a respeito antes.
- ( ) Eu consigo, com sucesso, aprender sobre segurança cibernética, somente tendo referência de informações.
- ( ) Eu consigo, com sucesso, aprender sobre segurança cibernética, mesmo sem ajuda de ninguém.
- ( ) Eu consigo, com sucesso, apender sobre segurança cibernética, se tiver muito tempo disponível.
- ( ) Eu consigo, com sucesso, aprender sobre segurança cibernética, mesmo sem ninguém me ensinar.

Na sua instituição existe uma pessoa dedicada a cuidar das questões de segurança da informação? Ex.: escritório, local de trabalho, estágio

- ( ) sim
- ( ) não

Existe uma verificação de vida pregressa em relação às pessoas que acessam e manipulam dados da instituição?

- ( ) sim
- ( ) não

A instituição possui uma política de segurança da informação?

- ( ) sim
- ( ) não

A instituição possui uma política que detalha as regras para formação de senhas?

- ◯ sim
- ◯ não

Os colaboradores recebem treinamento sobre cuidados com segurança da informação?

- ◯ sim
- ◯ não

Você acredita que o nível de segurança, da informação, em seu ambiente de trabalho/estágio/estudo é elevado?

- ◯ Sim
- ◯ Não
- ◯ Não sei responder
- ◯ Outra resposta

Já houve ou Existe compartilhamento de usuário/senha para acesso a informações?

- ◯ sim
- ◯ não

Existe controle de acesso aos sistemas que armazenam dados sensíveis?

- ◯ sim
- ◯ não

No seu ambiente de trabalho ou estágio existe controle de uso das portas USB dos equipamentos?

- ◯ sim
- ◯ não

No seu ambiente de trabalho ou estágio existe controle de acesso a sites da Internet?

- ◯ sim
- ◯ não

No seu ambiente de trabalho ou estágio, para acessar a rede Wi-Fi, existe uma senha única para todos?

- ◯ sim
- ◯ não

Todos possuem responsabilidade de proteger seu computador de ataques e garantir que as informações estão armazenadas de forma segura?

- ( ) Concordo completamente
- ( ) Concordo
- ( ) Nem discordo, nem concordo
- ( ) Discordo
- ( ) Discordo completamente

Se um atacante quer roubar informações, eu posso fazer muito pouco para impedi-lo?

- ( ) Concordo completamente
- ( ) Concordo
- ( ) Nem discordo, nem concordo
- ( ) Discordo
- ( ) Discordo completamente

Segurança da informação é algo que deve ser foco somente das equipes técnicas?

- ( ) Concordo completamente
- ( ) Concordo
- ( ) Nem discordo, nem concordo
- ( ) Discordo
- ( ) Discordo completamente

Eu consigo, com segurança, reconhecer uma mensagem de e-mail suspeita?

- ( ) Concordo completamente
- ( ) Concordo
- ( ) Nem discordo, nem concordo
- ( ) Discordo
- ( ) Discordo completamente

Esta pesquisa está sendo conduzida por Frank Ned Santa Cruz de Oliveira. Caso tenha qualquer sugestão ou dúvida gentileza enviar para: ppcafn@gmail.com Caso tenha interesse em receber o resultado da pesquisa gentileza informar seu e-mail: