MAKING SENSE OF **BIG DATA** THROUGH ANALYTICS

DECEMBER 2016

COCIR

SUSTAINABLE COMPETENCE IN ADVANCING HEALTHCARE

European Coordination Committee of the Radiological, Electromedical and Healthcare IT Industry

EXECUTIVE SUMMARY

Big Data is commonly characterised by the so-called 5Vs:

VOLUME, VARIETY, VELOCITY, VERACITY & VALUE

New technology, enhanced connectivity and mobility make it possible to combine diverse and widely distributed data sources into a combined "data lake". This offers huge potential for healthcare organisations, which can now capture, share, integrate and analyse data from all relevant stakeholders. Insights from these analyses of large-scale health data have multiple applications in healthcare and medical research. However, the uptake of health data analytics in Europe still faces important barriers. These include; the risk of a fragmented enforcement of the General Data Protection Regulation; the lack of trust in IT systems poor data quality and reliability; limited integration between health IT and other data-collecting systems, unmatched workforce skills; low uptake of Electronic Health Records; the lack of sustainable business models and technological challenges linked to the capacity of storage systems or analytical engines.

COCIR RECOMMENDATIONS

1. ACTIONS BY HEALTH IT VENDORS AND HEALTHCARE PROVIDERS:

- **Invest in digital security measures** that safeguard health data confidentiality. This implies not only investing in physical and digital security assets, but also establishing standard operating procedures for secure data management and training staff to apply them.

2. ACTIONS BY MEMBER STATES:

- **Build public confidence in information security** and develop information campaigns that make the public case for sharing health data and address public concerns around the misuse of health data. Awareness-raising programmes will have a greater impact if designed and run at local/regional level;
- **Encourage the adoption of international standards** for clinical information by implementing policy levers, including legislative measures and financial incentives;
- **Enable EHR data extraction** for secondary purposes, incentivising healthcare professionals to register clinical data in a standardised manner. Countries that lag behind should accelerate the deployment and use of EHRs systems.

3. ACTIONS BY THE EUROPEAN COMMISSION AND MEMBER STATES:

- **Launch a European health data quality initiative** that complements and reinforces existing national initiatives for improving data reliability.
- **Support R&D and innovation** programmes that investigate the opportunities of data analytics in care settings.

4. ACTIONS BY EUROPEAN DATA PROTECTION AUTHORITIES:

- **Ensure a harmonised implementation of the General Data Protection Regulation** (GDPR) through a structured dialogue with industry and other stakeholders, that will enable secondary uses of data under the same conditions across the EU.
- **Develop and implement Health Data Governance Frameworks** that ensure efficient, transparent and fair health data management practices and access to health data while safeguarding patients' privacy.

5. JOINT ACTIONS BY HEALTH AUTHORITIES, PAYERS AND HEALTHCARE PROVIDERS:

- **Create a new generation of ICT skilled** workforce and clinical data scientists for the health sector.
- **Accelerate the development of value-based reimbursement** models for health ICT and implement new approaches that recognise and reward the value of ICT solutions.

1. INTRODUCTION

Big Data refers to datasets with sizes is greater than standard databases. New technology, enhanced connectivity and mobility makes it possible to combine diverse and widely distributed data sources into a combined “data lake”. This offers huge potential for healthcare organisations, which can now capture, share, integrate and analyse data from all relevant stakeholders.

Around 700.000 scientific articles are published each year; the amount of medical information available doubles every five years. Additionally, a single patient medical record is rarely greater than a few Gigabytes, nothing compared to the estimated 1,1 Petabytes of data collected over a lifetime when including so-called exogenous data.

Big Data is characterised by the so-called ’5Vs’: Volume, Variety, Velocity, Veracity and Value [1] [2]. It can also be defined as, “a large amount of different types of data produced with high velocity from types of sources and which must be processed through novel approaches to bypass processing limitations extending from current management tools methods” [3].

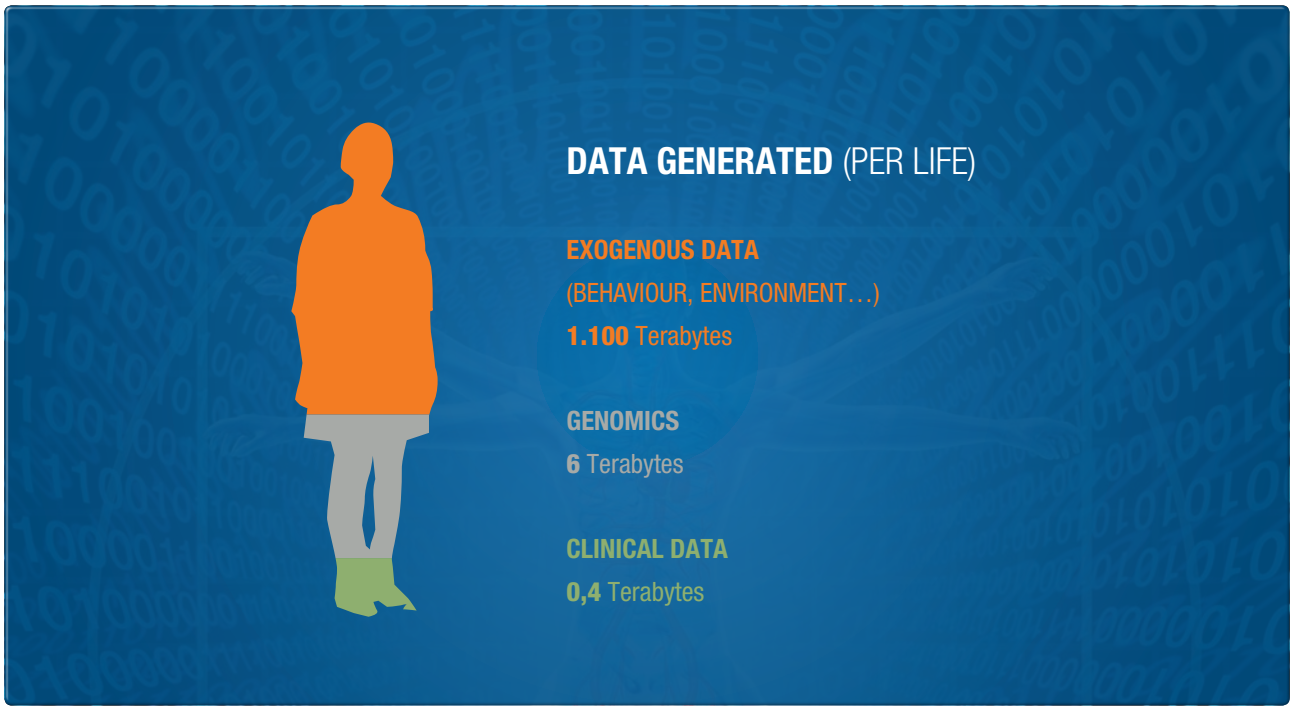
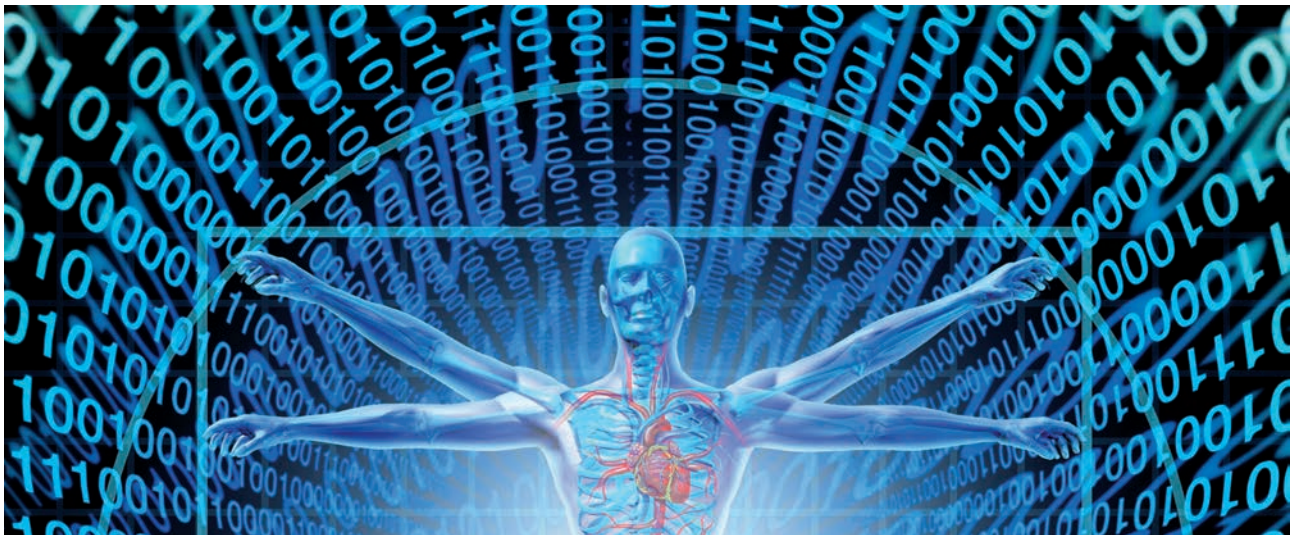


FIGURE 1. CATEGORIES OF PATIENT DATA GENERATED (per life) *Source: IBM*

The value of Big Data lies in the information and knowledge it can generate. There are a wide range of technologies for aggregating, analysing and visualising Big Data. The most important types of analytic technologies used in the healthcare sector are:

- **Descriptive analytics**, used for ad-hoc analysis and Clinical and Business Intelligence reporting. It addresses the questions ‘what happened?’ and ‘why did it happen?
- **Predictive analytics**, which help anticipate ‘what will happen’?
- **Prescriptive Analytics**, which help anticipate what will happen, when and why. Furthermore, it will propose decision options and will indicate the potential implications of each option.
- **Streaming Analytics** allows to process data in real-time, allowing real-time decision-making. This can be valuable in intensive care units or emergency departments. Used in combination with predictive analytics, streaming analytics techniques have enormous potential in situations requiring instantaneous decision-making based on large amounts of data or combinations of several data types.



A number of factors are converging to drive the demand for Big Data in healthcare:

- **Fiscal sustainability** calls for more efficient and cost-effective healthcare services and a shift to outcome-based healthcare models. These rely on accurate and timely measurement of health outcomes and evidence on effectiveness, efficiency or cost savings.
- The drive towards **more evidence-based medicine** requires systematic reviewing of, and timely access to, clinical data to make treatment decisions on the best available information. In addition, healthcare providers are increasingly looking to leverage data to improve patient safety and reduce hazards, particularly those that are less predictable.
- In parallel with these societal factors, recent technical advances make it easier to **collect and analyse information from multiple sources**, allowing the healthcare and research community to link and extract knowledge out of data.
- Finally, by bringing together **different types of information** (e.g. financial + operational + clinical) the whole will be greater than the simple sum of its parts.

Over the last decade, pharmaceutical companies have been aggregating years of research and development data into medical databases. There is also a wealth of patient information stored in Electronic Health Records (EHRs) and soon, as a result of the Clinical Trial Regulation, summary reports of clinical trials' results will be publicly available. However, we are still not able of capturing the full potential of processing all these datasets.

A recent WHO report [4] identified three important barriers to the uptake of health data analytics in Europe, namely:

1. LACK OF DATA PRIVACY AND SECURITY LAWS
2. LIMITED INTEGRATION BETWEEN HEALTH IT SYSTEMS AND OTHER SYSTEMS FOR COLLECTING DATA
3. LACK OF SUPPORT FOR NEW ANALYTICAL METHODS

In addition, regulatory fragmentation across different countries, but even within the same country, further hinders the development of Big Data analytics at national and EU level. It is not clear whether the new EU General Data Protection Regulation (GDPR) will succeed in reducing this variability. Also, the WHO report highlights that only six European countries (13% of respondents) have a national policy or strategy regulating the use of Big Data in the health sector.

European policy makers are beginning to shape EU policy on Big Data in the healthcare sector. Notably, the European Commission (DG Santé) is currently developing a study on Big Data in Public Health, Telemedicine and Healthcare. This document will contain recommendations that will hopefully lay the ground for a coherent EU-wide policy in this sector.

This paper briefly summarises the opportunities data analytics presents to the healthcare sector, further develops the barriers and makes industry recommendations to overcome these challenges.

2. BIG DATA AND ANALYTICS: OPPORTUNITIES AND CHALLENGES

2.1 OPPORTUNITIES

Insights from the analysis of large-scale health data have multiple applications in healthcare and medical research [5]. They can provide valuable information on the interactions between therapies, individuals and the environment. Potential benefits from health analytics for healthcare systems include:

1. IMPROVE QUALITY OF HEALTHCARE DELIVERY AND HEALTH OUTCOMES

Data analytics have the potential to improve healthcare delivery and may also reduce costs [6]. **Descriptive analytics** allow healthcare providers to monitor the efficacy of treatments, compare the safety and effectiveness of drugs and treatments, monitor adverse events in medication and medical devices, compare the efficiency of different pathways and where necessary adapt standard care protocols.

For example, analysing the continuous flow of data coming from remote patient monitoring services can help observe treatment adherence, improve future drug and treatment options and better understand how chronic conditions evolve. This data-based 'intelligence' allows the healthcare community to define optimal care pathways and move to evidence-based management of healthcare systems.

Predictive and **prescriptive analytics** go further than passive monitoring, providing healthcare providers with support in choosing the right preventive, diagnostic and therapeutic options based on existing data. Predictive and prescriptive analytics may also aid development of personalised medicine. Combining data from multiple sources (clinical, lifestyle, genomic data) healthcare professionals can assess the predisposition of an individual to a given disease, tailor the right therapeutic strategy at the right time and deliver timely and targeted prevention. With the support of health ICT, personalised medicine can improve prevention; enhance treatment efficiency and may also improve patients' quality of life. From a care pathway perspective, personalised medicine, underpinned by data analytics, will allow the delivery of patient-centred care, as opposed to the same standard care for all, irrespective of their specific profile and needs.

FIGURE 2. FROM DESCRIPTIVE TO PRESCRIPTIVE ANALYTICS. THE VALUE OF BIG DATA *Source: Agfa HealthCare*

2. ENHANCED MEDICAL RESEARCH

Data analytics also brings substantial advantages for biomedical research. In particular, it may help identify target patients for clinical trials, using data from EHRs, and it also offers robust structures for tracking progress of trial participants in the long term. Likewise, data analytics will play an important role in supporting real world evidence studies that evaluate the efficacy and cost-effectiveness of therapies in routine medical practice.

Analytics will also aid post-market surveillance in medicines, providing evidence on the effectiveness and safety of drugs in real-life situations as opposed to the (artificial and highly controlled) clinical trials environment. Moreover, Big Data analytics can support early detection of adverse events. Lastly, advanced analytics of huge datasets are likely to detect previously unknown patterns accelerating medical research and knowledge.

3. FOSTER MEDICAL EQUIPMENT DEVELOPMENT AND MAINTENANCE

Development, testing, maintenance and post-market surveillance of medical devices demands continuous data processing. Data analytics allows technicians to model potential scenarios in a healthcare episode in order to develop safe, hazard-free equipment.

4. SUPPORT PUBLIC HEALTH POLICIES AND MANAGEMENT OF EPIDEMIC OUTBREAKS

Prescriptive analytics can help anticipate possible pandemics and define the best prevention approaches. Public health authorities may leverage data analytics to detect and model epidemic waves. They can evaluate the efficacy and efficiency of treatments and correlate risk factors for the emergence of a given disease, define the population at risk and articulate informed-based prevention and public health policies.

5. FACILITATE INTERNATIONAL COMPARISONS TO IMPROVE HEALTH SYSTEMS PERFORMANCE

Analysing health data can help countries make informed decisions on planning and resource allocation, enhancing patient safety and improving population health outcomes. In addition, internationally comparable population level health data may provide powerful ways to benchmark and contrast how different health systems perform.

2.2 CHALLENGES

Although health data analytics offers clear and widely acknowledged opportunities, a number of challenges remain before the healthcare sector can exploit data analytics to their full potential.

1. RISK OF FRAGMENTED ENFORCEMENT OF THE GENERAL DATA PROTECTION REGULATION

Capturing value from Big Data depends on access to data, often beyond organisations' operational borders. Healthcare stakeholders do not routinely share existing datasets because of legal constraints, usually related to data privacy safeguards. The current fragmented EU data protection legal environment is further hindering development of data driven healthcare.

The General Data Protection Regulation (GDPR), which will be enforced nationally as of May 2018, establishes a common data protection regulatory framework across the EU. However, it foresees that Member States may maintain or introduce additional conditions - including limitations - on the lawful grounds for processing genetic data, biometric data or data concerning health. They may include additional requirements, such as mandating specific safeguards of limiting the processing of these data to certain healthcare professionals. Therefore, the harmonising effect of the GDPR could be weaker for health data.

2. LACK OF TRUST IN IT SYSTEMS AND CYBERTHREATS

Another drag on developing health analytics is the reluctance of many stakeholders to share health data due to concerns over patient privacy and security. Indeed, securing public trust and support will be one of the key success factors for Big Data projects in the health sector [7].

There is a lack of trust in the ability of state-of-the-art ICT solutions to prevent intrusion, data theft, data alteration etc. COCIR acknowledges that these fears need to be addressed effectively. eHealth, Cloud Computing, Big Data and other promising technologies will not develop without the trust of citizens, professionals and healthcare providers.

Big Data companies and other ICT infrastructure providers (e.g. hosting) have invested in cutting edge security measures as well as policies and controls to improve data security and privacy. However, 'zero risk' is impossible to achieve. In addition, COCIR notes that, for certain aspects such as hosting, regulatory discrepancies between Member States leads to fragmented and uncertain legal environments.

3. DATA QUALITY AND LOW READINESS OF ELECTRONIC HEALTH RECORDS FOR ANALYTICAL USES

A major challenge is the reliability and quality of the data and its value and usability in a given context. Striving for high data quality is both an important requirement and challenge; even the best data cleansing methods cannot remove all inherent unpredictability and detect all inaccuracies and errors when linking different datasets.

Another significant challenge is the lack of integration of existing data pools. Data analytics builds upon heterogeneous data sources. Indeed, the fragmented nature of data repositories (epidemiological and registry data, clinical data, claims data, pharmaceutical and research data, patient behaviour data etc.), the sheer quantity of unstructured data or the multiplicity of incompatible data formats, present major hurdles. Yet for Big Data to create value, datasets will need to be integrated. It will be also imperative that IT service providers and healthcare organisations overcome data quality and reliability barriers to realise Big Data's full potential.

On a different note, Electronic Health Records offer the foundation for pooling single patient-related data in standardised formats. However, although most European countries have invested in developing EHR systems, only few are using data for secondary purposes [8].

4. UNMATCHED WORKFORCE SKILLS

Another significant constraint in realising the value from Big Data is the shortage of IT and adjacent skills in healthcare organisations. Using, analysing and understanding Big Data requires a multidisciplinary approach, bringing together clinicians, IT engineers, statisticians, researchers, bio-informaticians, mathematicians, data analysts, etc.

5. LACK OF A SUSTAINABLE BUSINESS MODEL

Applying data analytics in the healthcare sector has a cost and therefore needs to deliver a "return in outcome". COCIR acknowledges that reimbursement mechanisms are complex and must be carefully considered. However, preventive diagnostic and therapeutic regimes that prove higher sensitivity and specificity leading to better health outcomes should be adequately rewarded. Current mechanisms are poorly suited to reward health IT solutions that contribute to reduce system waste, enhance quality care delivery or improve health outcomes.

6. TECHNOLOGICAL CHALLENGES

The ability and capacity of software and hardware to cope with the ever-increasing volume and variability of data has advanced remarkably. However, these challenges will remain important as data quantities continue to expand. Storage technologies, databases, communication, analytical engines and reasoning engines will need increased capacities and capabilities beyond the current state of the art.

3. COCIR RECOMMENDATIONS

The use of Big Data analytics in the health sector will certainly enable the delivery of value-based healthcare and evidence-based medicine. Public health interventions, as well as biomedical research and patient engagement, will also greatly benefit from the insights gained through data analytics. However, this will only happen if we pave the way to foster the uptake of Big Data in the health sector through strong political push. In particular, COCIR recommends undertaking the following actions:

ACTIONS BY HEALTH IT VENDORS AND HEALTHCARE PROVIDERS

1. Invest in digital security measures that safeguard health data confidentiality. This implies not only investing in physical and digital security assets but also establishing standard operating procedures and training staff to apply them.

ACTIONS BY MEMBER STATES

1. Build public confidence in information security and develop informational campaigns that make the public case for sharing health data and address public concerns around the misuse of health data. Awareness-raising programmes will have more impact if they are designed and run at local/regional level.
2. Encourage the adoption of international standards for clinical information by implementing policy levers, including legislative measures and financial incentives.
3. Enable EHR data extraction for secondary purposes, incentivising healthcare professionals to register clinical data in a standardised manner. Countries that lag behind should accelerate the deployment and use of EHRs systems.

ACTIONS BY THE EUROPEAN COMMISSION AND MEMBER STATES

1. Launch a European Health Data Quality initiative that complements and reinforces existing national initiatives for improving data reliability.
2. Support R&D and innovation programmes that investigate the opportunities of data analytics in the care setting.

ACTIONS BY EUROPEAN DATA PROTECTION AUTHORITIES

1. Ensure a harmonised implementation of the General Data Protection Regulation (GDPR) through a structured dialogue with industry and other stakeholders, that will enable secondary uses of data under the same conditions across the EU.
2. Develop and implement Health Data Governance Frameworks that ensure efficient, transparent and fair health data management practices and access to health data, while safeguarding patients' privacy.

JOINT ACTIONS BY HEALTH AUTHORITIES, PAYERS AND HEALTHCARE PROVIDERS

1. Create a new generation of ICT skilled workforce and clinical data scientists for the health sector.
2. Accelerate the development of value-based reimbursement models for health ICT and implement new approaches that reward the value of ICT solutions that reduce system waste, enhance quality care delivery or contribute to deliver better health outcomes.

REFERENCES

[1] Morley-Fletcher, E, "Big data healthcare: an overview of the challenges in data intensive healthcare," 2013. [Online]. Available: https://ec.europa.eu/digital-single-market/en/news/big-data-what-it-and-why-it-important. [Accessed 08 August 2016].

[2] COCIR, "eHealth Toolkit: Data Driven Healthcare," 2014.

[3] Sebastian Salas-Vegas, Adria Haimann, Elias Mossialos, "Big Data and Health Care: Challeges and Opportunities for Coordinated Policy Development in the EU," Health Systems & Reform, vol. 1:4, pp. 285-300, 2015.

[4] World Health Organisation, "From innovation to implementation: eHealth in the WHO European Region," 2016.

[5] OECD, "Data Driven Innovation for Growth and Well-Being," OECD, Paris, 2015.

[6] Manyika J, Chui M, Brown B, Bughin J, Dobbs R, Roxburgh AH, "Big data: the next frontier for innovation," 2011. [Online]. Available: http://www.mckinsey.com/insights/business_technology/. [Accessed 08 August 2016].

[7] Van Staa TP, Goldacre B, Buchan I, Smeeth L, "Big health data: the need to earn public trust", British Medical Journal, vol. 354, pp. 1-3, 2015.

[8] OECD, "Strengthening Health Information Infrastructure for Health Care Quality Governance: Good Practices, New Opportunities and Data Protection Challenges," OECD, Paris, 2013.