Digital Health Europe

Better Utilisation of Data Infrastructures to Support Secondary Uses of Health Data

White Paper

March 2020

Better Utilisation of Data Infrastructures

to Support Secondary Uses of Health Data

This white paper was developed for the European Commission by

Digital Health Europe

Lead Authors

Dipak Kalra, Zoi Kolitsi (EuroRec)
Oliver Zobell, Stefan Bujok (FJZ)
Veli Stroetmann (empirica)

Acknowledgments

DigitalHealthEurope has received funding from the European Union's, Horizon 2020 Research and Innovation programme under the Grant Agreement No. 826353.

DISCLAIMER

The information and views set out in this publication are those of the authors and do not necessarily reflect the official opinion of the European Commission. The Commission does not guarantee the accuracy of the data included in this study. Neither the Commission nor any person acting on the Commission's behalf may be held responsible for the use which may be made of the information contained therein.

1. Introduction

There is a rapidly growing interest in making greater use of routinely collected health data (Real World Data - RWD). This interest is emphasised in Priority 2 of the European Commission's Communication on the Digital Transformation of Health Care: "Better Data to Promote Research, Disease Prevention and Personalised Health and Care". Corresponding investments in this field are being made across the whole of Europe. What is to be pursued is the use of real-world data (RWD) by healthcare professionals, public authorities, and industry, to ensure that healthcare products, innovative technologies and therapies meet a patient's needs and lead to favourable health outcomes. Such outcomes include improvement of understanding of health and disease; better anticipation of disease outbreaks; faster diagnosis and development of more effective preventive measures and treatments.

The vision and main characteristics of the common European data spaces have been outlined in two EC Communications, both supported by the "Directive on open data and the re-use of public sector information"[1], the Artificial Intelligence (AI) Strategy[2] and other relevant EU legislation. Specifically, COM(2018)232 "Towards a common European data space"[3] introduced the vision of a common European Data Space across several sectors; the strategy for its implementation has been further defined in the recently released COM(2020) 66 "A European strategy for data"[4]. The latter annexes health specific policies and legislation underpinning the creation of the common European Health data space and in particular the role of the eHDSI and the application of the Electronic Health Record Exchange Format for scaling up cross-border exchange of health data for care and for Regulatory purposes.

European Commission (EC) planned action in this area foresees policy support through facilitating voluntary coordination of authorities and other stakeholders to share data and infrastructure for prevention and for personalised medicine research. It will also focus on fostering the further use of standards and the development of technical specifications for secure access and cross-border exchange of genomic and other health datasets in the EU (for research purposes and through the launch of pilot actions, so as to pool data and resources across the Union). Support of these priorities will lead to "increase the quality of data, standardise data collection, promote interoperability of European disease registries and advance the analysis of data using of high-performance computing and modelling"[1].

The intended use of this DHE White Paper is therefore to support the dialogue focusing on the needs of the health industrial innovation and research sectors, that would complement the currently defined scope of the EHDS. This will be pursued through a series of DHE workshops during the spring and summer of 2020 which should lead to a common understanding of what actions are needed by the EC and the key industry players, including of financial commitments, in order to support the implementation of a number of use cases in priority areas such as accelerated drug and device development, AI and personalised medicine genomics through re-using electronic health record data for research on a Europe-wide scale.

This **DigitalHealthEurope** (DHE) Report is the first of two documents that explore the evolving European and national landscapes of health data sharing for care and for research. The report aims to consolidate knowledge and lessons learned in this field, collect insights from key stakeholders, and explore the success factors for the acceptance and good use of health research data infrastructures. These issues are relevant to shaping the success of the recently proposed **European Health Data Space**.

1. Directive (EU) 2019/1024 on open data and the re-use of public sector information
2. White Paper on Artificial Intelligence: a European approach to excellence and trust
3. https://eur-lex.europa.eu/legal-content/en/TXT/?uri=CELEX:52018DC0232
4. https://ec.europa.eu/info/sites/info/files/communication-european-strategy-data-19feb2020_en.pdf

2. The Context and the Opportunity

Over the last decade, European policies and legislation, brought together with advances made by large-scale pilots, have encouraged the implementation of cross-border digital infrastructures, supporting the exchange of health data across borders in the EU for care purposes and for research.

There is sufficient knowledge on the role and the functioning of federated interrogation networks in supporting the re-use of real-world data, collected for health care purposes, in research and innovation. EU, regional and national approaches share common architectures, governance considerations, elements of infrastructure design, success factors and bottlenecks that need to be further addressed. Our current context creates the opportunity for building upon these experiences to accelerate progress fueled by a fast-developing health data economy, within a transparently governed, common European Health Data Space supporting EU innovators and researchers for the benefit of European citizens.

Cross border health data Infrastructures

Health data infrastructures are information technology (IT) infrastructures used for storage, access, and analysis of health data. They typically support research activities (e.g. at research-performing organisations such as pharmaceutical companies) or healthcare activities (e.g. at medical practices and hospitals). Whereas cost-optimisation has driven a major shift towards public or private cloud-based infrastructures, for security reasons many health data remain in traditional, in-house datacentres. This is especially true of data generated during healthcare delivery, in which handling of personally identifiable information (PII) is the norm.

The **eHealth Digital Service Infrastructure** (eHDSI)[5] is an open interoperability architecture that supports the cross-border exchange of Patient Summaries and ePrescriptions for the purpose of supporting unplanned and emergency care abroad.

The supported Cross Border Health Information Services (CBeHIS) is a closed network of government to government (G2G) services, available only to public entities (eHealth competence centres) that have the mandate to implement and operate nationally services enabling the sharing of a citizen's health data among authorised users. This network of national organisations has become "federated" by means of a governance mechanism encompassing both policy and operational co-ordination. It includes a legal agreement covering aspects not covered by EU legislation, common implementation standards, guidelines on all implementation aspects of the services, interoperability testing for technical readiness and compliance verification by means of formal audit, monitoring, and change management. The recent EC Recommendation for the definition of a **Common Data Exchange Format for Electronic Health Records** (EHRs)[6], under the Digital Single Market strategy[7], has introduced a vision for enlarging the scope of use cases that could support national as well as cross-border shared care, beyond emergency situations.

In parallel with the launch and operation of the eHDSI, the EC has made available an infrastructure for sharing data in support of the operation of the **European Reference Networks** (ERNs)[8], established also under the Cross-Border Healthcare Directive. ERNs are networks of centres of expertise and healthcare providers that enable the collaboration of clinicians and researchers who share expertise, knowledge and resources across the EU, under a clear governance structure, so as to improve access to diagnosis and treatment and provide high-quality healthcare for patients.

European Regulators, such as EMA, have been exploring the potential created by real world data (from electronic health records, registry data and claims data, pooled clinical trials data, reported suspected adverse drug reaction reports, and genomics, proteomics and metabolomics datasets) to complement clinical

5. eHealth DSI Operations Home: eHDSI Mission, Governance and Communities: https://ec.europa.eu/cefdigital/wiki/display/EHOPERATIONS/eHealth+DSI+Operations+Home, eHDSI Service Offering: https://ec.europa.eu/cefdigital/wiki/display/EHOPERATIONS/eHDSI+Service+Offering
6. European Commission Recommendation on a European Electronic Health Record exchange format (C(2019)800); https://ec.europa.eu/digital-single-market/en/news/recommendation-european-electronic-health-record-exchange-format
7. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52015DC0192
8. ERNs are part of the legal framework of Directive 2011/24/EU, the so-called Cross-Border Healthcare Directive

trials and improve the evidence upon which Regulators make decisions on the safety, efficacy and use of medicines and verify the validity of claims made by the industry. In their recent DARWIN HMA-EMA Joint Big Data Taskforce Phase II report: 'Evolving Data-Driven Regulation'[9] they have prioritized the **Business case for an EU platform for accessing and analysing healthcare data** (**D**ata **A**nalysis and **R**eal **W**orld **I**nterrogation **N**etwork, **DARWIN**).

> One set of goals to be pursued through the common EHDS and described in COM (2020) 66, is to "enable the exchange of electronic patient summaries and ePrescriptions between 22 Member States participating in the eHealth Digital Service Infrastructure (eHDSI) by 2022 and start cross border electronic exchanges through eHDSI of medical images, laboratory results and discharge reports and enhance the virtual consultation model and registries of European Reference Networks; support big data projects promoted by the network of regulators."

Federated Health data Infrastructures for Research and Innovation

Beyond supporting data exchange for cross border healthcare and regulatory purposes, the second priority of the DTHC Communication, "Better data to promote research, disease prevention and personalised health and care" focuses on certain key areas (including but not limited to rare, infectious, and complex diseases). It relates to data collection and utilisation for purposes other than the initial intended use. The federated infrastructure model is also used by EU initiatives including the European Open Science Cloud (EOSC)[10], ELIXIR supporting life sciences research, but also regional and national initiatives promoting the development of regional or national federated infrastructures. DHE has studied a number of representative initiatives- some Europe-wide, and others are being conducted nationally in France, Germany, and collaboratively in the Nordic countries. Some European research infrastructure projects funded under the IMI Joint Undertaking, such as EHDEN[11], are now in large-scale deployment.

→ The Nordic e-Infrastructure Collaboration NeIC–Tryggve, connects health data infrastructures across the **Nordic countries**[12]. The initiative was triggered by a shared aspiration to link **genomic data**, which are already available at substantial scale in the Nordic countries, with **phenotypic data** collected in each country (typically in the course of healthcare or public service activities e.g. EHRs or citizen registries);

→ The European Medical Informatics Framework (EMIF) developed an infrastructure, platform and tools to scale up federated access to **EHRs** and **longitudinal cohort study data**[13]. EMIF was sponsored through the EU's Innovative Medicines Initiative (IMI)[14], a public-private research and development partnership between the EC and the European Federation of Pharmaceutical Industries and Associations (EFPIA). The EMIF Platform provided an efficient integrated framework for the **large-scale re-use of health and life sciences data;**

→ The European Health Data & Evidence Network (EHDEN) is an IMI2[15] consortium, using many of the architectural results and tools developed in EMIF, with the mission to reduce the time to provide an answer in real-world health research. It will connect diverse data sources across Europe, with the ambition of providing federated access to data from over 100 million patient records. It will also grow **a community of ICT companies** with skills to map legacy data repositories to the OMOP data model.

→ The Electronic Health Records for Clinical Research (EHR4CR)[16] was one of the largest of the IMI public-private partnerships (PPPs) in this area. It developed a **robust and scalable platform** that utilised de-identified data from hospital EHR systems, in full compliance with the ethical, regulatory, and data protection policies and requirements of each participating country. The EHR4CR platform supported distributed querying to assist in **clinical trials feasibility assessment** and **patient recruitment**.

Registries are important enablers for clinical research. In the RD field, they have permitted the collection and pooling of patient data in a way that, by combining data on as many patients as possible at the regional, national, European or global levels, has increased knowledge on RD and further developed clinical research.

9. https://www.ema.europa.eu/en/news/ten-recommendations-unlock-potential-big-data-public-health-eu
10. A cloud for research data in Europe, promoted by the European Commission to provide all researchers, innovators, companies and citizens with seamless access to an open-by-default, efficient and cross-disciplinary environment for storing, accessing, reusing data, tools, publications and any EOSC Resource for research, innovation and educational purposes. https://ec.europa.eu/research/openscience/index.cfm?pg=open-science-cloud#
11. The European Health Data and Evidence Network. https://www.ehden.eu
12. https://neic.no/tryggve/

Better Utilisation of Data Infrastructures

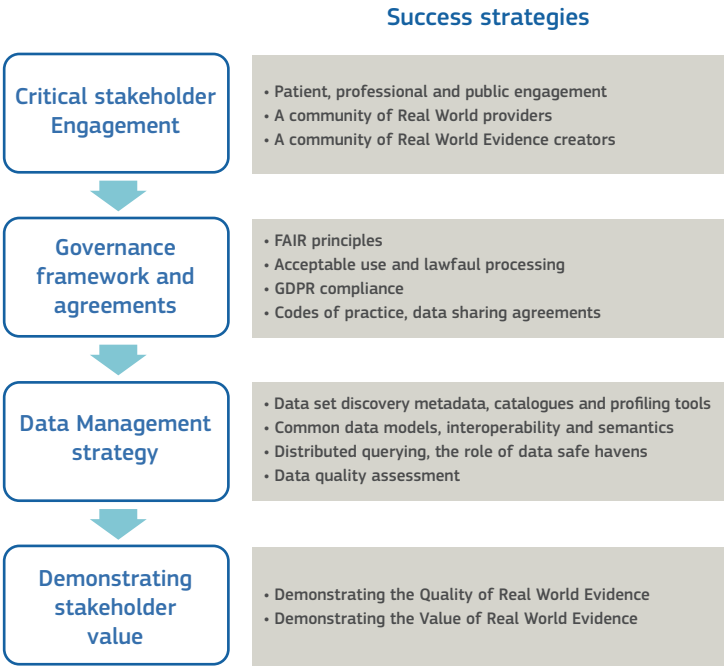
to Support Secondary Uses of Health Data

The European Platform on Rare Diseases Registration (EU RD Platform)[17] was launched in December 2013 by the EC's Joint Research Centre (JRC), in collaboration with DG SANTE to address the high degree of fragmentation of RD patient data held in hundreds of different registries across Europe. The platform aims to provide a single point of access to information on RD patient data with a large transnational coverage for all RD. Today, it has achieved to make RD patient data searchable, queriable, and findable across rare disease patient registries. →

It is based on the development of the European RD Registry Infrastructure (ERDRI). The Platform sets also EU-level standards for RD data collection and data exchange and provides training on the use of the tools and services offered by the platform.

Although much of the data held in registries is also collected in medical records there are no systematic efforts reported to automate the extraction of these data into the registries, thereby minimising effort duplication and registration errors.

A common feature of all initiatives investigated by DHE is the establishment of an overarching interoperability and governance layer that sets out common access rules and common interfaces for mapping diverse datasets to common data models, while the federated health data ecosystems that they support share common ingredients that are indispensable for their success.



Designing a successful big data infrastructure

A high level of maturity of federating research data and infrastructure has allowed successful EU level approaches to federating data sources and data governance models. This is also reflected in the increasing number of countries that are taking legislative initiatives and are implementing infrastructures to support data re-use projects and communities for health system, policy support research and for clinical research. There is a need for the EC to support the implementation of use cases of data re-use, that will furthermore leverage on common transversal governance, interoperability and data quality frameworks, sustainable business models and common infrastructures, collaboratively developed in high priority areas in the EU, such as:

- → Cross-border Surveillance of Antimicrobial Resistance (including its cross sectorial relevance to Agriculture)
- → Electronic Health Records to Electronic Data Capture for Rare Disease Registries
- → Reusing electronic health records to optimise and accelerate clinical trials
- → Collating and comparing health outcomes, in order to deliver more value-based care
- → Accelerating big health data research into neurodegenerative brain diseases
- → Achieving the necessary data scale to conduct accurate, trustworthy, AI research

Beyond care and research, data is an invaluable resource for boosting innovation and through it the growth of the economy of the EU. Two interlinked areas of EU policy are of particular relevance in boosting data-driven innovation: Artificial Intelligence, and the common European Data Space.

The EU is developing approaches to high performance computing, data analytics and artificial intelligence which can help design and test new healthcare products and provide faster diagnosis and better treatments. Succeeding in these endeavours depends on the availability of vast amounts of high-quality data and appropriate regulatory frameworks that will safeguard the rights of individuals and society as well as stimulate innovation.

In this context, the Commission has adopted a "Coordinated Plan on Artificial Intelligence"[18], and has recently launched the "White Paper on Artificial Intelligence: a European approach to excellence and trust"[19] which outlines the EU AI strategy for policy and legislative action. On the infrastructure side the EC has launched the AI4EU platform[20], which is connecting and opening up European

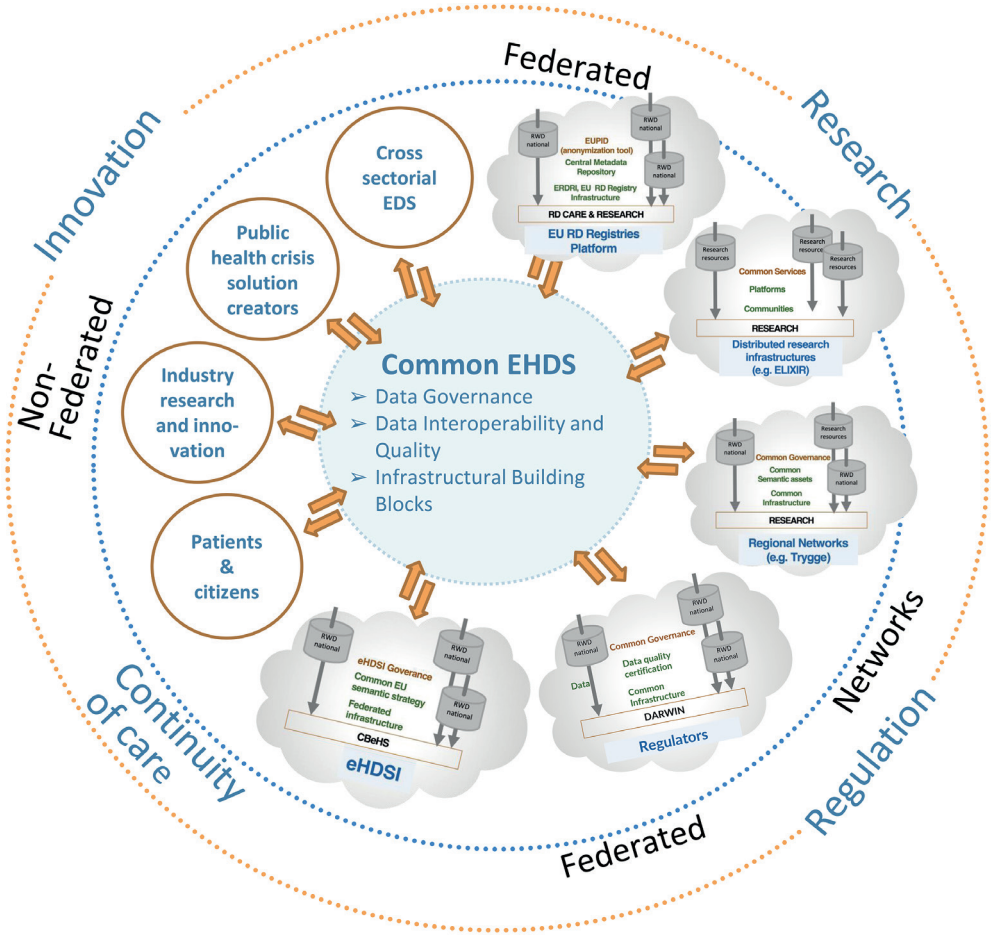
13. http://www.emif.eu
14. http://www.imi.europa.eu
15. IMI2 is the second phase (2014-2020) of the Innovative Medicines Initiative (IMI), the public-private partnership (PPP) between the European Union (represented by the European Commission) and the European pharmaceutical industry (represented by EFPIA, the European Federation of Pharmaceutical Industries and Associations). https://www.imi.europa.eu/
16. http://www.ehr4cr.eu/
17. https://eu-rd-platform.jrc.ec.europa.eu/

AI resources in the form of an AI-on-demand platform supporting the EU business sector and SMEs. The 'Open Data and re-use of PSI Directive', is expected to also stimulate the publishing of dynamic data, rationalise charges for the re-use of publicly held data, and enlarge the scope of the PSI Directive[21] (including to research data resulting from public funding promoting policies for open access to publicly funded research data, and facilitating the re-usability of research data that is already contained in open repositories).

In addition, the Open Data Directive requires the adoption by the EC (through implementing acts) of a list of high-value datasets that are to be provided free of charge. These datasets will have a high commercial potential and can speed up the emergence of value-added EU-wide information products. They will also serve as key data sources for the development of Artificial Intelligence and the Internet of Things (IoT). Although not currently in focus, such data sets are very relevant to healthcare innovation efforts.

18. https://ec.europa.eu/digital-single-market/en/news/coordinated-plan-artificial-intelligence
19. https://ec.europa.eu/info/files/white-paper-artificial-intelligence-european-approach-excellence-and-trust_en
20. https://cordis.europa.eu/project/id/825619
21. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32019L1024

3. Towards a European Health Data Space



In its Resolution on enabling the digital transformation of health and care in the Digital Single Market; empowering citizens and building a healthier society, adopted on 18 December 2019[22], the European Parliament (EP) also noted that "the creation of a European Health Data Space is included in the mission letter to the Commissioner for Health, with a view to promoting health data exchange and support research on new preventive strategies, as well as on treatments, medicines, medical devices and outcomes". The Resolution provides more than 50 recommendations and calls on the European Commission to include these in the proposal for a European Health Data Space.

22. European Parliament resolution of 18 December 2019 on enabling the digital transformation of health and care in the Digital Single Market; empowering citizens and building a healthier society (2019/2804(RSP)), http://www.europarl.europa.eu/doceo/document/TA-9-2019-0105_EN.html

Better Utilisation of Data Infrastructures

to Support Secondary Uses of Health Data

While an initial scope of the common European Health Data Space (EHDS) has been communicated in COM(2020)66, the Space as such is much more extended and encompasses the needs and expectations of the health research and innovation communities.

Our research so far has led us to an initial conceptualization of this EHDS, incorporating existing and emerging research infrastructures, usually in the form of federated interrogation networks, where health data are not transferred and pooled but are made temporarily accessible for remote analysis by researchers. This is known as "health data on demand": i.e. health data remain where they are, behind the firewall of the institution entrusted with protecting the privacy of the data subject, whereas the algorithms are sent to and analytical results sent back from the querying of the data. The transition to federated health data infrastructures is accelerated by EU initiatives to exploit datasets in a Findable, Accessible, Interoperable and Reusable (FAIR) manner.

Outside such networks, EU citizens contribute data through apps linked to wearable devices and fitness, nutrition and in general healthy lifestyle programs. This data has the potential to boost innovation, leveraging on big data and AI resources and computing power. Sets of Public Sector information including through other sectors such as climate change, food and agriculture are also relevant to the EHDS. These circles are currently scarcely illustrated and will need to be defined through a series of workshops engaging industry, innovators and citizens. The discussions should provide the general direction as to what are potential areas that should be supported in the immediate future which leverage on existing high value data sets such as EHRs, genomic data etc.

In its H2020 Research and Innovation programme and the Industrial Leadership programme the EC has been supporting businesses to develop and launch value added products and services in the healthcare sector. Despite the financial commitments made over the years to support R&I projects, when it comes to tackling the most important bottlenecks for scaling up infrastructures for research and innovation, three potential areas that need further investigation are identified:

- → The challenges of health data governance,
- → The need to re-focus interoperability efforts on high value data sets, and
- → The requirement to articulate clear business models for the reuse of health data.

The respective three activity tiers need to be addressed cohesively and through wide multi-stakeholder engagement. DHE discussions with stakeholders have almost universally, identified the main, common, challenge/bottleneck to reusing health data for research was perceived to be legal: obtaining consent from the citizen or patient that her/his health data can be re-used for another purpose or agreeing that an alternative GDPR legal basis may be used.

Major bottlenecks for scaling up health data sharing

Bottlenecks to be addressed cohesively and through multi-stakeholder engagement, in three areas

1. **Health data governance**
   - → Policy environment: EU and MS level's
   - → Building and assuring trust across stakeholders
   - → Agreeing on the safeguards to achieve that trust
2. **High re-use value data sets**
   - → Co-defining these across stakeholders and re-use cases
   - → Prioritising interoperability standards adoption
   - → Prioritising data quality efforts
3. **Business models linked to use of re-use**
   - → To engage stakeholder support
   - → For sustainability

Further complicating this challenge is a general lack of trust among citizens with regard to the reuses of health data (which has much to do with their general lack of understanding of how data can usefully be reused). Feeding information on the impact of re-used health data back to citizens (the data subjects) could effectively promote trust. In the end, it would also generate a better, general consciousness about the benefits of sharing health data.

Data Interoperability and data quality aspects remain also a major challenge, in part because of the vast scope of health data and its semantics. Attention is now being focused on tackling sub-domains of health data rather than its entirety. The sub-domains of priority are those for which interoperability can deliver the greatest value across stakeholders (i.e. interoperability is of high value).

Better Utilisation of Data Infrastructures

to Support Secondary Uses of Health Data

Before specifying the interoperability standards, clinical models, and term lists needed for focused data sets, it is important that the data sets themselves are defined or ratified through a transparent, evidence-based, multi-stakeholder process. Such a process would ensure that the standards adoption is well targeted, before the standards are implemented at scale. This therefore requires an investment in multi-stakeholder consensus-building, preferably at a European level, prior to large-scale ICT investments.

Exploitation of health data for research purposes takes place in a range of businesses and institutes, from SMEs to (national) data providers. There is a growing body of European SMEs whose business model is to collect data from citizens, or to collate data about citizens (through their consent) from data sources such as their health care provider records, fitness information, or even banking information. These companies hope to sustain and expand their business models around the commercial value of that data, for example, by selling access to the data to the pharma industry. This market is at too early a stage of maturity to judge whether it will be scalable and sustainable, and what its success factors would be. The added value of these data providers is the curation, cleaning, data harmonisation, and anonymisation efforts they undertake to make the data suitable for research use, e.g. by the pharma industry.

There is a risk that the hype surrounding the use of big health data for research will cause a reduction in the interest shown in this field, unless a portfolio of viable business use cases and models is developed and widely shared. It is also important that these business models embody FAIR principles so that data can be widely used, rather than treated as a secret resource through protective practices around data access. Transparency about these business models, on the part of industry, will help to reduce public anxiety about excessive profiteering from health data, and help to highlight where the societal benefits could arise from health data use.

Business models are needed to scale at least three categories of investment relating to health data:

- → Attracting pump-priming investments for implementation and early operation
- → Developing a medium and long-term business model for sustainability, including return on investment
- → Investing in the actual data e.g. in data quality improvements, data provision resources

In addition to the financial aspects of these business models, it is important that societal value is demonstrated from the uses made of the data, so as to justify the public portion of any (research) infrastructure funding. This will notably apply to the European Health Data Space, which is in a formative stage of development at the time of writing this report. Rather than initially focusing explorations on what data should be and could be put into the health data space, it is first important to identify the candidate users and use cases that ideally this space might initially support. Only then will it make sense to explore the possible data feeds that should be attracted into the space.

There is an outstanding need to further investigate with representatives of the EU industry sector (ICT, AI, Pharma and Medical Devices)

- → The potential of RWD in EHRs, to drive industrial innovation in the health care sector
- → The high priority use cases and projects of high relevance to the competitiveness of European industry
- → Explore common principles for governance, data quality and Interoperability as well as common infrastructural elements needed
- → Work with the relevant research and innovation players, and especially industry, to define gaps and opportunities to be catered for by the EHDS.

DigitalHealthEurope.eu has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 826353.

Editor & Design: empirica Gesellschaft für Kommunikations- und Technologieforschung mbH
Oxfordstr. 2 · 53111 Bonn, Germany
https://www.empirica.com