



theLisboncouncil
think tank for the 21st century

**Update of the European Data
Market
SMART 2016/0063**

**Story 6 – The Secondary Use of
Health Data and Data-driven
Innovation in the European
Healthcare Industry**

21st January 2020

| | |
|-------------------|--|
| Author(s) | Silvia Piai, Adriana Allocato, Giulia Besana (IDC) |
| Deliverable | D 3.6 Quarterly Stories (Story 6) |
| Date of delivery | 21 st January 2020 |
| Version | 3.0 |
| Addressee officer | Katalin IMREI Policy Officer European Commission - DG CONNECT Unit G1 – Data Policy & Innovation EUFO 1/178, L-2557 Luxembourg/Gasperich katalin.imrei@ec.europa.eu |
| Contract ref. | N. 30-CE-0835309/00-96 |

Table of contents

| | |
|---|-----------|
| TABLE OF CONTENTS | 3 |
| TABLE OF FIGURES | 4 |
| LIST OF TABLES | 4 |
| EXECUTIVE SUMMARY | 5 |
| 1. INTRODUCTION | 8 |
| 1.1 <i>MAIN OBJECTIVES AND SCOPE</i> | 8 |
| 1.2 <i>METHODOLOGY AND STRUCTURE</i> | 8 |
| 2. THE DIGITAL PATHWAY TO HEALTHCARE DATA-DRIVEN INNOVATION ... 10 | |
| 2.1 <i>HARNESSING THE VALUE OF HEALTH CARE DATA</i> | 10 |
| 2.2 <i>A DATA GOVERNANCE STRATEGY AT THE FOUNDATION OF A DATA DRIVEN APPROACH IN HEALTHCARE</i> | 14 |
| 2.2.1 DATA GOVERNANCE FOR SECONDARY USE OF HEALTH DATA..... | 14 |
| 2.2.2 DATA DONATION IN EUROPE: STATE OF THE ART AND KEY CHALLENGES | 15 |
| 2.3 <i>THE DIGITAL TECH FOR SECONDARY USE OF HEALTH DATA</i> | 19 |
| DATA LAKES: THEIR DATA QUALITY SECURITY AND GOVERNANCE IMPLICATIONS | 19 |
| AI, BIG DATA AND ADVANCED ANALYTICS INVESTMENT TRENDS | 20 |
| 3. THE CASE STUDIES - WHAT WE FOUND SO FAR | 26 |
| 3.1 <i>SECONDARY USE OF DATA IN FINLAND</i> | 26 |
| ISAACUS – THE DIGITAL HEALTH HUB PROJECT | 26 |
| 3.2 <i>HEALTH DATA HUB IN FRANCE</i> | 28 |
| BACKGROUND: FRENCH TRANSFORMING ITS HEALTHCARE SYSTEM..... | 28 |
| HEALTH DATA HUB: A SINGLE PLATFORM FOR MANY PLAYERS..... | 29 |
| 3.3 <i>THE DEVELOPMENT OF DATA POLICIES IN PORTUGAL</i> | 31 |
| TELE-DERMATOLOGY SCREENING..... | 32 |
| IDENTIFYING AND PREDICTING EMERGENCY ADMISSIONS | 32 |
| 3.4 <i>ARIA’S HEALTH DATA WAREHOUSE AND BUSINESS INTELLIGENCE COMPETENCY CENTER (ITALY)</i> 32 | |
| BACKGROUND | 32 |
| "predictive BDA" PROOF OF CONCEPT FOR PREDICTING THE INSURGENCE OF CARDIOVASCULAR CHRONIC CONDITIONS | 33 |
| "predictive BDA": PILOT ON DETERMINING THE “CLINICAL NEXT BEST ACTION” IN ALZHEIMER PATIENTS’ JOURNEY | 33 |
| predictive BDA: “GEOGRAPHIC MACHINE LEARNING” FOR RESPIRATORY CHRONIC CONDITIONS..... | 34 |
| THE CREATION OF A DATA INFORMATION HUB FOR HEALTHCARE RESEARCH | 34 |
| LESSONS LEARNED FROM ARIA’S DW-BICC’S PROJECTS..... | 34 |
| 4. PRELIMINARY POLICY CONCLUSIONS | 36 |
| IDC RESEARCH REFERENCES | 39 |

Table of Figures

| | |
|--|----|
| Figure 1 European healthcare providers' approach to DX efforts | 11 |
| Figure 2 European healthcare providers approach to establishing a DX roadmap | 12 |
| Figure 3 Top five digital capability needs | 12 |
| Figure 4 DX budget priorities in 2018..... | 13 |
| Figure 5 Top three investment priorities in data analytics..... | 13 |
| Figure 6 Artificial Intelligence adoption and investment plans in the European healthcare industry | 20 |
| Figure 7 Top three use case for AI among European healthcare providers | 21 |
| Figure 8 Big Data Analytics adoption and investment plans in the European healthcare industry | 22 |
| Figure 9 Top three use cases for BDA among European healthcare providers | 23 |
| Figure 10 Health data collaboration platform adoption and investment plans in European healthcare industry | 24 |
| Figure 11 Patient Portal adoption and investment plans in the European healthcare industry .. | 25 |

List of Tables

| | |
|---|----|
| Table 1 Pilot projects of the French HDH Initiative | 30 |
| Table 2: Horizontal Overview of the 4 Case Studies..... | 36 |

Executive summary

A growing number of European Healthcare Systems are embarking on long-term reforms to improve outcomes and foster innovation with the ultimate goal of benefiting patients, while, at the same time, ensuring long-term sustainability of healthcare services provision. An un-locked potential is clearly there as recent evidence has highlighted that the majority of healthcare providers (59%) has not adopted a Digital Transformation roadmap yet and only the 6% has established a unique roadmap for Digital Transformation and general business strategy¹ (see Figure 1 and 2 of this report).

This unexplored potential of the use of Big Data and Analytics (BDA) in healthcare is eliciting a new wave of interest in data-driven value creation, which, in the medium to long run, will enable to reward performance rather than just volume. According to the IDC DX Sentiment survey 2018², over 60% of the **European healthcare providers reported that developing data management** is a priority. Indeed, being able to analyze and use data for process automation and decision support, in a granular, accurate, safe and context-relevant way, is key to long-term competitiveness and sustainability.

AI is still in its infancy and only the 30% of European healthcare providers are already using/testing or have immediate plans for the technology, adding to another 23% evaluating AI use cases (Figure 6)³. The top three BDA use cases that European healthcare providers are working on (please refer to Figure 6 for the extensive list) are **Clinical decision support (16%)**, **Illness progression (15%)** and **Patient engagement (14%)**⁴. Patient engagement is a key priority for almost 40% of European healthcare providers, particularly among countries that are experimenting with value-based reimbursement models structured around patient experience and outcomes, while the most significant AI use cases that European healthcare providers are working on (Figure 7) are: **Personalization of clinical pathway (10%)** **Clinical decision support (9%)** and **Patient risk (9%)**⁵.

As the ability to use, share and manage data becomes a critical enabler of digital transformation, the need to define an appropriate framework to safely manage data as key assets becomes an imperative. According to IDC European Vertical Market survey, 2018, over 70% of European healthcare providers adopt or plan to invest in Healthcare data collaboration platform. This is particularly true for hospitals that have to comply with new safety and data protection regulations and align their business models with the new requirements of integrated care. The ultimate goal is to create secure, standardized, interoperable and accessible pools of data that can be used for enhancing decisions and automate processes to improve care quality, patient experience, organizational efficiency and innovation. However, the value of patient data comes from the value of insights and innovation that can be unlocked from the combination of multiple pieces of information.

Further research by IDC reveals that 39% of European healthcare providers⁶ consider improving patient experience a key priority in the next 12 months. To better understand how European health systems are approaching and implementing this concept with the use of Big Data and Analytics (BDA) and value-based care (VBHC), the present research features a series of real-life case studies among some of the most significant health systems in Europe and illustrates some of the most widespread frameworks and best practices for health data collection, sharing, access and usage in Europe's healthcare sector today and how these frameworks and practices can be

¹ Source: IDC DX Executive Sentiment Survey, May 2018 (n=66)

² Source: IDC DX Executive Sentiment Survey, May 2018 (n=66)

³ The IDC European Vertical Markets Survey is an annual landmark study of IT solutions, investment priorities, and emerging technologies. In the 2018-2019 version, the sample covers over 77% of the European economy (in terms of GDP of the 40 countries). Respondents are distributed across Western Europe (UK, France, Germany, Italy, Spain, Netherlands, the Nordics) and Central & East Europe (Russia, Poland, and Czech Republic). The survey was conducted in the native language of each country, using either telephone interviewing (CATI) or web interviewing (CAWI) systems. Eligible respondents are individuals primarily involved in IT and/or business decisions at their companies and ranked director level or above. Results are analyzed by vertical market and company size and represent a basis for a series of demand-side reports published by IDC.

⁴ Data retrieved from IDC European Vertical Markets Survey, 2018–2019 (n= 290 [WE = 232, CEE = 58])

⁵ ibidem

⁶ Source: IDC European vertical Market survey, 2018

fostered to further benefit the wider community of healthcare providers and users. This story therefore presents four different case studies:

- The Secondary Use of Data in Finland - in the health and social welfare sector there has been an extensive work between the public and private sectors to promote the secondary use of health and well-being data, which has led the creation of a new ecosystem through a national development project, leading to the development of a groundbreaking new legislation and the establishment of a permit authority for secondary use of health data. Interviews conducted with Jaana Sinipuro, Hannu Hämäläinen e Antti Kivelä (SITRA).
- The Health Data Hub in France - In 2018, president Macron launched this initiative to establish a nationwide data platform. The project is today strongly pushed forward from the French Health Act "Ma Santé 2022", approved last July, and including the creation of the "Espace Numérique de Santé" (ENS – e-health personal space) with the aim to establish a more efficient and patient oriented healthcare system by leveraging the power of data and artificial intelligence. The project goal is to enrich and enhance the National Health Data System (SNDS - Système National des Données de Santé) by including the wider French heritage of health data in one place, for open use by researchers, healthcare professionals, care institutions, start-ups, insurers, etc.
- The development of Data Policies in Portugal - In August 2019, the Portugal Health Ministry has made available the strategic document "From big Data to smart data: putting data to work for the public's health", that outlines the vision, key areas and principles for secondary use of data, advanced analytics and artificial intelligence to improve Portuguese population's health. There are several initiatives and pilot projects, that are testing the capabilities of AI and providing evidence to support the development of new data management and governance policy.
- ARIA's Health Data Warehouse and Business Intelligence Competency Center (Italy) – In July 2019, Regione Lombardia (Lombardy Region, Italy), has established a new Regional company called ARIA (Agency for Innovation and Procurement) from the merge of ARCA (regional agency for procurement) and Lombardia Informatica (the regional in-house digital company). Within its mandate, ARIA has the specific aim to enhance the value of all regional health data assets. Interviews conducted with Giuseppe Preziosi (Aria)

The case studies presented in this research highlight several benefits obtained by the organizations adopting BDA technologies, coupled with Artificial Intelligence and Machine Learning technologies (AI/ML):

- The easy and convenient access to intelligent solutions for clinicians and patients offers more opportunities to advance decision making and enhance clinical process efficiency at the point of care. Portugal skin cancer screening solution is an example of how technology supports a clinical collaborative framework and enables the integration of information to serve population health management,
- More advanced predictive capabilities, allowing greater control over disease-specific variables impacting health outcomes, as well as costs and resources utilization associated with care. This approach enables to more efficiently target population segments at risk of developing chronic and long-term conditions by putting in place initiatives aimed at promoting health and preventing or delaying the development of risk factors. Predictive BDA by ARIA is an example of development of a predictive model able to effectively target cardiovascular conditions and offer an accurate estimate of the number of future cases in a specific geographical area.

However, these benefits come at a price and a series of challenges remain to be addressed if the full potential of BDA in the healthcare sector is to be reaped.

- The establishment of a regulatory framework for the use of and access to data. This involves multiple players in the public and private healthcare arena to establish partnerships and create a collaborative environment. It requires all stakeholders to agree on the value of data as a shared asset, and to actively promote initiatives where common standards and a one-stop-shop approach to data access are brought forward. This is the case of Portugal, where the Health Ministry authority is seeking to establish a new healthcare ecosystem based on a data-driven approach towards the delivery of healthcare services across the nation.

- The collection, processing, storing and access of complex data coming from different structured (e.g. National health records) and unstructured (e.g. wearables) sources, along with semantic, geographical and time complexities. This is the case of ARIA and its collection of over 10 years of healthcare data stored in different locations, as well as Sitra's project working at establishing a common framework and developing metadata descriptions. Additionally, data collected require intelligent solutions, capabilities and skills to extract value from data and provide actionable areas for the deployment of information (i.e. population risk stratification, clinical decision support, personalization of clinical pathways, etc.).
- Maintaining trust and ensuring security. The high sensitiveness of healthcare data requires an attentive approach to identifying and enforcing regulatory frameworks and solutions that ensure information is treated in compliance with regional and country-specific policies. In addition, authorities need to establish a high level of public trust for the ethical and secure use of healthcare and social population data for the public good. In this example, Finnish citizens are informed about how their data are used for secondary purposes out of primary care. The adoption of GDPR strengthens and unifies data protection for individuals within the EU, regulating how data integration happen safely. It gives individuals key control over the usage, processing and transfer of their personal data held by healthcare organizations. The transition to a VBHC model, in which care plans should be personalized and stakeholders should integrate their activities, must be underpinned by consistent information management governance that enables patient data integration across providers, processes and IT systems. GDPR is expected to provide a patient-centric ecosystem.

1. Introduction

Today's healthcare leaders in Europe and elsewhere need to quickly identify areas of synergy with adjacent sectors to create new value for patients. At the same time, they need to better understand how innovation will impact their operations and processes and translate in the more specific objectives of their organizations - whether these are faster and more successful clinical trials, or an optimized usage of available resources, or more effective public health initiatives, sustainable management of chronic disease population, more performant medical devices, drugs and treatments, etc.

Several use cases of data-driven innovation in the health sector are presented in this study; from population health management (in particular population risk stratification and healthcare services planning), public health (e.g. the use of healthcare data to address antimicrobial resistance), clinical research (for example the adoption of RWE - real world evidence - to accelerate clinical trials and improve our understanding of human biology at individual level) to the issues of data governance (including data security and protection, technical and organizational pre-conditions necessary for the secondary use of health data and the management of granting access to such data) and interoperability in the healthcare industry. This story investigates the European context providing the background of such use cases, as well as a series of real-life case studies in Europe. A careful examination of these case studies could indeed unveil some of the most significant interactions between healthcare providers and policy makers, suggesting the synergies that need to be developed to obtain highly innovative and pioneering solutions in future healthcare reforms, starting from the ongoing best practices.

1.1 Main objectives and scope

The aim of this quali-quantitative story is to investigate how the power of data and of data-related technologies is triggering the process of digital transformation in Europe, thus radically changing the healthcare industry. The story illustrates emerging trends such as health data donation practices or health data cooperative models (such as HealthBank and Midata), with the specific aim to investigate further the voluntary data donation phenomenon and the reasons behind the difficulties that it currently encounters to be scaled up in Europe. Real-life case studies and references have been selected so to ensure a geographical balance (when possible) and, more importantly, with the aim to intercept some of the most representative examples of data-driven transformation at play in today's Europe's healthcare sector.

1.2 Methodology and structure

To better understand how European health systems are approaching and implementing the use of Big Data Analytics (BDA) and value-based care (VBHC), we have conducted extensive desk research across a multitude of publicly available sources and IDC existing research material. Through literature review, IDC sources and a select number of in-depth interviews with key experts and healthcare industry stakeholders, the story illustrates current frameworks and best practices for health data collection, sharing, access and usage in the broader healthcare sector and how these data benefit the wider community of healthcare providers and users.

The interviews with relevant stakeholders and the research effort led to the realization of distinct case-studies across different European Member States. In particular:

- The Secondary Use of Data in Finland - in the health and social welfare sector there has been an extensive work between the public and private sectors to promote the secondary use of health and well-being data, which has led the creation of a new ecosystem through a national development project, leading to the development of a groundbreaking new legislation and the establishment of a permit authority for secondary use of health data. Interviews conducted with Jaana Sinipuro, Hannu Hämäläinen e Antti Kivelä (SITRA).

- The Health Data Hub in France - In 2018, president Macron launched this initiative called. This initiative is today strongly pushed forward from the new Health French Act, approved last July, to establish a more efficient and patient oriented healthcare system. The project aims to enrich and enhance the National Health Data System (NSDS) by including the wider French heritage of health data in one place
- The development of Data Policies in Portugal - In August 2019, the Portugal Health Ministry has made available the strategic document "From big Data to smart data: putting data to work for the public's health", that outlines the vision, key areas and principles for secondary use of data, advanced analytics and artificial intelligence to improve Portuguese population's health. There are several initiatives and pilot projects, that are testing the capabilities of AI and providing evidence to support the development of new data management and governance policy.
- ARIA's Health Data Warehouse and Business Intelligence Competency Center (Italy) – In July 2019, Regione Lombardia (Lombardy Region, Italy), has established a new Regional company called ARIA (Agency for Innovation and Procurement) from the merge of ARCA (regional agency for procurement) and Lombardia Informatica (the regional in-house digital company). Within its mandate, ARIA has the specific aim to enhance the value of all regional health data assets.

The current document is structured along three main sections.

- The first section (Chapter 2) describes the unexploited potential of the use of a data-driven innovation in health care. Moreover, it explores the Data Governance strategy introducing key aspects and challenges on Data Donation and on the secondary Use of Data.
- The second section (Chapter 3) is devoted to an overview of the real-life case studies that formed the bulk of the primary and secondary research underpinning this story.
- The final section (Chapter 4 and 5) presents the common themes and main lessons learnt from the case studies and provides a few conclusive remarks with possible policy directions for the years to come.

2. The digital pathway to healthcare data-driven innovation

2.1 *Harnessing the value of health care data*

A growing number of European Healthcare Systems is embarking on long-term reforms to improve outcomes and foster innovation to benefit patients, and, at the same time, to ensure long-term sustainability of healthcare services provision. Through a restructuring of healthcare systems based on a combination of personalized medicine and integrated care pathways, these new models are designed to ease the pressure on health services demand. Such pressures have indeed been made more urgent by rising costs, new regulations governing patient rights and quality of care provision, coupled with shrinking financial resources and skills to be allocated to healthcare delivery. In recent years a significant number of reforms have been focused on implementing new healthcare business models inspired by the "value-based healthcare" paradigm (VBHC). Value-based healthcare's central tenet is that the overarching principle in redesigning healthcare delivery systems must be "value for patients"⁷. Value is measured by analyzing multiple factors that affect the effectiveness of healthcare services, including clinical outcomes, patient safety, patient experience (outcomes that matter to patients) and the overall costs encountered across the system to deliver those outcomes and their sustainability. Outcomes and costs are measured for every patient, along entire care cycles. VBHC, therefore, demands to recognize and understand where, and how, "patient value" is created, and how it can be maximized across the different healthcare processes, by:

- rewarding performance rather than volume, aligning payments and incentives with value creation;
- promoting better collaboration between healthcare providers and organizing care around a patient medical condition or of a specific segment of the population, matching the correct provider, treatment, and setting along patient journeys;
- incorporating the views and preferences of patients, to ensure the best possible experience.

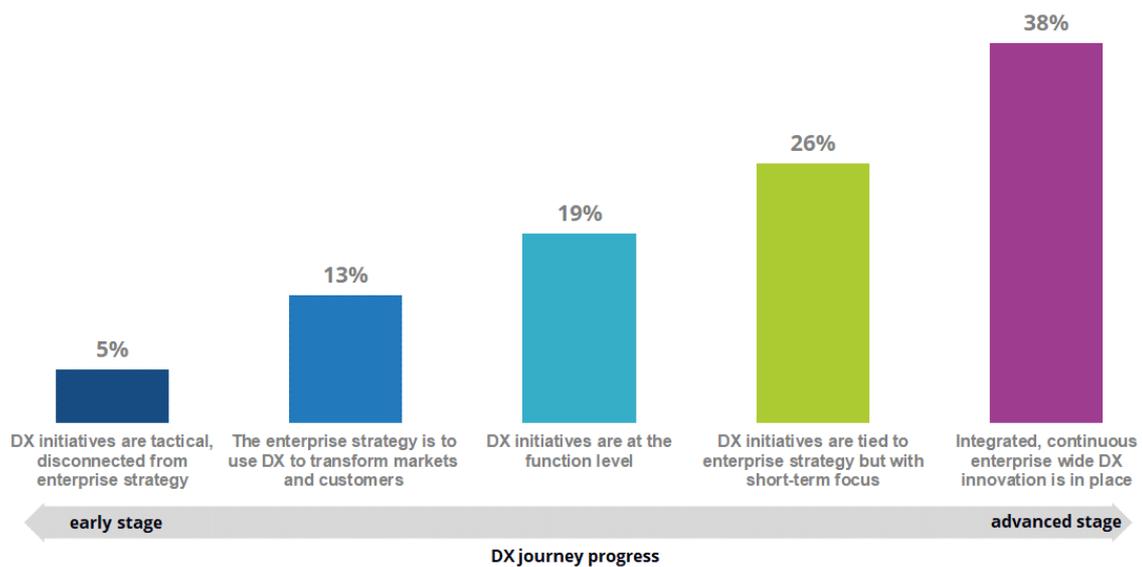
This concept has changed the perspective of policy makers when addressing the question of efficiency and effectiveness policies in healthcare. While traditional business models are still predominant in Europe, there is a stronger focus in aligning commissioning and provision of healthcare services to a variety of patient value indicators in the latest healthcare reforms. Therefore, the most innovative organizations taking part to healthcare systems (healthcare providers, payers, medical researchers and, as well, life science organizations) are working toward scenarios based on stronger convergence of the different value chains and on patient outcome-driven closed-loop collaborative models. Today's healthcare leaders need to quickly identify areas of synergy with adjacent sectors to create new value for patients. At the same time, they need to understand better how innovation will impact their processes and will translate in the more specific objectives of their organizations, whether it's faster and more successful clinical trials, an optimized usage of available resources, more effective public health initiatives, sustainable management of chronic disease population, more performant medical devices, drugs and treatments, etc.

To succeed in these fast pacing scenarios, healthcare organizations need to develop new capabilities to rapidly produce the evidences required to assess value. This implies developing the technical, regulatory, cultural and organizational setups to leverage the value of fast growing and varied data produced by healthcare systems. The need to create the technical and operational conditions to better leverage data is driving fundamental changes in the sector's use of digital technologies. Digital should be seen no longer just a tool to support a system of records. Digital, today, is an enabler for using data to get insights that feed into a system of engagement

⁷ Value Based Healthcare has been put on the healthcare policy makers' agenda thanks to the work of Porter, Kaplan and Teisberg at Harvard Business School and Medical School. <https://www.isc.hbs.edu/health-care/value-based-health-care/Pages/default.aspx>

within the single organization and across its ecosystem. A more **granular and automated level of access to relevant data is required and it can be achieved only through the digital transformation** of the entire healthcare value chain. To enable new services, transform the patient experience, optimize operating models, collaborate and orchestrate internal and external resources, many healthcare organizations seek to better leverage the value of health information by putting digital transformation at the top of their agenda. However, most of European healthcare providers are still mired in a digital deadlock. Only the 38% of European healthcare providers have claimed their Digital Transformation (DX) initiatives to be an integral part of enterprise wide strategies⁸. In particular, the majority of healthcare providers (59%) has not adopted a digital transformation roadmap yet and only the 6% has established a unique roadmap for digital transformation and general business strategy⁹.

Figure 1 European healthcare providers' approach to DX efforts



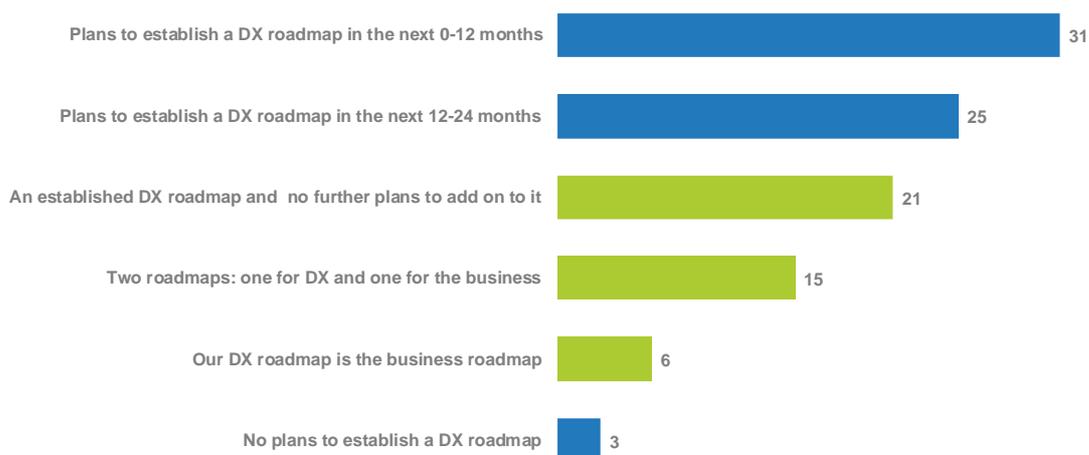
Source: IDC DX Executive Sentiment Survey, 2018, May 2018 (European Healthcare Providers = 66)

Digital transformation and data-driven innovation may be tactical, siloed and, even when goals are aligned, they can often be on short-term outcomes. Truly digitally determined healthcare organizations that see digital transformation as a foundational component of their business strategy are few, thus preventing the business from exploiting the potential of technology in making process and decisions more evidence based.

⁸ The IDC DX Executive Sentiment Survey, 2018 is a landmark study of worldwide organizations' DX (Digital Transformation) strategy road maps, DX budgets, IT investments, strategic priorities, business initiatives, and emerging technologies across the globe. The present figures refer to the survey data from healthcare providers in Europe (N=32). Eligible respondents were high-level executives and managers involved in IT decisions made at their respective organizations or leaders or part of the leadership team responsible for digital transformation. Survey results-at global level- were weighted based on country-specific GDPs.

⁹ Source: IDC DX Executive Sentiment Survey, May 2018 (n=66)

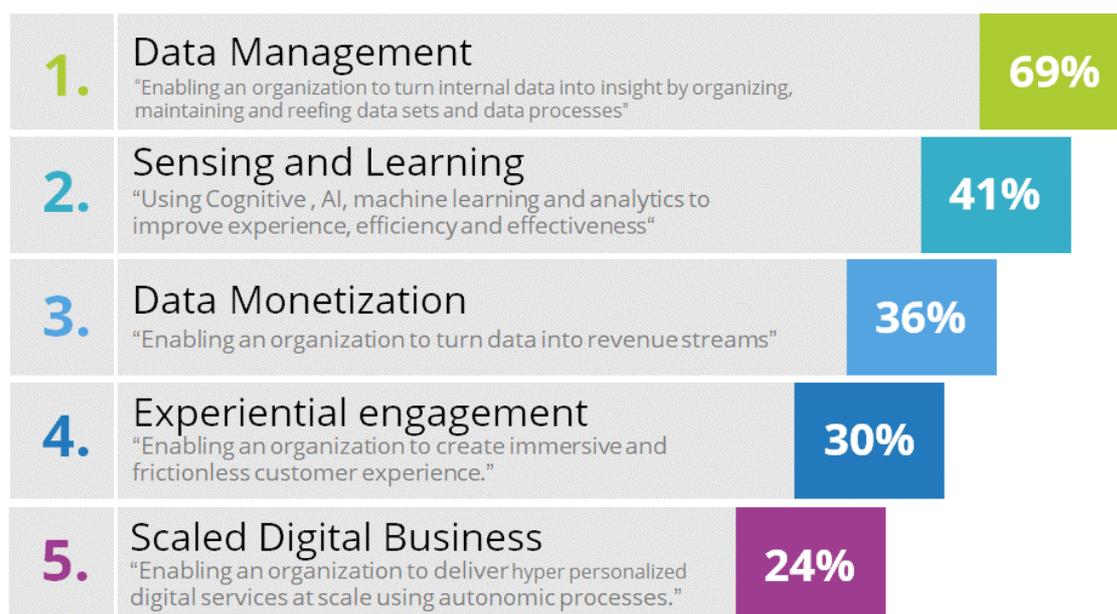
Figure 2 European healthcare providers approach to establishing a DX roadmap



Source: IDC DX Executive Sentiment Survey, 2018, May 2018 (n = 66)

The transformation of healthcare is a massive undertaking, so it is not surprising that organizations are still digitally immature. **The industry is highly fragmented and highly regulated**. Systems are disconnected and data is treated in isolation, thus limiting collaboration. But technology itself is not the major problem. Only the 20% of European healthcare providers consider the lack of right infrastructure as a barrier to digital transformation. One of the top issues holding back digital transformation plans is the limited digital expertise (45%), followed by the organizational structure (40%)¹⁰. Therefore, to pick up the pace of their digital transformation and realize care delivery and cost-efficiency benefits, healthcare providers need to **build new digital capacity**. According to the IDC DX Sentiment survey 2018¹¹, over 60% of the **European healthcare providers reported that developing data management** is a priority, followed by the closely related ability to leverage data through AI and advanced analytics technologies for sensing and learning.

Figure 3 Top five digital capability needs



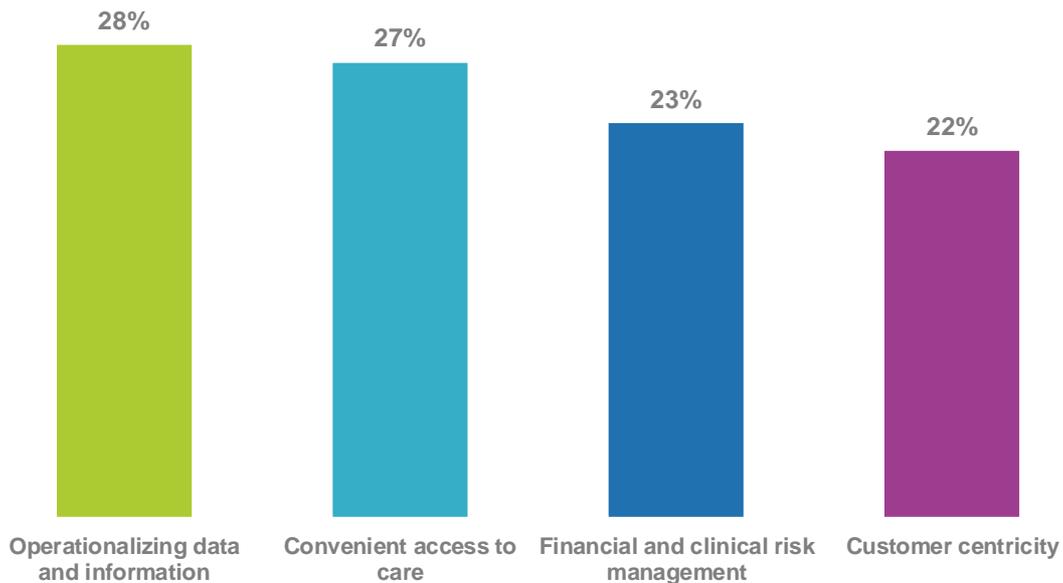
Source: IDC DX Executive Sentiment Survey, 2018, May 2018 (n = 66)

¹⁰ Source: IDC DX Executive Sentiment Survey, May 2018 (n=66)

¹¹ Source: IDC DX Executive Sentiment Survey, May 2018 (n=66)

In order to develop these digital capability needs; European healthcare providers have already planned to allocate the majority of their DX budget to turning data into value (27%). They recognize the value of analyzing multiple types of health data (clinical, administrative, public health, social care, and clinical research) from multiple sources (hospital information systems, patient reported outcomes, medical, and wearable devices) as data insights are at the heart of the transition from a fee-for-service to a value-based care model. Assembling data to identify high-risk patients, enable predictions, and deploy evidence-based medicine help organizations become more efficient in their day-to-day businesses.

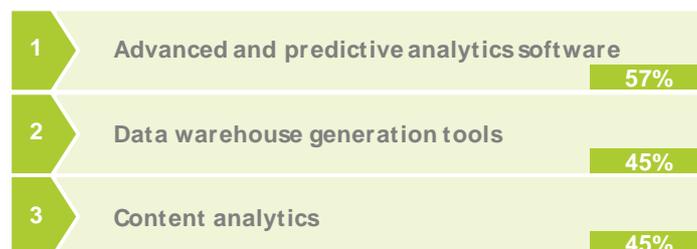
Figure 4 DX budget priorities in 2018



Source: IDC DX Executive Sentiment Survey, 2018, May 2018 (n = 66)

While data analytics represents one of the areas of major interest on the road to digitalization, the investments into predictive analytics software and data warehouse generation tools are the leading applications to operationalize data and information within the European healthcare ecosystem. Artificial Intelligence (AI) applications are used to operationalize the insights coming from Big Data to optimize resources allocation, ensure compliance with care quality standards and better engage patients in their personal health journeys. For healthcare information systems, this means that Big Data and AI will have a major impact on key applications such as EHR, order management, prescription support, and population management platforms.

Figure 5 Top three investment priorities in data analytics



Source: IDC DX Executive Sentiment Survey, 2018, May 2018 (n = 66)

The data above strongly suggest that the digital transformation journey happens through a data-driven approach which is aligned to a value-based approach. **Unlocking insights from patient's**

record would increase speed and improve the quality of diagnosis and treatment while taking waste out of the system and lowering costs through evidence of cost effectiveness and outcomes to inform value-based payments. Additional advantages, such as clinical trial design for target populations and reduced time to market for new therapies, would even benefit to pharmaceutical companies.

Therefore, to leverage a data driven culture, healthcare providers are progressively adopting:

- An information governance policies and best practices so that data is formally, consistently and securely managed across the organization. This helps to connect all data, irrespective of source and format, including health-related data beyond the clinical environment.
- Innovative technologies, which are open to be integrated with existing systems, scalable to accommodate increasing datasets and real-time to maximize the impact on the patients.

2.2 A data governance strategy at the foundation of a data driven approach in healthcare

2.2.1 Data Governance for secondary use of health data

Being able to **analyze and use data for process automation and decision support**, in a granular, accurate, safe and context-relevant way, is key to long-term competitiveness and sustainability. As the ability to use, share and manage data becomes a critical enabler of digital transformation, the need to define an appropriate framework to safely manage data as key assets becomes an imperative. The ultimate goal is to create secure, standardized, interoperable and accessible pools of data that can be used for enhancing decisions and automate processes to improve care quality, patient experience, organizational efficiency and innovation. However, the value of patient data comes from the value of insights and innovation that can be unlocked from the combination of multiple pieces of information. That is how data becomes actionable information for decision-making. The understanding of the links and the trends within data is already being used to study health both at the population level and the personalized, individual one. For example, in population health management, healthcare organizations analyze vast array of data collected from multiple sources, to segment population based on specific health needs, to study the causes and evolution patterns of diseases affecting specific segments, and to define most appropriate and economically effective policies, services and treatments. In population health, healthcare organizations rely heavily on big data derived from both their own health IT systems and from the other stakeholders within the healthcare ecosystem. The growing availability, standardization and integration of data across connected systems support diseases prevention, the improvement of health outcomes and a more personalized healthcare for better patients' outcomes. The access to patient demographics, vital signs, laboratory results, progress notes, diagnoses, procedures codes, admission, discharge and transfer (ADT) data, environment and socioeconomic data (such as employment and education), offer an extremely rich playground for analytics thanks to their high level of detail and relevance. Increasingly, health authorities and health research institutions are mandating interoperability requirements and developing different data repositories (registries) and biobanks that serve as platforms for future research and analysis.

As traditional areas of research and statistics expand, secondary use of healthcare data (being able to reuse existing data for different purposes, other than the one for which it was originally collected) is becoming strategic. Secondary use of personal health information enables a range of organizations (healthcare providers, researchers and policy makers), to research and innovate while improving patients' outcomes and experiences, enhancing system's productivity and promoting efficient use of resources. By bringing real-world evidence in research and decision-making process, secondary use of health data can drive multiple benefits:

- enhancing health care experiences for individuals and ensure continuity of care;
- streamlining healthcare service delivery;
- expanding knowledge about diseases and appropriate treatments;
- reducing the "time to market" for innovative products and services;

- strengthening understanding about effectiveness and efficiency of health care system;
- mitigating healthcare costs' growth, with more targeted standardized evidence-based; treatment being provided to the patients;
- informing health system policies.

Yet, complex ethical, political, technical, and social issues surround the secondary use of health data. These issues play increasingly critical and complex roles given current public and private sector activities not only expanding health data volume, but also enhancing access to data. While GDPR has started providing a clearer framework for defining common principles and rules for health data processing, for secondary use, the definition of coherent policies, specific legislation standards and good practices is not homogeneous across European countries. In Finland, the regulation for secondary use of data is already into force. On 7 November 2019, the German House of Parliament passed new rules supporting digital innovation in Germany, the Digital Supply Act, extending the use of social and healthcare data for research purposes to public authorities, research institutes and university hospitals. To make national healthcare data more accessible to research and public health planning, the government is planning for a National Trust Center and Research Data Center that will anonymize and analyze data, while ensuring full compliance with GDPR and data security regulations.

The evolving regulatory framework raises new challenges for the healthcare industry players with respect to managing the data in a way that comply with existing privacy, data security and intellectual property safeguards requirements. Healthcare and life sciences organizations will need to invest on comprehensive data management and information governance policies to address many of these challenges within their organizations, and across their partner ecosystems, to break down both internal and external data silos and process bottlenecks. While an information governance policy addresses data processing, analysis and protection, a data management framework defines the rules regarding data format and their consistency with the technical aspects of mapping and combining different data sources. Setting a rigorous data management and information governance strategy is essential in healthcare, especially for secondary uses, to determine the level of detail that may be released while still protecting patients' information rights. This transform the way in which administrative and clinical data are collected, changing approaches to consent management, data sharing practices as well as interoperability and IT infrastructures and solutions, as they might represent as a potential limitation and may compromise the reliability of any resulting model for secondary use.

2.2.2 Data Donation in Europe: state of the art and key challenges

Bietz, Patrick and Bloss define data donation as an act where "people voluntarily contribute their own personal data that was generated for a different purpose, to a collective dataset."¹² Healthcare data voluntary donation for research and healthcare systems policy making purposes in Europe is still a developing concept that is receiving increasing attention from policy makers, research institutions and players in the life science and healthcare ecosystem. Data donation enables the creation of data sets and archives that can serve multiple purposes, such as for example, scientific research in identifying more precise treatment for a range of conditions tailored to the genetic makeup of an individual, as well as helping to improve the prediction and prevention of disease for a wider and more complex population. This definition entails that healthcare data are voluntary donated by a data subject that agrees to share their personal information and opt in to support scientific research studies. Data donation research does not only include citizen science (i.e. the collection and analysis of data relating to the natural world by members of the general public, typically as part of a collaborative project with professional scientists), but it embraces the broader spectrum of research studies performed by players such as academics institutions as well as private corporations, given that data subjects provide their consent for the use.

According to Bonney and colleagues¹³ there are three levels of public participation in research:

- Contributory projects: data subjects primarily contribute data;

¹² Bietz, M., Patrick, K., & Bloss, C. (2019). Data Donation as a Model for Citizen Science Health Research. *Citizen Science: Theory and Practice*, 4(1), 6. DOI: <http://doi.org/10.5334/cstp.178>

¹³ Bonney, R, Ballard, H, Jordan, R, McCallie, E, Phillips, T, Shirk, J and Wilderman, CC. (2009). Public Participation in Scientific Research: Defining the Field and Assessing Its Potential for Informal Science Education. A CAISE Inquiry Group Report. Available at: <https://eric.ed.gov/?id=ED519688>

- Collaborative projects: data subjects may assist with research design, analysis, or dissemination;
- Co-created projects: data subjects and scientists collaboratively work together.

Collected and donated data can be generated from a broad range of digital technologies such as wearable devices, traces of online activity such as social media posts, or a patient's medical tests or electronic health records that are voluntarily donated. Citizens can also apply a secondary use of their previously collected data (e.g. through, as an example, wearables technology) to then contribute to a collective research project that they voluntarily choose, as they deem it important for themselves or for others. There are several drivers of individual data donation, the most common lying in perceived benefit for the individual donating data (e.g. who may be able to access some sort of data analysis, predictions of health status and other health-related results) or the perceived public good benefits.

Despite the existence of country-specific ethical frameworks and regulations for tissue and organ donation, there is still no framework addressing medical and healthcare-related data donation at scale. Krutzinna and colleagues have made some efforts in researching the different implications of posthumous medical data donation to propose an ethical framework that should enable widespread adoption while ensuring the respect of individuals' rights¹⁴.

Examples of efforts towards the creation of healthcare data banks out of voluntary donation to support and foster scientific research and innovation (mostly managed within citizen science frameworks) are to be found across Europe. The following initiatives are encouraging and experimental examples to promote standards for storage, control, access and use of data to unlock the potential that the data donation holds for healthcare.

- **UK Biobank project.** Started in 2006, the biobank study is investigating the respective contributions of genetic predisposition and environmental exposure to the development of disease. The aim is to improve the prevention, diagnosis and treatment of a wide range of serious and life-threatening illnesses. It recruited 500,000 volunteer participants aged between 40-69 years in 2006-2010 from across the country and provide health information, which does not identify them, to approved researchers in the UK and overseas, from academia to university. In September this year, a £200M public-private collaboration was announced, with the aim to sequence the genomes of the more than 500,000 people that have taken part in the project. Funding for the sequencing project comes from the government's research and innovation agency, UK Research and Innovation (UKRI) with £50M through the Industrial Strategy Challenge Fund, £50M from Wellcome and a further £100M in total from Amgen, AstraZeneca, GlaxoSmithKline (GSK) and Johnson & Johnson. The latest update reports that, between March 2012 and January 2019, a total of 1,463 access applications for studies using UK Biobank data has been submitted, with the greatest focus being on genetics and genotyping research.¹⁵ To date (October 2019), 989 papers have been published in leading health journals-providing new evidence about a wide range of diseases that moves research forward. For example, in 2018 researchers have found that people exposed to air pollution well within UK guidelines have changes in the structure of the heart, similar to those seen in the early stages of heart failure¹⁶.
- **100,000 Genomics England project.** This NHS' flagship project sequences 100,000 whole genomes from patients with a rare disease and their families and common cancers. For rare disease patients, Genomics England aims to provide genomic diagnoses for the first time. So far, there are some examples of how the project has provided new diagnostic tools for rare diseases in UK directly to patients who took part in the initiative: pioneering researchers at the University of Manchester found out a DNA defect in a 16-years old, providing a complete diagnostic frame to a rare pulmonary condition. Similarly, doctors at the Great Ormond Street Hospital in London have identified a rare genetic mutation on

¹⁴ Krutzinna, J., Taddeo, M., & Floridi, L. (2019). An Ethical Code for Posthumous Medical Data Donation. In *The Ethics of Medical Data Donation* (pp. 181-195). Springer, Cham.

Krutzinna, J., Taddeo, M., & Floridi, L. (2019). Enabling Posthumous Medical Data Donation: A Plea for the Ethical Utilisation of Personal Health Data. In *The Ethics of Medical Data Donation* (pp. 163-180). Springer, Cham.

¹⁵ <http://www.ukbiobank.ac.uk/wp-content/uploads/2019/02/research-jan-2019-big.jpg>

¹⁶ Association between ambient air pollution and cardiac morpho-functional phenotypes. Insights from the UK Biobank population Imaging study. <https://www.ahajournals.org/doi/full/10.1161/CIRCULATIONAHA.118.034856>

a single gene for two 4-years old, who now have a precise diagnosis. For patients with cancer, there is the potential for more personalized treatments and in time. The project will lead to the development of new or more effective diagnostics and treatments and, to this goal, it is expected that – by 2030 – 123,768,000 cancer patients will have their genomes sequenced. According to the timeline and objectives set by the Health and Social Care Secretary, Matt Hancock, the vision is to sequence 5 million genomes by 2023, with new, broader guidance on inclusion criteria for young patients and adults. To date, the Project website reports that actionable findings have been found for 1 in 5 rare disease patients, and around 50% of cancer cases contain the potential for a therapy or a clinical trial.

- **Scotland's Health Research Register (SHARE).** It offers an example of voluntary population involvement in scientific research. The SHARE initiative consists in establishing a register of citizens interested in taking part in health research, who voluntarily give their consent to include their health records into the register and be contacted in case their profile is suitable for a study. 273.140 people have registered so far. In the same framework, the GoSHARE initiative (Genetics of the Scottish Health Research Register) asks participants consent to store and spare blood samples for genetics studies. To date, the SHARE initiative counts 12 projects kicked off across Scotland to investigate various diseases and health conditions, ranging from asthma to diabetes, Alzheimer's disease, obesity and epilepsy. As an example, SHARE is participating in the EHPAD (European Prevention of Alzheimer's Dementia) consortium, a European interdisciplinary research initiative that aims to improve the understanding of the early stages of Alzheimer's disease and deliver new preventative treatments. To date, they have just released a baseline data for the first 500 research participants, representing the first formal public data release from the EPAD project for use by academic researchers and institutions worldwide.
- **Health Data Cooperatives.** Health Data Cooperatives are non-profit entities that enable citizens to securely store, manage and control access to their personal data while also offering the opportunity to contribute with their health information to academic-led studies. Cooperatives, as opposed to other data entities, act as fiduciaries of their account holder's data and have as core principles a cooperative approach, citizen-centeredness, not-for-profit structure, where only data subjects have access to their personal data through the platform (not even IT administrators or cooperative managers) and the cooperatives negotiate access to data by industrial partners (e.g. pharma and data companies). Citizens joining these initiatives are active contributors in the digital society and tend to value the benefits of data sharing for public health as well as for their own health. Some Cooperatives offer citizens the opportunity to access data services and mobile apps to better manage their chronic condition. However, incentive schemes in this framework are deliberately not based on financial rewards, as this would hinder the mission of cooperatives and the intrinsic motivation of data subjects to join. An example is the launch of the project "Ally Science", supported by the Swiss MIDATA platform, that allows citizens' control over their own data. Ally Science is a smartphone app launched in 2018 by The Bern University of Applied Sciences BFH and the University Hospital Zurich (USZ) aimed at contributing to the largest study on pollen allergy in the country to date. The project has already reached more than 8,000 participants contributing to improve early warning systems and therapies for pollen allergy sufferers in the future. The data collected via the app are stored in personal accounts on the secure MIDATA platform developed by ETH Zurich and BFH. The app available in all four national languages as well as in English, allows users to document their experience. They also receive pollen forecasts and an exclusive presentation of the current development of allergy symptoms in the various regions. A second example is a Dutch data cooperative (Holland Health Data Coöperatie, HHDC) established by the nonprofit MD/OG in collaboration with the City of Rotterdam, Medical Delta, the IJsselwijs Foundation and TNO Netherlands Organization for Applied Scientific Research, which endorses the same principles as MIDATA. In Spain, Salus Coop is a citizens-led cooperative of health data aiming to legitimize citizens' rights to control their own health records while facilitating data sharing to accelerate research innovation in healthcare.
- **The i-Share cohort project** aims to evaluate important health aspects among 30.000 university students (voluntarily enrolling and donating information) over the course of 10 years. i-Share has been initiated in 2013 by the Universities of Bordeaux and Versailles

Saint-Quentin (France) and aims to create a significant base of longitudinal data on young adults' health. The novel aspect of i-Share resides in an innovative modular structure that allows several targeted sub-studies (in the key research areas as migraine, mental health, infections and vaccinations, risk behaviors addiction and accidents) in which detailed information can be collected and interventions implemented. iShare has the following objectives:

- Public health objectives: enabling research for the evaluation of the frequency and consequences of diseases affecting young adults. Additional goals are to establish a platform that will allow the development of intervention trials and to test strategies for sampling, prevention, and health coverage.
- Biomedical objectives: to test specific scientific hypotheses about the physiopathological or psychopathological mechanisms of these diseases. This part relies on interdisciplinary collaborations (epidemiology, human & social sciences, molecular biology, neurosciences, and others) and specific biological measurements (cerebral Magnetic Resonance Imaging - MRI, biochemistry, genetics, vascular physiology, etc.)
- An update of November 2019 reports that the number of recruited students is higher than 20300, with 2000 MRI and 2000 blood samples already collected. An update of November 2019 reports that the number of recruited students is higher than 20300, with 2.000 MRI and 2000 blood samples already collected.
- **Sitra's IHAN project** aims to build the foundation for a fair and functioning data economy¹⁷ by creating a method for data exchange and to help Finland in supporting the definition of European level rules and guidelines for ethical use of data. The objective is to create the legal, technological, cultural and organizational premises for using personal data, without compromising on privacy and citizens' rights. In this framework data sharing based on consent and portability is the key and within the IHAN project (started in late 2018 and running until the end of 2021). The IHAN projects build on the health and social care data secondary use legislation entered into force on 1 May 2019 in Finland (see section 3.1) and aims to create new concept and to test solutions to enrich existing data sources with consent-based broader sets of data. Sitra is working on a series of pilot projects¹⁸, that starting from real business needs will create new services and solutions by using individuals' personal data with their permission. Some of these proof of concepts projects is within the healthcare sector and the most relevant one within the realm of data donation/ data sharing include
 - The Patient-Centric Real-World Data project which is testing the feasibility and the opportunity for pharmacies to act, with the patient's consent and using strong identification, as trusted collectors of information that flows between patients, data controllers and the pharmaceutical industry. In this pilot project, multiple sclerosis patients have been invited to provide information about their disease using an electronic form. The objective for this project is to allow combination of this real-world information and the clinical data found in national health registries to yield more detailed material that will serves as a basis for creating new and improved care services and treatments. Projects' results and finding are not yet publicly available. Key project partners include Finnish pharmaceutical wholesaler Oriola and its pharmaceutical services branch Farenta, Vastuu Group (previously known as Suomen Tilaajavastuu), cybersecurity services company Nixu, and and pharmacy chain Kronas Apotek.
 - Self-collected healthcare and wellbeing data to Kanta service (the national electronic health record in Finland) The main objective is to ensure that patients are able to share personal health data they generate, in a way where they are in full control. At the end of the project patient should be able to control how they share their data with healthcare organizations and under which terms it can be used there. They also get to choose when, how, and by whom their data is being used for research and other purposes. The project will end in 2021, and the key projects participants are: Sensotrend (a Finnish startup developing automated diabetes diary that gathers data from products like insulin pumps, blood glucose meters and meal app, wearables, etc.), Kela-the national social

¹⁷ <https://www.sitra.fi/en/topics/fair-data-economy/>

¹⁸ <https://www.sitra.fi/en/projects/ihan-pilot-projects/#what-is-it-about>

2.3 The digital tech for secondary use of health data

Health systems are looking for solutions to support the transition to value-based care and mine insights from myriad data sources. These include data warehouses and data lakes¹⁹ to make better use of data internally; cloud computing to provide the computing capabilities for analyzing and managing large data sets; AI and advanced analytics solutions to make better more accurate, faster analysis and support decisions; and healthcare data-sharing platforms, patient portals and API (application programming interface) based systems to facilitate the data workflow to and from devices, patients and other stakeholders of the health ecosystem; and the data security solutions that will support safe and compliant exchange and analysis of data.

The growing need to create value out of the increasing volumes of healthcare data and its fragmented nature across different sources has pushed European National Healthcare Systems, providers, organizations and other players in the industry to start investing in particular in data lakes, to enable then the use of AI , big data and advanced analytic

Data lakes: their data quality security and governance implications

The concept of data lakes is attractive to the healthcare sector, as it allows to retain a variety of information in its rawest form with no hierarchy or organization among individual pieces of information, allowing streams of data to gather in a single environment from different sources. As opposed to data warehouses, data lakes can serve current and future purposes without the need to extract, transform, and load data in a conforming, pre-defined way. Being highly agile, flexible and configurable solutions, data lakes are better able to answer complex questions, such as population health queries, by leveraging the full range of data available. From structured data (e.g. electronic health records) to semi-structured and unstructured data (e.g. coming from mobile apps, wearables, internet of things and connected devices), they allow to uncover variables and metrics that support predictive and prescriptive analysis with greater agility. The complexity and potentially countless scopes of data lakes in healthcare make them a suitable environment for data and clinical researchers, who can mine data and analyze statistical relevance between genome markers, patient information, treatment and outcomes. Advanced analytics and BI allow to drive large-scale initiatives focused on population health, risk management, personalized patients' pathways, as well as on efficiency, resource utilization and performance.

However, the agility of data lakes' environment requires a maintenance and a level of metadata, terminology and ontology management to provide researchers with an environment where they can easily experiment and learn from the data. Moreover, skepticism and challenges in adopting data lakes solutions are grounded in data quality, governance, security and privacy. Ensuring that the data lake is populated with high-quality data is key to drive analytics able to bring real, actionable value. With highly sensitive patients and population data available in one place, the establishment of a solid security framework including network isolation, data protection and document-level security is paramount to ensure that patients data are protected, and vulnerability addressed. In the current growing ecosystem and network of players that are increasingly adopting open source approaches to facilitate research and boost scientific advancement, data governance is escalating in complexity and needs to adapt to the requirements of a changing and ever-connected environment. Poorly anonymized and pseudonymized data in a data lake can put data privacy at risk and result in breaking GDPR while also providing means to directly identify people by their name, address, medical appointments and image. By using basic combination

¹⁹ A data lake is not a direct replacement for a data warehouse. They are supplemental data storage that serve different use cases with some overlap. While a data warehouse is a repository for structured, filtered data that has already been processed for a specific purpose, a data lake is a storage repository for large amount of structured, semi-structured and unstructured data the purpose for which is not defined yet. In the data lake, all data is kept irrespective of the source and its structure and it is transformed when it is ready to be used.

techniques that use different sources of information contained in a data lake and in external sources, it is also possible to re-identify individuals from pseudonymized data.

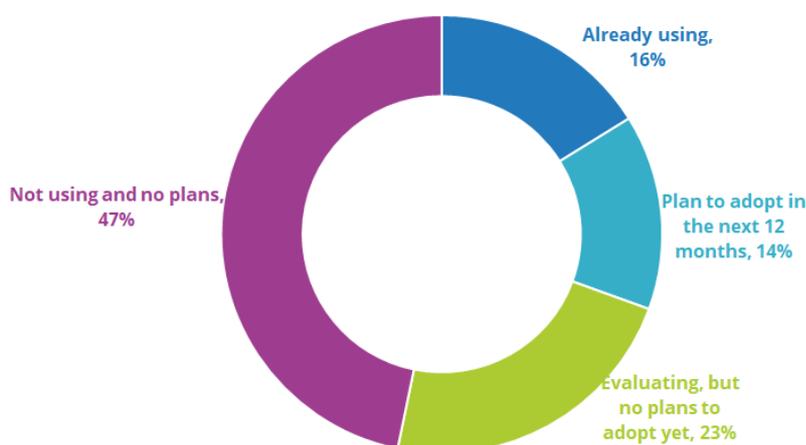
AI, Big Data and advanced analytics investment trends

Despite the current hype, the full potential of analytics, machine learning and other advanced technologies are not yet entirely understood or sufficiently leveraged. The quality of data inputted determines the reliability of outputted information. Advanced data mining, such as artificial intelligence (AI) and natural language processing, help reveal patterns and trends useful to healthcare providers and other stakeholders, to improve diagnoses, treatments and operational effectiveness. Many data analytics tools currently used focus on descriptive analytics, which provide insights into the past. Leveraging AI allows the analytics to move toward prescriptive analytics which make recommendations to advise on possible outcomes.

According to the IDC European Vertical markets Survey²⁰, 2018, AI is still in its infancy, the 30% of European healthcare providers are already using/testing or have immediate plans for the technology, adding to another 23% evaluating AI use cases (Figure 6). The last two years have seen a rise in the interest toward artificial intelligence, also driven by the development of new digital national plans. The European Commission has in fact requested member states to have AI plans in place by mid-2019. While the level of maturity of the strategies differ from country to country, these national plans have started to provide more defined frameworks and resources for AI implementation by healthcare organizations.

The definition of application potential and national budget allocation for AI and Big data Analytics (BDA) related projects in healthcare has boosted research plans, academic programs and capability building along with the definition of national policies and codes of conduct for the use of AI.

Figure 6 Artificial Intelligence adoption and investment plans in the European healthcare industry



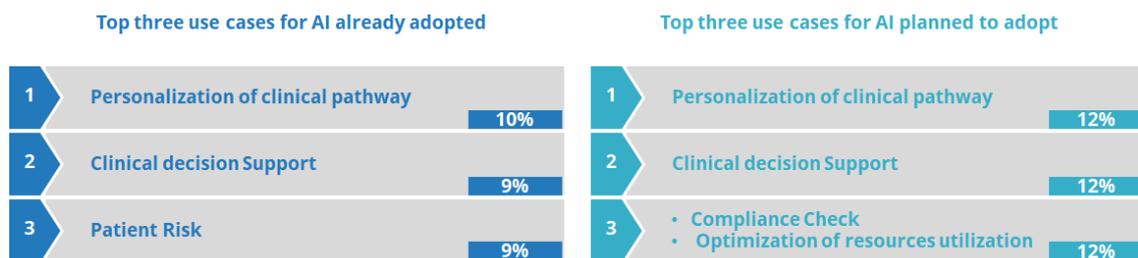
Source: IDC European Vertical Markets Survey, 2018–2019 (n= 290 [WE = 232, CEE = 58])

The most significant AI use cases that European healthcare providers are working on (Figure 7) are:

²⁰ The IDC European Vertical Markets Survey is an annual landmark study of IT solutions, investment priorities, and emerging technologies. In the 2018-2019 version, the sample covers over 77% of the European economy (in terms of GDP of the 40 countries). Respondents are distributed across Western Europe (UK, France, Germany, Italy, Spain, Netherlands, the Nordics) and Central & East Europe (Russia, Poland, and Czech Republic). The survey was conducted in the native language of each country, using either telephone interviewing (CATI) or web interviewing (CAWI) systems. Eligible respondents are individuals primarily involved in IT and/or business decisions at their companies and ranked director level or above. Results are analyzed by vertical market and company size and represent a basis for a series of demand-side reports published by IDC.

- **Personalization of clinical pathway.** As PHM is transforming the traditional hospital focused approach into a value care to deliver more efficient outcome using fewer resources, AI and machine learning algorithms can derive actionable insights from the untapped datasets essential for population health programs. AI can, for example, elaborate real-time data to pinpoint specific demographics where health issues exists and target them precisely with ad-hoc treatment program.
- **Clinical decision support.** A machine learning system can provide a high level of clinical accuracy and a coverage of a broad range of conditions. Symptoms can be entered via natural language and can be used to drive diagnoses and the level of care direction.
- **Patient risk.** AI can predict the future of patient's health with better accuracy, as the risk of contracting specific diseases. AI system can predict the outcomes of hospital visits to prevent readmissions and shorten the amount of time patients are kept in hospitals.
- **Compliance check.** The healthcare industry is highly regulated but maintaining compliance within evolving patchworks of national and regional regulations can be strain on providers' limited resources. The adoption of AI in this area optimize administrative procedures and frees experienced caregivers from routine tasks.
- **Optimization of resource utilization.** Machine learning and AI have the potential to provide the front line with the real time wisdom to improve the speed and the quality of the hundreds of decisions they make each day in order to improve the flow of patients through the various clinical services involved in delivering appropriate care.

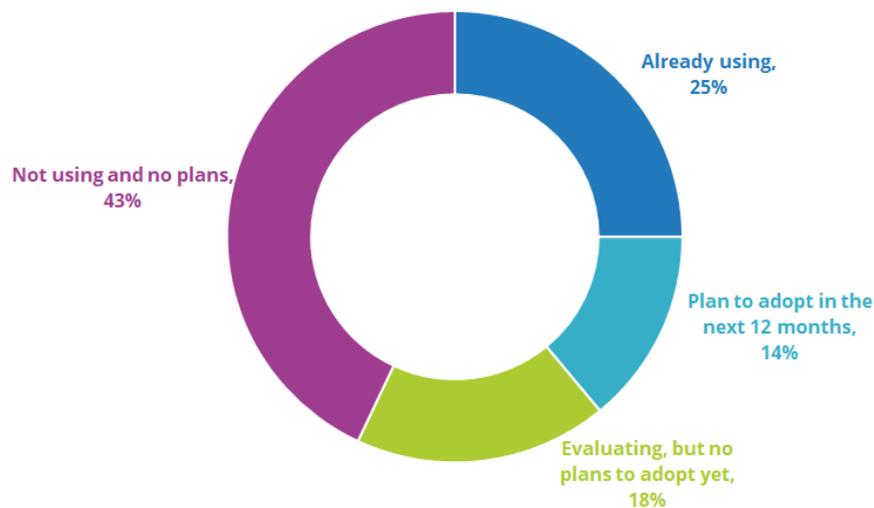
Figure 7 Top three use case for AI among European healthcare providers



Source: IDC European Vertical Markets Survey, 2018–2019 (n= 290 [WE = 232, CEE = 58])

European healthcare providers' adoption and investment plans for BDA vary depending on the readiness of the market in terms of data availability, integration, security as well as by the support of national policies. However, roughly half of healthcare providers use BDA to pull together data from diverse sources to diagnose and treat individual patients (Figure 8).

Figure 8 Big Data Analytics adoption and investment plans in the European healthcare industry

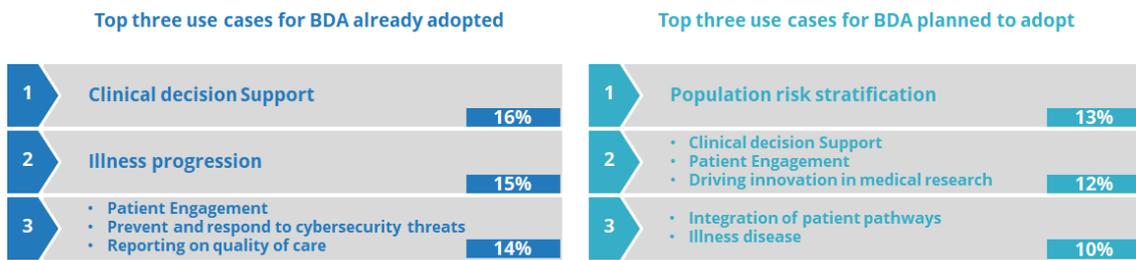


Source: IDC European Vertical Markets Survey, 2018–2019 (n= 290 [WE = 232, CEE = 58])

The most significant BDA use cases that European healthcare providers are working on (Figure 6) are:

- **Clinical decision support.** It's the most common area for BDA adoption in healthcare. BDA is the foundation for clinical decision support as it enhances decision-making in the clinical workflow.
- **Illness progression.** European countries are investing in modernizing national health registries on specific diseases, on prescriptions and leveraging BDA to improve their population risk management strategies. Enhanced analysis of healthcare needs and their evolution also offers more opportunities to identify, predicts and manage patient risk for complication and proactively tailor treatments and care paths.
- **Patient engagement.** Patient engagement is a key priority for almost 40% of European healthcare providers, particularly among countries that are experimenting with value-based reimbursement models structured around patient experience and outcomes. BDA are leveraged to draw useful conclusions and ensure patients actively participate in their own care.
- **Prevent and respond to cybersecurity threats.** Insights from BDA tools can be used to detect cybersecurity threats as they make the data available to analysts more easily and quickly. BDA also enables analysts to build predictive model that can issue an alert as soon as it sees an entry point for a cybersecurity attack.
- **Reporting on quality of care.** European hospitals are recognizing the value of BDA in managing care quality reporting to ensure their organizations is aligned with best practices and key guidance on patient safety, clinical procedures and protocols.
- **Population risk stratification.**
- **Driving innovation in medical research.** The recent wave of innovation in medical research around genomics is leveraging huge amounts of data for advanced DNA sequencing techniques and computational biology to find variations that affect health, disease or drug response.
- **Integration of patient pathways.** In the context of value-based care delivery, BDA is serving healthcare providers in offering targeted, personalized care through the integration of different sources of information, and to design pathways able to engage patients and improve outcomes.

Figure 9 Top three use cases for BDA among European healthcare providers



Source: IDC European Vertical Markets Survey, 2018–2019 (n= 290 [WE = 232, CEE = 58])

Healthcare data-sharing platforms

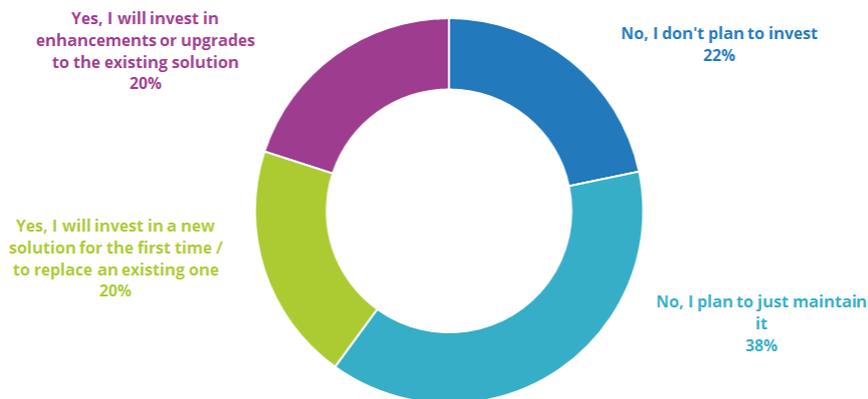
The transition to a VBHC reimbursement model is pushing healthcare organizations to provide cost-effective, high-quality accessible healthcare services, where payers begin to see providers as partners. Adopting this paradigm requires the integration of stakeholders and the development of an overall healthcare ecosystem. For this purpose, healthcare organizations should be able to handle large amounts of data and information across distributed, heterogeneous, and disparate information systems to make it available at the point of care. The data should be accurate, timely, and understandable for every stakeholder in the ecosystem to reduce medical errors and drive down health delivery costs. In this context, it is also crucial to maintain a service-oriented architecture so integrations can be shared and reused, thereby reducing the cost and maintenance of the architecture. To fully leverage the potential of a health integration exchange (HIE) architecture, each stakeholder must agree on what data to share and implement shared data models, semantic interoperability, and normalization.

To achieve data integration, healthcare providers need to address the following challenges:

- Sensitive, personally identifiable information must be protected. The flow and visibility of health-related information among people have to be controlled and monitored to ensure that the information will not be misused. Patients must trust that their sensitive information is safe from being disclosed to unwanted parties.
- The data has to be accurate, timely and understandable to every stakeholder across the health ecosystem.
- Technical solutions must be affordable to enable sharing of an increasing amount of data in different formats, available through legacy and new sources, thereby facilitating seamless collaboration among key stakeholders.

According to IDC European Vertical Market survey, 2018, over 70% of European healthcare providers adopt or plan to invest in Healthcare data collaboration platform. This is particularly true for hospitals that have to comply with new safety and data protection regulations and align their business models with the new requirements of integrated care.

Figure 10 Health data collaboration platform adoption and investment plans in European healthcare industry



Source: IDC European Vertical Markets Survey, 2018–2019 (n= 290 [WE = 232, CEE = 58])

Patient portal

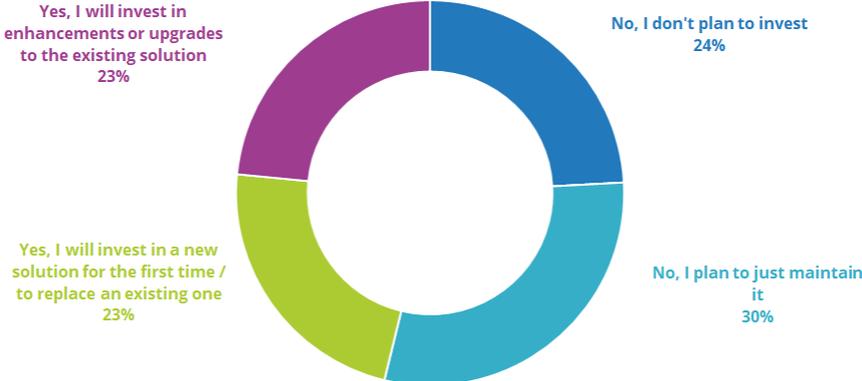
The 39% of European healthcare providers²¹ consider improving patient experience a key priority in the next 12 months. For patients, it isn't about single encounters at hospitals and other providers — it is about the entire care journey. Patients can access their personal health records through a portal that also connects them with doctors. They take the time to navigate the complexities and challenges of their health status to discuss their concerns and conditions with their doctors. The adoption of new patient portals for care services provision, as well as for prevention and healthy lifestyle management, is high on the agenda of many healthcare providers (76%)²².

Data created in one instance must be made available in others to save lives and increase quality by enabling patient-centric pathways. With more control over their healthcare choices and increasing deductibles, patients are putting pressure on providers to give them high-quality care. Moreover, considering the transition toward value, they're using digital tools to measure care efficacy. Patient engagement has become a key strategy to promote informed decision making and behavior to facilitate improved health outcomes. Providers are paid not only on the volume of services provided, but they are also rewarded on outcome-based indicators valued by patients such as quality, efficiency, wellness, care coordination and prevention.

²¹ Source: IDC European vertical Market survey, 2018

²² Source: IDC European vertical Market survey, 2018

Figure 11 Patient Portal adoption and investment plans in the European healthcare industry



Source: IDC European Vertical Markets Survey, 2018–2019 (n= 290 [WE = 232, CEE = 58])

3. The Case Studies - What We Found So Far

3.1 Secondary use of data in Finland

Knowledge-based decision-making supported by digitisation, experimentation, openness and integration of services, has become a central approach by which Finland is seeking more efficient and effective ways of providing public services and foster the country competitiveness. In 2017, Business Finland (national funding agency for research and technology development and related business growth) estimated that by 2030, a platform economy based on the exploitation of data and integrating artificial intelligence into processes and service provisions might account for 30% of Finland's GDP. In the health and social welfare sector there has been an extensive work between the public and private sectors to promote the secondary use of health and well-being data, which has led the creation of a new ecosystem through a national development project, leading to the development of a groundbreaking new legislation and the establishment of a permit authority for secondary use of health data.

Isaacus – the Digital health HUB project.

The Isaacus project, carried out between 2015 and 2018 was led by Sitra²³, with the aim to support the shared national objective to make Finland an internationally renowned pioneer in evidence-based health and well-being²⁴ research and business. Its main objective was to set the technical, operational and organizational foundations for a new permit authority (the new data service operator) and new legislation for the secondary use of data. Isaacus implemented a new one-stop-shop operating model, the Digital health HUB, for a secure and GDPR compliant coordination of all the data kept in the Finnish social welfare and healthcare registries. The Digital Health Hub allowed for the permit and guidance services for the users of data to be available in one place, paying special attention to privacy protection, data security and the right of individuals to decide who the information concerning them is used by and for what. Besides the preparation of the prototype for the one-stop-shop service model and its related infrastructures, one of the project main accomplishments was the development of a greater expertise in the use of the new technologies and multi-stakeholder collaboration. Within Isaacus, Sitra launched 8 pre-production projects, whose lesson learned, and outcomes have been then integrated into an action plan for the new permit authority. Among those projects:

- A project steered by the National Institute for health and Welfare (with Statistics Finland and the Finnish Social Science Data Archive as key co-operative partners) has developed a solution for creating and managing common metadata descriptions, which are integrated into a public web service where they can be also used by researchers²⁵. The developed metadata description solution is now used at national level for statistical purposes
- Data lakes project by the Hospital District of the City of Kuopio, the Hospital District of Helsinki and Uusimaa (HUS) and the Hospital District of Southwest Finland (VSSHP) collecting well-being data that has previously been stored into separate places and operational systems. Additional pilots at Hospital District of Southwest Finland and City of Kuopio are looking at refining data and make them available for different users and care setting. Three data lakes were built:
 - HUS data lake: A cloud-based, open-source code data lake integrating different patient data systems and collecting the metadata on the transferred data. The project provided HUS with know-how on the technologies used and an understanding of how this know-

²³ **Sitra is the Finnish Innovation Fund**, an independent public foundation which operates directly under the supervision of the Finnish Parliament. Its objective is to promote stable and balanced development in Finland, qualitative and quantitative economic growth and international competitiveness and cooperation.

²⁴ "**Health Sector Growth Strategy for research and Innovation Activities, Roadmap for 2016-2018**". The strategy identified the health industry as a new important focus area, which has eased the process of getting funding for health R&D organizations. Finland's Ministry of Social Affairs and Health has continued to implement the growth strategy with its eHealth and eSocial Strategy 2020, designating the health sector as one of the core industries for Finland's future.

²⁵ <https://thl.fi/sv/web/thlfi-sv/forskning-och-expertarbete/projekt-och-program/nationellt-system-for-beskrivning-av-digitala-datamaterial>

how can be used in clinical research and biobank research. To date, the HUS data lake covers a population of 3.5 million participants and includes 785 million laboratory tests, 22 million radiology images and 37 million EMR records

- Kuopio data lake: a data lake integrating some of the social welfare and healthcare data systems of the City of Kuopio. The data lake offers a researcher's workspace with several tools for analyzing and visualizing data. The project used the data to examine the functioning of the social services care paths and conducted a text analysis on the written narratives of social services, allowing researcher to leverage more comprehensive sets of data (for example also those related to socio-economic health determinants) for their research, adding new evidence to better support healthcare services planning and policy making. The City of Kuopio has now transferred the maintenance and further development of the data lake to Kuopio University Hospital.
- VSSHP data lake: A regional data lake tested in different use cases, creating views for different types of users at hospitals. The project has provided a large amount of new expertise in data lake technologies, data quality and data management for the secondary use of healthcare data.
- A project by Statistics Finland, THL, the Biobanking and Biomolecular Resources research Infrastructure (BBMRI) and the Institute for Molecular medicine Finland has developed a data-secure environment for the use of health data. A specific project by BBMRI.fi is also creating a common process and tools for all biobanks for the formulation of research data. The pilot combined digital samples collected by the biobanks from breast cancer patients and the data related to the samples, and information from pharmaceutical reimbursement statistics and the Finnish Cancer Registry with clinical treatment information.
- A project led by the national Archives of Finland which has created the operational features of the permit service operator, through which researchers can apply for authorisation to use the data and material in the social welfare and healthcare register. The project has been implementing a portal that offers information about the data access limitations, the prerequisites for using data and the various intended uses of the data.

Sitra's involvement in the project has come to an end, and the Ministry of Social Affairs and Health (which has been a key partner during the whole Isaacus project) will now be responsible for developing the new services in co-operation with the National Institute for Health and Welfare. Findings and experiences developed within project Isaacus have been instrumental in providing guidance for the development of the new legislation on secondary use of health data and for the creation of a national health data permit authority (Findata) and its supporting digital health hub (service provider taking care of needed infrastructures and network of experts for data use and knowledge management).

Act on the secondary use of health and social data²⁶.

The Ministry of Social affairs and Health began its preparations for drawing up the Act in 2015 that finally entered into force on 1 May 2019. The goals of the act are

- to provide a better-defined legal framework for healthcare data secondary uses (such as steering, supervision, research, statistics and development) in a GDPR compliant way. In this sense, the Act gives clear legal grounds for using register data in innovation and development activities as well as guaranteeing individuals' legitimate expectations, rights and freedoms when personal data is processed.
- to set the requirements and to define the basis for the new data permit authority.

Secondary uses referred to in the Act include:

- scientific research
- statistics

²⁶ Secondary use of health and social data https://stm.fi/sote-tiedonhyodyntaminen?p_p_id=56_INSTANCE_7SjjYVdYeJHp&p_p_lifecycle=0&p_p_state=normal&p_p_mode=view&p_p_col_id=column-2&p_p_col_count=4&_56_INSTANCE_7SjjYVdYeJHp_languageld=en_US

- development and innovation activities
- steering and supervision of authorities
- planning and reporting duties by authorities
- teaching
- knowledge management

The Act is expected to create the environment to drive new opportunities for research of clinical and register data and international scientific cooperation, as well as for health technology and life-science companies to see opportunities and expand their R&D activities in Finland.

A data permit authority, Findata²⁷, is being established for the implementation of the new act. Findata operates within the National institute for health and welfare, but as a separate entity. The main task of Findata is to collect, and co-ordinate data received from different registers and provide services related to data utilization and knowledge management (building on the vision and the know-how of the Isaacus' Digital Health Hub) . Its services include:

- managing the request processes;
- obtaining the required data from different registries;
- combining the data with personal IDs;
- pseudonymization and anonymization;
- aggregation of the data and transfer of the data to a safe and secure environment.

Three key technological systems have been prepared to support operations of Findata (when it will be operational in 2020):

- Permit and information portal: to direct users in the permit process and describe available material and conditions
- National Data description system (metadata) that serves as a centralised system and practice for saving the metadata of available materials. The solution includes also a metadata editor for editing and updating description data.
- Data collection, processing and remote desktop: the actual data management system, which also includes a secure remote user environment and embeds data protection tools such as pseudonymisation, and anonymization.

3.2 Health data hub in France

Background: French transforming its healthcare system

The French Health Data Hub (HDH - Plateforme nationale des données de santé) is born from the political willingness to promote AI in health, identified by the French government as a priority domain where AI technologies could provide a strategic advantage for the nation and for Europe. In 2018, president Macron launched the creation of HDH. This initiative is today strongly pushed forward from the new Health French Act, approved last July, to establish a more efficient and patient oriented healthcare system. The project aims to enrich and enhance the National Health Data System (NSDS – Systems National des Données de Santé) by including the wider French heritage of health data in one place, where it can be used by researchers, health professionals, care institutions, insurers, start-ups and medtech companies. While technology and data are already there, the effective implementation of a platform that facilitates interactions between producers of healthcare data, users and citizens depends on the collection and location of data, and on the frameworks regulating the exchange, storage and use of that data.

In the fragmented healthcare ecosystem of France, organizing resources and capabilities to enable the effective analysis of healthcare big data and the extraction of value in a secure governance framework is paramount to launching and supporting an efficient and patient-oriented national system.

²⁷ <https://thl.fi/en/web/thlfi-en/statistics/data-and-services/data-permit-authority-findata>

The time frame and the objectives are the following²⁸:

- At the end of 2019: launch of the first version of the platform and open to all the offer of Hub services
- At the end of 2020: improvement of the organization, processes and tools, enrichment of the catalog data and creation of the first Local Hubs
- At the end of 2021: deployment of a network of local Hubs throughout the territory

Health Data Hub: a single platform for many players

The Health Data Hub (HDH) has been created to broaden the NSDS to a larger number of data sources (in particular, clinical data, collected by health insurance such as results of biology analysis, imaging, medical records, etc.) with harmonized rules for data access, and use across the healthcare ecosystem of players, from research organizations and healthcare providers to citizens. From May to October 2018, a specific Committee worked on the definition of a roadmap for this platform and the health minister entrusted the creation of HDH to the director of the Directorate of Research, Studies, Evaluation and Statistics of the Ministry of Solidarity and Health.

It was born officially with the publication on 30 November 2019 of a ministerial decision. More precisely, this ministerial decision approves a rider to a previous convention that created the INDS (Institut National des Données de sante, the national Institute of Health Data) since the Health data Hub is the successor of the INDS.

The Health Data Hub would offer public and private entities a standardized service structured around four axes:

- Provide access to health data
- Support the collection and consolidation of health data
- Drive the evaluation and analysis of health data
- Support the ecosystem and ensure a connection with the community and society's good

The platform's unique interface will guarantee shorter access times and better readability of the ecosystem for all stakeholders, and it will pave the way to the creation of a unified governance and a standard data format opening the door to international databases, while staying in a secured and regulated environment. Data sensitivity requires the intervention of a unique Ethics and Scientific Committee (comité éthique et scientifique - CES) evaluating data access requests and mobilizing one or more experts depending on the nature of the project (e.g. patients, ethicists, experts from the industry, technology experts, etc.). For users and data providers, rules for transparent and non-discretionary access are defined, with clear and fixed deadlines and rejections explained in detail. Moreover, the project will expand the scope of data available and provide the means needed to analyze these data (tools, team of experts, etc.). This single and unique data platform will also be a place to share knowledge, as users, data providers and citizens will be able to interact with each other and benefit from each other's expertise.²⁹

The governance rules for the main stakeholders of the Health Data Hub is formalized in three key areas:

- Entities producing data: requirements around the quality and structuring of data, documentation and common identifiers, following the principles of easy, accessible, interoperable and reusable data
- Entities accessing and using data: the frame of reference and general principles of use and shared data processing in the Hub. It includes data protection and intellectual property compliance
- Citizens: the commitments of transparency, ethics and respect for fundamental rights that all the actors of the Hub need to consider for citizens, the community and society

²⁸ Health Data Hub – Mission de Préfiguration. https://solidarites-sante.gouv.fr/IMG/pdf/181012_-_rapport_health_data_hub.pdf

²⁹ Source : <https://www.health-data-hub.fr/>

The Health Data Hub will be composed of 18 databases, including:

- key existing data sources (e.g. regional data registries on patients' pathways, genomics data)
- mature data sources that are ready to be included in the Hub (e.g. national health registries)
- mature data sources requiring a prior selection procedure (e.g. epidemiological registries)

This structure offers the opportunity to connect healthcare and social security data with environmental data and other social determinants of health (e.g. environmental data, quality of life, etc.).

Through the "ENS" ("Espace Numérique de Santé" – ehealth personal space), citizens and users will be able to access and manage their health data in their personal health space, from their personal records to mail exchange with healthcare professionals, to data collected by smart connected devices (e.g. wearables), with the aim to support a privileged and direct access to the healthcare system and to personalized services. The service will be delivered through a central structure supported by centers, operating according to geographical proximity. To ensure the overall sustainability of the model, access to services can be billed to private actors in the form of a subscription or pay-per-use.

The secure collection of a wide variety of patient data in a single point of access opens up research and innovation perspectives for the wider ecosystem of public and private players in the French healthcare system. There are a number of use cases applicable to the Health Data Hub:

- Clinical Trials: healthcare detailed information can support a more accurate and efficient enrollment of patients to trial treatments and support a faster development of care solutions and pharmaceutical products.
- Real-World Data (RWD): the combination of structured and unstructured data collected from different sources (e.g. medical visit vs. wearables) offers the opportunity to easily access and analyze RWD.
- Rare disease diagnosis: research on rare diseases can only be performed at its best by training algorithms and systems on population big data, allowing more precise diagnostics and efficient detection frameworks, which all support population health prevention strategies.

In 2019, France allocated about 10 million euros to HDH conceptual phase. Pilot data sharing projects will be carried out mainly with academic partners. From 2020 to 2022, 76 million euros have been earmarked to complete HDH and to support projects with the whole ecosystem.

In April 2019, the Health Data Hub jury has launched a call for private and public-led projects with various focuses ranging from breast cancer to health surveillance, using predictive tools enabling predictions and insights able to drive decision making.³⁰ (Table 1)

Table 1 Pilot projects of the French HDH Initiative

| Pilot Projects lead by public sector partners and academic partners | | |
|---|-----------------------------------|--|
| SMPP | Oncology/urology | Association of prostatic cancer and metabolic syndrome |
| ENEAS | Public health/Economy | Access to resources for people with disabilities |
| SMPP | Infectious diseases | Antibiotic resistance during bacteremia |
| ENEAS | Pharmacology | Drug interactions during breast |
| BACTERIEMIES | Clinical decision support systems | |
| COMBICANCER | Care assessment | |
| NHANCE | | |

³⁰ Cuggia, M., & Combes, S. (2019). The French Health Data Hub and the German Medical Informatics Initiatives: Two National Projects to Promote Data Sharing in Healthcare. Yearbook of medical informatics, 28(01), 195-202.

| | | |
|---|--|---|
| FREISA MULTIMO HUGOSHARE BNDMR | Oncology/Ear-nose-throat Medication monitoring Epidemiology | cancer treatment Automatic detection of digestive organ lesions Care effectiveness following low-trauma fracture Automatic detection of HPV infection in epidermoid tumors Adverse effect detection and medication reconciliation Enrichment of the national rare disease registry using SNDS |
| Pilot projects lead by private-sector partners and start-ups | | |
| DEEPSARC PARCOURS IDM HYDRO NS-PARK ARAC ORDEI OSCOUR DEEP.PISTE REXETRIS PIMPON | Public Health/Oncology Public Health/Cardiology Cardiology Public Health/Neurology Health economics Pharmaco-epidemiology Emergency medicine Oncology Pharmaco-epidemiology Drug decision support | Enrichment of the national sarcoma registry using SNDS information Care pathways of patients with heart failure Prevention of congestive heart failure crisis Care pathway prediction for Parkinsonian patients Calculation of the amount to be paid after reimbursement Drug adverse effect detection Study of stroke events in an emergency unit AI for breast cancer screening Effects of long-term exposure to immunosuppressive drugs Drug interaction alert system |

3.3 The development of data policies in Portugal

In August 2019, the Portugal Health Ministry has made available the strategic document "From big Data to smart data: putting data to work for the public's health", that outlines the vision, key areas and principles for secondary use of data, advanced analytics and artificial intelligence to improve Portuguese population's health. This strategy will be part of the wider eHealth national strategy for 2020 till 2022, that includes a strong focus on the implementation of AI in public and private activities within healthcare.

Currently, Portugal has a Digital National Health Service (eNHS, e-National Health Service) that coexists with the traditional NHS. The eNHS includes a network of telehealth services and health information systems that collect and communicate personal healthcare data and information on healthcare service delivery. Portugal is investing in the potential that big data and artificial intelligence hold to transform healthcare, public health and medical practice to become more efficient and of higher quality through the intelligent operation of information systems. The evolution of the current digital national health service, e-NHS, to an intelligent National Health Service, i-NHS, that is data-driven by default and measures performance on the basis of what it delivers to peoples' health is the core principle of this transformation. According to the 2019 strategic document, strengthening the collection, storage and reuse of health data will improve the capacity of the Ministry of Health to develop health policies informed by the best available evidence and strengthen national public health surveillance capacity and response so that the Portuguese population can live long and healthy lives. There are several initiatives and pilot projects, that are testing the capabilities of AI and providing evidence to support the development of new data management and governance policy.

Tele-dermatology Screening

Portugal has 1320 new skin cancer cases each year, and despite the advances in telemedicine and other technological solutions to address this specific issue, there is still a significant potential unexplored. In partnership with Portugal's NHS (SPMS – Serviços Partilhados do Ministério da Saúde) under the Data Science and Artificial Intelligence program, Fraunhofer AICOS has launched a pioneering project, called DERM.AI, that explored the use of artificial intelligence to foster teledermatological screening for skin cancer. The key objective of the project is the reduction of skin cancer mortality in the country. Through the DERM.AI smartphone application, GPs can take pictures of skin lesions or suspected skin cancer and send them for screening. The project started in January 2019 and will last for 3 years, with the aim to improve the existing Teledermatology processes between Primary Care Units and Hospital Dermatology Departments in the National Health System for skin lesions diagnosis through the usage of Artificial Intelligence (AI). The planned outcome will change processes by assisting both general practitioners through a computer vision-based mobile application integrated with the eReferral system and dermatologists through an AI-powered Risk Prioritization and Decision Support platform, to be included in the eReferral system, which will serve future risk predictions and assessment for population health management.

Identifying and predicting Emergency Admissions

Emergency care units are subject to a very variable flow of patient and turnout, which can easily affect the efficient and timely use of resources to provide care for all. Therefore, timely information regarding possible variations in patient inflow is fundamental for adequate planning and service quality. In partnership with Fundação para a Ciência e a Tecnologia (Science and Technology Foundation) and Instituto Gulbenkian de Ciência (Gulbenkian Institute of Science), the project focused on identifying the top drivers of emergency care units demands by using data science and machine learning approach to study variations in emergency peaks and the factors that may be able to predict them. The project, named "Data2Help", aims to create new tools to optimize resource allocation and thus improve the quality and response time to medical emergencies in Portugal. Launched in 2018, it has received a funding of €294 million. The key goals in terms of contribution to the population are the optimization of the emergency services and the increase of know-how and skills to scale the project to other areas at national level.

3.4 ARIA's Health Data Warehouse and Business Intelligence Competency Center (Italy)

Background

In July 2019, Regione Lombardia (Lombardy Region, Italy), has established a new Regional company called ARIA (Agency for Innovation and Procurement) from the merge of ARCA (regional agency for procurement) and Lombardia Informatica (the regional in-house digital company). Within its mandate, ARIA has the specific aim to enhance the value of all regional health data assets. This objective has been entrusted to the Data Warehouse and Business Intelligence Competence Center (DW-BICC), that since 2017 (when still in Lombardia Informatica), it has been supporting the regional digital transformation strategy, defining and delivering a value proposition driven by innovative usages of regional datasets.

The center has been developing different initiatives, also leveraging AI and machine learning, to support the regional healthcare service in areas such as: planning and management, costs rationalization, evaluation of safety and efficiency of clinical pathways and integrated patient journeys, as well as health risk prevention. In the last two years, the center has been focusing its activity on

- the integration on the Data Warehouse single platform of more than 10 years of health data – collected especially from different administrative data streams such as hospital admissions, prescriptions, outpatient services- of the 10 million people living in the region, to create 60 different information pillars, clustering information and delivering insights for example on clinical governance and management, epidemiology, management audit,

resources management , financial sustainability, etc. including safeguards, security measures and mechanisms to ensure the protection of personal data and to guarantee compliance with the General Data Protection Regulation

- the setting up of the organizational and technological resources to move the healthcare and welfare data management approach from a pure system of record to a system of insights to inform decisions and processes

Key pilot projects that the center has developed in recent years to test AI and advanced analytics predictive capabilities for population health management include:

- Proof of Concept for Predicting the insurgence of Cardiovascular Chronic Conditions
- Pilot on determining the “clinical next best action” in Alzheimer patients’ journey
- “Geographic Machine Learning” for Respiratory Chronic Conditions

The Center has also started to look at the opportunity to enhance the cooperation with the broader research and life science industry space and started developing a safe Data Information Hub (DIH) for healthcare research.

All these initiatives have been developed following the strategic guidance described in the white paper published by AgID, the Italian Digital Agency, promoting the experimentation and adoption of artificial intelligence.

"Predictive BDA" Proof of Concept for Predicting the insurgence of Cardiovascular Chronic Conditions

The prevalence and incidence of chronic conditions in the Lombardy population is growing: currently it is estimated that chronic conditions are affecting 30% of the Lombardy population and have an associated cost of 70% of the available resources³¹. To better address the population healthcare needs, the center has launched a “Predictive BDA” PoC, using data mining and machine learning techniques to analyze information within the regional health databases (“*Banca Dati dell’Assistito*”) and predict patients’ outcomes. In 2017, the Competence Center has started the proof of concept phase under the scientific coordination of the Epidemiological Observatory of the Lombardy Region, with a focus on cardiovascular conditions, such as ischemic heart disease, high blood pressure and heart failure, as they represent the most diffused chronic conditions in the region and the most costs and resources intensive.

Key results: With the help of clinician, analyzing 5 years’ worth of data about hospital admission, prescriptions and other services provision of 2 million individuals, the center has developed a model able to predict the number of healthy citizens that would have developed a cardiovascular condition in the following year. According to the Competency Center, the estimated number of cases was 25.065, while the actual cases were slightly less (25.055), providing high accuracy levels (0,04% error). The same model has been applied backwards considering previous year, demonstrating a prediction accuracy between 96,22% al 97,96%.

“Predictive BDA”: Pilot on determining the “clinical next best action” in Alzheimer patients’ journey

Through data and processes mining techniques, the Competence center has been developing a model able to determine the optimal patient journey for patients affected by Alzheimer diseases. The optimal patient journey can be predicted year on year, for the first 5 years, since the diagnosis. To determine this optimal journey, the model learns from the records of episodes, services and pharmaceutical consumptions, collected twice a year within the epidemiology registry. The initial pilot has been focusing on a limited geography, taking in consideration the complexity of the disease and the intense consumption of services and medications for the average patient.

³¹ https://www.ariaspa.it/wps/wcm/connect/42ca5684-1be8-4e49-af7b-eb05afe10d26/IA+ARIA_OfficeAutomation_giugnoluglio2019.pdf?MOD=AJPERES&CACHEID=ROOTWORKSPACE-42ca5684-1be8-4e49-af7b-eb05afe10d26-mMEI7D6

Key results: The experimentation produced a clinical decision support system. While clinicians' decision independence from the system is paramount, first results shows that 93% of physicians generally agree on the classifications suggested by the system.

Predictive BDA: “Geographic Machine Learning” for Respiratory Chronic Conditions

By using machine learning, the Competency Center has been able to assess different determinants of chronic obstructive pulmonary disease (COPD) and asthma and predict how these will impact the population in the future in a specific geographical area. The center identified minimum geographical data aggregation units (using for example Voronoi diagrams) leveraging data collected from the regional data warehouse, as well as from other sources (e.g. environment, climate or social-related determinants) to develop a predictive model that assessed the correlation between territorial characteristics and the incidence of respiratory conditions in that clustered population in two years-time.

Key results: The project has demonstrated that this predictive approach can offer institutions, regional and local bodies better insights to plan and enforce preventive and targeted solutions in terms of care provision and resources' planning, including a specific methodology to enrich the aggregated health data base with other sources, a sort of “super-patient” characterized by health features common to an homogenous group of people without the possibility of tracing the individual back, according to the GDPR requirements.

The creation of a Data Information Hub for Healthcare research

Under the regional decree XI/491/2018, ARIA has been developing a Data-as-a-Service platform and a technical-operational model that allows a controlled access to sets of healthcare data along with the tools and solutions to analyze them to universities and research centers, for the regional governance purpose of planning, managing, controlling and evaluating health care pathways. . In this way ARIA created a data sand-box, in a separated environment from the regional health registries and data bases, where research organizations can test and develop new hypothesis, leveraging real world evidence and achieve data-driven benefits, in full compliance to data privacy regulation.

The key objective is to create an ecosystem of information, a sort of "Data Information Hub", that foster innovation through collective intelligence and new international and national partnership with the different healthcare and research organizations.

Lessons learned from ARIA's DW-BICC's projects

The factors determining the positive results so far achieved by ARIA's DW-BICC projects are mostly down to a change in the approach and the culture about data. The work of the center testifies a paradigm shift: from merely collecting and recoding the data, to actively using the data to proactively and dynamically inform processes and decisions.

AI is seen as a tool that enables to derive from data **not only cognitive intelligence, but also collective intelligence**. While the technical competence of the center is a key enabler, the experiences from ARIA highlights how creating an ecosystem that includes, clinicians, public health experts along with data analysts and IT is fundamental for achieving meaningful results. Applying the CRISP-DM (cross-industry standard process for data mining and knowledge discovery projects³²), **DW-BICC has as established practice the involvement of clinical and business users across the critical phases of business understanding, data understanding, data preparation and evaluation**. These phases absorb more than the 80% of the typical project timeline and are critical in determining the odds of data analysis projects success as they determine the business questions, the selection of variables to consider and the data quality requirements.

³² see also CRISP-DM: Towards a Standard Process Model for Data Mining (Wirth, Hipp, 2000) <https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.198.5133&rep=rep1&type=pdf>

Another lesson that can be learned from ARIA experience is the **importance of close collaboration with the national data privacy authority**. The collaboration is key to determine how to ensure that data processing is lawful across different use cases and respects, key data governance principles such as for example that of data minimization. Moreover, as using AI entails, almost intrinsically, the possibility of “unknown unknowns”, it is essential to address the related concerns for the data subjects as they appear. Therefore ARIA is working with the DPO authorities to finetune and adapt the regulatory framework to the emerging needs, while at the same time maintain accountability and respect individuals’ rights.

4. Preliminary Policy Conclusions

The table below summarizes the common themes and findings across the four case studies featured in this piece of research. For the clarity of the analysis we propose a horizontal overview taking into account the specific challenges related to data-driven innovation at country-level, areas of deployment of AI and BDA technologies to realize data-driven initiatives, as well as the operational benefits that were possible to observe.

Table 2: Horizontal Overview of the 4 Case Studies

| Country | Challenges/ business needs | Key initiatives and application areas for secondary use of health data | Benefits (current or expected) | Output to date |
|---------|--|---|---|---|
| Finland | <ul style="list-style-type: none"> • Making use of the huge amount of health and social care data collected from different sources and registers • Offering to researchers and organizations looking to leverage health data for innovation, a coherent, and simple environment for accessing and using the data in a legally compliant way • Maintaining public trust in government capability to manage data towards common good and respect of individual rights | <ul style="list-style-type: none"> • Creation of a data access permit authority and the related organizational, infrastructural, and legal conditions • Key application areas: innovation in clinical and public health research, as well as in the private sector (pharma and life sciences) R&D to promote economic growth | <ul style="list-style-type: none"> • Simple yet comprehensive permit procedure to get access to data for R&D scopes • Strong knowledge base and experience on managing a complex ecosystem of stakeholders, sometimes with conflicting interests, and to enable sustainable cooperation • Establishment of a single authority that work as a one stop shop for permits regarding several registers | <p>Sitra launched 8 pre-production projects, whose lesson learned, and outcomes have been then integrated into an action plan for the new permit authority once in operation.</p> |
| France | <ul style="list-style-type: none"> • Fragmentation of the healthcare ecosystem • The need for a secure governance framework enabling the ethical use of data for analysis and research • The need for establishing a regulated platform for data access and use as well for facilitating the interaction of multiple stakeholders (collecting, producing and using data) | <ul style="list-style-type: none"> • Establishment of a Health Data Hub (HDH) to manage a large number of data sources with harmonized rules for data access, and use. • Key application areas for secondary use of research include the development of RWE research, support to clinical trials, development of precision and personalized medicine, predictive clinical | <ul style="list-style-type: none"> • A harmonized standard of data access at national and international level • Patients/citizens access and control over the use of their data • Accelerated innovation and personalized services | <p>19 pilot projects have been launched projects with various focuses ranging from breast cancer to health surveillance, using predictive tools enabling predictions and insights able to drive decision making</p> |

| | | | | |
|-------------------------|--|---|---|--|
| | | decision support solutions | | |
| Portugal | <ul style="list-style-type: none"> The need to transform the National Health Service into an intelligent, data driven NHS to drive efficiency and better patient outcomes The need to define a framework for the secure collection, storage and reuse of health data | <ul style="list-style-type: none"> Establishment of a national strategy for health data management and for secondary use Funding of AI enabled research programs and projects aimed at the development of clinical decision support; personalization of clinical pathways; patient risk management; optimization of resources utilization | Creation of <ul style="list-style-type: none"> A data infrastructure that supports population health management Predictive solution aimed at reducing skin cancer mortality Predictive solution aimed at optimizing the delivery of emergency services | There are several initiatives and pilot projects, that are testing the capabilities of AI and providing evidence to support the development of new data management and governance policy. |
| Italy (Lombardy region) | <ul style="list-style-type: none"> Need to integrate and use >10 years of collected health data in the regional Data Warehouse Development of infrastructural, organizational and policy framework for the use of and access to data | Predictive BDA models and Data health hub <ul style="list-style-type: none"> Patient risk Population risk management Illness progression and clinical decision support RWE driven research | <ul style="list-style-type: none"> Enablement of new data driven research streams and collaboration at national and international level Establishment of a PoC framework to predict cardiovascular risk with accuracy between 96,22% and 97,96% Launch of a pilot determining the clinical best next action for Alzheimer patients | Several Initiatives to support the regional healthcare service in areas such as: planning and management, costs rationalization, evaluation of safety and efficiency of clinical pathways and integrated patient journeys, as well as health risk prevention |

In terms of the triggering **challenges**, security and data privacy concerns are on top, due to the highly regulated environment of the healthcare industry. However, the common theme emerging from the case studies are threefold:

- **The establishment of a regulatory framework for the use of and access to data.** This involves multiple players in the public and private healthcare arena to establish partnerships and create a collaborative environment. It requires all stakeholders to agree on the value of data as a shared asset, and to actively promote initiatives where common standards and a one-stop-shop approach to data access are brought forward. This is the case of Portugal, where the Health Ministry authority is seeking to establish a new healthcare ecosystem based on a data-driven approach towards the delivery of healthcare services across the nation.
- **The collection, processing, storing and access of complex data coming from different structured (e.g. National health records) and unstructured (e.g. wearables) sources,** along with semantic, geographical and time complexities. This is the case of ARIA and its collection of over 10 years of healthcare data stored in different locations, as well as Sitra's project working at establishing a common framework and developing metadata descriptions. Additionally, data collected require intelligent solutions, capabilities and skills to extract value

from data and provide actionable areas for the deployment of information (i.e. population risk stratification, clinical decision support, personalization of clinical pathways, etc.).

- **Maintaining trust and ensuring security.** The high sensitiveness of healthcare data requires an attentive approach to identifying and enforcing regulatory frameworks and solutions that ensure information is treated in compliance with regional and country-specific policies. The adoption of GDPR strengthens and unifies data protection for individuals within the EU, regulating how data integration happen safely. It gives individuals key control over the usage, processing and transfer of their personal data held by healthcare organizations. The transition to a VBHC model, in which care plans should be personalized and stakeholders should integrate their activities, must be underpinned by consistent information management governance that enables patient data integration across providers, processes and IT systems. GDPR is expected to provide a patient-centric ecosystem. In addition, authorities need to establish a high level of public trust for the ethical and secure use of healthcare and social population data for the public good. In this example, Finnish citizens are informed about how their data are used for secondary purposes out of primary care.

The case studies presented in this research highlight several **benefits** obtained by the organizations adopting AI/ML and BDA technologies:

- **The easy and convenient access to intelligent solutions for clinicians and patients offers more opportunities to advance decision making and enhance clinical process efficiency at the point of care.** Portugal skin cancer screening solution is an example of how technology supports a clinical collaborative framework and enables the integration of information to serve population health management,
- **More advanced predictive capabilities, allowing greater control over disease-specific variables impacting health outcomes, as well as costs and resources utilization associated with care.** This approach enables to more efficiently target population segments at risk of developing chronic and long-term conditions by putting in place initiatives aimed at promoting health and preventing or delaying the development of risk factors. Predictive BDA by ARIA is an example of development of a predictive model able to effectively target cardiovascular conditions and offer an accurate estimate of the number of future cases in a specific geographical area.
- **The adoption of BDA as part of their business intelligence strategy, help healthcare providers to improve the overall operational efficiency.** Advanced predictive analytics model supports the definition of admission rates along with attrition rate to help with staff allocation. In Finland, for example, the new Act on secondary use of data is expected also to support the planning and the reporting duties of authorities. The aim is to reduce the healthcare costs and focus more resources on the delivery of better healthcare.
- **Big data can help in uncovering unknown correlations, hidden patterns, and insights by examining large sets of data.** By applying machine learning, big data can study human genomes and find the correct treatment or drugs to treat cancer or other rare disease. Clinical studies are long and expensive to implement. They provide results on the drug administration in a specific controlled frame that is not real-life conditions. Moreover, it can be difficult to compare results between different clinical studies (indirect comparison of different new treatments). Real World Data (RWD) analysis could provide exhaustive and real-life analysis. Aware of RWD potential, health authorities are developing new evaluation frames with RWE and artificial intelligence. The HDH in France is an example of how a secure and a regulated environment offer the access to relevant data for health actors but also to the means needed to analyze these data, facilitating clinical trials initiatives.

IDC Research References

Big Data and Analytics in European Hospitals

<https://www.idc.com/getdoc.jsp?containerId=EUR145532419>

The pulse of IT in the European Healthcare Provider sector: Key Deals and Initiatives, April-June 2019 <https://www.idc.com/getdoc.jsp?containerId=EUR145406919>

Artificial Intelligence Strategies in European Healthcare: Adoption and Investment Plans

<https://www.idc.com/getdoc.jsp?containerId=EMEA44354519>

Are European Hospitals Ready for Digital Determination?

<https://www.idc.com/getdoc.jsp?containerId=EMEA44972919>

Priorities, Strategies and Investments: Overcoming DX Challenges in European Healthcare

<https://www.idc.com/getdoc.jsp?containerId=EMEA44355219>