FÓRUM
GESTÃO DE DADOS
DE INVESTIGAÇÃO

Grupo de Trabalho
FORMAÇÃO E COMPETÊNCIAS PARA GESTÃO E DADOS FAIR

Café com Dados

Q&A sobre Proteção de Dados, Segurança e Dados Sensíveis

Utilização de dados sensíveis para investigação

Rita Rb-Silva
MD, PhD

Instituto Português de Oncologia do Porto FG, EPE (IPO-Porto)
Magalhães & Taveira-Gomes (MTG)

2021-11-26

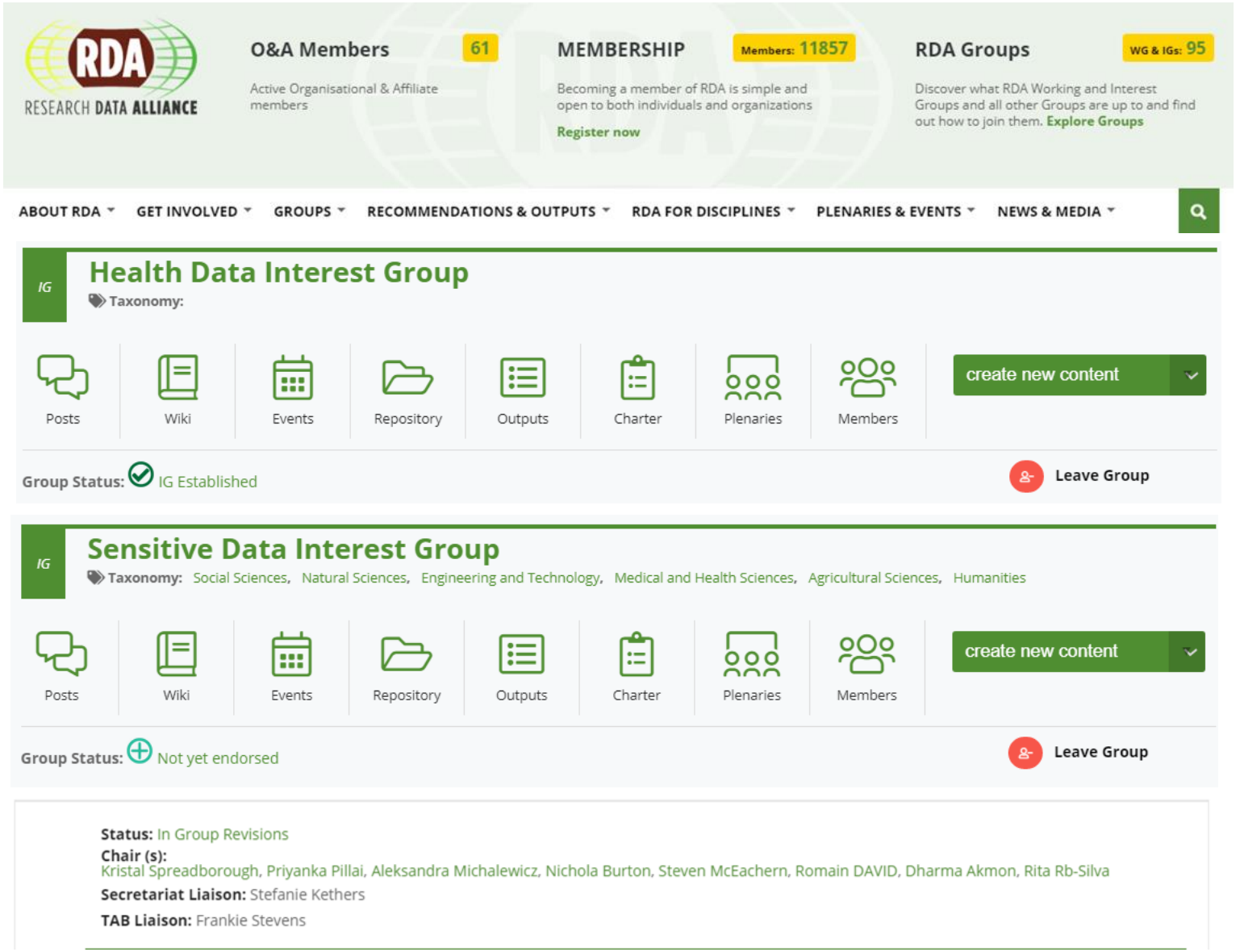
Porto, Portugal

Notas:

O conteúdo desta apresentação tem carácter informativo e reflete a perspetiva individual da autora e não a visão ou valores institucionais das organizações nas quais exerce a sua actividade profissional.

A autora não garante que a informação contida nesta apresentação seja completa nem o mais actualizada possível, pelo que outras referências deverão ser consultadas para o desenvolvimento de projetos de investigação envolvendo dados sensíveis.

A autora declara a ausência de conflitos de interesse.


RESEARCH DATA ALLIANCE
O&A Members
61
Active Organisational & Affiliate members
MEMBERSHIP
Members: 11857
Becoming a member of RDA is simple and open to both individuals and organizations
Register now
RDA Groups
WG & IGs: 95
Discover what RDA Working and Interest Groups and all other Groups are up to and find out how to join them. Explore Groups
ABOUT RDA
GET INVOLVED
GROUPS
RECOMMENDATIONS & OUTPUTS
RDA FOR DISCIPLINES
PLENARIES & EVENTS
NEWS & MEDIA
IG
Health Data Interest Group
Taxonomy:
Posts
Wiki
Events
Repository
Outputs
Charter
Plenaries
Members
create new content
Group Status: IG Established
Leave Group
IG
Sensitive Data Interest Group
Taxonomy: Social Sciences, Natural Sciences, Engineering and Technology, Medical and Health Sciences, Agricultural Sciences, Humanities
Posts
Wiki
Events
Repository
Outputs
Charter
Plenaries
Members
create new content
Group Status: Not yet endorsed
Leave Group
Status: In Group Revisions
Chair (s):
Kristal Spreadborough, Priyanka Pillai, Aleksandra Michalewicz, Nichola Burton, Steven McEachern, Romain DAVID, Dharma Akmon, Rita Rb-Silva
Secretariat Liaison: Stefanie Kethers
TAB Liaison: Frankie Stevens

RDA
RESEARCH DATA ALLIANCE

VIRTUAL PLENARY 18

3-11 November 2021

Wednesday 10 November 2021

Towards a common understanding of sensitive data across regions and disciplines

06 AUG 2021

By Kristal Spreadborough

Group(s) submitting the application: Sensitive Data Interest Group

Meeting objectives:
This will be the third Plenary meeting of the Sensitive Data Interest Group (yet to be formally endorsed). In previous meetings, we have scoped the interest and needs of the RDA community. At RDA18, we begin to focus on specific challenges associated with working with sensitive data. Namely, understanding how sensitive data is defined in different regions and disciplines and looking forward to how we might develop a shared language around sensitive data (e.g. developing an understanding of how different community agreed vocabularies relate to each other).
At this meeting, attendees will:

1. Hear latest discussions about sensitive data classifications and how different levels of sensitivity are defined
2. Contribute examples of how sensitive data classifications and levels are defined in their region/discipline
3. Contribute examples of how differing definitions of sensitive data have presented challenges to their work
4. Co-design the next steps for developing a shared understanding of sensitive data classifications and levels, how these might be harmonised across contexts, and consider what outputs could be developed to address these opportunities and challenges

The meeting objective is to focus the discussion on developing a shared understanding of sensitive data by:

1. Facilitating the discussion on how sensitive data is currently defined
2. Working towards a common language for discussing sensitive data classifications and levels across contexts/disciplines/localities
3. Assessing interest in developing resources and outputs for sensitive data access and management

1- O que são dados sensíveis?

2- Como se classificam estes dados / como se definem níveis de sensibilidade?

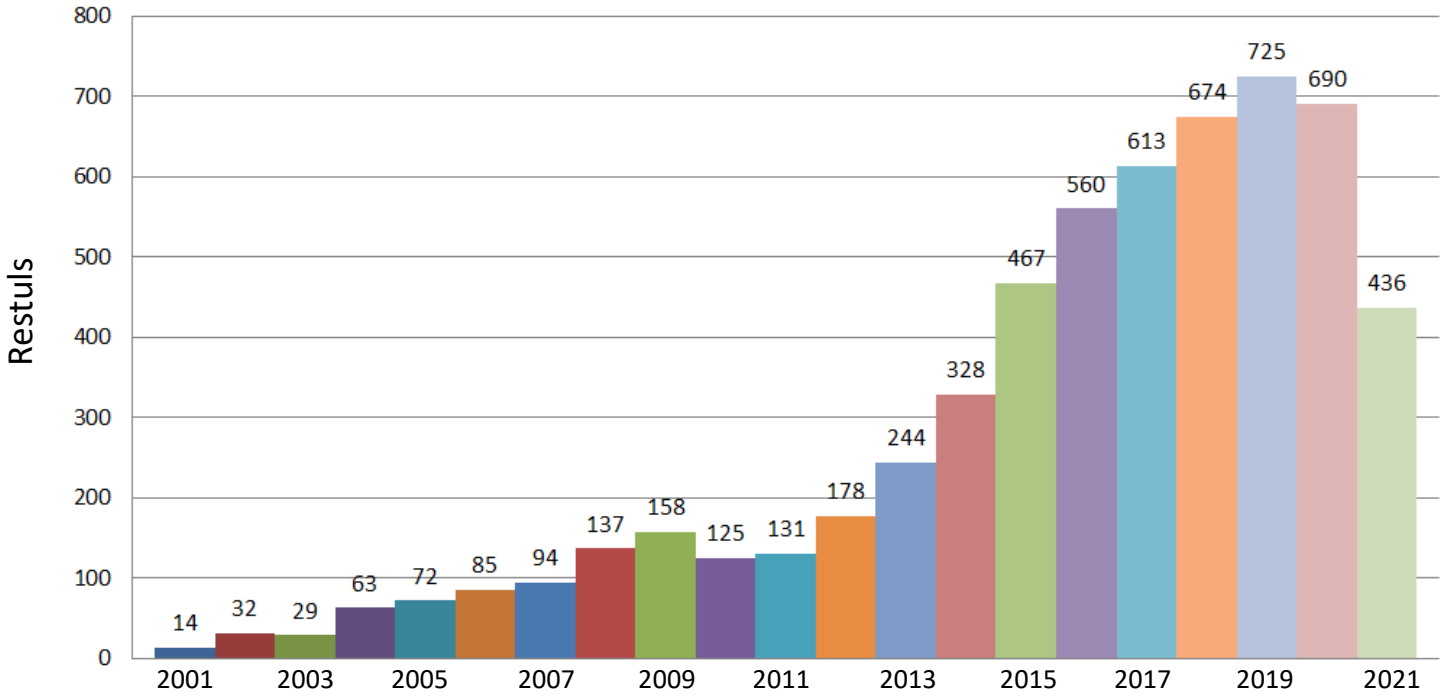
3- Existe variabilidade de definições entre instituições/organizações? Entre diferentes setores e disciplinas?

4- Quais os desafios impostos pelos dados sensíveis?

5- Como encontrar soluções para esses desafios?

Dados sensíveis

- Utilização crescente em diferentes áreas: ***marketing*, indústria, investigação**
- Necessidade de regulamentação para **SEGURANÇA** e **PROTEÇÃO**
- **Interesse crescente no ESTUDO de dados sensíveis**



Search: TOPIC: ("sensitive data") *from Web of Science Core Collection* (2021-11-24)

Dados sensíveis – o que são?

Dados = Informações num formato que facilita o seu processamento

Brahma, o criador hindu.

Informações referentes a um indivíduo, espécie, objecto, processo ou localização, que criam um risco potencial de discriminação, dano ou atenção indesejada.[a]

- Indivíduo(s)
- Espécie(s)
- Objecto(s)
- Processo(s)
- Lcalização(ões)

- Discriminação
- Dano
- Atenção indesejada

Informação regulamentada por lei devido ao possível risco para plantas, animais, indivíduos e/ou comunidades e organizações públicas ou privadas.[b]

- Regulamentação/Lei

- Plantas
- Animais
- Indivíduos
- Comunidades
- Organizações

[a] Australian National Data Service, 2018

[b] David *et al. 2020,* "Templates for FAIRness evaluation criteria - RDA-SHARC IG"

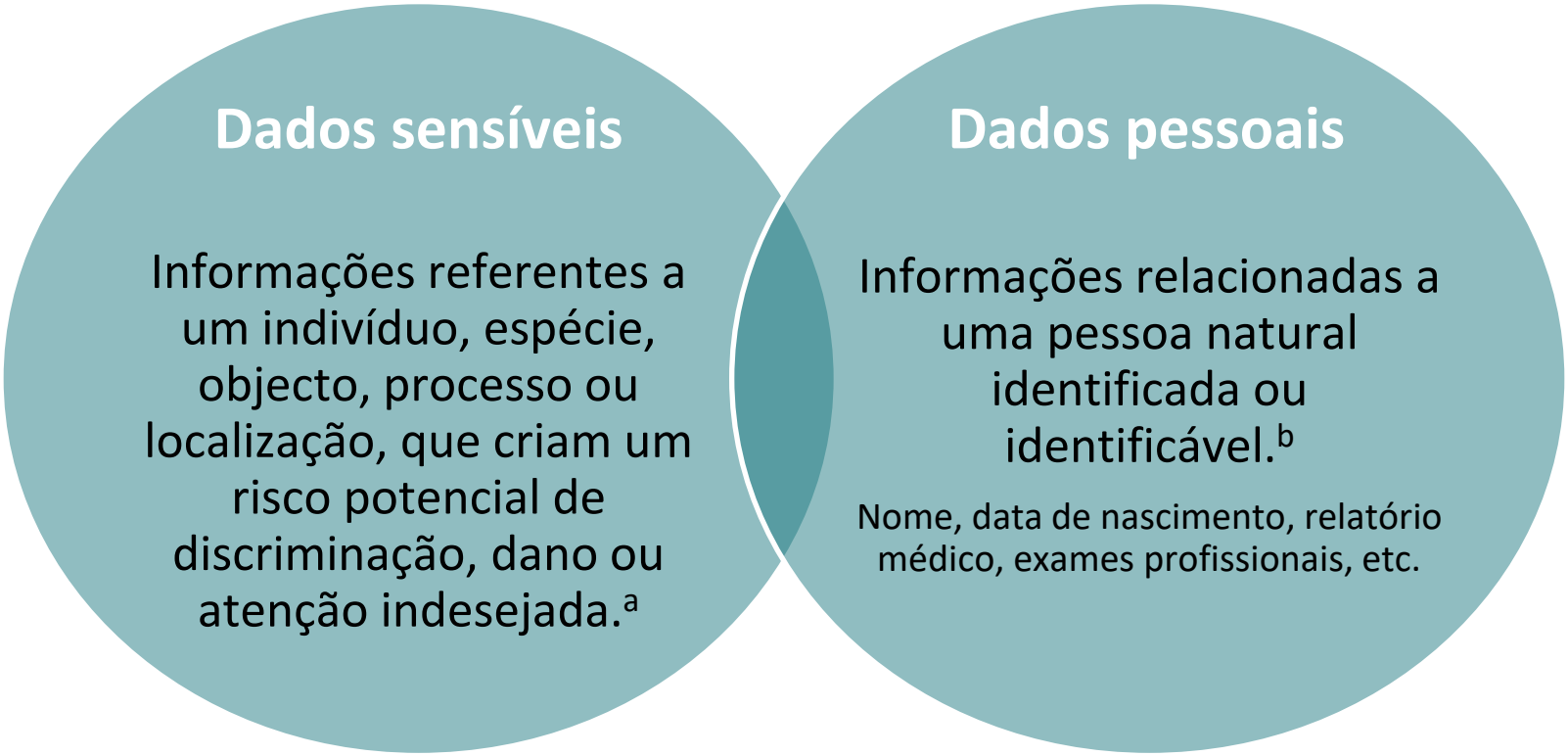
Dados sensíveis

Riscos incluem, por exemplo:

- Discriminação
- Riscos financeiros/económicos
- Interferência com programas/ ferramentas de segurança
- Apropriação ilegítima de dados ou conhecimento
- Terrorismo
- ....

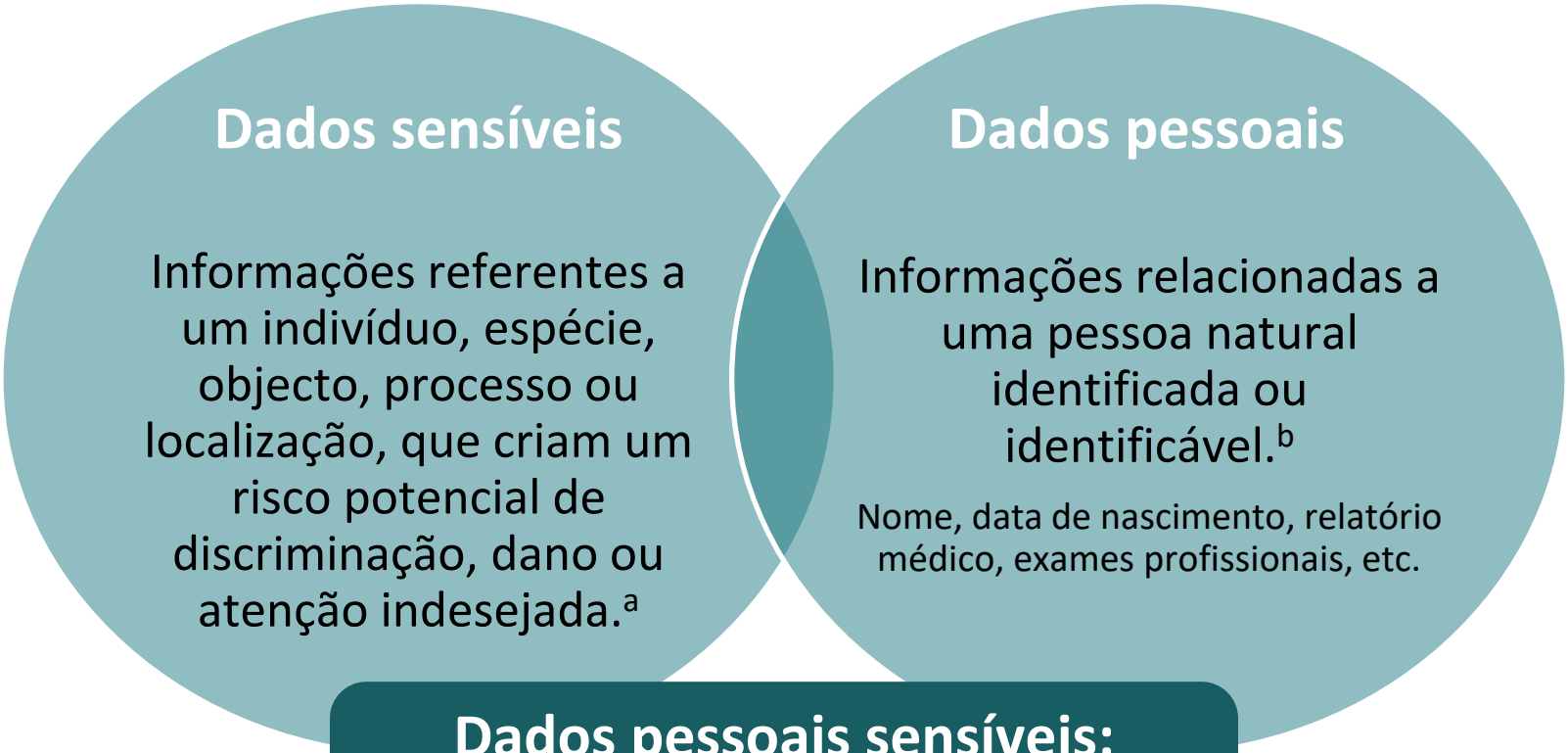
EXEMPLOS:

- Dados pessoais sensíveis
- Dados sobre recursos naturais escassos ou preciosos
- Dados sobre espécies em vias de extinção
- Dados internos de empresas concorrentes
- Dados sobre agentes patogénicos com elevado risco para a saúde humana ou de outras espécies
- ...



[a] Australian National Data Service, 2018
[b] Regulamento Geral de Proteção de Dados.

.



[a] Australian National Data Service, 2018

[b] Regulamento Geral de Proteção de Dados.

.

European Commission

What personal data is considered sensitive?

PAGE CONTENTS

Answer

The following personal data is considered 'sensitive' and is subject to specific processing conditions:

- personal data revealing racial or ethnic origin, political opinions, religious or philosophical beliefs;
- trade-union membership;
- genetic data, biometric data processed solely to identify a human being;
- health-related data;
- data concerning a person's sex life or sexual orientation.

References

- Article 4(13), (14) and (15) and Article 9 and Recitals (51) to (56) of the GDPR

https://ec.europa.eu/info/law/law-topic/data-protection/reform/rules-business-and-organisations/legal-grounds-processing-data/sensitive-data/what-personal-data-considered-sensitive_en

Classificação

Informação:

- proibída

- limitada

- confidencial

- pública

Low Risk

- Research data (at data owner's discretion)
- SUNet IDs
- Information authorized to be available on or through Stanford's website without SUNet ID authentication
- Policy and procedure manuals designated by the owner as public
- Job postings
- University contact information not designated by the individual as "private" in StanfordYou
- Information in the public domain
- Publicly available campus maps

Moderate Risk

- Unpublished research data (at data owner's discretion)
- Student records and admission applications
- Faculty/staff employment applications, personnel files, benefits, salary, birth date, personal contact information
- Non-public Stanford policies and policy manuals
- Non-public contracts
- Stanford internal memos and email, non-public reports, budgets, plans, financial info
- University and employee ID numbers

High Risk

- Health Information, including Protected Health Information (PHI)
- Health Insurance policy ID numbers
- Social Security Numbers
- Credit card numbers
- Financial account numbers
- Export controlled information
- Driver's license numbers
- Passport and visa numbers
- Donor contact information and non-public gift information

https://uit.stanford.edu/guide/riskclassifications

Stanford University IT

Low Risk

Data and systems are classified as Low Risk if they are not considered to be Moderate or High Risk, and:

1. The data is intended for public disclosure, or
2. The loss of confidentiality, integrity, or availability of the data or system would have no adverse impact on our mission, safety, finances, or reputation.

Moderate Risk

Data and systems are classified as Moderate Risk if they are not considered to be High Risk, and:

1. The data is not generally available to the public, or
2. The loss of confidentiality, integrity, or availability of the data or system could have a mildly adverse impact on our mission, safety, finances, or reputation.

High Risk

Data and systems are classified as High Risk if:

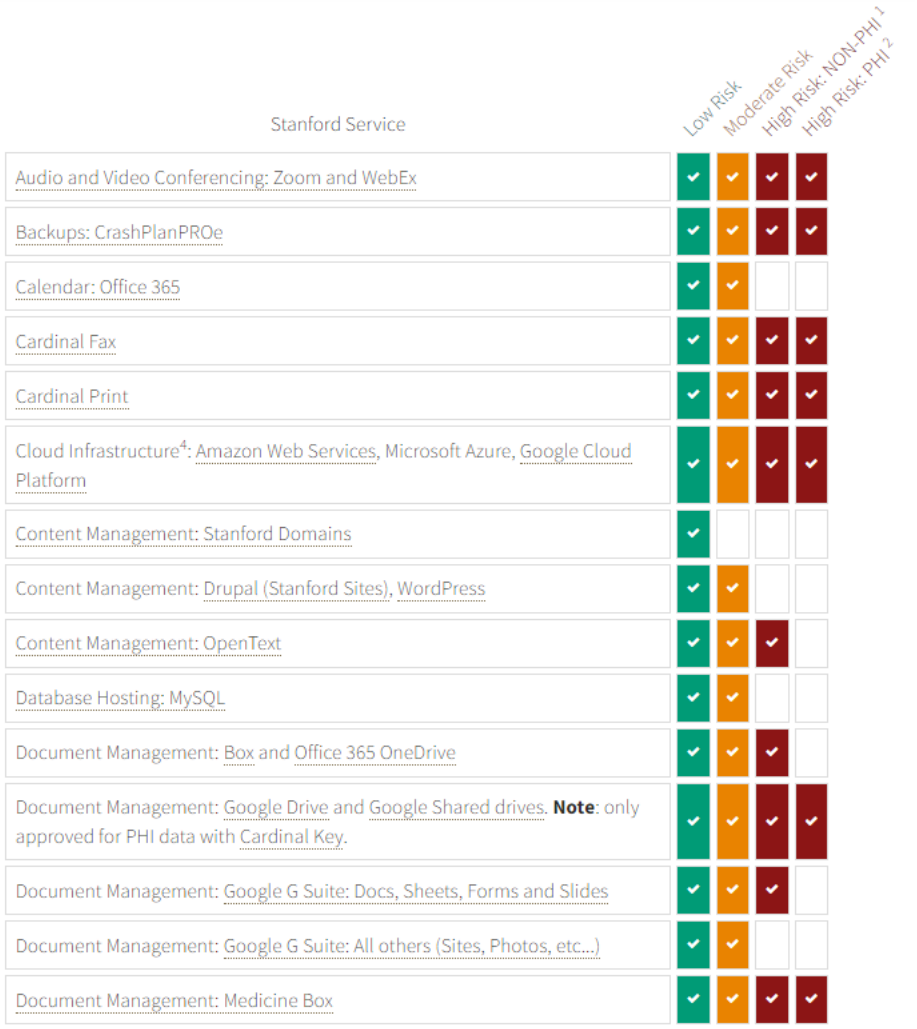
1. Protection of the data is required by law/regulation,
2. Stanford is required to self-report to the government and/or provide notice to the individual if the data is inappropriately accessed, or
3. The loss of confidentiality, integrity, or availability of the data or system could have a significant adverse impact on our mission, safety, finances, or reputation.

https://uit.stanford.edu/guide/riskclassifications

Stanford University IT

Approved Services

This table indicates which classifications of data are allowed on a selection of commonly used



Stanford Service	Low Risk	Moderate Risk	High Risk: NON-PHI [1]	High Risk: PHI [2]
Audio and Video Conferencing: Zoom and WebEx	✓	✓	✓	✓
Backups: CrashPlanPROe	✓	✓	✓	✓
Calendar: Office 365	✓	✓		
Cardinal Fax	✓	✓	✓	✓
Cardinal Print	✓	✓	✓	✓
Cloud Infrastructure[4]: Amazon Web Services, Microsoft Azure, Google Cloud Platform	✓	✓	✓	✓
Content Management: Stanford Domains	✓			
Content Management: Drupal (Stanford Sites), WordPress	✓	✓		
Content Management: OpenText	✓	✓	✓	
Database Hosting: MySQL	✓	✓		
Document Management: Box and Office 365 OneDrive	✓	✓	✓	
Document Management: Google Drive and Google Shared drives. **Note**: only approved for PHI data with Cardinal Key.	✓	✓	✓	✓
Document Management: Google G Suite: Docs, Sheets, Forms and Slides	✓	✓	✓	
Document Management: Google G Suite: All others (Sites, Photos, etc...)	✓	✓		
Document Management: Medicine Box	✓	✓	✓	✓

Segurança dos dados

Privacidade/Proteção:

- Refere-se à colheita, utilização e disseminação da informação de acordo com a autorização e vontade do proprietário dessa informação.

Confidencialidade:

- A possibilidade de aceder ou modificar a informação é apenas concedida a utilizadores autorizados e para fins específicos autorizados.

Integridade:

- A informação é considerada correcta e suficientemente confiável na representação da realidade.

Disponibilidade:

- A informação pode ser acedida e utilizada para determinados fins.

https://www.pexels.com/photo/two-person-standing-under-lot-of-bullet-cctv-camera-374103/

Sensitive Data Toolkit for Researchers

Part 1: Glossary of Terms for Sensitive Data used for Research Purposes

Prepared by the Portage Network Sensitive Data Expert Group on behalf of the Canadian Association of Research Libraries (CARL)

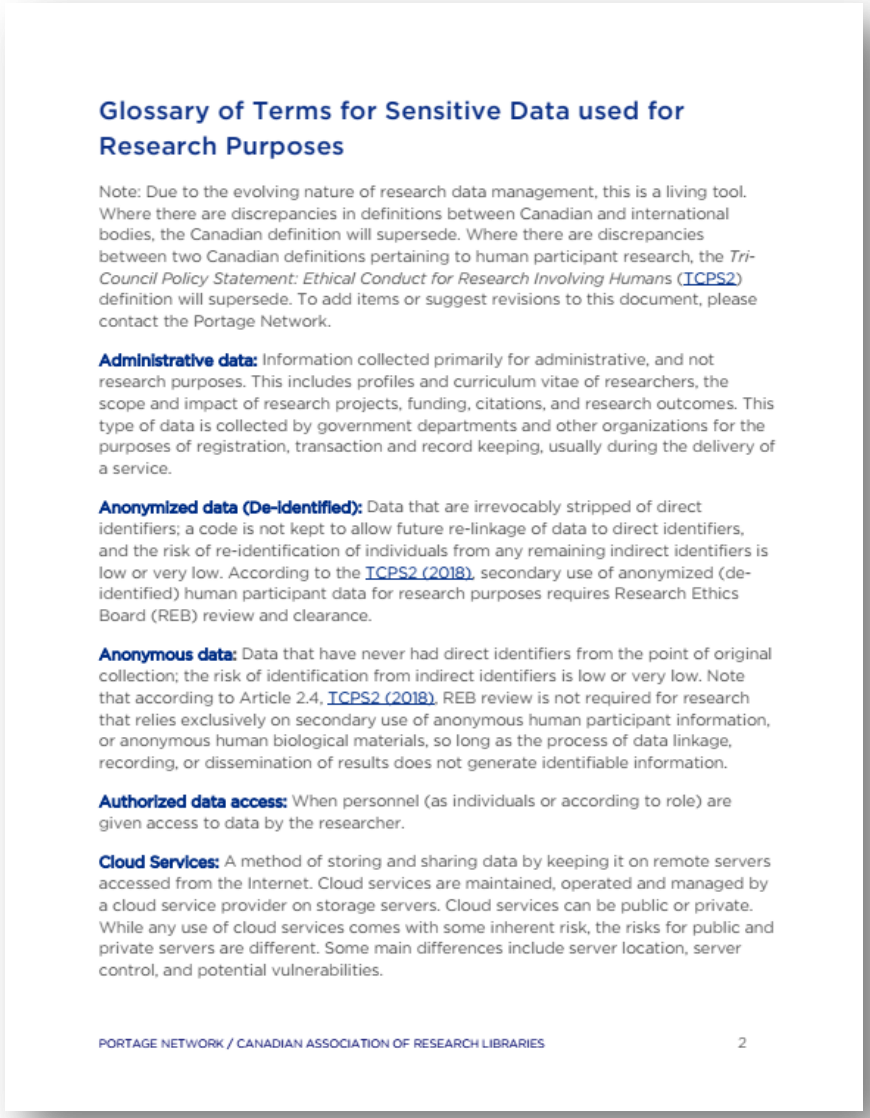
SEPTEMBER 2020

Portage Network
Canadian Association of Research Libraries
portage@carl-abrc.ca

www.carl-abrc.ca

portage
SERVICES PARTAGÉS POUR LES DONNÉES DE RECHERCHE
SHARED STEWARDSHIP OF RESEARCH DATA

CARL ABRC
CANADIAN ASSOCIATION OF RESEARCH LIBRARIES | ASSOCIATION DES BIBLIOTHÈQUES DE RECHERCHE DU CANADA



Glossary of Terms for Sensitive Data used for Research Purposes

Note: Due to the evolving nature of research data management, this is a living tool. Where there are discrepancies in definitions between Canadian and international bodies, the Canadian definition will supersede. Where there are discrepancies between two Canadian definitions pertaining to human participant research, the *Tri-Council Policy Statement: Ethical Conduct for Research Involving Humans* (TCPS2) definition will supersede. To add items or suggest revisions to this document, please contact the Portage Network.

Administrative data: Information collected primarily for administrative, and not research purposes. This includes profiles and curriculum vitae of researchers, the scope and impact of research projects, funding, citations, and research outcomes. This type of data is collected by government departments and other organizations for the purposes of registration, transaction and record keeping, usually during the delivery of a service.

Anonymized data (De-identified): Data that are irrevocably stripped of direct identifiers; a code is not kept to allow future re-linkage of data to direct identifiers, and the risk of re-identification of individuals from any remaining indirect identifiers is low or very low. According to the TCPS2 (2018), secondary use of anonymized (de-identified) human participant data for research purposes requires Research Ethics Board (REB) review and clearance.

Anonymous data: Data that have never had direct identifiers from the point of original collection; the risk of identification from indirect identifiers is low or very low. Note that according to Article 2.4, TCPS2 (2018), REB review is not required for research that relies exclusively on secondary use of anonymous human participant information, or anonymous human biological materials, so long as the process of data linkage, recording, or dissemination of results does not generate identifiable information.

Authorized data access: When personnel (as individuals or according to role) are given access to data by the researcher.

Cloud Services: A method of storing and sharing data by keeping it on remote servers accessed from the Internet. Cloud services are maintained, operated and managed by a cloud service provider on storage servers. Cloud services can be public or private. While any use of cloud services comes with some inherent risk, the risks for public and private servers are different. Some main differences include server location, server control, and potential vulnerabilities.

Direitos do Titular dos Dados Pessoais

- Confirmação da existência do tratamento de dados
- Acesso aos dados
- Correção de dados incompletos, inexatos ou desatualizados
- Anonimização, bloqueio ou eliminação de dados desnecessários, excessivos ou em desconformidade
- Portabilidade dos dados
- Eliminação e revogação do consentimento
- Informação sobre compartilhamento de dados
- Informação sobre não fornecimento do consentimento e consequências

https://www.pexels.com/photo/judgement-scale-and-gavel-in-judge-office-5669602/

Declaração de Helsínquia

(1964, 1975, 1983, 1989, 1996, 2008, 2013)

8. *Embora o objetivo primário da investigação médica seja gerar* ***novo conhecimento****,* ***essa finalidade nunca prevalece sobre os direitos e interesses individuais*** *dos participantes (...)*

25. *A participação de pessoas capazes de dar* ***consentimento informado*** *para serem participantes sujeitos de investigação médica tem de ser voluntária.*

30. *(...) [em situações específicas]* ***<u>o estudo pode prosseguir sem consentimento informado</u>*** *desde que as razões específicas (...) estejam expressas no protocolo de investigação e o estudo tenha sido aprovado por uma comissão de ética para a investigação. O consentimento para permanecer na investigação deve ser obtido logo que possível (...)*

http://www. hopitalmontfort.com/en/new-edition-declaration-helsinki

HIPAA Privacy Rule, 1996

Health Insurance Portability and Accountability Act

The 18 identifiers that make health information PHI are:

1. Names
2. Dates, except year
3. Telephone numbers
4. Geographic data
5. FAX numbers
6. Social Security numbers
7. Email addresses
8. Medical record numbers
9. Account numbers
10. Health plan beneficiary numbers
11. Certificate/license numbers
12. Vehicle identifiers and serial numbers including license plates
13. Web URLs
14. Device identifiers and serial numbers
15. Internet protocol addresses
16. Full face photos and comparable images
17. Biometric identifiers (i.e. retinal scan, fingerprints)
18. Any unique identifying number or code

The NEW ENGLAND JOURNAL of MEDICINE

PERSPECTIVE
HISTORY OF MEDICINE

HIPAA at 25 — A Work in Progress

A.L. Allen

HIPAA is best viewed as a framework of evolving regulation that's revised periodically in response to demands of biomedical innovation and public health in the digital age. That capacity for adaptive modification is among the greatest strengths of HIPAA and its rules.

JUN 05

PERSPECTIVE

HIPAA and the Leak of "Deidentified" EHR Data

K.D. Mandl and E.D. Perakslis

Although patients (and their physicians) still have difficulty obtaining complete medical record information in a timely fashion, HIPAA policies permit massive troves of digital health data to traverse the medical–industrial complex unmonitored and unregulated.

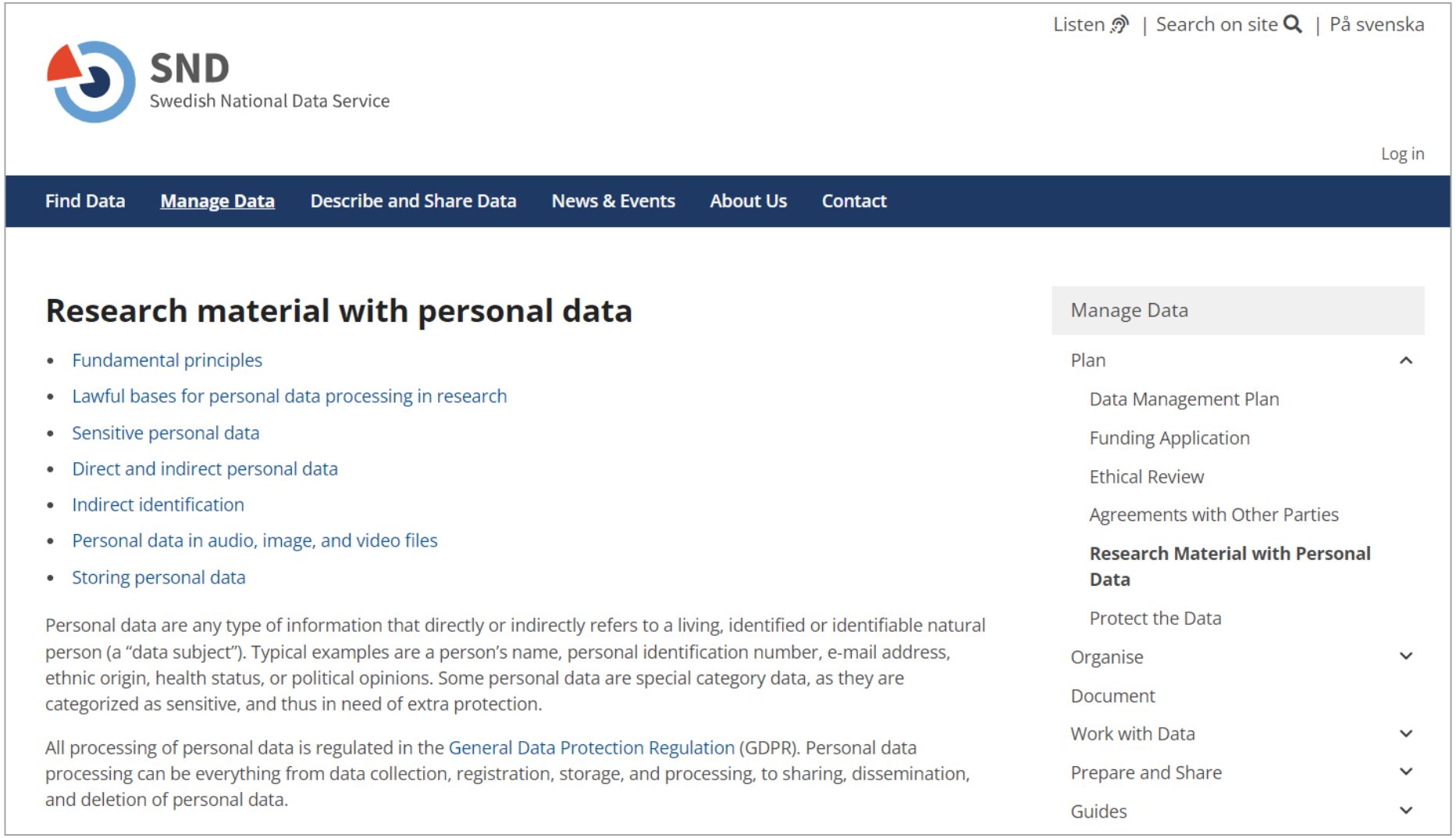
JUN 05

Regulação Geral de Proteção de Dados, RGPD

(General Data Protection Regulation, GDPR)

25 de maio de 2018

FÓRUM GESTÃO DE DADOS DE INVESTIGAÇÃO



https://snd.gu.se/en/manage-data/plan/research-material-with-personal-data

OpenAIRE

Guides for Researchers

How to deal with sensitive data

Learn how to preserve your sensitive data safely

What is Sensitive data

What is the best way of managing access to sensitive data? This is not a straightforward question as it involves ethical, legal and technical issues to be tackled. This guide will help you on your way to preserve your sensitive data safely. It explains the different types of sensitive data, how to prepare them for storage and the possible cost involved in the process.

What is Sensitive data

Sensitive data is data that must be protected against unwanted disclosure. Access to sensitive data should be safeguarded. Protection of sensitive data may be required for legal or ethical reasons, for issues pertaining to personal privacy, or for proprietary considerations.

The EU has strong regulations regarding personal (e.g. General Data Protection Regulation - GDPR) and sensitive non-personal data. In Horizon 2020 an Ethical review is required when applying for a grant.

Examples of sensitive data are:

- **Personal data:** identifiers such as names or identification numbers, physical, physiological, genetic, mental, economic, cultural or social characteristics, it also includes location data from GPS or mobile phones
- **Confidential data**: trade secrets, investigations,data protected by intellectual property rights Security: passwords, financial information, national safety, military information...
- **Combination of different datasets** that can be combined into sensitive or personal data
- **Biological data**: endangered (plant or animal) species, where their survival is dependent on the protection of their location data (biodiversity community)
- **Personal and sensitive metadata**

https://www.openaire.eu/sensitive-data-guide

ANDS Guides

Publishing and sharing sensitive data

Key messages from the ANDS Guide to Publishing and sharing sensitive data

- The advantages of publishing your sensitive data will probably far outweigh any potential disadvantages when simple and appropriate steps are taken.
- Publishing your data, or just a description of your data (that is the metadata), means that others can discover it and cite it.
- You can publish a description of your data without making the data itself openly accessible.
- You can place conditions around access to published data.
- Sensitive data that has been de-identified can be shared.

https://www.ands.org.au/__data/assets/pdf_file/0010/489187/Sensitive-Data-Guide-2018.pdf

JMIR Publications
Advancing Digital Health & Open Science

COVID-19 Research: Navigating the European General Data Protection Regulation

Regina Becker [1]; Adrian Thorogood [2]; Johan Ordish [3]; Michael J.S. Beauvais [2]

Published on 27.8.2020 in Vol 22, No 8 (2020): August

Preprints (earlier versions) of this paper are available at https://preprints.jmir.org/preprint/19799, first published May 01, 2020.

Abstract

Researchers must collaborate globally to rapidly respond to the COVID-19 pandemic. In Europe, the General Data Protection Regulation (GDPR) regulates the processing of personal data, including health data of value to researchers. Even during a pandemic, research still requires a legal basis for the processing of sensitive data, additional justification for its processing, and a basis for any transfer of data outside Europe. The GDPR does provide legal grounds and derogations that can support research addressing a pandemic, if the data processing activities are proportionate to the aim pursued and accompanied by suitable safeguards. During a pandemic, a public interest basis may be more promising for research than a consent basis, given the high standards set out in the GDPR. However, the GDPR leaves many aspects of the public interest basis to be determined by individual Member States, which have not fully or uniformly made use of all options. The consequence is an inconsistent legal patchwork that displays insufficient clarity and impedes joint approaches. The COVID-19 experience provides lessons for national legislatures. Responsiveness to pandemics requires clear and harmonized laws that consider the related practical challenges and support collaborative global research in the public interest.

J Med Internet Res 2020;22(8):e19799

doi:10.2196/19799

Revista do Ministério Público 163 : Julho : Setembro 2020 [pp. 203-226]

Acesso e reutilização de registos clínicos para fins de investigação no âmbito da pandemia por COVID-19

Rui Guimarães
Professor Auxiliar da Faculdade de Medicina da Universidade do Porto
Responsável pelo Acesso à Informação, Centro Hospitalar Universitário de São João, Porto

Miguel Guimarães
Bastonário da Ordem dos Médicos
Assistente graduado de Urologia, Centro Hospitalar Universitário de São João, Porto

Nuno Sousa
Presidente da Escola de Medicina da Universidade do Minho
Diretor do Centro Clínico Académico (2CA), Braga

Fernando Araújo
Professor Auxiliar da Faculdade de Medicina da Universidade do Porto
Presidente do Conselho de Administração do Centro Hospitalar Universitário de São João, Porto

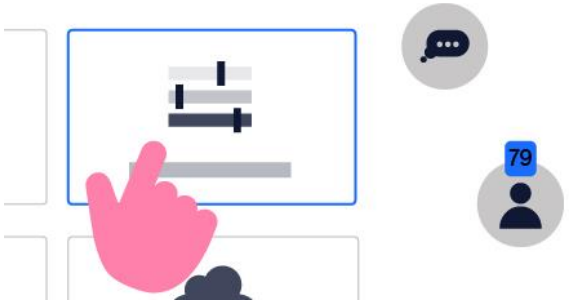
Filipe Almeida
Presidente da Comissão de Ética do Centro Hospitalar Universitário de São João e Faculdade de Medicina da Universidade do Porto
Membro do Conselho Nacional de Ética para as Ciências da Vida

Altamiro Pereira
Diretor da Faculdade de Medicina da Universidade do Porto
Membro do Conselho das Escolas Médicas Portuguesas

André Lamas Leite
Professor Auxiliar das Faculdades de Direito da Universidade do Porto e da Universidade Lusíada (Porto)
Investigador do Centro de Investigação Jurídico-Económica da Faculdade de Direito da Universidade do Porto

Daniel Gonçalves
Membro do Conselho Nacional de Ética para as Ciências da Vida
Presidente da Associação Portuguesa Direito e Medicina

Rui Guimarães *et al.*
Revista do Ministério Público, 2020



Mentimeter

Aceda ao URL: www.menti.com

Introduza o código indicado.

Participe anonimamente na discussão!

Agradecimentos

Acknowledgements

RDA Sensitive Data Interest Group

Special thanks to Kristal Spreadborough and Romain David

IPO-Porto

António Carvalho (*DPO*)

MTG

Tiago Taveira Gomes

A todos os participantes!

IPOPORTO
INSTITUTO PORTUGUÊS DE

MTG DATA SCIENCE MACHINE LEARNING REAL-WORLD EVIDENCE
Resilient End-to-end Healthcare Data Services

rita.silva@mtg.pt

Rb-Silva, Rita

rita-rbsilva