Nome: Cíntia Rosa Pereira de Lima

Título do projeto em português: A dicotomia entre dados pessoais e dados anônimos: desafios e perspectivas regulatórias no Brasil.

Título do projeto em e inglês: *The dichotomy between personal data and anonymous data: regulation challenges and prospects in Brazil.*

Período: 6 meses, 1º semestre de 2020

Resumo:
Os modelos regulatórios sobre proteção de dados estão estruturados sobre dois pilares: o consentimento livre e informado - que exige a manifestação consciente e voluntária do indivíduo cujos dados serão coletados e tratados - e a anonimização, ou seja, a aplicação de técnicas que impedem a identificação do indivíduo a quem os dados dizem respeito. Entretanto, afirma-se que o uso de algoritmos que analisam o relacionamento entre dados poderia resultar na reidentificação precisa do titular dos dados. A recente Lei Geral de Proteção de Dados (LGPD) brasileira considera dado anonimizado o dado relativo a uma pessoa que não possa ser identificada, considerando a utilização de meios técnicos razoáveis e disponíveis na época do tratamento dos dados. E, além disso, assegura aos titulares o direito à anonimização dos seus dados pessoais. Entretanto, o fundamento para o exercício desse direito é compreender as medidas técnicas utilizadas para inviabilizar a reidentificação desta pessoa, que impõe uma análise interdisciplinar do tema. Portanto, o projeto de pesquisa pretende estabelecer um diálogo entre direito e tecnologia para responder se a dicotomia entre dados pessoais e anônimos se sustenta no atual estado da arte, e, com isso, rediscutir os modelos regulatórios, sugerindo soluções de *lege ferenda* e tecnológica, bem como oferecer sugestões sobre boas práticas aos controladores e operadores de tratamento de dados pessoais, que garantam a efetiva proteção dos dados pessoais, tida como fundamental para o pleno desenvolvimento de uma pessoa.
Palavras-chave: anonimização, pseudoanonimização, Lei n. 13.709/2018 (Lei Geral de Proteção de Dados Pessoais - LGPD), Medida Provisória n. 869/2018 (Autoridade Nacional de Proteção de Dados – ANPD), Código de Boas Práticas.

Abstract:
Regulatory models of data protection are structured on two pillars: informed consent, which implies the conscious and voluntary acceptance from the individual whose data is being collected and manipulated, and anonymization, meaning the employment of techniques that prevent the identification of the individual whom the data is related to. However, it is suggested that the use of algorithms to analyze data relationships could result in the precise identification of the data subject. The new Brazilian General Law on Personal Data Protection (LGPD) states that anonymous data the information that are not attributed to an identified or identifiable person, considering technical and organizational measures existing on the time of the processing of personal data. And beyond that, the LGPD guarantees to the data subject the right to data anonymization. However, the core of such right is to comprehend such technical and organizational measures in a way not to re-identify the data subject, which imposes a multidisciplinary analysis. This project aims at establishing a dialogue between law and technology in order to verify if the personal / anonymous data dichotomy can still be held in light of the current state of art and, thus, to evaluate regulatory models, advancing legislative proposals and new technologies, as well as to offer suggestion on good practices for processors and controllers of data processing, that ensure an effective protection of personal data, understood as a requirement for the full development of a human being.
Keywords: anonymization, pseudonymisation, Brazilian Act n. 13.709/2018 (General Law on Personal Data Protection - LGPD), Brazilian Presidencial Act n. 869/2018 (Personal Data Protection Authority – ANPD), Good Practices Code.

Áreas do conhecimento: Direito Civil (Direitos Especiais – Direito e Internet)

Sumário

1 Introdução:

Atualmente, a economia informacional destaca-se porque a própria informação é o produto e a prestação dos serviços. A informação é um valor em si mesma e não como um meio para criar bens e prestar serviços (CASTELLS 2000, p. 77). Curioso notar que na gênese da *World Wide Web,* a "gratuidade" dos *softwares* disponibilizados foi o que impulsionou a disseminação da rede em escala global[1]. Entretanto, esta aparente gratuidade deve-se ao fato de que as informações coletadas são valiosíssimas, impulsionando o fenômeno da monetização dos dados.

Portanto, o advento da Internet e, fundamentalmente, dos serviços via *web* prestados, as relações sociais, comerciais, judiciais, financeiras e mesmo as governamentais, passando pelas relações de pesquisa científica e lazer, adquiriram vínculos de dependências tecnológicas sem precedentes. Da mesma maneira com que estas relações cresceram em compasso impetuoso, multiplicaram-se os repositórios de dados que hoje contém basicamente todos os dados envolvidos neste vasto número de transações. Nestes repositórios, hoje, estão documentos, comprovantes, fotos, relatórios, entre outros, muitos dos quais os indivíduos ignoram ou esqueceram. Para muitos, este fenômeno marca o início da *Web 2.0* e para outros o início da era "*homo digitas*" (OLSEN 2005).

Esta grande quantidade e variabilidade de dados permitiu a instalação de métodos de mineração de dados que relacionam dados comerciais e governamentais, além de dados de perfis sócio-econômico-culturais. Estes relacionamentos de dados alimentam desde mecanismos de prevenção à evasão fiscal até sugestões de produtos a preços e condições competitivas que podem ser adquiridos pelo comércio via *web*. É inegável que exemplos como estes trazem benefícios à sociedade, tal como trazem malefícios, por exemplo, o uso do CPF

[1] Disponível em: <https://tenyears-www.web.cern.ch/tenyears-www/>, acessado em 12 de outubro de 2015.

que se tornou uma banalidade no nosso sistema, um identificador único usado amplamente e sem os devidos cuidados quanto à proteção dos dados do titular, entre outros; e, também, como dados processuais disponibilizados via *web* por meio de diversas varas de muitos tribunais de justiça, federais e estaduais. Por outro lado, toda esta tecnologia de processamento e mineração de dados pode ser vista como uma grande ameaça à privacidade e à proteção de dados pessoais, como citado por Jändel e colaboradores (JÄNDEL 2014).

Portanto, surge a necessidade de uma lei geral para a proteção dos dados pessoais a fim de reequilibrar os interesses econômicos e da pessoa humana, considerado esse direito como um direito fundamental. No Brasil, essa lei foi recentemente aprovada, Lei n. 13.709, de 14 de agosto de 2018, sendo complementada pela Medida Provisória n. 869, de 27 de dezembro de 2018, esta última criou a Autoridade Nacional de Proteção de Dados, tendo sido aprovada no Senado Federal, sua conversão em lei, estando pendente a sanção presidencial, que deve ocorrer em julho de 2019.

A fonte inspiradora da lei brasileira foi a Dir. 95/46/CE da União Europeia, que disciplinava a proteção de dados, e estabelecia (no art. 3º) que a normativa se aplica aos dados pessoais, considerados aqueles que digam respeito a uma pessoa identificada ou identificável. Assim, a Diretiva 95/46/CE afastava de seu âmbito de aplicação os dados anônimos, vez que o art. 2º alínea "a", desta normativa comunitária, conceituava "dados pessoais" como informações que possam identificar, direta ou indiretamente uma pessoa, nomeadamente por referência a um número de identificação ou a um ou mais elementos específicos da sua identidade física, fisiológica, psíquica, econômica, cultural ou social. O atual Regulamento Geral Europeu (2016/679, de 27 de abril de 2016 – *General Data Protection Regulation - GDPR*) substituiu a antiga Dir. 95/46/CE, trazendo um modelo regulatório mais detalhado e completo à luz dos avanços tecnológicos, e, portanto, também exerceu importante influência na LGPD brasileira.

O GDPR manteve, no art. 4º, "1", o mesmo conceito de dado pessoal estabelecido anteriormente pela diretiva; mas, trouxe o conceito de "*pseudonymisation"*, no item "4" desse mesmo artigo, entendido como o tratamento de dados pessoais de forma que deixem de poder ser atribuídos a um titular de dados específico sem recorrer a informações suplementares, desde que essas informações suplementares sejam mantidas separadamente e sujeitas a medidas técnicas e organizativas para assegurar que os dados pessoais não possam ser atribuídos a uma pessoa singular identificada ou identificável.

Observe-se que os atuais padrões regulatórios sobre proteção de dados estão fundados na dicotomia entre dados pessoais e dados anônimos. No entanto, segundo Barrocas e Nissenbaum (BARROCAS 2014) apesar de todos estes avanços tecnológicos, a proteção a estes dados ainda se concentra em dois frágeis pilares, que são o consentimento livre e esclarecido e a anonimização. O primeiro pilar, o consentimento, enseja a conotação que a formação da coleção de dados, sua manipulação e seu eventual processamento é uma escolha individual; enquanto que o segundo pilar, o da anonimização, pretende atingir esta proteção na medida em que implementa mecanismos que possam desconectar a informação de interesse dos dados que realmente identificam as pessoas.

Nesse sentido, Paul Ohm (2010, p. 1.705) alertava para o mito da anonimização, pois as técnicas de reidentificação criam novos riscos à proteção dos dados pessoais e à privacidade,

e agravam os já existentes. Isso porque tais práticas combinam dados mantidos em banco de dados distintos, resultando na reidentificação bem-sucedida do titular dos dados. Exemplo, quando o usuário da Netflix avalia um filme ou uma série, informação aparentemente não prejudicial, essa informação é associada a outras mantidas em outras bases de dados, criando um perfil sociocultural bem preciso sobre o titular.

Em linhas gerais, existem duas estratégias para a anonimização. A primeira é baseada na aleatorização e a segunda parte da ideia da generalização (*European Commission Justice*, Fung 2007). A aleatorização altera atributos com o objetivo de remover as ligações entre os dados e o indivíduo. A aleatorização é obtida por meio (i) da adição de ruído e (ii) da permutação dos dados entre si, ambos os métodos modificando de maneira irreversível o perfil da distribuição dos dados. A generalização altera os atributos modificando a escala ou a ordem de magnitude dos dados. A agregação ou *k*-anonimidade (Jiuyong 2011) junta uma pessoa com *k* outros indivíduos, de forma que o grupo resultante de indivíduos compartilham o mesmo valor de atributos. Por exemplo, atributos como salários ou dose de um medicamento podem ser generalizados para intervalos de valores (0-10, 10-20, etc.). A *l*-diversidade (Sun 2012) é uma técnica de generalização que procura garantir que cada classe resultante tenha ao menos l diferentes valores. A pseudoanonimização é uma técnica mais simples em que o valor de um atributo é substituído por outro. Por exemplo, o nome de uma pessoa é trocado por um identificador numérico.

Assim, a atual legislação europeia fala em "pseudoanonimização" e não mais em "anonimização". A LGPD brasileira prefere esse último termo, porém para que seu conceito possa se concretizar deve-se atentar às medidas técnicas e organizacionais utilizadas pelos controladores e operadores do tratamento de dados pessoais que efetivamente garantam a desvinculação da pessoa a estas informações. Consequentemente, o tema não pode se afastar de uma análise interdisciplinar. Por isso, esse projeto será desenvolvido em conjunto com o Professor Evandro Eduardo Seron Ruiz, do Departamento da Computação e Matemática da Faculdade de Filosofia, Ciências e Letras de Ribeirão Preto – FFCLRP/USP, que submeteu um projeto ao Edital Ano Sabático 2020, intitulado "*Aspectos computacionais do tratamento de dados pessoais no âmbito da Lei Geral de Proteção de Dados Pessoais*", conforme será detalhado quando se tratar do potencial de interdisciplinariedade.

Analisados e delimitados os conceitos de anonimização e de pseudoanomização sob um estudo interdisciplinar, ao final, pretende-se oferecer subsídios à Autoridade Nacional de Proteção de Dados Pessoais para que se possa definir padrões técnicos e organizacionais para que se garanta segurança na dissociação de determinadas informações ao titular dos dados pessoais. Ademais, a LGPD sugere, no art. 50, que os controladores e operadores do tratamento de dados pessoais formulem regras de boas práticas e de governança que estabeleçam as condições de organização, o regime de funcionamento, os procedimentos, normas de segurança e padrões técnicos. Assim, pretende-se oferecer às empresas, às organizações públicas e privadas, bem como à Autoridade Nacional de Proteção de Dados, sugestões para serem implementadas a fim que se garanta a efetiva anonimização dos dados pessoais, representando uma diminuição de riscos aos quais as empresas que trabalham com dados pessoais estão sujeitas.

2 Objetivos:

Os objetivos do presente projeto de pesquisa dividem-se em: objetivo geral e objetivos específicos, abaixo descritos.

2.1 Objetivo geral:

Investigar se a conceituação mutualmente excludente entre dados anônimos e dados pessoais está (des)ajustada frente ao estado da arte da tecnologia, tendo em vista a prática de agregação de dados e algoritmos de reidentificação, colocando em xeque as técnicas de anonimização de dados e, por conseguinte, a irreversibilidade de tal processo.

Trata-se, pois, de se constatar se essa categorização ainda se sustenta ou se é uma ficção jurídica. Com esse diagnóstico, clarear-se-á qual é a melhor estratégia regulatória a ser adotada para se garantir uma efetiva proteção dos dados pessoais.

2.2 Objetivos específicos:

(i) Investigar a literatura sobre a (in)eficiência das técnicas de anonimização dos dados pessoais, de modo interdisciplinar associando às pesquisas técnicas para trazer novas evidências sobre tal questão, a fim de se fomentar, sobretudo, o seu desenvolvimento em âmbito nacional.

(ii) Na medida em a atual LGPD brasileira trouxe a anonimização como um direito do titular dos dados pessoais, pretende-se qualificar o debate legislativo e acadêmico com um estudo que traga maiores evidências sobre o tema, seja do ponto de vista teórico, seja da perspectiva de evidências empíricas.

(iii) Empreender uma pesquisa interdisciplinar em que o estado atual da arte de tecnologia irá orientar a formulação das estratégias regulatórias para a proteção dos dados pessoais, indagando-se, especificamente, a respeito do alargamento ou restrição do escopo da lei com a inserção/exclusão dos dados anônimos. Trata-se, sobretudo, de um estudo informativo para as políticas públicas de regulação da Internet, mais especificamente em torno da temática de proteção de dados pessoais.

3 Justificativa (escopo acadêmico e científico):

A LGPD brasileira foi recentemente aprovada, trazendo muitos desafios às organizações empresariais, que trabalham com dados pessoais, bem como ao Governo e aos operadores do Direito. Destaca-se a Autoridade Nacional de Proteção de Dados Pessoais, criada pela Medida Provisória n. 869, de 27 de dezembro de 2018, cuja conversão em lei foi aprovada pelo Senado Federal em maio de 2019, pendente de sanção presidencial, que ocorrerá em julho deste mesmo ano, cuja competência regulatória impõe a normatização de técnicas de anonimização para a efetiva proteção dos dados pessoais. Isso porque um dos pontos

fundamentais trazido pela lei foi a anonimização, cujo conceito estabelecido no art. 5º, inc. III, correlaciona-o às medidas técnicas e organizacionais atualmente existentes que inviabilizam a reidentificação do indivíduo, trazendo o premente questionamento: que medidas seriam estas?

O problema de pesquisa a ser enfrentado neste projeto impõe um diálogo constante entre o Direito e a Tecnologia, pois o direito parte da dicotomia entre dados pessoais e dados anônimos para a proteção dos primeiros. No entanto, a tecnologia pode demonstrar se tal dicotomia se sustenta tendo em vista os avanços tecnológicos e as ferramentas de reidentificação, auxiliando, portanto na definição de padrões técnicos e organizacionais.

Parte-se da seguinte hipótese de pesquisa:

a) se ainda se sustenta a categorização polarizada entre dados pessoais e dados anônimos e a consequente dicotomia, isto é, como conceitos mutuamente excludentes;

b) se dados anonimizados são dados pessoais em razão do seu potencial de identificar um sujeito, e, por isso, estariam ou não dentro do escopo de qualquer lei de proteção de dados pessoais; sem olvidar, no entanto, como dever ser encorajado esse padrão de segurança à proteção aos dados pessoais dos usuários.

Trata-se, portanto, de tema relevante e pertinente em vista das discussões travadas no debate público do Anteprojeto de Lei que resultou na atual Lei Geral de Proteção de Dados Pessoais brasileira, as quais demonstraram haver forte demanda sobre essa temática. Em âmbito nacional, há pouco material produzido acerca do tema.

Em linhas gerais, a proteção de dados parte de uma premissa, qual seja, que existem dados referentes a pessoas determinadas ou determináveis e dados que não podem identificar o indivíduo (dados anônimos). Sendo que a lei de proteção de dados restringe seu âmbito de aplicação apenas aos dados pessoais. Portanto, a definição de dados anônimos tem grande importância porque sobre estes não incidirá a regulação legal. No entanto, para que se possa definir se tal dado pode ou não identificar uma pessoa, depende da compreensão da arquitetura do conjunto de dados, ou seja, das estruturas de dados relacionadas, e outros sistemas de informação, justificando a necessidade de uma investigação interdisciplinar.

Por isso, o problema de pesquisa levantado neste projeto é: A dicotomia conceitual legal e tecnológica entre dados pessoais e dados anônimos é elusiva? Para responder essa questão, deve-se responder às seguintes indagações:

(i) Dado o estado atual da arte permeado pela prática de agregação de dados e algoritmos de reidentificação, possíveis, principalmente, por meio de metodologias que utilizam tecnologias de *Big Data*, como se poderia avaliar a eficiência das técnicas de anonimização dos dados pessoais?

(ii) Como essa (in)eficiência deve impactar o arranjo jurídico/regulatório de proteção de dados pessoais?

Com essas respostas, pretende-se elucidar o debate tanto para as empresas que poderão adotar as sugestões oferecidas pelo projeto para o respectivo Código de Boas Práticas e Governança, bem como para a Autoridade Nacional de Proteção de Dados, a quem compete, nos termos do art. 55-J da LGPD, inciso II, “editar normas e procedimentos sobre a proteção de dados pessoais”.

4 Razões para desenvolver o projeto no IEA:

O Instituto de Estudos Avançados da USP, caracterizado pela "ativação de um espaço de reflexão onde se cultivem os estudos avançados conduzidos por mestres de excelência nacional e internacional, no interior da instituição", é o ambiente propício para o desenvolvimento da presente pesquisa por várias razões.

1ª) o IEA é considerado um centro de pesquisa avançado, ou seja, congrega professores e pesquisadores em adiantado nível acadêmico e científico; a Docente completou dez anos de doutoramento em abril de 2019, tendo já defendido a livre-docência em maio de 2016, e tem se dedicado à linha de pesquisa "Direito e Internet", reconhecida como tal no cenário nacional e internacional, tendo sido convidada para participar em debates e audiências públicas quando se discutia a proposta que resultou na atual Lei Geral de Proteção de Dados Pessoais, bem como tem sido convidada como palestrante e professora visitante em renomadas instituições brasileiras e estrangeiras;

2ª) a Docente defendeu, em maio de 2016, a sua tese de livre docência, intitulada "*A imprescindibilidade de um órgão independente para a efetiva proteção dos dados pessoais no cenário futuro do Brasil*". Ribeirão Preto: FRDP/USP, 2015. 487 páginas. Nessa oportunidade, a Docente defendeu um modelo para a ANPD brasileira, apresentando na Câmara dos Deputados os resultados da tese, influenciando a criação do órgão brasileiro pela Medida Provisória n. 869/2018. Assim, o presente projeto está intimamente relacionado ao tema, pois pretende oferecer subsídios para a ANPD adotar normas sobre anonimização em prol da efetiva proteção dos dados pessoais.

2ª) a Docente tem doutores (3), mestres (8), especialização (34), trabalhos de conclusão de curso (23), orientados em IC (13), demonstrando uma capacidade de formar discípulos, sendo que o desenvolvimento da pesquisa no IEA poderá oferecer melhores condições para que a Docente continue multiplicando os conhecimentos adquiridos na atividade de pesquisa;

3ª) a interdisciplinaridade que é fundamental ao tema do presente projeto de pesquisa, justifica o desenvolvimento desta investigação no IEA, que se destaca pelas pesquisas atuais, socialmente relevantes e que concilie diversas ciências, como o tema ora em análise, na medida em que se impõe um estudo jurídico e tecnológico sobre as diversas questões levantadas.

5 Potencial de interdisciplinaridade:

Como ficou demonstrado não se pode compreender o que são dados anônimos sem recorrer à tecnologia para explicar em que consistem os padrões técnicos e organizacionais para evitar a reidentificação do titular dos dados pessoais. Portanto, esse projeto será realizado em conjunto com o Departamento de Matemática e Computação da FFCLRP da USP, em parceria com o Prof. Dr. Evandro Eduardo Seron Ruiz, que também submeteu um projeto para o Edital Ano Sabático 2020. Assim, enquanto o pesquisador da computação dedicar-se-á a elucidar os padrões técnicos, a pesquisadora do direito, que ora submete esse projeto de pesquisa, irá elaborar sugestões para a implementação regulatória da LGPD (Lei n. 13.709/2018), cuja

competência é da ANPD editar normas para a efetiva proteção dos dados pessoais. Além disso, a pesquisadora do Direito, com base nos resultados tecnológicos, irá oferecer uma sugestão de Código de Boas Práticas empresariais para os controladores e operadores do tratamento de dados pessoais.

6 Impactos científicos e sociais:

A luz do que foi dito *supra*, esse projeto de pesquisa poderá ter um forte impacto científico, pois o tema é atual e pouco estudado na doutrina brasileira, mesmo porque a LGPD foi aprovada somente em 14 de agosto de 2018. Assim, os resultados que serão apresentados oferecerão subsídios à ANPD, aos operadores do Direito bem como às empresas que lidam com dados pessoais.

Portanto, o impacto social do presente projeto consiste em preparar a sociedade de um pode em geral para a entrada em vigor da LGPD, que ocorrerá em 2020.

7 Metodologia:

Na primeira etapa da pesquisa, pretende-se coletar os dados bibliográficos para aprofundamento, tais como, livros, capítulos em livros, artigos publicados em periódicos sobre o tema que é muito recente. Em seguida, pretende-se concentrar na leitura e na documentação dos dados bibliográficos levantados consoante os métodos abaixo mencionados, esta fase preliminar serve de base para a construção lógica do trabalho e redação de artigos científicos. Na segunda etapa, será feito um levantamento das atuais tecnologias de anonimização, destacando possíveis formas de reidentificação.

Para o desenvolvimento desta pesquisa serão utilizados os métodos dedutivo e indutivo sob uma perspectiva dialética e o método comparativo. O método dedutivo será utilizado na análise geral sobre o modelo europeu de proteção dos dados pessoais para destacar seus pontos positivos e negativos quanto à dicotomia entre dados pessoais e dados pseudoanônimos, dada a forte influência que a LGPD sofreu pelo GDPR.

O método indutivo será utilizado na elaboração das conclusões desta pesquisa, posto que tais conclusões terão como ponto de partida premissas particulares obtidas por meio de análise da efetividade das técnicas de anonimização para a efetiva proteção dos dados pessoais. E, então, fazer proposta para que a ANPD possa normatizar o tema, bem como sugestões para as empresas adotarem como boas práticas.

A perspectiva dialética mostra-se eficaz à medida que proporciona o confronto dos dados teóricos, obtidos por meio da análise crítica da pesquisa bibliográfica, com os dados práticos, obtidos na realização da pesquisa jurisprudencial e pesquisa técnica na área da ciência da computação com a finalidade de alcançar os objetivos, inicialmente, propostos.

Por fim, o método comparativo irá auxiliar no desenvolvimento da pesquisa na medida em que integrando a pesquisadora na realidade socioeconômica de diversos países,

notadamente europeus, poderá compreender de maneira integral as deficiências e insuficiências do sistema brasileiro, que se inspirou naquele.

8 Plano de trabalho a ser executado pelo pesquisador:

A Docente irá realizar uma pesquisa bibliográfica, por meio de revisão de literatura especializada no tema, notadamente estrangeira, e analisar referenciais teóricos divergentes sobre a eficiência do processo de anonimização dos dados pessoais, priorizando-se estudos empíricos para captar as diversas metodologias utilizadas para tais estudos. Além disso, após a revisão de literatura, serão realizados experimentos, com a colaboração do pesquisador da matemática, para testar a eficiência da de-identificação da base de dados pessoais, verificando-se o grau de eficiência das técnicas de anonimização. Procurar-se-á combinar as metodologias já utilizadas em pesquisas anteriores para se extrair um método mais completo a assegurar a qualidade das amostras obtidas da pesquisa empírica.

9 Cronograma:

O projeto de pesquisa terá duração de 06 (seis) meses, com início previsto para 1º de fevereiro de 2020 a 1º de agosto de 2020, quando o projeto será completado. Durante este período, a pesquisadora irá desempenhar o cronograma que ora se apresenta:

Mês/ano	Atividades a serem realizadas
Fev./2020 a abr./2020	Levantamento e análise de bibliografia de aprofundamento sobre o tema.
Mar./2020 a jun./2020	Reuniões quinzenais do Grupo de Pesquisa a ser desenvolvido no IEA.
Jun./2020	Seminário interdisciplinar sobre “Padrões técnicos e organizacionais para a anonimização e/ou pseudoanonimização: orientações para as boas práticas empresariais”.
Jul./2020	Redação e submissão de artigo científico para publicar os resultados de pesquisa.
Ago./2020	Elaboração de sugestão de Código de Boas Práticas para as empresas e governo quanto à anonimização de dados pessoais.
Fev./2020 a ago./2020	Participação com apresentação de trabalho em eventos científicos nacionais e/ou internacionais.

10 Elaboração de trabalhos científicos (*papers*, livros e outros):

Durante o 1º semestre de 2020 pretende-se submeter um *paper* sobre o tema para publicar os resultados desse projeto de pesquisa em periódico bem avaliado pelo sistema *Qualis* CAPES, especialmente A1, A2 ou B1. Além disso, durante o mesmo período, buscar-se-á

participar de eventos científicos nacionais e/ou internacionais para apresentar os resultados da pesquisa.

11 Previsão de organização de seminários, simpósios ou atividades assemelhadas:

A Docente é líder de dois grupos de pesquisa cadastrados e avalizados pela USP no Diretório CNPq de grupos de pesquisa, a saber: (a) **Observatório do Marco Civil da Internet no Brasil** (http://dgp.cnpq.br/dgp/espelhogrupo/2215582162179038), ativo desde 2014; e (b) **Tutela jurídica dos dados pessoais na internet** (http://plsql1.cnpq.br/buscaoperacional/detalhegrupo.jsp?grupo=0067601HS29JDN), ativo desde 2012. Assim, pretende-se conciliar as atividades destes dois grupos de pesquisa para solicitar a criação de um grupo de pesquisa no IEA, do qual o Professor Evandro Eduardo Seron Ruiz (Depto. Comutação e Matemática da FFCLRP) e a Professora Cíntia Rosa Pereira de Lima (Depto de Direito Privado e Processo Civil da FDRP) serão os coordenadores.

No final do 1º semestre de 2020, pretende-se realizar um seminário interdisciplinar sobre "Padrões técnicos e organizacionais para a anonimização e/ou pseudoanonimização: orientações para as boas práticas empresariais". Pretende-se abordar neste evento, bem como nas publicações os temas relevantes para o desenvolvimento deste projeto de pesquisa, a fim de gerar conhecimento e inovação afinada com os grandes problemas na efetiva proteção dos dados pessoais.

12 Referências bibliográficas:

BAROCAS, Solon; NISSENBAUM, Helen. **Big data's end run around procedural privacy protections**. *Communications of the ACM* 57.11 (2014): 31-33.

BEDAGKAR-GALA, A.; SHAH, S. K. (2014). **A survey of approaches and trends in person re-identification**. *Image and Vision Computing*, *32*(4), 270-286.

BELL, Daniel. **The coming of post-industrial society**: a venture in social forecast. Basic Books, 1973.

BIANCHI, G.; TEOFILI, S.; POMPOSINI, M. **New Directions in Privacy-preserving Anomaly Detection for Network Traffic.** *Proceedings of the 1st ACM Workshop on Network Data Anonymization, ACM,* 2008, 11-1.

BURKHART, M.; BRAUCKHOFF, D.; MAY, M.; BOSCHI, E. **The Risk-utility Tradeoff for IP Address Truncation** *Proceedings of the 1st ACM Workshop on Network Data Anonymization, ACM,* 2008, 23-30.

CALLANAN, C.; JERMAN-BLAZIC, B. **User Understanding of Privacy in Emerging Mobile Markets** *Technology and Society Magazine, IEEE,* 2014*, 33*, 48-56.

CASTELLS, Manuel. **End of millenium**. 2. ed. The information age: economy, society and culture. vol. 3. Massachusetts: Blackwell, 1998.

_____. **The power of identity**. 2. ed. The information age: economy, society and culture. vol. 2. Massachusetts: Blackwell, 2004.

_____. **The rise of the network society**. 2. ed. The information age: economy, society and culture. vol. 1. Massachusetts: Blackwell, 2000.

The rise of the network society. 2 ed. Vol. I. Oxford: Blackwell, 2000.

CAVOUKIAN, A.; CASTRO, D. **Big Data and Innovation, Setting the Record Straight: De-identification Does Work**, 2014.

CAVOUKIAN, Ann. **Privacy by design: The 7 foundational principles**. *Information and Privacy Commissioner of Ontario, Canada* (2009).

DALENIUS, T. **Towards a methodology for statistical disclosure control.** *Statistik Tidskrift,* 1977*, 15*, 429-444.
DANCIU, I.; COWAN, J. D.; BASFORD, M.; WANG, X.; SAIP, A.; OSGOOD, S.; HARRIS, P. A. (2014). **Secondary use of clinical data: the Vanderbilt approach**. *Journal of biomedical informatics*, *52*, 28-35.
DATTA, A.; SHARMA, D.; SINHA, A. **Provable de-anonymization of large datasets with sparse dimensions**. *Principles of Security and Trust, Springer,* 2012, 229-248.
DONEDA, Danilo. **Da privacidade à proteção de dados pessoais**. Rio de Janeiro: Renovar, 2006.
DUVALL, S. L.; KERBER, R. A.; THOMAS, A. (2010). **Extending the Fellegi–Sunter probabilistic record linkage method for approximate field comparators**. *Journal of biomedical informatics*, *43*(1), 24-30.
DWORK, C. **A Firm Foundation for Private Data Analysis** *Commun. ACM, ACM,* 2011*, 54*, 86-95.
EDWARDS, Lilian; WAELDE, Charlotte. **Law and the Internet**. 3. Ed. Oregon, Oxford and Portland: Hart, 2009.
EUROPEAN COMMISSION JUSTICE, **Opinions and Recommendation - Opinion 05/2014 on Anonymization Techniques**, disponível em <https://www.pdpjournals.com/docs/88197.pdf>, acessado em 27 de junho de 2019.
FINOCCHIARO, Giusella. **Privacy e protezione dei dati personali:** disciplina e strumenti operativi. Bologna: Zanichelli, 2012.
FUNG, B.C.M.; KE WANG; YU, P.S., **Anonymizing Classification Data for Privacy Preservation**, IEEE Transactions on Knowledge and a Data Engineering, v.19, no. 5., p. 711-725, 2007.
GATTANI, S.; DANIELS, T. E. **Reference Models for Network Data Anonymization**. Disponível em: <http://ec.europa.eu/justice/data-protection/article-29/documentation/opinion-recommendation/files/2014/wp216_en.pdf>, acesso em 27 de novembro de 2015.
GEIST, Michael. Law, **Privacy and Surveillance in Canada in the Post-Snowden Era**. Ottawa: University of Ottawa Press, 2015.
HUGHES, R.L. David. **Two concepts of privacy**, *Computer Law & Security Review*, v. 31, no. 4, p. 527-537, 2015.
JANDEL, M. **Decision support for releasing anonymized data**. *Computers & Security,* 2014*, 46*, 48 – 61.
JÄNDEL, Magnus. **Decision support for releasing anonymized data**. Computers & Security 46 (2014): 48-61.
JIUYONG, Li; JIXUE, Liu; MUZAMMIL, Baig; WONG, Raymond Chi-Wing, **Information based data anonymization for classification utility**, *Data & Knowledge Engineering*, v.70, no. 12, p. 1030-1045, 2011.
KABUDULA, C. W.; CLARK, B. D.; GÓMEZ-OLIVÉ, F. X.; TOLLMAN, S.; MENKEN, J.; RENIERS, G. (2014). **The promise of record linkage for assessing the uptake of health services in resource constrained settings: a pilot study from South Africa**. *BMC medical research methodology*, *14*(1), 71.
KELLY, D. J.; RAINES, R. A.; GRIMAILA, M. R.; Baldwin, R. O.; MULLINS, B. E. **A Survey of State-of-the-art in Anonymity Metrics**. *Proceedings of the 1st ACM Workshop on Network Data Anonymization, ACM,* 2008, 31-40.
KRUMM, J. **A Survey of Computational Location Privacy**. *Personal Ubiquitous Comput., Springer-Verlag,* 2009*, 13*, 391-399
LANE, N. D.; XIE, J.; MOSCIBRODA, T.; ZHAO, F. **On the Feasibility of User De-anonymization from Shared Mobile Sensor Data**. *Proceedings of the Third International Workshop on Sensing Applications on Mobile Phones, ACM,* 2012, 3:1-3:5.
LI, H.; MURALIDHAR, K.; SARATHY, R.; LUO, X. R. (2014). **Evaluating Re-Identification Risks of Data Protected by Additive Data Perturbation**. *Journal of Database Management (JDM)*, *25*(2), 52-74.
LI, J.; LIU, J.; Baig, M.; WONG, R. C.-W. **Information based data anonymization for classification utility**. *Data & Knowledge Engineering ,* 2011*, 70*, 1030 – 1045.
LIMA, Cíntia Rosa Pereira de; BIONI, Bruno Ricardo. **A proteção dos dados pessoais na fase de coleta: apontamentos sobre a adjetivação do consentimento implementada pelo artigo 7, incisos VIII e IX do Marco Civil da Internet a partir da Human Computer Interaction e da Privacy by Default**. *In*: DE LUCCA, Newton; SIMÃO FILHO, Adalberto; LIMA, Cíntia Rosa Pereira de. Direito & Internet III: Marco Civil da Internet (Lei n. 12.965/2014). Tomos I e II. São Paulo: Quartier Latin, 2014. pp. 263 – 290.
_____ (coord.). **Comentários à Lei Geral de Proteção de Dados Pessoais (Lei n. 13.709/2018).** São Paulo: Almedina, 2019. (no prelo)

_____. **A imprescindibilidade de um órgão independente para a efetiva proteção dos dados pessoais no cenário futuro do Brasil**. Tese de Livre Docência. Ribeirão Preto: FRDP/USP, 2015. 487 páginas.
_____; DE LUCCA, Newton; SIMÃO Filho, Adalberto; (coords.). **Direito & Internet: Marco Civil da Internet, Lei n. 12.965/2014.** 2 volumes. São Paulo: Quartier Latin, 2015.
_____; DE LUCCA, Newton; SIMÃO Filho, Adalberto; DEZEM, Renata Mota Maciel Madeira (coords.). **Direito & Internet: sistema de proteção de dados pessoais, Lei n. 13.709/2018**. 2 volumes. São Paulo: Quartier Latin, 2019. (no prelo)
MARCO-RUIZ, L., MONER, D., MALDONADO, J. A., KOLSTRUP, N., & BELLIKA, J. G. (2015). **Archetype-based data warehouse environment to enable the reuse of electronic health record data.** *International journal of medical informatics*, *84*(9), 702-714.
MELO, Marco Aurélio Vilaça de. **Aspectos Técnicos e Legais da Coleta e Anonimização de Tráfego de Redes IP**. *Dissertação (mestrado), Universidade Federal de Minas Gerais.* 2009.
MURRAY, Andrew. **Information Technology Law**: the law and society. Oxford: Oxford University Press, 2010.
MUSAFIR, V. E. N.; de FREITAS, C. S. (2015, June). **Brazilian e-Government Strategies**. In *Proceedings of the 15th European Conference on eGovernment 2015: ECEG 2015* (p. 187). Academic Conferences Limited.
NARAYANAN, A.; SHMATIKOV, V. **Robust De-anonymization of Large Sparse Datasets**. *Proceedings of the 2008 IEEE Symposium on Security and Privacy, IEEE Computer Society,* 2008, 111-125.
OECD. **Guidelines governing the protection of privacy and transborder flows of operational data** – Revisited Guidelines, 2013. Disponível em: < http://www.oecd.org/sti/ieconomy/oecd_privacy_framework.pdf>, último acesso em 26 de novembro de 2015.
OHM, P. Pottle, D. *(Ed.)* **Broken Promises of Privacy: Responding to the Surprising Failure of Anonymization.** *UCLA Law Review, Law School at UCLA,* 2010*, 57*, 1701-177.
RANDALL, S.; BOYD, J.; FERRANTE, A.; BAUER, J.; Semmens, J. 2014. **Use of graph theory measures to identify errors in record linkage**. *Computer Methods and Programs in Biomedicine*. 115 (2): pp. 55-63.
ROSSETTI, José Paschoal. *Introdução à Economia.* 19 ed. São Paulo: Atlas, 2002.
SAFRAN, C.; BLOOMROSEN, M.; HAMMOND, W. E.; LABKOFF, S.; MARKEL-FOX, S.; TANG, P. C.; DETMER, D. E. (2007). **Toward a national framework for the secondary use of health data: an American Medical Informatics Association White Paper.** *Journal of the American Medical Informatics Association*, *14*(1), 1-9
SCHWARTZ, P. M.. (2004). **Property, Privacy, and Personal Data**. *Harvard Law Review*, *117*(7), 2056–2128. http://doi.org/10.2307/4093335
SCOTT, S. V.; ORLIKOWSKI, W. J. (2014). **Entanglements in practice: performing anonymity through social media**. Disponível em: < http://eprints.lse.ac.uk/57603/1/Entanglements%20in%20Practice.pdf>, acessado em 20 de maio de 2019.
SLAGELL, A.; YURCIK, W. **Sharing computer network logs for security and privacy: a motivation for new methodologies of anonymization.** *Security and Privacy for Emerging Areas in Communication Networks, 2005. Workshop of the 1st International Conference on, Security and Privacy for Emerging Areas in Communication Networks, 2005. Workshop of the 1st International Conference on,* 2005, 82-91.
SUN, X.; SUN, L.; WANG, H. **Extended K-anonymity Models Against Sensitive Attribute Disclosure.** *Comput. Commun., Elsevier Science Publishers B. V.,* 2011*, 34*, 526-535.
_____; _____; ZHANG, Y. **Injecting purpose and trust into data anonymization.** *Computers & Security ,* 2011*, 30*, 332 – 345.
_____; _____; _____; WANG, H.; LI, J.. **Satisfying Privacy Requirements Before Data Anonymization.** *The Computer Journal,* 2012*, 55*, 422-437.
SUN, Xiaoxun; WANG, Hua; LI, Jiuyong; ZHANG, Yanchun, **Satisfying Privacy Requirements Before Data Anonymization**, *The Computer Journal*, v. 55, no. 4., p. 422-437, 2012.
ZITTRAIN, Jonathan. **The Future of the Internet and How to Stop It**. Yale University Press: 2008.