Big Data in Healthcare

Opportunities, emerging risks and potential liabilities

Willis Towers Watson

Dive into Big Data

Undeniably, all digital health innovations rely on Big Data – they collect, store, exchange, synthesise and/or produce substantial amounts of health data.

The promise of Big Data to transform healthcare is identical to the promise of digital health innovations: reduced costs, increased efficiency and improved outcomes and access.

Therefore, it is important for digital health companies to explore and consider the emerging areas of concerns and potential liability related to Big Data within their organisations.

This whitepaper was written by:
Kirsten Beasley
Head of Healthcare Broking, North America
Willis Towers Watson
Doris Fischer-Sanchez
Clinical and Enterprise Risk Management Leader
Willis Towers Watson

What is Big Data?

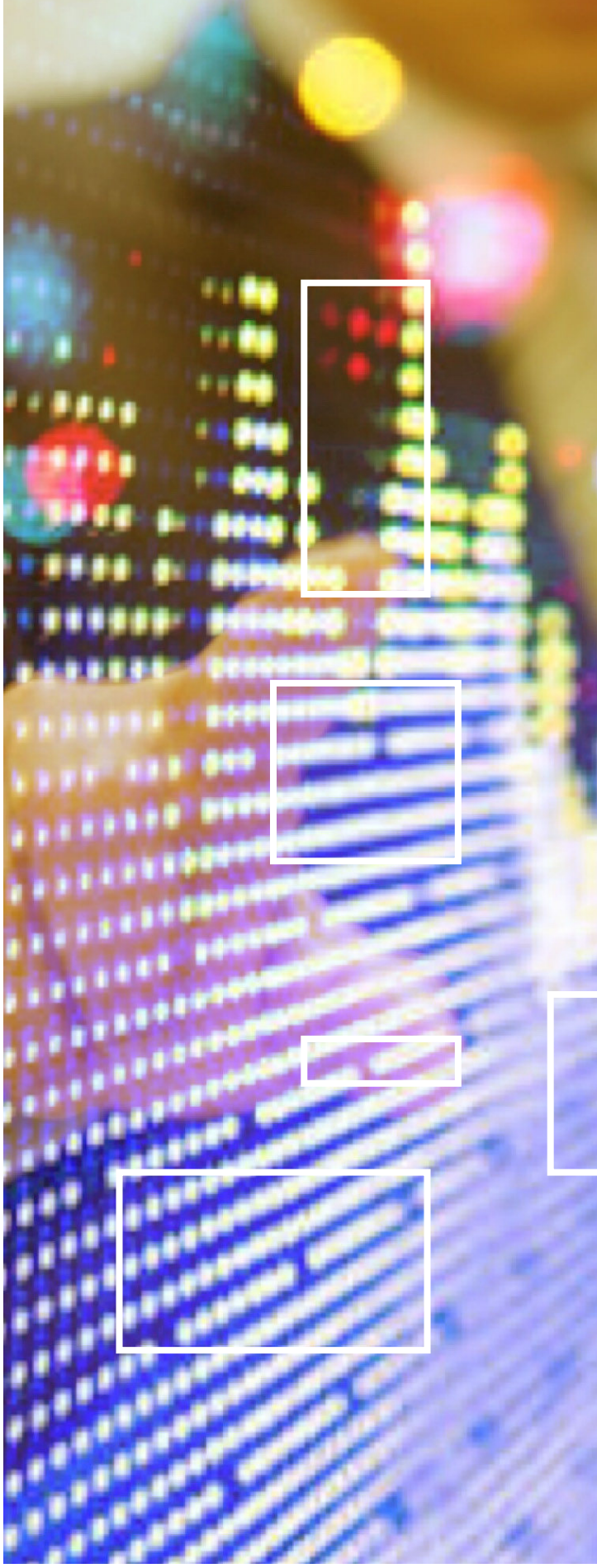
There is no universal agreement on the definition of 'Big Data'; the term is generally used to describe the growth and availability of large datasets. Big Data is frequently characterised in terms of the 7Vs: volume, variety, velocity, validity, value, volatility and veracity[1].

Big Data is data whose scale, and complexity require new architecture, techniques, algorithms, and analytics to manage it and extract value and hidden knowledge from it[2].

In a healthcare context the term often has a multidimensional meaning that incorporates:

1. the volume and diversity of data available from disparate sources with
2. the efficient real-time linking and analysis of those data in order to
3. provide actionable insights and enable informed decision making.

Consequently, 'Big Data' in healthcare is not wholly focused on the flood of data; rather, the emphasis is on the analysis, parsing and synthesis of the data into knowledge and understanding.



[1]Khan, M., Uddin, M., and Gupta, N. (2014) Seven V's of Big Data Understanding Big Data to Extract Value. Accessed at: http://www.asee.org/documents/zones/zone1/2014/Professional/PDFs/113.pdf
[2]Bellazzi R. Big Data and biomedical informatics: a challenging opportunity. Yearb Med Inform. 2014;9:8–13. doi: 10.15265/IY-2014-0024. https://pubmed.ncbi.nlm.nih.gov/24853034/

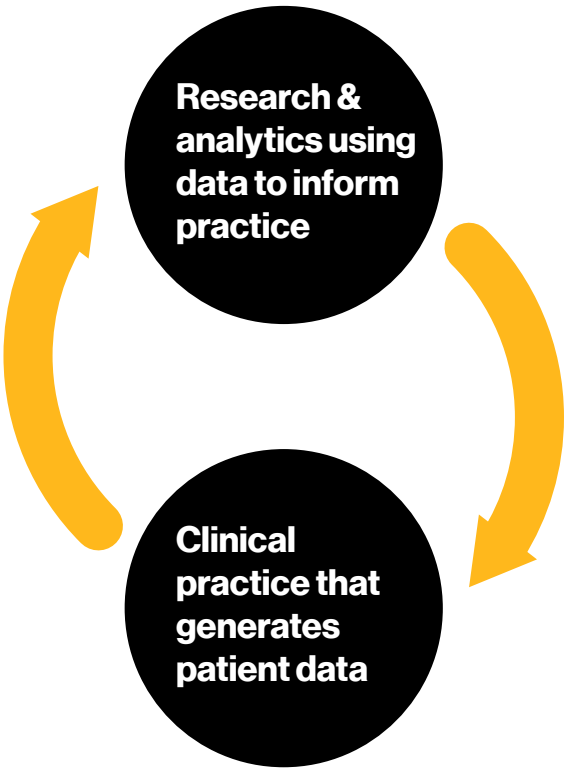
Why does it matter?

Many healthcare systems, though often data rich, do not properly utilise existing datasets to generate a better understanding of how to improve access to better quality care and to reduce waste.

Such missed opportunities result in unnecessary patient harm and serve to increase the gap between the cost of healthcare and the outcomes achieved.

These system limitations could be overcome by the development of a continuous learning healthcare system that harnesses Big Data to 'fuel' a virtuous cycle in which research informs and influences clinical practice and clinical practice informs and influences research.

Virtuous cycle of continuous learning



In such a system, Big Data can help facilitate a more empirically driven healthcare system, ideally, free from bias, to drive lowered costs, improved quality of care and patient safety and ultimately better outcomes. This is the promise of 'Big Data'.

Yet, the pathway to these potentially transformational changes is littered with challenges.

Key challenges in the use of data in healthcare

Interoperability challenges

Health data is unstructured, from variant sources (i.e. financial, billing, pharmaceutical, durable medical, laboratory, etc.). They do not utilise standardised terminology and require great collaboration among all stakeholders.

Accuracy and quality of data needs to be established and issues addressed. Improving the interoperability of healthcare data and systems is a top priority in many countries.

Data governance

Once data is validated, global privacy regulations often require that policies and procedures protect health information.

From a compliance perspective, there are more limitations on data disclosure than on data collection.

There is need for access control, authentication and security during transmission; cloud storage can solve some of these issues.

Policy and procedures regarding how devices, employee access/use and data are managed are critical to the modern healthcare system.

Data storage

The volume of healthcare data is massive and growing unimaginably as more and more electronic information is captured.

On premises storage is almost antiquated as the volume, hardware and power requirements to store the data on site has become very costly to organisations.

On site storage is vulnerable and upkeep of security is precarious as many organisations do not have sufficient IT staff.

It is not easy to add physical space to increase storage capacity and the backup batteries that it requires.

Data accessibility/mobile competency

Downtime or inability to access data presents a substantial risk to care in the modern, and increasingly digitised, healthcare system.

Proactive risk management strategies are essential to protect data. One mitigation strategy is cloud storage; a scalable less capital-intensive way to transfer and provide patient/provider access to complex data without compromising security – even as more mobile devices are used.

Data ownership

It is important to help both providers and patients understand that the medical record, whether paper or electronic has levels of ownership.

The individual data (i.e. vital signs, laboratory values, radiologic images) is typically owned by the patient.

However, the media in which the information is recorded and stored are typically considered the property of the organisation or individual provider (who is the legal custodian and has a duty to protect that information). Simply stated, medical Big Data is somewhat of a shared ownership[3].

Data mining liability

Within healthcare there is a continuously (and necessarily) changing standard of care.

Will the wide availability of large datasets and the proliferation of technological capabilities to synthesise data lead to an expansion on our understanding of the standard of care – one that will eventually evolve to include Big Data analytics?

Management challenges/need for healthcare analytics talent

The healthcare system needs data scientists and IT staff with healthcare knowledge in order to run meaningful analytics.

Healthcare risk managers need to recognise new and different risks that are emerging due to the transformational influence of Big Data.

Further, medical professionals and administrators alike require education about the clinically relevant input that Big Data analytics can generate[4].

Cybersecurity

Cybersecurity encompasses a myriad of potential hazards, including: malware/phishing attempts, ransomware, vendor selection, unsecured mobile devices that may not meet security standards (leaving networks vulnerable), lost and stolen devices, online medical devices (potential for interception and manipulation of data), unrestricted access to computers and inadequate disposal of old hardware[5].

The need for continuous monitoring and ongoing improvements

Organisations can feel like they are drowning in data. The challenge however is not just volume but also the integrity and quality of data.

Strong organisational data governance, with continuous monitoring, is integral to creating ongoing improvements with Big Data analytics.

End-users should receive regular training and reminders about optimal data integrity, data entry practices, and organisations should conduct frequent internal audits and assessments to ensure they are maintaining a high level of data quality and appropriate usage of the data[6].

[3]Shama, R. (April, 2018) Who really owns your health data? Forbes Technology Council. Accessed at: https://www.forbes.com/sites/forbestechcouncil/2018/04/23/who-really-owns-your-health-data/?sh=6da592c16d62
[4]Ibid.
[5]Security Threats in Healthcare Systems. (March, 2019) Accessed at: https://consoltech.com/blog/security-threats-healthcare-systems/
[6]Bresnick, J. (July, 2016) The Role of Healthcare Data Governance in Big Data Analytics. Health IT Analytics. Accessed at: https://healthitanalytics.com/features/the-role-of-healthcare-data-governance-in-big-data-analytics

The health data ecosystem and data reuse

Despite these challenges, health-related Big Datasets hold great promise in the eyes of multiple stakeholders that seek to harness their power.

Each want to use health data generated through digital health innovations in varying ways and to differing purposes, including research, public health surveillance, system analysis and/or marketing and commercial activities.

Through these activities, stakeholders hope to achieve some of the well recognised benefits of secondary uses within the healthcare industry: reductions in systems costs, improvements in system quality, better health outcomes and overall advancements in clinical care and practice.

Whatever the impetus, use of data beyond their originally intended use (typically personal medical care) is often referred to as secondary use or data reuse. Two notable data sharing projects include: Electronic Health Records for Clinical Research (EHR4CR)[7] and eMERGE Consortium[8].

Core to the desire to derive better insights from large healthcare data sets is data linking – a practice distinct from data reuse, but closely related to it.

Data linking entails the aggregation of different data related to the same individual, family, place or event from multiple diverse sources.

The main goal of data linking is to create richer datasets that enable the user to gain deeper insights through the synthesis and aggregation of data.

Data being linked with health records:

- Population registries
- Census data
- Education system data
- Prison service data
- Genomic data (direct to consumer genetic tests)
- Social media
- Credit card data

[7]i-HD. EHR4CR: Electronic Health Records for Clinical Research [Internet]. 2016 [cited 2020 Sep 24]. Available from: https://www.i-hd.eu/index.cfm/resources/ec-projects-results/ehr4cr
[8]National Human Genome Research Institute. Electronic Medical Records and Genomics (eMERGE). Network [Internet]. Genome.gov. 2020 [cited 2020 Sep 24]. Available from: https://www.genome.gov/Funded-Programs-Projects/Electronic-Medical-Records-and-Genomics-Network-eMERGE

Shifting payment models, legislative incentives, health system policies and goals and commercial goals are all driving a surge in secondary use and data linking projects.

Some key focus areas of these projects include:

Research

- **Public health research –** monitoring Big Data to identify disease trends and health strategies based on demographics, geography and socioeconomics.[9]
- **Medical research –** data-driven medical and pharmacological research to cure disease and discover new treatments and medicines

The vast promise of data reuse and data sharing has prompted many healthcare bodies to release statements advocating data sharing as a method to accelerate and advance clinical knowledge. In certain circumstances medical and research institutions require results sharing as a stipulation of funding.

Clinical care

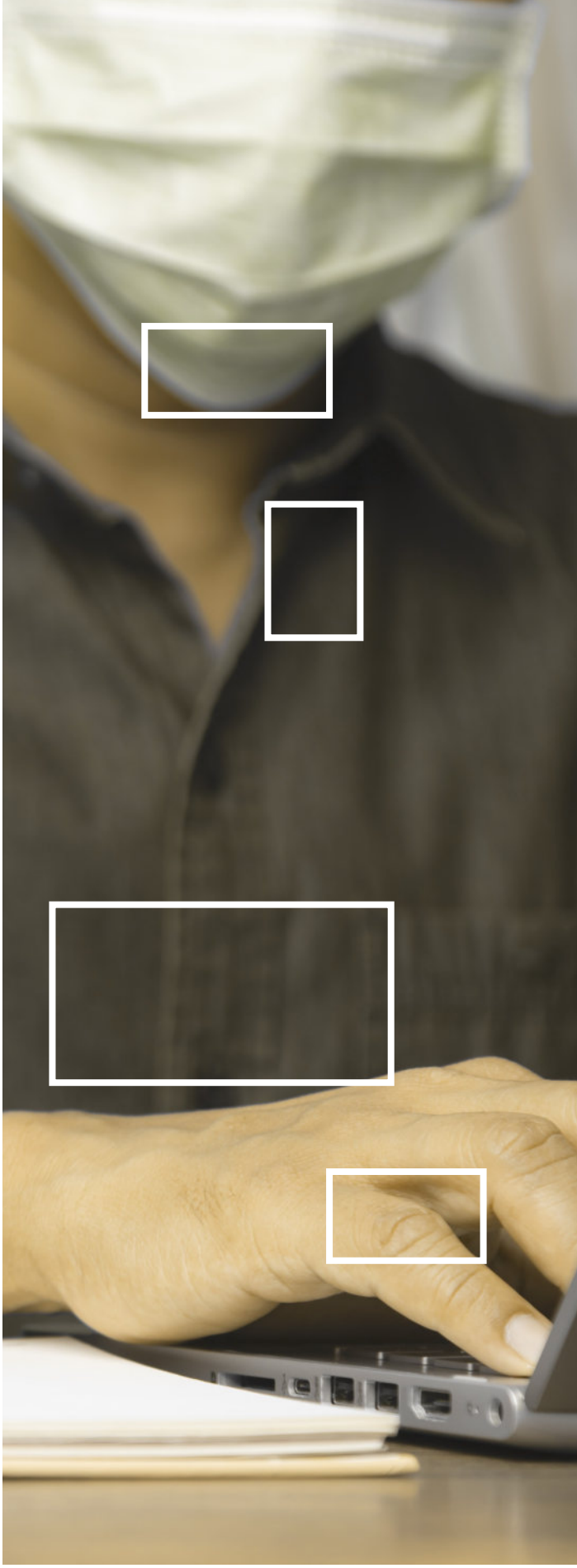
- **Diagnostics –** mining data to analyse and identify causes of illness.
- **Preventative medicine –** predictive analytics and data analysis of genetic, lifestyle, and social circumstances to prevent disease.
- **Precision medicine –** leveraging aggregate data to drive hyper-personalised care.

Risk reduction and compliance

- **Reduction of adverse events,** e.g. medication errors – Big Data can be used to spot medication errors and flag potential adverse reactions.
- **Prevention of opioid abuse –** Big Data can be used to identify risk factors that can predict whether someone is at risk for abusing opioids.
- Big Data can also be used to determine what **healthcare employees may need more support** or training and to encourage continuous learning.

Finance and operational efficiencies

- **Cost reduction through optimised business decision support –** e.g., fraud reduction within the healthcare system or identification of value that drives better patient outcomes for long term savings.
- **Improved supply chain –** leverage Big Data to track supply chain performance and drive better data-informed operational decisions.



[9]New England Journal of Medicine Catalyst (January, 2018) Healthcare Big Data and the Promise of Value-Based Care. Accessed at: https://catalyst.nejm.org/doi/full/10.1056/CAT.18.0290

Health data ecosystem

It is imperative to understand the health data ecosystem as a complex network with numerous participants that utilise and find value in health data in diverse ways.

However, the potential benefits and harms of data reuse are not universally agreed and may shift according to one's perspective and context.

What — Types of data

- Lab data
- Pharmacy data
- Personal health records
- Social media
- Mhealth (apps, wearable)
- Provider data
- Research data
- EHR data
- Regulators
- Direct-to-consumer genetic testing (DTCGT)
- Government agencies
- Genetic data
- Biospecimens
- Payor data
- Biobanks

Why — Data uses

- Improving care and treatment
- Product development
- Market research
- Quality improvement
- Law enforcement
- Health system planning and improvement
- Revenue generation
- Public health monitoring
- Population health mangement
- Fraud detection
- Patient safety
- Disease management
- Illicit activities (eg opioids)
- Sales

Who — Data users

- Academics and universities
- Patients/individuals
- Data brokers
- Medical technology/device companies
- Pharmacies
- Community based organisations (CBOS)
- Biobanks
- Governments and agencies
- Private and public research institutions
- Insurance companies
- Regulators
- Analytics and consulting companies
- Precision medicine projects
- Genetic testing companies
- Labs
- Registries
- Hospitals and health systems
- Medical professionals and providers
- Patient organisations
- IT and technology sector
- Pharmaceutical companies
- Charities

Often, within such practices (reuses and linking of datasets) there is inherent tension between the vast prospective benefits for the community and potential for individual harms – including the risk to privacy and possible misuse of sensitive data.

This tension is evidenced by the growing number of public controversies related to data sharing initiatives and the growing global discussion surrounding the ethical challenges of COVID-19 related contact tracing apps.

Secondary use project	Nature of data share	Objections
A data program	A program aimed at extracting data from GPs for a central database.	Lack of 1) patient awareness of the program and 2) clarity around opt-out options.
Social network company	The social network's personal data was harvested for political advertising.	Lack of consent and transparency.
Consulting company and a government agency	180,000 lung cancer patients' anonymised data was shared with the consulting firm for a study on lung cancer trends.	Consent of the patients was not obtained.
Artificial intelligence (AI) company	1.6M patients' data was transferred to the AI company to test an acute kidney injury altering system (streams).	Inadequate public engagement, awareness and lack of transparency.
A healthcare organisation and a technology company	Identifiable patient data was shared with a technology company to pilot an electronic health record (EHR) search tool.	Lack of notification to patients; a government inquiry to ensure compliance legislation.
A healthcare company and a technology corporation	Agreement to share identifiable patient data to develop cancer algorithms.	Identifiable data utilised; how has consent been obtained?
Online pharmacy	Names and addresses of >20,000 customers were sold to a marketing company.	Breach of data protection rules by not seeking customers' consent.

There are a few critical questions to contemplate:

Do patients know how much of their health data are being used, by whom and for what purpose(s)?

What liability does 'Big Data' and corresponding sharing, linking and reuse practices pose to an organisation?

Some liability considerations of health data reuse:

- **Privacy concerns:** some privacy regulations rely heavily on deidentification as the main mechanism to protect patient privacy.

  With the continued progression of reidentification science and data linking projects, standards and regulations that rely on deidentification provide an increasingly inadequate privacy protection.

  Alongside the reidentification risks, are increasing public concerns surrounding the commercial exploitation of data.

Global privacy related laws

US
HIPAA's Privacy rule (as expanded by the HITECH Act) and various state laws

UK and EU
General Data Protection Regulation (GDPR)

Australia
Federal Privacy Act 1988 and Amendment Bill and various state privacy bills

Canada
Personal Information Protection and Electronic Documents Act (PIPEDA) and provincial

- **Lack of transparency:** consumers don't understand or are unaware of the full extent of the collection and reuse of their health data. This makes it difficult for people to assess the risks and benefits of sharing their data.

  This limited understanding must be addressed in order to have a transparent system that fosters truly informed consent. Notably, not all countries require informed consent for secondary use of patient data – particularly deidentified data.

  For example, while GDPR sets a high threshold for achieving consent, consent is not an absolute requirement and there are several legal bases for exceptions.

  There is great potential for abuse and misuse of patient data in any system that lacks transparency and public awareness.

[10]Mazor KM, Richards A, Gallagher M, Arterburn DE, Raebel MA, Nowell WB, et al. Stakeholders' views on data sharing in multicenter studies. J Comp Eff Res. 2017 Sep;6(6):537–47 https://pubmed.ncbi.nlm.nih.gov/28805448/
[11]Vezyridis P, Timmons S. Resisting Big Data exploitations in public healthcare: free riding or distributive justice? Sociol Health Illn. 2019 Nov;41(8):1585–99. https://onlinelibrary.wiley.com/doi/abs/10.1111/1467-9566.12969

- **Data monetisation: a word of caution about monetisation and profit-making within the health data ecosystem.**

There is widespread concern and caution about the involvement of commercial companies or profit-related motives within the health data ecosystem. Public trust and willingness to consent decrease when societal benefits are murky, or profit may be derived.

Many members of the public are concerned that profit-making, particularly covert profit-making, potentially incentivises companies to misuse data for their own gain and to the disadvantage of the patient/public[10/11].

Within the health data ecosystem, data brokers are one segment that is laser focused on how to monetise health data.

The definition of what a data broker is varies between the different regulatory authorities, but in its broadest sense a data broker is an operation/company that aggregates disparate information (about consumers) from a wide variety of sources; often for the purpose of analysis and reselling/licensing that data to a third party.

An in-depth discussion on data brokers is beyond the scope of this whitepaper, however it is important to note that, in general, the data broker industry has had little oversight and lacks transparency around their data collection, aggregation and selling practices.

Internationally, there is increasing regulatory attention to data brokering practices, but current regulations vary greatly from country to country and even the robust protections that GDPR offers provide several legal bases for processing a consumer's personal data.

What is important to note is that data is increasingly being viewed as a commodity to be sold and traded.

Consider the recent news that Israel has traded the data generated from their health system to Pfizer in return for rapid and continuous access to Pfizer's COVID-19 vaccine[12].

Past examples of sensitive data sets sold:
- Lists of rape victims
- Seniors with dementia
- HIV patients
- Patients with erectile dysfunction[13].

Additionally, some members of the public, (particularly minorities, the socially disadvantaged and vulnerable populations) are concerned about the potential for commercial companies to use these health datasets to stereotype and discriminate.

Financial discrimination (e.g., denial of mortgages), insurance (life and health) discrimination (e.g. raised premiums) and denial of employment are some top concerns related to commercial use of health data. The evidence shows that the public has limited awareness and understanding of commercial access and use of health-related data[14].

Furthermore, the presence of commercial companies in the data reuse equation amplifies public desire for greater control over their data reuse preferences. This presents a challenge that the digital health industry (alongside every other stakeholder) must address.

[12]https://www.pbs.org/newshour/science/israel-trades-pfizer-vaccine-doses-for-medical-data
[13]https://www.forbes.com/sites/kashmirhill/2013/12/19/data-broker-was-selling-lists-of-rape-alcoholism-and-erectile-dysfunction-sufferers/?sh=5a3a68201d53
[14]Beasley, K. (September 2020) Patient and public attitudes towards secondary use of health-related data, with particular focus on themes of commercialism: a two-phased scoping review

Conclusion

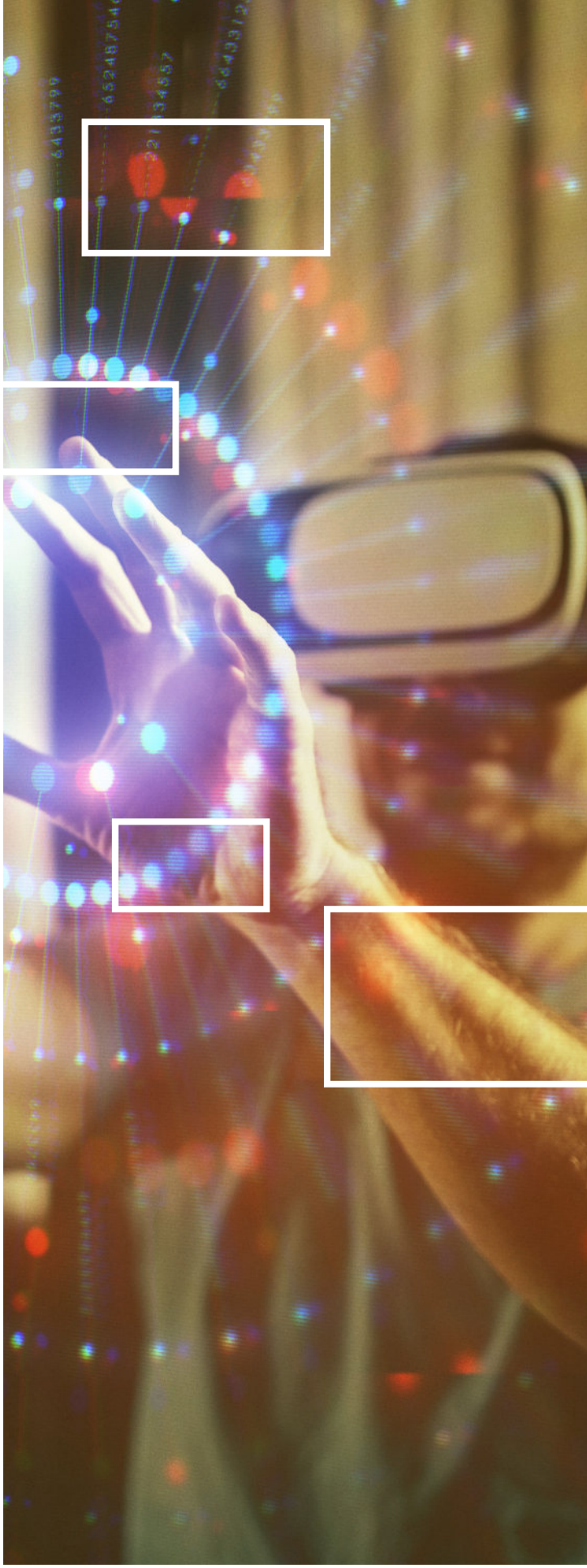
Clearly, health-related longitudinal datasets hold great promise in the eyes of multiple stakeholders seeking to harness their power in varying ways and to differing purposes.

However, organisations must recognise that when tapping into that potential there is a responsibility to ensure that the public's health-related data is not abused.

There should be a concerted effort to be proactively transparent about health data reuse in order to increase public awareness and understanding.

The individual is at the core of the health data ecosystem, thus prioritising their views enables the construction of data reuse policies that will better reconcile the inherent frictions that exist within data sharing.

Improved public awareness about data sharing and linking in conjunction with more transparency about commercial involvement will lay a path towards equitable, sustainable and community-centric secondary use strategies.



For more information please contact:

Tracy Grant
Healthcare Industry Leader – Australasia
Willis Towers Watson
Direct: +61 292 854 052
Mobile: +61 478 305 955
tracy.grant@willistowerswatson.com

Disclaimer

Willis Towers Watson offers insurance-related services through its appropriately licensed and authorised companies in each country in which Willis Towers Watson operates, for example:

In Australia:
- Willis Australia Limited | ABN 90 000 321 237 | AFSL No. 240600
- CKA Risk Solutions Pty Ltd | ABN 33 109 033 123 | AFSL No. 276915

In New Zealand:
- Willis New Zealand Limited | Company No. 111584 | FSP No. FSP37782

For further authorisation and regulatory details about our Willis Towers Watson legal entities, operating in your country, please refer to our Willis Towers Watson website.

It is a regulatory requirement for us to consider our local licensing requirements prior to establishing any contractual agreement with our clients.

This whitepaper offers a general overview of its subject matter. It does not necessarily address every aspect of its subject or every product available in the market. It is not intended to be, and should not be, used to replace specific advice relating to individual situations and we do not offer, and this should not be seen as, legal, accounting or tax advice. If you intend to take any action or make any decision on the basis of the content of this whitepaper you should first seek specific advice from an appropriate professional. Some of the information in this whitepaper may be compiled from third party sources we consider to be reliable, however we do not guarantee and are not responsible for the accuracy of such. The views expressed are not necessarily those of Willis Towers Watson.

About Willis Towers Watson

Willis Towers Watson (NASDAQ: WLTW) is a leading global advisory, broking and solutions company that helps clients around the world turn risk into a path for growth. With roots dating to 1828, Willis Towers Watson has 45,000 employees serving more than 140 countries and markets. We design and deliver solutions that manage risk, optimise benefits, cultivate talent, and expand the power of capital to protect and strengthen institutions and individuals. Our unique perspective allows us to see the critical intersections between talent, assets and ideas – the dynamic formula that drives business performance. Together, we unlock potential. Learn more at willistowerswatson.com.

willistowerswatson.com/social-media

0521

willistowerswatson.com

Willis Towers Watson