Big data - the big promise of the new digitised world

Big data is a widely used buzzword in today's information era. The use of big data in the digital world presents both an opportunity and a risk. Mass data is now used and analysed in almost all areas of life. Even the healthcare sector is undergoing extensive digitisation.

Searching for the term "big data" on the internet results in around 100 million hits within a fraction of a second. But what does it mean? "Big data" usually has two meanings. Firstly, it reflects a current-day problem, namely the ever-increasing flood of data, which long ago surpassed the capacity of human imagination and is close to exceeding the limits of traditional analysis methods (Klein). Secondly, the term also encompasses the solution to the problem. When people refer to "big data", they usually also have in mind the techniques and analytical methods used to capture, evaluate and store this flood of data.

Dazzling ambivalence

Informational self-determination is no longer enough in the era of big data. The German Ethics Council is calling for data sovereignty, hoping that this will allow individuals to interactively develop their personality whilst being able to maintain privacy in a networked world. The Ethics Council sees it as the duty of institutional and governmental players to set the necessary non-legal, legal and technical framework.
© German Ethics Council

The term "big data" is relatively new. As recently as 2008, the term decorated the title page of a special issue of the science magazine "Nature", highlighting an urgent problem in the world of research, i.e. how to deal with the growing amount of data generated by experiments and simulations that exceeds all previously known and manageable orders of magnitude. Two years later, the business magazine "The Economist" looked at the data deluge in terms of its economic value, debating its opportunities and challenges.

In late 2017, the German Ethics Council published a statement – "Big Data und Gesundheit – Datensouveränität als informationelle Freiheitsgestaltung" ("Big data and health – data sovereignty as informational freedom") - which offers a "working definition" of big data. Based on international discussions and findings, the German Ethics Council concluded that there was no uniform definition of the term (cf. de Mauro). It also attributed to big data a property that is especially important for the healthcare sector, i.e. the possibility of exchanging and recombining data.

According to the German Ethics Council, big data is about using large amounts of data to identify patterns and gain new insights from the latter. Moreover, due to the vast quantity and diversity of data and the speed with which the data is collected, analysed and recombined, using big data requires the application of innovative IT approaches that are continually progressing (German Ethics Council, 2017, p. 36).

Suddenly, all data can become relevant to health

Data that is more disparate, produced more quickly and in ever-increasing quantities. Using big data requires not only artificial intelligence, but also people. Experts agree that data science will become a key discipline in the digital age.
© European Commission

It has been around three years since the association "Deutsche Hochschulmedizin" (position paper on IT solutions, p. 2) claimed that one third of the data collected and exchanged worldwide will in future be collected and exchanged in the healthcare sector. In its 2017 statement, the German Ethics Council goes one step further, stating that the logic of big data is to constantly recombine personal data from an increasing number of sources, so that almost anything can become relevant to health, including data from so-called wearables, sensors or statements in social media (German Ethics Council, 2017, p. 4ff., 54).

In the healthcare sector, big data comes from a variety of sources: clinical and cohort studies, electronic health records (which are still under development in Germany, but already offered by private companies), registries, biobanks and panomics (i.e. the combined use of all omics data generated with high-throughput molecular technologies). The UK Biobank is an impressive illustration of where this development will take us. In 2020, the UK Biobank intends to make available for research the genomic data of half a million British citizens. Already at present, pharmaceutical companies are paying for access to this data treasure trove, hoping that it will boost drug development and biological research (Technology Review, 8th January 2018).

Big IT companies are beginning to focus their business on human health

However, the necessary infrastructure is still not in place in many healthcare areas, thus preventing the systematic use of the big data that would be required for bringing together, analysing and evaluating health-related data such as ECGs, blood values, images, billing and –omics data of all the patients in a particular hospital or hospital chain.

eHealth is the goal of the German healthcare system, but for the time being, this remains still a pipe dream. Before big data is able to exploit its full potential, a digital database for individual patient models as well as information and counselling systems needs to be put in place. This infographic only takes into account information related to the primary healthcare market, and does not depict the flow of data from the field of research.
© acatech

In addition, the age of digitisation affects virtually all areas of life, and is characterised by data generated by the internet, mobile communications companies, published in scientific papers, recorded in medical settings or with medical assistance and medical devices, as well as data from so-called wearables and similar activity-tracking devices. The large number of shares, investments, research projects and collaborations suggest that IT companies such as Alphabet, Google's parent company, have high expectations for the healthcare data business (Gigerenzer, 2016, p 34f).

The smart combination of health data is expected to increase dramatically and contribute to securing medical knowledge in the future. Computers can independently generate new knowledge from these data with methods such as knowledge modelling, machine learning and cognitive computing (see article entitled "Eliciting reliable information from big data with classifiers and multimodal data fusion"). Stakeholders in the healthcare sector are hoping that the integration and analysis of these data will help identify previously unknown patterns and generate new insights and decision-making tools. This opens up opportunities for improving diagnostics (e.g. the panomics-driven identification of biomarkers), therapy, prevention and for reducing costs.

The field of oncology is pioneering the use of big data

It is in the field of oncology in particular where big data from research is combined with a broad range of clinical data under what is known as precision, individualised or stratified medicine (acatech, p. 48f). Precision medicine is expected to make possible more accurate diagnoses and personalised therapies based on the molecular profiles of individual patients (see article entitled "Big data make therapy work better") However, the necessary conditions will first have to be put in place at university hospitals, both as regards patients as data providers and (non-) medical researchers. With the "Cornerstones for a Heidelberg practice of genome sequencing" (Marsilius-Kolleg of the University Heidelberg and the EURAT group, 2013) researchers from the University of Heidelberg and the Heidelberg-based DKFZ have drawn up their own self-regulation in the era of big data that seeks to achieve a balance between patient well-being, freedom of research and clinical progress.

In the healthcare sector, big data is used by a large number of stakeholders with partially conflicting interests (German Ethics Council, 2017, p. 11). Biomedical researchers hope to gain a better understanding of scientifically relevant relationships and processes through the application of imaging and molecular biology techniques. Healthcare (the so-called primary market) stakeholders expect the utilisation of big data to lead to more personalised treatments and greater effectiveness and efficiency. Insurers and employers expect big data to provide them with accurate profiles of individuals and groups of people. Global IT and Internet companies (the so-called secondary market) strive for better commercial exploitation by linking and analysing health-related data with numerous other pieces of information. Affected people or patients hope that the growing use of tracking systems will provide them with relevant information on healthy lifestyles and personal well-being (German Ethics Council, 2017, p. 11ff).

Will evidence-based medicine come under attack?

While big data creates opportunities, it is also associated with risks (dual use). Experts believe that in the highly sensitive healthcare sector, the increasing use of big data will challenge traditional data protection rights, especially as far as patient data is concerned. A number of people (see article entitled "More data does not automatically mean more information") have warned against the danger of replacing theories and laws with correlations and data patterns, thus insidiously abandoning evidence-based medicine as the gold standard and attaching greater importance to correlations than causalities (e.g. science theoreticians such as Prof. Klaus Mainzer, 2015). The German Ethics Council concludes that big data analyses need to be

complemented with other data and be verified (German Ethics Council, 2017, chairman Prof. Steffen Augsberg, press conference of the German Ethics Council on 30th November 2017).

The Ethics Council further states that the potential benefit of big data depends on the expertise and integrity of individuals and institutions that generate, select, link and interpret data (German Ethics Council, 2017, p. 45). In light of the hype about big data fueled by various interests, the Ethics Council concludes that it would be misunderstanding the issue to believe that more data automatically leads to more knowledge about causal effects (German Ethics Council, 2017, p. 47).

References:

Acatech (Ed..): Individualisierte Medizin durch Medizintechnik, Munich 2017

Auffray, C. et al. : Making sense of big data in health research: Towards an EU action plan, in: Genome Medicine (2016) 8:71, DOI: 10.1186/s13073-016-0323-y

BIO Deutschland (Ed..): Position paper of Bio Deutschland on the topic of big data in the healthcare sector, Berlin, 9th December 2016. https://www.biodeutschland.org/de/positionspapiere/positionspapier-zur-anwendung-von-big-data-im-gesundheitswesen.html

Bitkom (Ed.) Big Data im Praxiseinsatz – Szenarien, Beispiele, Effekte, Berlin 2012, p. 7, p. 19ff.

German Pharmaceutical Industry Association (Ed..): Whitepaper. Digitalisierung in der Pharmaindustrie. Die Revolution ist Realität, Berlin June 2017.

De Mauro, A. et al.: What is big data? A consensual definition and a review of key research topics, AIP Conference Proceedings 1644, 97 (2015); doi: 10.1063/1.4907823, https://nowherevilleblog.files.wordpress.com/2015/05/what-is-big-data.pdf

Deutsche Hochschulmedizin: Positionspapier Medizinischer Fortschritt braucht leistungsstarke IT-Lösungen, Berlin, 1st July 2014 https://www.uniklinika.de/fileadmin/user_upload/pdf/2014-07-09_Deutsche_Hochschulmedizin_e.V._IT-Positionspapier_Final.pdf

German Ethics Council (Ed.): Big Data und Gesundheit – Datensouveränität als informationelle Freiheitsgestaltung. Stellungnahme, Berlin, 30th November 2017. http://www.ethikrat.org/dateien/pdf/stellungnahme-big-data-und-gesundheit.pdf

German Ethics Council (Ed.): Die Vermessung des Menschen – Big Data und Gesundheit, Jahrestagung, 21st May 2015, Simultanmitschrift. http://www.ethikrat.org/dateien/pdf/jt-21-05-2015-simultanmitschrift.pdf

European Parliament: Report on fundamental rights implications of big data: privacy, data protection, non-discrimination, security and law-enforcement (2016/2225(INI)), 20th February 2017 http://www.europarl.europa.eu/sides/getDoc.do?pubRef=-//EP//NONSGML+REPORT+A8-2017-0044+0+DOC+PDF+V0//EN

European Commission: Study on Big Data in Public Health, Telemedicine and Healthcare, December, 2016, DOI:10.2875/734795

Floridi, L.: Big Data and Their Epistemological Challenge, Philos. Technol. (2012) 25: 435. https://doi.org/10.1007/s13347-012-0093-4

Fraunhofer Institute for Intelligent Analysis and Information Systems IAIS (Ed.): Big Data – Vorsprung durch Wissen. Innovationspotentialanalyse, Sankt-Augustin 2012.
https://www.iais.fraunhofer.de/content/dam/iais/gf/bda/Downloads/Innovationspotenzialanalyse_Big-Data_FraunhoferIAIS_2012.pdf

Gigerenzer, G. et al.: Digitale Welt und Gesundheit. eHealth und mHealth – Chancen und Risiken der Digitalisierung im Gesundheitsbereich, in: Sachverständigenrat für Verbraucherfragen, Berlin, January 2016. http://www.telemedallianz.de/pdf/studien/01192016_Digitale_Welt_und_Gesundheit.pdf

Klein, D. et al.: Big Data, in: Informatik-Spektrum, June 2013, , DOI 10.1007/s00287-013-0702-3.

Krumholz,H. Big Data and New Knowledge in Medicine: The Thinking, Training, and Tools Needed for a Learning Health System, in: Health Affairs, 33: 7, 2014, p. 1163–1170. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5459394/

Leimbach, T./Bachlechner, D.: Big Data in der Cloud, in: Büro für Technikfolgenabschätzung beim Deutschen Bundestag (Hg.): Big Data in der Cloud (2014), p. 39ff., Berlin 2014 (Hintergrundpapier Nr. 19).

Mainzer, K. Die Berechnung des Menschen. http://www.ethikrat.org/dateien/pdf/jt-21-05-2015-mainzer.pdf

Marquardt, W.: Was ist Big Data? Versuch einer Positionsbestimmung, Jahrestagung des Deutschen Ethikrates, 2nd June 2015. http://www.ethikrat.org/dateien/pdf/jt-21-05-2015-marquardt.pdf

Marsilius-Kolleg of the University of Heidelberg and EURAT Group. Cornerstones for an ethically and legally informed practice of whole genome sequencing. https://www.uni-heidelberg.de/md/totalsequenzierung/informationen/mk_eurat_stellungnahme_2013.pdf

Nature: Science in the petabyte era
http://www.nature.com/nature/journal/v455/n7209/covers/

PWC: Weiterentwicklung der eHealth-Strategie. Studie im Auftrag des Bundesministeriums für Gesundheit, Berlin, October 2016. https://www.bundesgesundheitsministerium.de/fileadmin/Dateien/3_Downloads/E/eHealth/BMG-Weiterentwicklung_der_eHealth-Strategie-Abschlussfassung.pdf

Ramge, T.: Wie groß ist Big Data, in: Brand Eins 07/2016,
https://www.brandeins.de/archiv/2016/digitalisierung/wie-gross-ist-big-data/

Regalado, Antonio, 8th January 2018, https://www.technologyreview.com/s/609897/500000-britons-genomes-will-be-public-by-2020-transforming-drug-research/

Rüping, S./Graf, N.: Big Data in der Medizin, Deutsches Ärzteblatt, Jg. 110, Issue 41, 11th October 2013.

Springer Gabler Verlag (Ed.), Gabler Wirtschaftslexikon, keyword: Big Data, online at:
http://wirtschaftslexikon.gabler.de/Archiv/-2046774198/big-data-v5.html

Vfa, 18th July 2017 , "Panomics" in der Personalisierten Medizin.
https://www.vfa-bio.de/vb-de/aktuelle-themen/branche/panomics-in-der-personalisierten-medizin.html

Zusammenführen verschiedener Datenformate im Nationalen Centrum für Tumorerkrankungen, Uni Heidelberg, mit SAP
https://www.nct-heidelberg.de/das-nct/vorstellung/innovationen/nct-datatherehouse.html

Dossier

06-Mar-2018
Walter Pytlik
© BIOPRO Baden-Württemberg GmbH

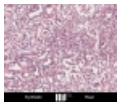
Other dossier articles



22.09.2026
The digital twin – shaping the future of neonatal medicine



04.05.2026
Fractures in focus: predicting rather than waiting



25.09.2025
Powerful AI systems using synthetic training data