

Data warehouse and medical research

Thiago Gonçalves dos Santos Martins¹, Flavia de Souza Rangel²

¹ Universidade de Coimbra, Coimbra, Portugal.

² Universidade Estácio de Sá, Rio de Janeiro, RJ, Brazil.

DOI: 10.31744/einstein_journal/2022ED6324

In the last 2 years, about 90% of the world's data has been generated. The great increase in data production in medicine has often exceeded the capacity of processing them. This large production has allowed the development of multicenter epidemiological studies with large numbers of participants, and large amounts of data generated. This facilitates data generation about the development and popularization of electronic medical records at hospitals and medical research centers around the globe.⁽¹⁾

The development of new technologies for the interpretation and selection of these data has been of great importance in medicine. In addition, this allows the advancement of research with a larger number of participants aligned to a lower cost and shorter time invested.⁽²⁾ Among the technologies used for this purpose is the use of data warehouse.

This data storage program enables the analysis of large amounts of information in a short period of time. The program makes it possible to store and organize the collected data and to search for large amounts of data ("data mining"). The results obtained by data mining enable the improvement of classifications and associations of the analyzed information.

The studies performed with this large amount of data provide better clinical decision making and statistically significant results in less time compared with traditional epidemiological studies. For this reason, the data warehouse is useful for processing large amounts of medical data, which can be used to support decision making based on valid and reliable information. In addition, these data can be stored for months or years.^(3,4)

The use of the data warehouse allows real-time data analysis and provides a better development of public policies in the area of healthcare. In the area of public management, the examination of big data has focused on improving the quality of services provided to individuals, and the optimization of public resources. For technologies to work, they need to be adapted to the context and culture of the environment in which they will be applied. To do this, we must consider the particularities of each location and action to encompass digital accessibility, digital inclusion, scalability, and technological sustainability. Furthermore, the issue of digital accessibility and inclusion is the major challenge to ensure that citizens, broadly speaking, to have access to all services provided by the government.

Considering this issue, we must pay attention to the protection of research data, according to the laws of the place of the study. In Brazil, the right to the protection of personal data in digital media is not yet expressly included in the Constitution as a fundamental right. However, there is an ongoing constitutional amendment proposal number 17/2019 that aims to insert Article 5, XII-A, CRFB/88, to ensure the fundamental right to the protection of personal data as well as safety in digital media. Despite the existence of

How to cite this article:

Martins TG, Rangel FS. Data warehouse and medical research. *einstein* (São Paulo). 2022;20:eED6324.

Corresponding author:

Thiago Gonçalves dos Santos Martins
Rua Botucatu, 822 - Vila Clementino
Zip code: 04023-062 - São Paulo, SP, Brazil
Phone: (55 21) 2571-2248
E-mail: thiagogsmartins@yahoo.com.br

Copyright 2022



This content is licensed under a Creative Commons Attribution 4.0 International License.

Articles 5, X and XII of the Brazilian Constitution, the guarantee of data protection is still insufficient in the current information age.⁽⁵⁾

In this context, the Federal Supreme Court recently recognized the data protection as an autonomous fundamental right, in a paradigmatic case regarding the request for sharing citizens' personal data by telephone companies with the *Instituto Brasileiro de Geografia e Estatística* (IBGE) during the coronavirus pandemic in 2020. The decision was historical, and its degree of importance is compared by experts to the German Constitutional Court in 1983, which brought to the country the concept of informational self-determination.^(6,7)

At the infra-constitutional level, the General Data Protection Law - LGPG (*Lei Geral de Proteção de Dados Pessoais* - law 13.709 of August 14, 2018) after a 2-year period and numerous discussions is currently in force in Brazil.⁽⁸⁾ However, there are exceptions such as the embryonic implementation of the National Data Protection Authority, whose amendment of the LGPD, through the law 13.853/2019, provided for its creation with the legal nature of a public body, and its structure is regulated by Ordinary 10.474 of August 26, 2020.⁽⁹⁾ This law recently entry into force after the publication of the appointment of the Chief Executive Officer of the Board of Directors of the National Data Protection Authority (Federal Official Gazette (*Diário Oficial da União*), November 6, 2020), as expressed in Article 6 of the mentioned ordinary. However, in this regard, there are still several difficulties to be faced in this new reality.

In European countries, such as Portugal, the fundamental right to data protection is not new, being a theme pertaining to its Constitution and present in the Fundamental Rights of the European Union.⁽¹⁰⁾ Of note is that, in a context of the information age, the General Data Protection Regulation (GDPR; regulation 2016/679 of the European Parliament and Council) was issued on April 27, 2016, and became effective from May 25, 2018. It is important to highlight that the Brazilian LGPD was greatly inspired by the European GDPR, although there are divergences. Portugal subsequently edited the law 58/2019,⁽¹¹⁾ replacing it with the law 67/98 on the Personal Data Protection Law, which is currently in force.⁽¹²⁾

The creation of artificial intelligence algorithms has allowed the development of data warehouses with autonomy for data extraction and analysis. This automatization can reduce the human errors often found in the manual documentation of data, and this has been an alternative for studies with larger amounts of collected information. The data warehouse can extract data from the cloud, reduce the costs and complexity of

the process, and facilitate the decision making in real time,⁽¹³⁾ which would compensate the high investments to implement these programs.

Finally, there is need to highlight that data warehouse is currently unable to conduct adequate analysis of unstructured data, such as images and graphs. In addition, possible errors in the documentation can hinder the final analysis of the results. Therefore, it is important to involve the physician during the whole process of making and running the program in an attempt to ensure the quality of the data generated, and avoid that the big data analyzed produces a “big noise”, and may generate false correlations between the investigated data.

AUTHORS' INFORMATION

Martins TG: <http://orcid.org/0000-0002-3878-8564>

Rangel FS: <http://orcid.org/0000-0001-6263-8518>

REFERENCES

- Martins TG, Costa AL, Martins TG. Big Data use in medical research. *einstein* (São Paulo). 2018;16(3):eED4087.
- Schneeweiss S, Avorn J. A review of uses of health care utilization databases for epidemiologic research on therapeutics. *J Clin Epidemiol*. 2005;58(4):323-37. Review.
- Karami M, Rahimi A, Shahmirzadi AH. Clinical data warehouse: an effective tool to create intelligence in disease management. *Health Care Manag (Frederick)*. 2017;36(4):380-4. Review.
- Fette G, Kaspar M, Liman L, Dietrich G, Ertl M, Krebs J, et al. Exporting data from a clinical data warehouse. *Stud Health Technol Inform*. 2018;248:88-93.
- Brasil. Senado Federal. Proposta de Emenda à Constituição N° 17, DE 2019. Brasília (DF): Senado Federal; 2020 [citado 2020 Set 24]. Disponível em: <https://www25.senado.leg.br/web/atividade/materias/-/materia/135594>
- Brasil. Supremo Tribunal Federal. Processo Eletrônico. Ação Direta de Inconstitucionalidade. Brasília (DF): Supremo Tribunal Federal; 2020 [citado 2020 Set 24]. Disponível em: <https://portal.stf.jus.br/processos/detalhe.asp?incidente=5895165>
- Mendes LS. Decisão histórica do STF reconhece direito fundamental à proteção de dados pessoais. São Paulo: Jota; 2020 [citado 2020 Out 3]. Disponível em: <https://www.jota.info/opiniao-e-analise/artigos/decisao-historica-do-stf-reconhece-direito-fundamental-a-protecao-de-dados-pessoais-10052020>
- Brasil. Presidência da República. Secretária-geral. Subchefia para Assuntos Jurídicos. Lei N° 13.709, de 14 de agosto de 2018. Lei Geral de Proteção de Dados Pessoais (LGPD). (Redação dada pela Lei n° 13.853, de 2019). *Diário Oficial da União*, Brasília (DF): Presidência da República; 2020 [citado 2020 Set 24]. Disponível em: http://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/L13709.htm
- Brasil. Presidência da República. Secretária-geral. Subchefia para Assuntos Jurídicos. Decreto N° 10.474, de 26 de agosto de 2020. Aprova a Estrutura Regimental e o Quadro Demonstrativo dos Cargos em Comissão e das Funções de Confiança da Autoridade Nacional de Proteção de Dados e remaneja e transforma cargos em comissão e funções de confiança. Brasília (DF): Presidência da República; 2020 [citado 2020 Out 03]. Disponível em: http://www.planalto.gov.br/ccivil_03/_ato2019-2022/2020/decreto/D10474.htm

10. Portugal. Universidade de Coimbra. Informação Administrativa e Proteção de Dados. Da Privacidade à Proteção de Dados. Coimbra, 16 de dezembro de 2019. (Projeto Especial por Maria João Carvalho e Paulo Simões Lopes). Coimbra: Universidade de Coimbra; 2022 [citado 2020 Set 24]. Disponível em: https://www.uc.pt/pt/pt/protecao-de-dados/protecao_dados_pessoais/da_privacidade_a_protecao_de_dados
11. Portugal. Assembleia da República. Lei n.º 58/2019 de 8 de agosto de 2019. Assegura a execução, na ordem jurídica nacional, do Regulamento (UE) 2016/679 do Parlamento e do Conselho, de 27 de abril de 2016, relativo à proteção das pessoas singulares no que diz respeito ao tratamento de dados pessoais e à livre circulação desses dados. Diário da República, 1.ª série, N.º 151, página 03-40. Portugal: Assembleia da República; 2019 [citado 2020 Set 24]. Disponível em: https://www.uc.pt/pt/pt/protecao-de-dados/legis/20190808_lei_58_2019_execucao_protecao_de_dados_e_livre_circ_de_dados
12. Portugal. Assembleia da República. Lei nº 67/98 de 26 de outubro de 1998. Lei da Protecção de Dados Pessoais (transpõe para a ordem jurídica portuguesa a Directiva n.º 95/46/CE, do Parlamento Europeu e do Conselho, de 24 de outubro de 1995, relativa à protecção das pessoas singulares no que diz respeito ao tratamento dos dados pessoais e à livre circulação desses dados). Diário da República — I Série-A N.º 247. Lisboa: Diário da República; 1998. Série-A N.º 247 [citado 2020 Set 24]. Disponível em: https://www.uc.pt/pt/pt/protecao-de-dados/legis/19981026_lei_67_98_lpd
13. Fernández-Caramés TM, Blanco-Novoa O, Froiz-Míguez I, Fraga-Lamas P. Towards an Autonomous Industry 4.0 Warehouse: A UAV and Blockchain-Based System for Inventory and Traceability Applications in Big Data-Driven Supply Chain Management. *Sensors (Basel)*. 2019;19(10):2394.