A Vision for Predictive Health in Canada

Canada's Big Data Consortium
May 2018

We would like to thank the following organizations for their leadership and in-kind contributions to Canada's Big Data Consortium, Canada's Big Data Talent Gap Study, and to this paper, *"A Vision for Predictive Health in Canada."*

A Vision for Predictive Health in Canada

Canada's Big Data Consortium
May 2018

Contents

Executive Summary

Predictive Health is to apply predictive analytics to improve health and healthcare. This includes innovative data sourcing, advanced methodologies, multiple industrial collaboration,strategic investment, creative entrepreneurship and comprehensive governance and policies.

Canadian health system is facing tremendous challenges to meet increasing demands and improve efficiency. The predictive health provides a new approach to address the issues. This means a strategic focus needs to provided to the following actions:

- Advance Canada's open access to data and analysis.
- Apply advanced predictive algorithms to improve efficiency within the healthcare system.
- Solidify an environment that is receptive to continuing innovation in health IT.

Health data sources have expended substantially in recent years. Individual life real time measures, e.g. data from wearables, IoT data from inplanted instruments, social media sentiment from patients and many other types of data have become important part of health data and influenced health related decisions of individuals and the society. we recommend Canada to develop a new framework of health data to increase assess in a privacy-sensitive manner and encourage innovation and entrepreneurship. This will help making predictive health, and heath industry overall, a major economic growth driver.

The Big Data Consortium was established by Ryerson University in 2014 and works to bring together government, industry, and universities. Our aim is to impact the efficiency of the healthcare system and spur investment in innovation.

1 Introduction

1.1 Objectives and Organization of Paper

Canada's Big Data Consortium Working Group on Predictive Health aims to provide a vision of the impact of predictive analytics on the future of the Canadian healthcare system . The hope for this white paper is to disseminate key knowledge on the role that predictive analytics will inevitably play.

This paper identifies the need for change in the current healthcare landscape, and describes the transformative changes that predictive analytics could bring in the coming years. Then, the key players – practitioners, policy makers, health IT startups, industry, and investors – are highlighted, and guidance is provided on the role they can play in achieving this vision.

The main message of the white paper leaves the audience with a solid strategy for achieving this vision. The hope is that leaders across government and the public and private sectors will succeed in the near future in the missions to:

- Advance Canada's open access to data.
- Apply advanced predictive algorithms to improve efficiency within the healthcare system.
- Solidify an environment that is receptive to continuing innovation in health IT.

The paper outlines the framework of tools and functioning of Canada's digital-health ecosystem. Further, many influential examples are presented that highlight tremendous growth in the applications of predictive analytics. These predictive-task applications touch on domains of varying nature, such as chronic disease, medical intervention and diagnoses, genetics, pharmaceuticals, proactive health measures, electronic health records, and "smart" technology that is evolving rapidly as part of the Internet of Things (IoT). The challenges behind these use cases are addressed, and the most recent suggestions for analytics capabilities, data management, data integration, security and privacy, as well as cloud services are brought forward for discussion.

We are living in an age where technological advancements are evolving at an extraordinary pace. As Canadians, it is our collective duty to seize this potential, allowing our country to remain a world leader in healthcare.

1.2 Target Audience

Our target audience consists of Canadians of all ages who are invested in the healthcare sector, through financial or physical means, or both. It also targets those who actively engage in promoting innovative ideas; who raise concern for health quality and sustainability; and who have a keen interest in advancing this vision. This includes the innovators, entrepreneurs, health authorities, healthcare leaders and general public.


Scotiabank
DR.JOHN HU
DENTIST

1.3 Vision

This white paper offers a blueprint for the ways in which big-data tools and predictive analytics can help solve many of Canada's health and healthcare challenges. The aim is to present a vision of what the system could look like very soon.

Canada has a $219 billion dollar healthcare industry [6] and makes significantly higher investments into healthcare relative to economic growth than the average investment made by other countries belonging to the Organisation for Economic Co-operation and Development (OECD) [20]. However, there are still many challenges in our system that have not been solved, regardless of our investments. These challenges, such as simple financial sustainability, pose a great need and opportunity for technological disruption. The catalysts to the disruption will include population aging; patient-driven demand; increasing desire for improving healthy living and wellness habits with wearable devices capturing real time personal health data; the need for innovation in areas such as biomedicine and chronic disease management; and a push towards more community-based care.

The healthcare industry stands to benefit greatly from the application of big-data and analysis. Predictive analytics can uncover actionable insights that would otherwise not be evident from traditional data collection and analysis. The problems to be addressed will largely consist of those relating to patient outcomes and inefficiency within the healthcare system. Specifically,

data analytics will play a key role in decision support systems, in improving the patient journey throughout the healthcare system, in preventative medicine that extends into the home and community, as well as in using evidence-based decision-making for addressing concerns such as resource allocation and health spending. The vast volume of data that is, in a sense traditionally not machine-readable, can be processed and mined in real time. This will promote and improve healthcare outcomes, and it will help sustain the healthcare system in the long term. In addition, personal wellness data that is generated by increasing number of wearables and other instruments will play an important role in supporting healthy lifestyle and improving population health.

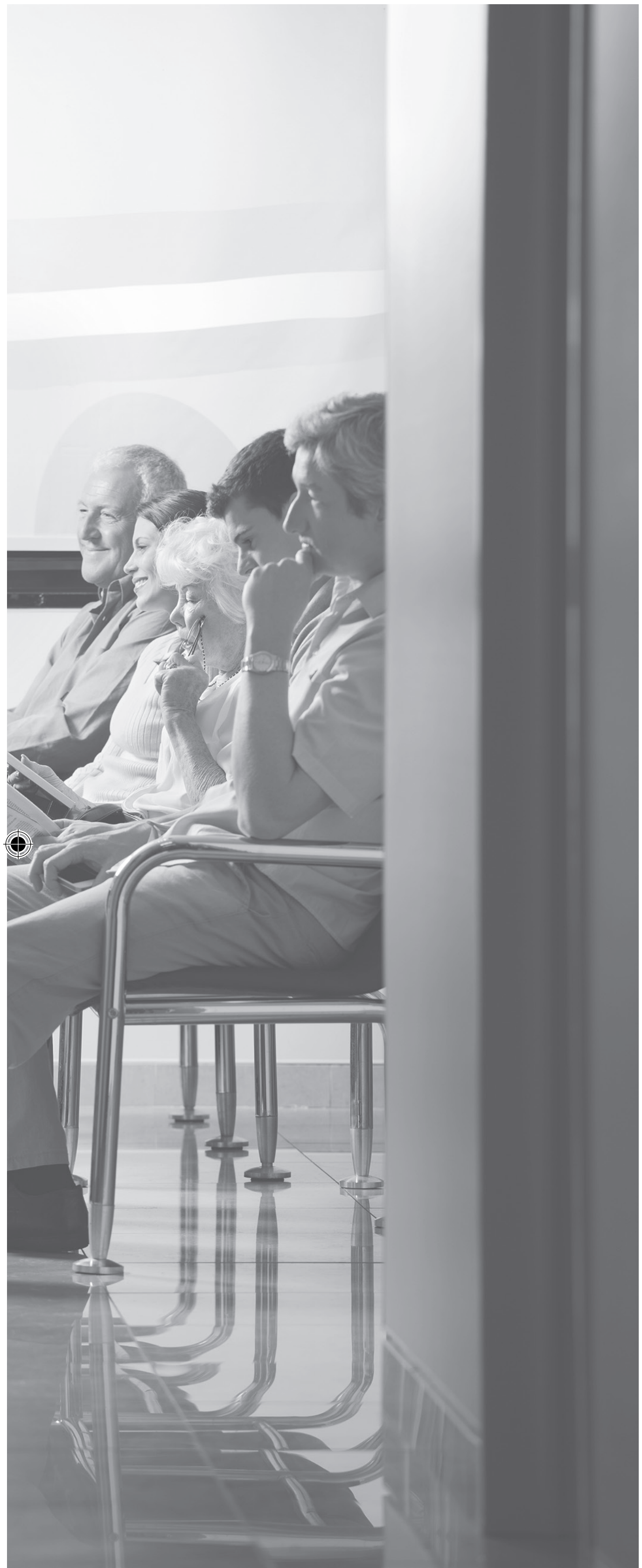
This white paper foresees a future with real-time predictive analytics delivered at all points of care by providers - both real and virtual - to better improve patient diagnoses, outcomes, quality of care, management of resources, and scientific discoveries. In a very short period of time, cloud technology will be universal, with each patient having a holistic digital representation of their entire health history. Interactions with the health data ecosystem will be at patient's fingertips, and the fingertips of healthcare providers. By this point, open data and advanced technological infrastructure should be prominent in allowing large-scale analytics and research initiatives to take place, and for innovations to flourish.

1.4 Evolving Healthcare System

A number of recent studies show that North American healthcare systems have inefficiencies that could be improved upon. A 2014 Canadian study by Allin et al., *Our health care system is inefficient. What can we do about it?*, estimates the inefficiency gap in healthcare to be between 18-35%, amounting to nearly 25,000 potentially avoidable deaths annually [1]. A 2015 report from the Harvard Business Review suggests that the United States could reduce annual waste in healthcare spending by $1 trillion [2]. A major challenge in reducing healthcare waste and improving patient outcomes, however, is identifying specific areas for improvement and developing data-informed approaches to eliminating errors and improving clinical care.

There is considerable promise for predictive analytics and big data to play a crucial role in reducing healthcare expenditures and improving patient outcomes. For example, a 2013 McKinsey report, *The big data revolution in US health care: Accelerating value and innovation*, suggests big-data analytics can enable more than $300 billion annually in healthcare savings in the United States [3].

Deep insights from data can present significant opportunities for improvement. Advances in computational infrastructure and approaches to analyzing data including machine learning are revolutionizing our ability to optimize efficiency and clinical care. The Canada Health Act (CHA) outlines the duty for provinces and territories to



provide medically necessary healthcare provider services. However, there are still problems in their accessibility, in redundant testing, in incorrect diagnoses, in a high risk of hospital infections, in a lack of preventative measures, and in long wait times. Additionally, Canada's healthcare landscape is challenged by healthcare costs, financial sustainability and a lack of integration among systems.

Timely access to healthcare services is a significant challenge in Canada. In 2015, out of 10 countries that participate yearly in the Commonwealth Fund Survey, 7 out of 10 Canadian primary-care physicians reported that their patients were subject to long wait times to see a specialist [4].

The use of electronic health records (EHRs), which are longitudinal records of a patient's interactions with the healthcare system, has grown considerably since Canada Health Infoway launched a mandate for its development and adoption in 2001 [5]. However, Canada still has a long way to go to provide easy access to health data that is standardized, consolidated and securely protected for innovation, healthcare improvement and economic development.

It is of utmost necessity that Canada does not lag behind in using digital health technologies for innovative transformation and that the digital health information is put to meaningful and actionable use.

1.4.1 Strengths in Technology, Big Data, Analytics, Innovation

Canada's tech sector is vibrant and booming, with potential for many entrepreneurs to break into the digital healthcare industry. The more healthcare technology and analytics are able to be disruptive and produce transformative change, the higher the adoption rates and positive impact they will bring to Canada's $219 billion healthcare industry [6]. The research community will continue to be of great importance in accelerating creative advances in predictive analytics and will continue to push for open access to data.

Many projects are already underway that take advantage of big-data capabilities; this is accomplished through collaboration between public and private sectors, open-source technology, and large domestic as well as foreign investment. Throughout this paper, there are descriptions and case studies describing highly successful projects.

Likewise, our respected research institutions continue to deliver the latest educational programs surrounding analytics, management, health informatics, engineering, and clinical training, which in turn produce skilled professionals that go off to serve the healthcare industry. Bright minds are essential drivers for innovation.

There are strategic research projects that have been and will be commissioned that support national collaborations for health innovation, plus new initiatives supported by the Canadian government to create social and economic value. A number of examples are outlined in this paper. Our publicly funded healthcare system is also striving to improve its efficiency, combat new health challenges that present themselves, and create new funding opportunities for innovation. For example, the Office of the Chief Health Innovation Strategist (OCHIS) was set up to help accelerate health technology commercialization efforts in Ontario. OCHIS[1], as an office of government, works with health technology innovators to remove barriers and improve access to Ontario's health care system.

1.5 Stakeholders in the Canadian Health System

1.5.1 Government

The Canadian medical system, also known as "Medicare", is made up of separate provincial and territorial insurance plans, rather than a single plan. It is the responsibility of each province and territory to provide medically necessary hospitalization as well as physician care to their residents, and to adhere to the values of the CHA [7].

This is why it is of utmost importance to engage all levels of government in order to strengthen the ability of Canada's publicly-funded universal health system to adapt to the challenging advances in technology and data analytics [8]. Decision and policy makers at all levels of government are key to guiding the national effort. However, the provincial and territorial focus goes beyond the primary obligation to deliver medically necessary services. It also includes the duty to promote and address public health issues and offer additional healthcare services targeting specific populations.

Addressing all of the shared responsibilities can be improved and supported by both small and large-scale data analytics. Having municipal, provincial, and federal policymakers involved, we ensure that there is a pan-Canadian approach to health innovation. These government bodies hold the key to financing, accelerating and delivering new implementations in health technology with support for strategic planning, project launching, statistical analysis, or driving innovation [9].

1.5.2 Clinical Sphere

The health care professionals, like physicians, surgeons, nurses, pharmacists, and community caregivers who deliver healthcare services, must be properly trained and ready to use new knowledge and technologies. The successful use of EHRs, development of information and communication technologies (ICTs), and the competency to work with different types of data is crucial for achieving the vision for the near future [10].

While it is important to have the front-line staff leading the way, patients must not be forgotten. A patient-centric model for healthcare allows for patient engagement in their own medical care and respects their health information, which ultimately advances the collective knowledge that is needed in applying predictive analytics. Partners such as the Canadian Foundation for Healthcare Improvement (CFHI) currently run a multitude of patient engagement initiatives, which are needed to ensure a holistic strategy is in place. Moreover, not-for-profit organizations such as Canada Health Infoway, Canadian Institute for Health Information (CIHI), COACH, CHIMA, Healthcare Information and Management Systems Society (HIMSS), and others provide strategy and guidance in the development of digital health solutions, national coding and data standards as well as in informed systems decision-making from evidence-based sources [11]. It is also critical to work with private sector partners beyond these organizations to support e-health investments.

1.5.3 Academic Community

Our vision also encompasses participants from the health informatics, information technology, and research communities. There is an increasing demand for skilled professionals to be able to understand big data, to use data analytics technology, and to have an interdisciplinary skill-set [12]. Post-secondary institutions that offer data analytics programs are combining disciplines in mathematics, engineering, computer science, statistics, and business to address this need. Ryerson University, for example, offers a professional Certificate in Data Analytics, Big Data, and Predictive Analytics. Recently, Ryerson University also launched a new master's program in Data Science and Analytics to address the growing need for these skills in the public and private sector alike [13]. Graduates will be prepared to pass the Institute for Operations Research and Management Services (INFORMS) Certified Analytics Professional (CAP) exam, certifying them in this field.

Another popular program is the Master of Science in Big Data at Simon Fraser University, which was introduced in 2014. Students in the program gain a deep foundation in data science, and real-world experience through co-op work. Other Canadian institutions such as University of Toronto, Carleton University, Schulich School of Business at York University and Queen's School of Business are also leading the way in filling the talent gap across North America.

Programs like these are garnering a tremendous amount of interest from individuals who wish to study the expanding field of data science. As an example in continuing education, Ryerson's Certificate in Data Analytics, Big Data, and Predictive Analytics attracted 361 new certificate candidates in its first year of offering in 2014-15. Enrollments grew substantially in the following two years, with 510 new certificate candidates in 2015-16 and 622 new certificate candidates in 2016-17. The demands of such education come from multiple groups of the population who are looking for opportunities to gain entry into the career of data analytics. Many of them are working in different sectors with varied professional backgrounds. To accommodate the high demand, the university has offered, in dedicated data analytics computer labs, the certificate in various delivery modes, including evening and weekend offerings allowing students to complete the certificate within 1-6 years while working full-time or parttime. In addition, students can choose to complete the daytime "fast track" program, within either a 8-month or 12-month time period. All lectures are streamed online to better meet the needs of today's learners.

Recently the Ontario Government announced a goal of increasing the number of AI-related Master's graduates by 1,000 per year by 2023. While the example above has been a success story, the post-secondary sector must continue to focus on the current demand for data analytics professionals as well as be able to adapt to the latest emerging technologies to stay competitive. To meet Canada's looming shortage of qualified analytics specialists, the expansion of post-secondary institution programs must include real hands-on experience and exposure to big-data tools.

1.5.4 Research Community

There are numerous ongoing and future research projects being executed and planned in Canada. At Concordia University, a new initiative in research computing, the Data Science Research Centre, was launched in 2016[2]. The Centre's aim is to promote open data in the academic community as well as in the government and business spheres where industry meets academia.

Other collaborations include the St. Michael's Hospital (Toronto) advanced analytical techniques and machine learning team, who are working with partners such as Google, IBM, Harvard, MIT, and Stanford to improve patient outcomes and reduce hospital costs within three years. Opportunities such as these help bridge the gap between academia and the real-life challenges of the healthcare industry.

From recent government investment in new genomic applications, infectious disease outbreaks, and cancer projects, the adoption of predictive analytics will continue to be well on its way [14]. Professional training programs, research projects, government funding, and industry support in this field are the building blocks to creating the foundation for this Canadian vision.

1.5.5 Health Informatics Professionals

In order to achieve the goal of improving health using data and analytics, the required skills are not only in data science, but in clinical informatics and health information management skills too. Health informatics (HI) professionals are heavily involved in the implementation of data governance frameworks, interoperability standards, e-health applications, and overall health system change [15]. In short, predictive analytics is as much about technology as it is about people. It is important that there are also change leaders with a deep understanding of the Canadian healthcare system is working towards achieving this vision.

There have been education programs developed to meet the needs for health informatics professionals in Canada. For example, Ryerson University's Certificate in Health Informatics is designed for professionals with some health or IT background, and develops skills in the interdisciplinary field of health informatics within a management context. The certificate also serves internationally educated professionals who need to understand health informatics within the Canadian context. The University of Victoria has both undergraduate and graduate programs in Health Informatics with a focus on the planning, design, implementation and evaluation of health information systems in diverse health care settings. The Université de Sherbrooke and Dalhousie University also have created similar programs.

[2] http://www.concordia.ca/cunews/encs/computer-science/2016/04/20/Launch-of-Data-Science-Research-Centre0.html

1.6 Adopting Predictive Analytics: Key Players

1.6.1 Practitioners

Predictive analytics will be of great benefit to practitioners by aiding in their decision making. Practitioners will spend less time planning for the delivery of care, and will also have access to evidence-based treatments and be able to avoid redundant testing. This in turn means that practitioners' workflows will be smoother and more patients can be seen, while being reassured that they are receiving the most up-to-date and high-quality care.

By using predictive analytics, complex datasets that normally take weeks, if not months, to analyze can be interpreted more quickly and greatly advance care. Powerful computational techniques can model and help forecast different outcomes such as mortality rate, risk of infection, and cancer prognosis.

Furthermore, the health system will be more patient-centric, linking up and integrating all of the patient's health information around them. Health professionals will be able to understand who is going to get sick and can take appropriate steps to prevent that from occurring, whether it be to reallocate resources, engage in preventative care, or adjust the current treatment plan. There will be greater ease of communication between healthcare providers and patients alike as the improved interoperability and facilitated use of data becomes more commonplace. It has long been a challenge for physicians to keep up with changing practices, newest clinical trial results, volumes of clinical notes, and new medical literature being published almost daily. It is paramount for physicians to be able to use predictive analytics to retain knowledge currency, be leading experts in their fields, and be confident in clinical decision-making. Better data analytics means greater accessibility to information, and the capability to help highly skilled professionals analyze patient cases with high precision and make better clinical decisions.

Developing trust of predictive analytics among practitioners is key, as they need to be confident in the accuracy of the resulting predictions. Providing practitioners with the ability to drill down into the analysis so that they understand the basis behind the prediction is necessary to establish this trust and ultimately, adoption of predictive analytics.

1.6.2 Startup Sector

Canada presents limitless potential for the health start-up industry, and their role in transforming healthcare is going to be (and already is) revolutionary. Many startups using predictive analytics tackle a wide range of health industry problems, ranging from monitoring patient outcomes, optimizing the patient journey throughout the system, personalized medicine, and resource allocation, to name a few.

There are major health innovation competitions being held to bring out the best ideas from health IT startups, including the IHX Challenge run by the Interface Health Society, and the HealthKick conference led by MaRS [16].

Startups can offer innovative ideas in applying predictive analytics in promoting healthy living as well as managing disease. For example, RightBlue Labs developed a user-friendly software suite that enables sports organizations, personal trainers, medical professionals, and military unit supervisors to monitor wellness and health risks of their athletes, clients, patients, and soldiers in near real-time. Moreover, there are also many potential applications for predictive analytics to enhance and innovate in the delivery of healthcare, specifically in risk prediction and evidence-based medicine. This poses a potentially prosperous future for startups to take the lead in using advanced technologies delivering seamless care. Their innovations can help connect healthcare systems, increase patient engagement, and move Canada towards more preventative health.

1.6.3 Industry

Without a doubt, the healthcare sector will continue to be disrupted with new digital health solutions. Specifically, information and communications technology (ICT) companies hold large potential to strengthen the Canadian health economy by offering secure cloud services, long term data management, mobile health applications, virtual care, electronic health records (EHRs), medical devices, and many other innovations.

Health ICTs are becoming more commercialized and embedded in the healthcare system, and with that comes revenue generation, a heavy investment for support, and a large demand for skilled workers. In fact, there are over 47,000 employees in the health ICT sector in Canada today with $3.4 billion in generated revenue per year [17].

1.6.4 Investors

Canada's venture capital funds are investing heavily into innovative ideas in the healthcare sector. In 2015, Northern Biologics, Profound Medical, and Clementia Pharmaceuticals received venture capital funding of $30 million, $24 million, and $60 million respectively [18]. Indeed there are many venture capital funds that are spurring innovation across the healthcare sector.

Venture capitalists are not the only ones that are stimulating growth in this sector; government agencies are as well. In early 2016, Jane Philpott, Canada's Minister of Health at the time, announced new funding totaling $13.8 million in support of 22 eHealth projects. These same projects will also include $32 million in funding from private, public, and non-profit-sector partners [19]. This announcement serves as a reminder that entrepreneurship is important for healthcare innovation, and that government, academia, and industry must work together to achieve success. Industry leaders are needed to ensure that Canada has proper IT infrastructure in place to be able to take on big-data analytics and offer viable solutions, and this is a time for Canadian industry leaders to step up and help improve the health of Canadians.

1.6.5 Policy Makers

All levels of the policy making community, both legislative and administrative branches, need to be aware of this great potential of opportunity for Canadian economic development and serving patients and population health. For policy to move towards the promotion and adaption of the new technology, champions in the government and policy community are needed. This is paramount, as it is the policy makers who need to lead community engagement by making the community aware of the potential of predictive analytics and promoting community discussion. The outcome of these efforts would be the creation of policies that support the predictive health and other adaptations, e.g. policy directions, government funding for innovation, tax credits etc., as well as improved health care integration and the breaking down of silos. However, to achieve this, other "key players" need to first inform policy makers what needs to be changed.

2 Data Ecosystem for Health and Healthcare

2.1 Ecosystem Overview

In a 2013 article published by McKinsey Quarterly, *Big Data in Healthcare, Hype and Hope*, primary research was conducted to assess how big data could help emerging companies improve healthcare. The article lists six ways, including facilitation of research, transformation of data to information, providing support for self-care and care providers, increasing awareness, and finally, pooling of data to expand the ecosystem [3]. As big data becomes more manageable, accessible and integrated, its uses and benefits will only increase in the coming years, enhancing the data ecosystem every step of the way.

Canada's healthcare network has always been plagued by a lack of open access to health data and minimal integration between systems. If we could ease the restrictions on healthcare data being shared openly, the clinical and research community could find new insights surrounding the causes of various diseases; discover cutting-edge and individualized treatments specific to sub-populations of patients; build province-wide and even nation-wide bio-banks; improve the standardization of medical terminology; and construct rich datasets for analyzing healthcare delivery challenges [21]. This would also allow for external players, particularly in the private sector, to aid in solving Canada's top challenges in its healthcare system by using powerful analytics technologies.

As the ownership of this data comes into question, patients must not be forgotten; trust requires transparency and patient involvement. Rather than having patients left out of the picture, we should actively engage them as open access is promoted. It is hoped that patient portals will soon become universal and that individuals will have the right to share their data for research purposes, in a sense, agreeing to become "data donors" [21]. Moreover, the other issue of concern is data integration.

Many systems currently in place across the nation do not support interoperability for data storage, capture, and exchange, with huge variation even at the provincial and territorial level. One problem area has been the transition from manual data collection to digitized health information and communication technologies (ICTs). Solutions for combating this include training clinical teams in using ICTs, integrating the technologies using high level implementation frameworks, incorporating technical and clinical expertise, and actively learning from past failures.

Focusing solely on introducing new technology tools to solve integration is not the answer; the proper strategy is to focus on the people involved in its use, understanding and streamlining the process, and then assessing the technology [22]. Once data standardization is in place, many rich sources can be used for analytics applications. Generally speaking, the push for open data and integration is largely in its infancy stages but should be integrated into the long-term vision for transforming Canada's healthcare system.

2.1.1 Ecosystem Stakeholders of Data

The main stakeholders in Canada stretch across the public and private sector. Due to the sensitive and personal nature of the data, the primary stakeholders are the patients and consumers of healthcare services and wellness practices.

National stakeholders of health research data include: Canadian Association of Research Libraries (CARL), CARL Portage Network, Canada Health Infoway, Consortia Advancing Standards in Research Administration Information (CASRAI), CIHI, Canadian National Committee for CODATA (CNC/CODATA), Compute Canada, Leadership Council for Digital Infrastructure (LCDI), and Research Data Canada (RDC) [23].

These organizations are making progress in tackling many issues that surround the health data ecosystem, such as funding, cyber-infrastructure, standards, interoperability, skills training, and policies [23]. Since clinical health data is exchanged between providers of different capacities and settings, the ecosystem environment spans them all, including the facilities; the clinical, administrative, and IT staff; vendor providers; payers and purchasers; and policy makers.

Finally, industry partners who provide technological services and support for housing, managing, and running analytics on the data are also key stakeholders.

2.1.2 Factors for Quality, Physical, Mental, and Emotional Health

There are many determinants to quality health for Canadians that can be broken down into five main categories: physical health, mental health, education, government, and necessities.

Physical Health

- Addressing medical issues and injuries through the stages of prevention, diagnosis, prognosis, treatment, recovery, rehabilitation, chronic care, and palliative care
- Staying fit
- Eating healthy
- Monitoring drug use, alcohol intake and smoking
- Getting enough rest

Mental Health

- Coping with mental illness
- Managing stress
- Healthy social behaviour and relationships

Education

- Understanding sexual health

- Education about drug use, risks, and available help
- Guidance for maintaining emotional and physical well-being

Government

- Receiving appropriate medical services and programs
- Having access to social support

Necessities

- Access to nutritional food, clean water and air, as well as sanitation
- Access to housing

These five domains contain the key data sources: insurance, public health, mHealth, Omics, environmental, familial, etc. It is important to distinguish the different domains that can directly affect one's health and in turn, identify opportunities for the application of predictive health analytics. The nature of health and well-being is complex and extends beyond the hospital walls.

The nature of health and well-being is complex and extends beyond the hospital walls. This is also the case for data sources and opportunities for applying predictive health analytics.

2.2 Understanding Data Diversity

2.2.1 Present and Emerging Data

The digital world is growing at an unprecedented rate. In fact, this growth is so explosive that industry research projects that there will be 44 zetabytes of digital data by the year 2020 [24]. The staggering volume of data is being generated by a diverse set of sources ranging from individual digital interactions with products and services, to sensors and machines.

To begin with, individuals are contributing an extraordinary amount of digital information about themselves every single day through social networking, creating online content, and engaging with systems, streaming video or other uses. On Twitter alone, users are sending more than 500 million Tweets a day [25]. The digital trail is continually growing on social media platforms, and is largely considered to be unstructured. However, this data is heterogeneous as it includes photos, videos, word documents, text files, service logs, e-mail communications, clinical notes, and of course, social media posts. This type of data is called human-sourced as it originates from human activity.

In contrast, machine-generated data originates from the technology itself rather than from a set of users. Examples can include logs from web servers or networks, call-detail records, and sensor recordings from devices found in everything from automobiles to security systems. Furthermore, wireless devices enable machine-to-machine communication, allowing the Internet of Things (IoT) to undergo rapid growth. By connecting devices and facilitating information exchange, real-time information such as motion detection, fuel levels, temperature, and pollution monitoring can give rise to smart cities, smart homes, smart workplaces, and so on. Similar to user-generated data, machine-generated data is voluminous in nature, the IDC forecasts it will account for up to 42% of all data present in 2020 [26].

Less voluminous in nature is structured data, which is considered to be predefined and able to fit neatly into a relational database. It can come from sources such as legal records, census data, financial records, lab results, library catalogues, and any kind of online web-forms users fill out online. This amount of structured data pales in comparison to the estimation that over 80% of the data in the world is unstructured [27].

In healthcare specifically, the diversity of data presents a challenge for healthcare professionals to take away actionable insights. This is seen in electronic health records (EHRs) that integrate genomic data, physiological data, demographics, diagnostic imaging, test results, written notes, and drug histories.

Moreover, the digital health footprint extends into wearable technology, telemedicine, and mobile health applications as well. It typically includes information about a patient's physical and mental well-being through the monitoring of heart rate, blood pressure, exercise patterns, blood sugar levels, chronic diseases, and calorie intake, amongst others. Biosensors are now being used in telehealth and medical equipment to promote improved quality of patient care, which is a form of machine-to-machine communication. The combination of this variety of structured and unstructured data in massive amounts in healthcare must be addressed in advancing the field of predictive medicine.

2.2.2 Standard Medical Terminology

An ontology is a way to express terminology, and has classes, entities, relations, and logic rules associated with it. Standardized ontologies in healthcare mainly pertain to medical terminologies. Ontologies organize concepts and terms, provide standards, and promote interoperability.

A medical terminology represents a related set of medical terms, whereas a controlled vocabulary concerns itself with organizing a collection of information in the form of words, providing non-redundant definitions. Medical terminologies are used to communicate medical information about procedures, diagnoses, patient symptomatology, and other data in a specific way for machines, such as health information systems, to interpret it in a standardized manner.

The goal is for terminologies to be universal so that each concept is represented by a corresponding code (or term) across all healthcare providers, regardless of which language or specific word was used to describe it. For example, it is important that 'heart attack', 'myocardial infarction', and 'HEART – infarction' are read as equivalent and point to the same code. The following are examples of internationally approved medical terminologies:

- SNOMED CT – clinical terms
- ICD-9 and ICD-10 – classifying diseases by diagnosis codes
- CPT – procedure codes
- LOINC – names and codes for laboratory and clinical data
- RxNORM – normalized drug notations
- DSM-IV – codes for mental health diagnosis
- NDC – codes for drugs assigned by the FDA

These medical terminologies are used throughout EHRs, computerized physician order entry (CPOE), clinical decision support (CDS), medical procedures, lab reports, and other clinical documentation systems.

Using them allows for accurate collection of data to be used for statistical reporting, billing information, public health monitoring, data sharing, interoperability, and improving or organizing medical knowledge.

2.2.3 Types of Data

When it comes to big data and healthcare, the different types of data that can be accumulated to perform healthcare predictive analytics has no bounds. Figure 2.1 shows different sources of data currently available in the healthcare realm. Data from one source can be used in stand-alone projects or can also be combined with other sources to discover patterns, associations and other predictive analytics. The opportunities are truly limitless and only bound by the creative and analytical minds of big-data analysts.

Although this is not an exhaustive list, this section lists some main categories of big data available in the healthcare realm that can be utilized to perform predictive analytics. Much of this data is already available and growing daily.

It is important to understand that although the data can be used directly from the source for healthcare analytics and predictive analysis, the real value lies when different types of data are integrated and studied in relation to each other [28]. Hence there is a clear need for ID service to allow data cross-linking, and the need for open access such that different databases can talk to each other, share information and stay connected in real time.

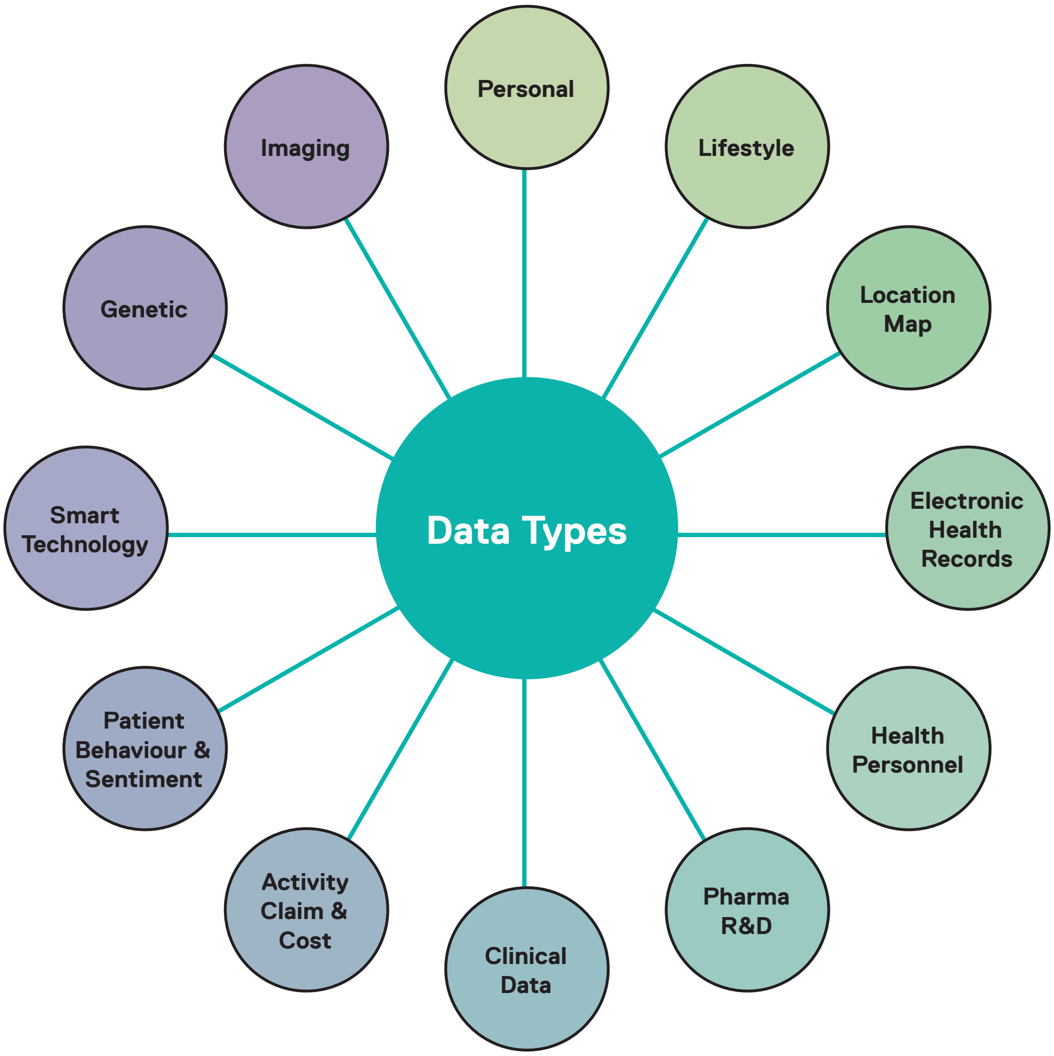
2.2.3.1 Personal Data

Examples of personal data are identity, gender, height, weight, ethnic background, etc. This information could easily be attached to all individuals' Electronic Medical Records or health cards as their primary required information.

2.2.3.2 Lifestyle Data

Lifestyle data is becoming increasingly available as we continue to embrace smart devices. Smart phones can already collect information on the amount of physical activity for a person using the sensors, and there are applications to record blood pressure, blood-glucose readings, calorie intake, etc. Figure 2.2 shows a screen shot taken from a Samsung Note 5's "Samsung Health" application. Some commercial companies also provide wearable devices to enhance their services to customers. For example, Manulife offered wearables to their life insurance customers to improve physical activities and healthy life style. Consumers can receive discounts for their insurance premium if they can reach their target. Data such as this could be grouped under lifestyle data.

Figure 2.1: List of Various Data Types

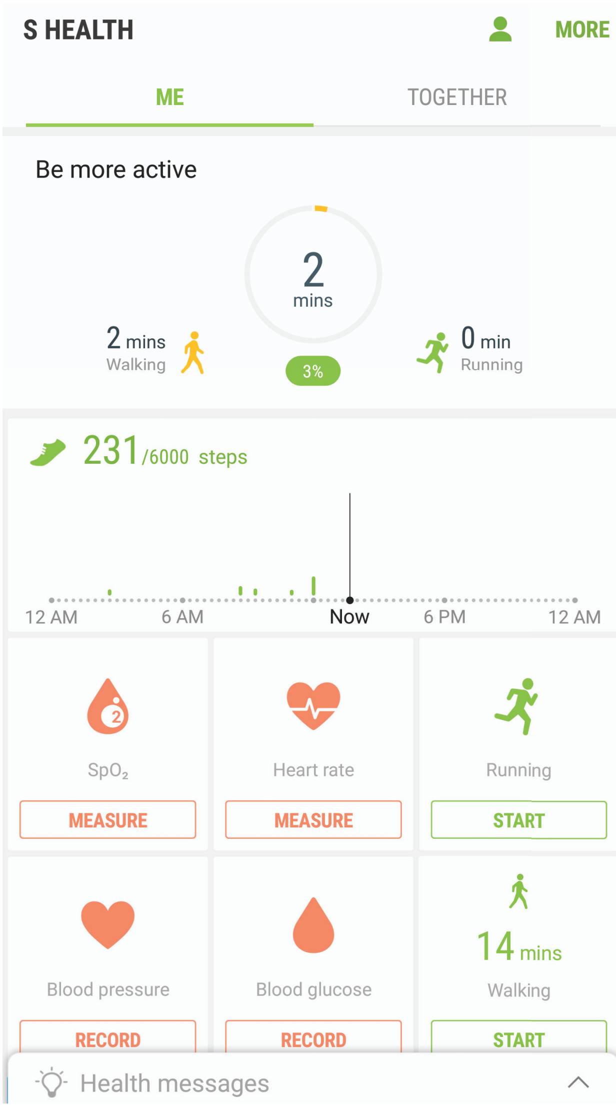


2.2.3.3 Location Data

Location data can be divided into two types. One can be categorized as static and the other dynamic. Static location data would include things like places of birth, residence, employment, travel etc. Dynamic location data can be defined as healthcare information tied to geographical coordinates and retrieved in association with a particular event or occurrence.

Smart phones can easily collect large amounts of data without us even knowing about it. If smart devices/applications are turned on, location data can be readily collected via GPS and combined with other kinds of data to make inferences and conduct predictive analytics in healthcare. For instance, in the event of a disease outbreak, location data from smart devices along with Twitter data with hash tags for the disease names and people's comments could be utilized to narrow down the area where the outbreak happened. It could also be utilized to identify areas with a higher density of a certain disease.

For example, consider a situation where a high percentage of the population in a given vicinity was suffering from the same disease, and the number of affected patients was increasing daily. This could happen in the case of an infectious disease, and location data of affected patients could help healthcare authorities assess risks, communicate directions and give timely warnings.



2.2.3.4 Health Record Data (Electronic Health Records)

Health Records include information including but not limited to doctor visits, tests, prescriptions, diagnoses, etc., along with commentary from medical personnel. Most of this information is readily available and included in Electronic Health Records (EHR), as we continue to embrace advancements in technology.

2.2.3.5 Health Personnel Data

Hospitals, doctors, pathology labs/pathologists and healthcare researchers can all be sources of data. For instance, hospitals can provide patient information, emergency ward records, demographic statistics of most-frequently visited patients, etc. Labs can provide details derived from samples tested for various diseases. Doctors, such as specialists, can provide medical records of all patients suffering with a particular disease. For instance, information retrieved from an endocrinologist's clinic can be used to study diabetes.

2.2.3.6 Pharma R&D Data

Pharmaceutical companies' research and development departments are a valuable source for clinical research information. This data can also describe a drug's therapeutic mechanism of action, target behaviour in the body, side effects and toxicity[28].

2.2.3.7 Clinical Data

It is expected that in the coming years, clinical data that today exists in hard copy, or at most scanned copies, will soon become electronic data. Having this information available in the proper electronic format will speed up information transfer, increase transparency and improve the ability of healthcare professionals to provide better service. It will also allow us to use this data for healthcare analytics.

To quote Ontario Genomics' 2016 paper, *Call for an Ontario Health Data Ecosystem*, [29], "At the moment, most clinicians in Ontario do not have ready access to their patients' comprehensive healthcare data. Basic information such as name, age and bare minimum clinical history must be re-recorded every time a patient enters a new healthcare establishment and critical diagnostic information (e.g. from a family physician to a specialist) is either not shared or is transferred with great difficulty."

2.2.3.8 Activity Claims and Cost Data

Insurance companies have medical records for patients claims, for which medicines and for how much. Data related to the claims and cost of medicines and medical equipment can be analysed to make predictions on future healthcare costs and trends.

2.2.3.9 Social and Sentiment Data

Patient behavior, preferences and sentiment data that describes patient activities such as patients' finances, buying preferences and other characteristics can be gathered from companies that sell consumer data [28]. Data related to patient exercise routines can be obtained via their smart devices, which come equipped with sensors. Financial data may not be directly used in healthcare, but it can be used to make certain inferences. For instance, a higher debt level could be a cause of depression, whereas, being financially stable could indicate that if a person is still depressed, then it could be due to some other reasons.

2.2.3.10 Smart Technology Data

Modern devices are more and more likely to include smart technology, whether it be a smart cell phone or a smart watch or even smart glasses. These devices, especially the wearable ones, can become the heart of our data ecosystem as most of them are equipped with sensors that can collect important healthcare information for predictive analytics.

Smart cars are also becoming increasingly popular, and it might be possible to equip these with sensors for weight check, vitals and eating habits. The idea of smart homes is also not too far-fetched: As real estate grows more expensive, architects have already started designing homes that are valued not by the square footage, but by the design and equipment installed in them. In the near future, it would be fair to expect that these smart homes will include built-in sensors that could gather lifestyle and health related data.

2.2.3.11 Genetic Data

Generation of the first human genome sequence required over a decade of work, and cost nearly $3 billion [30]. Advances in sequencing technologies have increased the capabilities for genome sequence generation by six orders of magnitude over the past several years, with the latest iteration of genome sequencing platforms capable of producing nearly 20,000 full genomes per year, at a cost of close to $1,000 per genome [31].

As a result, large-scale genome sequencing efforts are becoming more commonplace [32], with the potential to incorporate the sequencing of an individual's three billion base pairs of genomic data (the length of the human genome) to be integrated into each patient's clinical record. One of the main research focuses of the post-genome era has been to determine the role of genetic mutations, or variants, in human disease. This is typically done using one of two genetic analyses: genome-wide association (GWA), or genome sequencing.

GWA studies are designed to simultaneously test presence or absence of hundreds of thousands of known mutations for association with disease states. Participants in a GWA study do not have their full genome sequenced, rather a DNA sample from these patients is tested (or genotyped) for a pre-defined set of known mutations. To date there have been 2,176 published GWA studies linking over 20,000 genetic variants to 1,532 disease traits [33].

Genome-sequencing studies use advanced sequencing machines to scan/read a patient's genetic code, either within a specific region, or in its entirety. Genome sequencing studies are far more expensive than their GWA counterparts. However, since the entire genetic sequence is read for each patient, new mutations can be identified and associated with disease. The cost of large-scale genome sequencing studies have dropped in recent years, and today several public and private efforts are developing large repositories of genome sequence data [32]. To put this into the context of computational representation, currently the 1000 Genomes project [34]has completed full genome sequencing of 2500 patients, with a collective file-size of 464 Terabytes [35].

2.2.3.12 Imaging Data

As computational algorithms become more efficient at identifying image characteristics, digital medical images are increasingly seen as a resource for automated diagnosis. This section will describe some common medical-imaging tools and their use in diagnosis through advanced analytics.

Magnetic resonance imaging (MRI) uses analysis of radio frequencies emitted by atomic nuclei under a magnetic field to generate anatomical images. MRI technology is steadily increasing in prevalence as a diagnostic tool in the US and Canada, primarily for diagnosis of inflammatory or infectious diseases, musculoskeletal disorders, and cancer. Neurological assessment can also be performed using MRI, typically through the use of a functional MRI (fMRI), which identifies changes in blood oxygen levels as a proxy for neuronal activity. Initial surveys of neurological disorders such as Alzheimer's, anxiety/ depression, dementia and autism used fMRI results to associate specific brain regions with defects in information processing [36] [37] [38].

Ultrasound uses high-frequency sound waves to produce images of parts within the body. It is a non-invasive, non-radioactive means of visualizing muscles, tendons, and internal organs. Ultrasound imaging can be highly operator-dependent, resulting in inconsistencies in screening results. Recent applications of machine learning in the field of ultrasound include real-time evaluation of the malignancy of identified masses [39].

Mammography functions similarly to X-rays, using ionizing radiation to generate images of breast tissue. Based on these images, a radiologist will determine if a patient has a tumorous growth and should be recommended for further treatment. Unfortunately, identification of tumours can be problematic due to the nature of the images produced, resulting in a high false-positive rate. This causes a large number of unnecessary follow-up exams, resulting in undue stress to patients [40]. The use of computational algorithms to automate and improve mass detection from mammograms has been the subject of intense research over the past several years [41] [42] [43]. In addition to these examples, other technologies e.g. CT scans, are also widely used.

2.2.4 Managing Data Flow and Use: Cloud Technology

With the large amount of healthcare data potentially available for analysis, it will be a challenge to manage the flow of data efficiently, and it is critical that the solution include the ability to feed analysis results to multiple devices. There will be huge amounts of data to process (volume), a mixture of structured and unstructured data (variety), new data that is generated extremely frequently (velocity), and the need to manage data quality so that it can be trusted (veracity).

An enterprise technology solution must solve these four "Vs", with cloud platforms offering several distinct advantages to ease solutions-development.

In the healthcare space the variety, volume and type of data available for analytics purposes presents complex challenges. Data can range from traditional transactional data captured by existing online transaction processing (OLTP) systems and EHRs, real-time streaming data from sensors and devices, vast volumes of imaging and video data created by diagnostic equipment, as well as the massive amount of critical information that exists in paper records offline. This variety and scale poses unique challenges for organizations looking to leverage this data.

Often at the initial stages of an analytical project, a large exploratory phase is required. This often involves ingesting a large set of data, often of multiple data types to start to narrow down and understand where to focus the analytical efforts. This can mean the initial data set is significant in size, and may in fact be orders of magnitude larger than the data set that will be eventually utilized for operationalization of the analytics; e.g., researchers need to start with the whole picture of the data before focusing on the precise piece required for a project.

For these types of use cases, the elasticity of cloud computing offers significant benefits, versus purchasing and maintaining physical infrastructure. If existing physical capacity was not sufficient, investment would be required to increase the scale, storage and computing capabilities of the IT infrastructure, but this investment may not be fully utilized again as the project proceeds. With a cloud-based solution, the size, scale and computation capabilities can be scaled up and down on demand, without a large upfront investment. This also means that analytics practitioners do not have to make difficult decisions around deciding which portion of the data set to analyze.

The variety of data types required for this type of analysis also offer challenges that are well suited for the capabilities of a cloud platform. A flexible cloud offering allows for agile spin up and spin down of different data storage technologies as required for project progression. For example, at the beginning of a project, a researcher might need to spin up a large Hadoop instance to store and process image data, but could then extract the insight from this large data set and then store the refined data set in a simpler format in a cloud SQL environment. This historical large-image data could then be stored in a low-cost archiving solution in the cloud, and the Hadoop cluster spun down to reduce ongoing costs.

This would mean that the customer would not have to pay upfront for perpetual licenses for each technology type that may only be used for a portion of the project. Instead, they would pay only for the usage of each cloud service as it was being utilized, which would allow for agility and costs savings across multiple varieties and types of data over a project lifetime.

Long-term data storage is also simplified and made extremely cost effective via a cloud platform. Long-term storage is a relatively inexpensive capability offered by most cloud providers, with stringent service-level agreements (SLA) around data retention, redundancy and access. Data-retention and disaster-recovery strategies are frequently offered as part of end-to-end cloud services, freeing the end user organization from concerns about off-site backups or a local storage and recovery investments. This significantly streamlines the process of long-term data storage for an organization, and offers more flexibility for the volume and range of data that can be archived securely for long-term retrieval.

2.2.5 Computational and Analytical Capabilities

Many types of analytics that are relevant for processing data in healthcare have complex computational requirements. Cloud-computing platforms offer compelling advantages for organizations looking to drive large-scale analytical efforts in an agile and cost-effective fashion.

Another key advantage that a cloud platform offers is centralization of key data in a single repository or set of repositories. This offers flexibility to an end-user organization to enable their different analytics teams to access a common dataset at scale and offers the freedom to leverage the specific analytical tool that is relevant for their use case without needing to provision multiple analytical data environments across an organization. Leading cloud providers offer a multitude of analytical toolsets, including advanced machine-learning capabilities, real-time streaming data analytics, statistical analysis capabilities (e.g. R or Python) as well as big-data storage and computational abilities such as Hadoop.

The usage-based models and elasticity of analytics services in a cloud platform enable organizations to realize cost benefits, by paying for tools and capabilities only as they are needed, which ensures that investments in analytical capabilities are being leveraged when the actual work is being done.

2.3 Basic Use of Electronic Health Records

2.3.1 Background: Electronic Medical Records and Electronic Health Records

Electronic Health Records (EHRs) are digital versions of a patient's chart and are becoming more and more mainstream in ambulatory care practices and in inpatient care. EMRs contain a patient's medical history, diagnoses, medications, treatment plans, immunization dates, allergies, radiology images, and laboratory and test results. While an EMR contains the medical and treatment histories of patients, an EHR system is built to go beyond standard clinical data collected in a provider's office and can be inclusive of a broader view of a patient's care.

EHRs allow the digital information to be accessible through security layers to authorized providers across a single or multiple health organizations. EHR systems can allow information to be shared with and added by other healthcare providers and organizations, including laboratories, specialists, medical imaging facilities, pharmacies, emergency facilities, and school and workplace clinics. EHRs have the potential to contain information from all clinicians involved in a patient's care.

This transition to digital records also represents an important change in the way patient data is organized and made accessible, in ways that paper records never could have been used. For example, EHRs offer access to longitudinal data that may be useful to predict future outcomes or diagnoses, opening opportunities to personalize decision making for a given patient with the help of big-data tools and analytics. In parallel, over the past decade, powerful, modern machine-learning techniques have evolved and offer the potential to rapidly extract information from large, high-dimensional data sets for population health studies.

2.3.2 Data Integration

Electronic Health Record systems often collect different data elements and store them in different formats, making their combined use for analytics difficult and sometimes impossible. For example, one system may collect specific patient demographics such as smoking, religion etc., in one format, while another software system may collect them in a different format or not at all. To address this issue, governments and payers have introduced a standard data set called "Meaningful Use" (MU) data.

In healthcare, MU defines minimum standards for using EHRs and for exchanging patient clinical data between healthcare providers, between healthcare providers and insurers, and between healthcare providers and patients.

While MU is voluntary, it is often referred to as a carrot-and-stick program in which there are penalties that provide a strong economic compulsion to participate. Penalties against Medicare and Medicaid reimbursements for skipping MU will increase in each successive year, expressed as a "payment adjustment" or reduction of a provider's reimbursement for care provided to Medicare or Medicaid patients. If fewer than 75% of eligible providers have become meaningful users of EHRs by 2018, the adjustment will change by 1% point each year to a maximum of 5%.

The standard format that is used for exchange of data between EHRs and other IT systems is known as "Health Level Seven International" (HL7). Both HL7-sanctioned, and XML-based Continuity of Care Document (CCD) formats are used to transmit data to Health Information Exchanges (HIE) in the US and Health Links in Ontario.

2.3.3 Long-term Preservation and Storage of Records

Once the EHR has collected the data, there must be a plan for the long-term preservation and storage of these records. The healthcare industry needs to come to consensus on the length of time to store EHRs, methods to ensure the future accessibility and compatibility of archived data with yet-to-be developed retrieval systems, and how to ensure the physical and virtual security of the archives. Even though standards like PHIPA and Privacy By Design exist, application of these standards to EHR is sometimes driven by the subjective interpretation of the authorities involved.

Storage considerations are complicated by the possibility that the records might one day be used longitudinally and integrated across sites of care, and possibly used to conduct big-data analytics. Key stakeholders of this data include primary care physicians, hospitals, insurance companies, and patients, and the decisions around preservation, storage and ownership will have a significant impact on the accessibility and privacy of patients' medical records [44].

Data governance and storage regulations will have to be developed, and to ensure success in supporting those regulations, it will be important to consider the needed technology architecture carefully. Ruotsalainen and Manning [66] found that the typical preservation time of patient data varies between 20 and 100 years. In one example of how an EHR archive might function, their research "describes a cooperative trusted notary archive (TNA) which receives health data from different HER systems, stores data together with associated meta-information for long periods and distributes EHR-data objects. TNA can store objects in XML-format and prove the integrity of stored data with the help of event records, timestamps and archive e-signatures."

Digital Health Records' life expectancy will likely exceed the average shelf-life of paper records. Given the rapid evolution in technology of EHRs, the systems used to input information will likely not be available to a user who desires to examine archived data. Some authors have proposed a standardized framework for information fields in a time-invariant way, such as with XML language. Olhede and Peterson [45] argued that "the basic XML-format has undergone preliminary testing in Europe by a Spri project and been found suitable for EU purposes. Spri has advised the Swedish National Board of Health and Welfare and the Swedish National Archive to issue directives concerning the use of XML as the archive-format for EHCR (Electronic Health Care Record) information."

2.3.4 Synchronization of records: Access and Analytics

When care is provided at different healthcare facilities, it may be difficult to update records at both locations in a coordinated fashion due to the differences in systems and location of the server stack that is storing the data. Two paradigms have been used to satisfy this problem: a centralized data server solution, and a peer-to-peer (P2P) file-synchronization program (as has been developed for other P2P networks).

Without the standardization of electronic record format, synchronization programs for distributed storage models will not be useful. This may be due to the inconsistencies of data types that cannot be combined into a single source of truth, or due to the complications created by aggregating data into a single database. Merging of already existing public healthcare databases is a common software challenge. The ability of electronic health record systems to provide this function is a key benefit and can improve healthcare delivery [46] [47] [48].

Merging of healthcare data also brings a higher need to protect the data, as any breach or misuse of the consolidated data would be in severe breach of patient privacy. A strong governance model helps assure that such merged healthcare datasets are shared securely with the appropriate users according to security and privacy policies, and a strong audit trail is maintained.

Moluptationem fugianimolor aute et minulle nd-untio. Uptium fugita veritat ionecatio deribus as im esseque di doles nis sam, ut rerat ut fugitae essiminctus ped ut abor audanti aectate invenih iligend electur, vera il ius andam eatiate et lab ium ex eium.

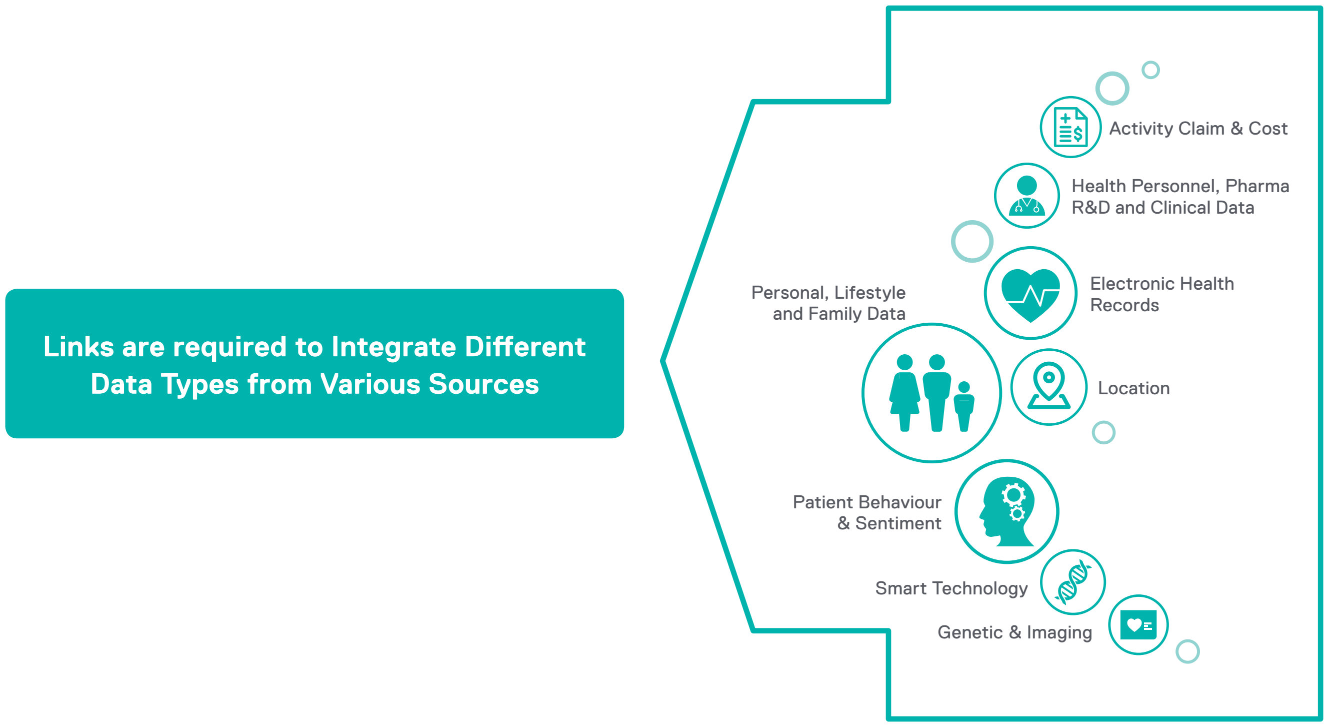
2.4 Links

Figure 2.3 shows different forms of data types available that can be included in the Data Ecosystem. The benefit of each is magnified when integrated with the others, hence the need for open access and the use of appropriate links to can connect them and create a hub of integrated healthcare data [28].

To facilitate communication and integration among all data sets, despite their differences and to being able to follow up efficiently with patients' conditions, Dieng-Kuntz et al., proposed a conceptual framework for managing data across a healthcare network [49]. As shown in Figure 2.4, the framework relies on semantic web technologies that analyze the database and decode it into an ontological framework known as the Resource Description Framework (RDF). RDF is a foundation for processing metadata, as it enables interoperability between applications that exchange information on the web [50].

Big data is more meaningful and useful if and when properly linked. This concept is crucial in order to understand and maintain a healthy data ecosystem. Using a human body as an example, "Humans have five vital organs that are essential for survival. These are the brain, heart, kidneys, liver and lungs. The human brain is the body's control center, receiving and sending signals to other organs through the nervous system and through secreted hormones" [51]. Although these five organs are essential on their own for survival, human bodies will not be able to talk, walk, smell, move and perform many other complex operations without a balanced coordination or connection amongst these organs.

Figure 2.3: Integration of Data Types via Meaningful Links



The same concept applies to the data ecosystem; having disconnected data types from disparate sources offers limited benefits.

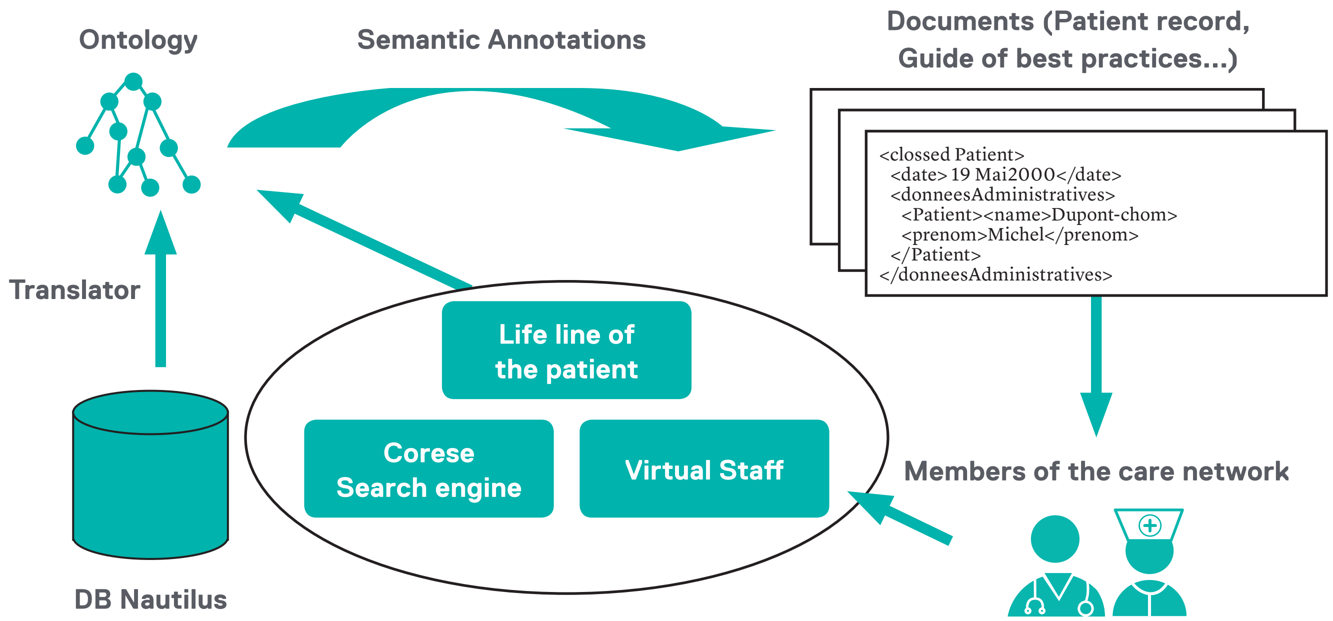
Patient data from different databases can be linked by a unique identifier, such as the Canadian health card. This would allow for quick lookups and fast retrieval of information. Similarly, there is also a need for a central database that stores the mapping of all commercial drug names to their chemical compounds. Keeping all linked compounds in one place would also help identify alternatives for drugs and perform analysis on different drug types and related outcomes with the use of those drugs. In a situation where only the drug's commercial name is available, the database can be queried to find the compound and all other drugs that have the same compound but a different commercial name.

Environmental factors can also be tied to location data to look for patterns while conducting predictive analytics. For instance, the outbreak of a disease can correlate or co-vary with certain environmental factors. And as with environmental factors, companies and their location data can be grouped together and tied to information available from work safety health records. This could help identify companies that do not follow rules and regulations, and trigger disease in their employees through exposure to unhealthy environments.

Having environmental and company information linked to personal and location data can enable predictive models to issue warnings promptly for locations at risk.

With data linked between different sources, predictive healthcare analytics can be performed to find patterns and associations, e.g. between sedentary work and certain types of employment and health factors.

Figure 2.4: A Semantic Web Framework for A Healthcare Network



2.5 Long-Term Data Management

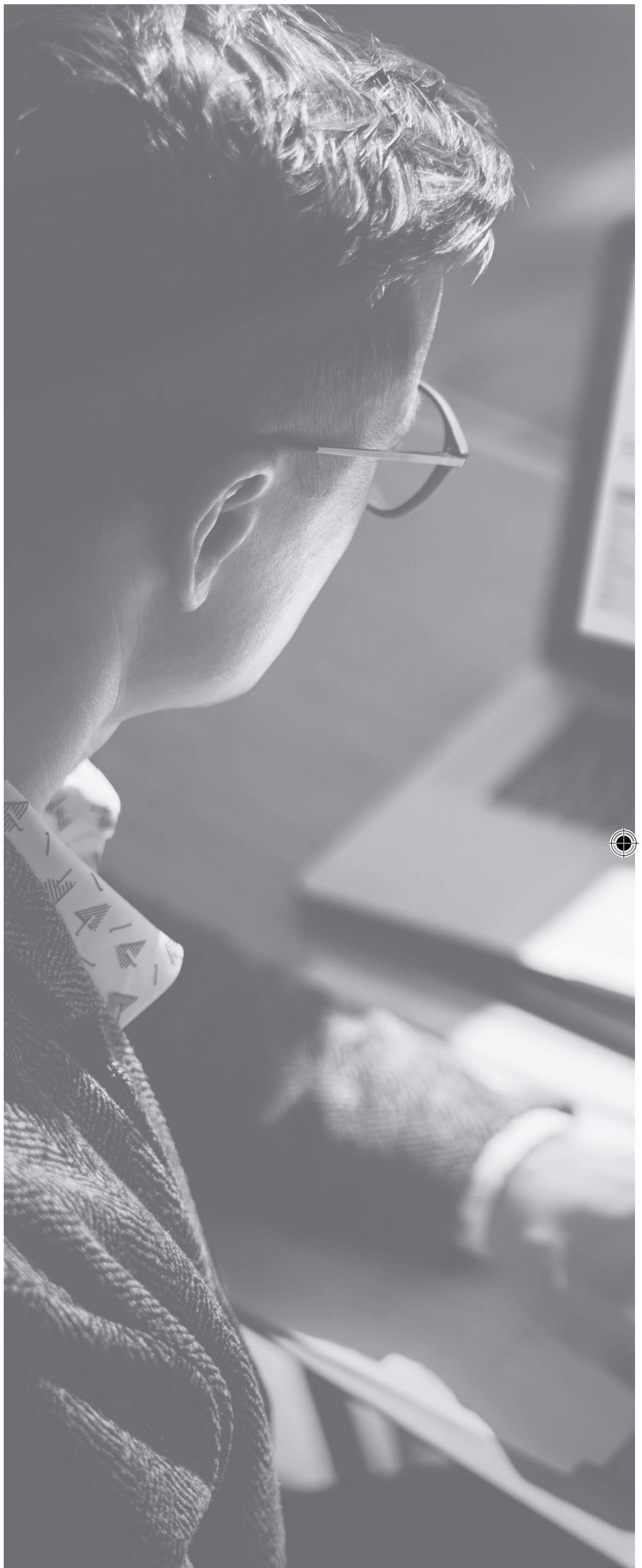
2.5.1 Technology Roadmap for Cloud Use in Healthcare

Cloud computing is still a term that may be unclear for many, however, organizations have been leveraging centralized online services for distributed services to millions of users for over 15 years, such as email services like Gmail. Over that time, there have been massive improvements in the scale, scope, depth, and reliability of cloud services to offer computing capability at scale for practically any workload.

At a high level, leading enterprise cloud platforms offer access to a massive, highly secure, centralized infrastructure on demand, and with complete control over the data. These platforms offer a wide range of services and typically fall into three general categories. Across all three categories an enterprise-grade cloud provider will allow the end user full control and access over the data, processes and applications in the cloud infrastructure.

2.5.2 Software as a Service (SaaS)

For this category, an organization or user accesses an application that is running on the provider's cloud infrastructure. This includes Microsoft Office 365, a productivity and collaboration application, or Gmail from Google. Organizations can also access cloud infrastructure to develop custom applications by leveraging building blocks offered by cloud providers to build customized solutions. A healthcare technology provider could easily leverage a cloud-based database service, application layers, analytical tools and visualization capabilities to build customized health system applications [52].



2.5.3 Infrastructure as a Service (IaaS)

Infrastructure as a Service allows users and organizations to directly access servers, networking and storage similarly to the way they would use physical hardware, and be able to further leverage these servers by running their own applications or storing their own data. Cloud computing across the healthcare space could allow for more effective collaboration for staff operating in multiple locations, for scalable applications to be delivered simultaneously to thousands of users with a consistent user experience and dataset, or even routine infrastructure needs such as affordable and secure options for long-term data storage and backup.

However, the one broader set of healthcare use cases especially suited for cloud computing is as a platform for advanced analytics, including predictive analytics.

2.5.4 Control: Privacy, Security, Access

One of the most critical issues with handling any kind of data in healthcare is the ability to safeguard patient privacy. When using healthcare data for analytical purposes, the patient's right to privacy is paramount throughout the entire data-use cycle, beginning with initial collection. This concern is a cornerstone to the delivery of care throughout traditional health delivery organizations (HDOs) but now also applies to new actors in the market, specifically decentralized providers of care, such as mobile applications and wearables manufacturers.

Privacy cannot exist without data security and access control, which are essential given the sensitive and personal nature of personal health information

(PHI). All HDOs and private players must abide by the appropriate privacy acts, legislation, and policies to protect PHI that is both in paper and electronic form. Steps that can be taken in a privacy-by-design approach for information systems include applying a proactive and preventative approach [53]. Specifically, privacy and security should be promoted from the very beginning through secure perimeters, proper access control, breach protocols, threat detection analytics, educational sessions for staff, continuous auditing/monitoring and the guidance of strong leaders within the organization. The adoption of good privacy and security practices from the start will ensure a vote of confidence and trust from patients and stakeholders, avoid damage to the organization's reputation, and avoid potentially disastrous incidents.

Moreover, the selection of a cloud provider for healthcare analytical workloads requires careful evaluation of the providers' security, access and data-handling procedures, as well as their compliance with governmental and legislative requirements. Leading cloud providers, such as Microsoft, IBM, and Amazon, have made significant investments to ensure their cloud offerings comply with all required privacy and regulatory requirements. As well, a significant investment has been contributed to the overall security infrastructure of their offerings across all dimensions.

Fortunately, cloud service providers can scale significant security measures across the entirety of their cloud, which offers additional savings to organizations, versus hosting their own infrastructure. For example, this may include ensuring all infrastructure software is up to date and patched against potential threats, and wrapped in multiple layers of network and access security measures. Centralization of key data stores in a single significantly protected cloud infrastructure offers other security benefits in terms of simplifying access to a single source of data versus multiple different, potentially less secure data stores.

Finally, leading cloud providers are making large scale investments in Canadian-specific data centre infrastructure, ensuring organizations to have the option of hosting their data on Canadian soil. Microsoft, for example, has launched two data centres in Canada to date, and other cloud providers have similar plans. This means Canadian healthcare organizations can host and store data with full control over the physical geographical presence of the data.

Centralizing key data storage and processing in a cloud infrastructure can offer further significant security upside versus storage of data across local on premise data centres.

2.5.4.1 Privacy in Predictive Health Analytics

In an era in which the potential for predictive analytics could hold tremendous promise to improve healthcare, the traditional thinking around collection, use and disclosure of personal health data has come into question. Current privacy laws were created when data collection occurred in a paper-based system, and therefore the type and amount of data, as well as the ease of sharing the data was much more limited than what is possible today. Although opportunities for game-changing data uses evolve rapidly, private-sector innovators encounter serious barriers because the legislation generally limits data use to health system management and research purposes [54].

Indeed, advances in digital technology have increased the number and variety of data sources that may be useful for predictive health analytics and have dramatically improved the ease of use. One obvious source is clinical data. Over the past several years, the healthcare sector has modernized by digitizing health records and creating the interoperable Electronic Medical Record (EMR).

Other data types anticipated to be of use in predictive analytics include: open data (e.g. the data released publicly by government and other organizations); personal data (e.g. monitoring done through handheld and sensor-based wireless or smart devices); and government data (e.g. detailed healthcare system costs). But dispersed data sources mean that governance structures, as well as interpretation and application of privacy legislation can vary significantly and therefore also pose access challenges [55].

Even in cases where efforts to harmonize data have occurred (e.g. EHR), access to the data is often limited to the patient and relevant healthcare providers, though researchers may gain access if consent is given at the time of collection [56]. However, for the private sector, this is challenging if the data has a secondary use that is unknown when the patient is asked for consent, and it is often impossible to obtain consent at a later date. De-identification of data is one strategy that may help improve access while protecting privacy, but because de-identification processes are not well harmonized, health data is still not readily accessible to private sector businesses in meaningful volumes.

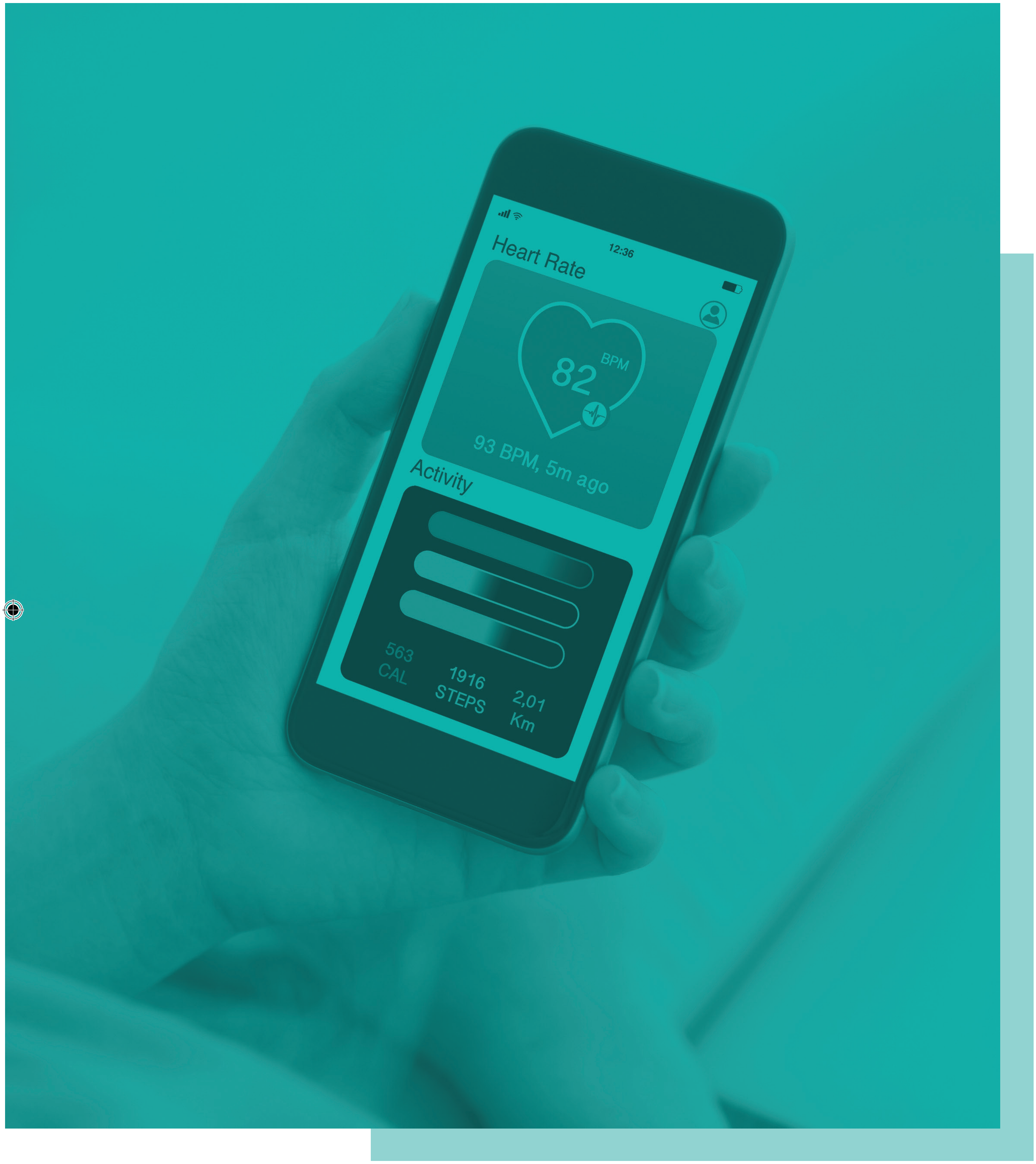
As reported in the "Health Analytics for Informed Decision Making: Health System Use Summit," supports and mentorship are required to help technology start-ups and innovators succeed and overcome structural barriers, security requirements and data sharing options [57].

In this white paper, we support the notion of systematic governing of data, but also believe that data access should be universal and linked, as long as organizations do their due diligence and meet legal, policy and ethical requirements. Facilitating open access to data by a wide variety of private-sector innovators will increase the potential of predictive health analytics to contribute to improving healthcare.

3 A Framework for Predictive Health

In the near future, technological advancement will be evident in the availability and breadth of health data. All this readily available information has immensely transformed our perspectives, especially when it comes to healthcare. Today, most of us wear some form of gadget or use devices such as smart phones that enable gathering of vital health data far more easily than it has ever been before.

If we keep up with the pace, then it is likely that, with the help of predictive analytics, we will have the power to send out timely alerts before the onset of health challenges, help patients monitor and manage their health in real-time, predict the likelihood of chronic diseases such as heart attack or stroke, or even inform individuals about good or bad food choices based on their recorded daily food intake information and physical activity details.


Heart Rate
12:36
82
BPM
93 BPM, 5m ago
Activity
563
CAL
1916
STEPS
2,01
Km

3.1 The Framework

As discussed in the previous chapter, this will not be possible without unobstructed availability of data and its proper linkage. In this chapter, we take a step further, and assume that the data will be readily available and the data ecosystem would have developed well with proper linkage and integration in the near future.

3.1.1 Data

1. Current State

   Canadian healthcare facilities are not yet universally using EHRs. Since 2001, Canada Health Infoway has been granted $2.1 billion in federal funding, used primarily to create an EHR for every Canadian, according to the original EHR solution (EHRS) blueprint [58]. In 2016, it is estimated 93.8% of records in Canada are electronic [59]. However, recent data published in 2015 about EMR use in Canada has shown that there is variability across provinces; the adoption rates by physicians range from a low 64% in New Brunswick to high rates of 83% in Ontario and 87% in Alberta [60].

   This variation is due to the number of non-interoperable systems, lack of required functionality, poor ease of use, and differences in provincial funding [61]. One key success has been the adoption of Picture Archiving and Information Systems (PACS); it is estimated that 99% of imaging in Canada is now digital [62].
   Predictive analytics applied in clinical settings is in its infancy stage. Data currently exists from a wide variety of healthcare sources but can be inaccurate, missing, and not integrated with other information. Further, there is a lack of standardization of health data, ranging from medical terminologies to clinical notes; this problem is not unique to Canada and is faced by healthcare systems universally.

2. Challenges

   - Interoperability between information systems
   - Open access and sharing
   - Not enough patient engagement through remote monitoring, mHealth solutions, participation in care
   - Patients lack access to their own data
   - Society has not realized true value of their own data
   - Lack of systematic and easy ways for health data capture

3. Opportunities

   The value of health data is widely recognized and can be put to meaningful use in an actionable way. Data will be generated by and captured from machines, medical services, the environment and by the patients/consumers of health services themselves. Ultimately, individuals are the owners of their own health data, and in this sense, the power balance will have been reshaped. Rather than individuals paying service providers, the service providers should be paying for access to people's data.

   Accessibility and data sharing will no longer be obstacles for research, clinical decision making, or for patients' right to their own data. The private

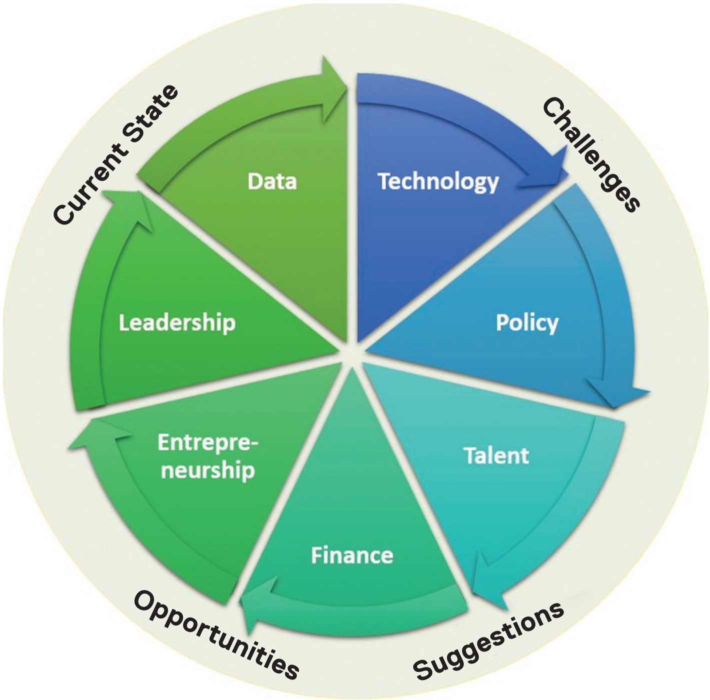


Figure 3.1: Proposed Framework

sector will be able to capitalize on this data accessibility and produce new products and analytics services.

4. Suggestions

- Ensure data quality.
- Link and integrate data from different sources.
- Implement healthcare enterprise data warehouses.
- Use data 'sandboxes'[1] that are specific for advanced analytics.
- Implement patient portals for access to own health information.
- Be open to sharing datasets with public, academic, and industry partners.
- Have strong advocacy for open science.
- Avoid pursuing patents on research discoveries.
- Protect privacy and confidentiality.
- Support standardization initiatives.
- Enforce adherence to standardization practices.
- Use data to make informed decisions in healthcare.
- Use data for measuring compliance, improving quality of care and patient safety, and for reporting.

[1] Sandboxes are infrastructures to isolate a system during testing in order to avoid impacting systems in use [63]

3.1.2 Technology

1. Current State

Provinces and territories have data warehouses that feature varying degrees of data integration. The HL7 v3 standard which focuses on the application layer in the OSI will soon be replaced by Fast HealthCare Interoperability Resources (FHIR) to facilitate greater system interoperability. Currently, there are different deployments of pilot and full-scale projects for health technologies ranging from EMRs, remote monitoring, to mHealth applications. Cloud technology is now being proposed for solutions and evaluation in the current healthcare landscape, especially for data storage, computing power, and providing applications.

Many hospitals and clinics are using information systems from different vendors and the degree varies across Canada and efforts are being made to set specifications for EMRs and data content. Moreover, Canada's IT infrastructure for research and compute-heavy applications is thriving with high-performance computing from Compute Canada, with high-bandwidth availability and low latency networks [64]. There has also been an immense reduction in the cost of human genome sequencing and a rise in data platforms that provide access to researchers and clinicians. In terms of predictive health analytics, big data is commonly talked about in the healthcare sector and applications for its use are being developed and applied (early stages).

Canada Health Infoway has recently developed a blueprint for health analytics which will be published in the coming year. Example applications have so far included: predicting high-cost/high risk health users, patient outcome prediction (preliminary stages), and healthcare forecasting for wait times, to name a few. Lastly, Canada Health Infoway is working towards integrating analytics capabilities within the EHR solutions to establish analytics in point of service.

In 2016, Canada Health Infoway proposed that there are 5 levels to analytics maturity in healthcare. Level 4, "maturity", is the stage when near real-time and predictive analytics could being used [65]. The fifth level, "advanced", was described as a state when prescriptive and predictive analytics are used and are patient centric; this is the end goal for Canada [65].

2. Challenges

- Lack of technological infrastructure
- Lack of funding to deliver and maintain technology solutions
- Security and privacy breaches

- Inadequate clinical training/education for technology use
- Workflow incompatibility
- Poor ease of use
- Lack of physician buy-in

3. Opportunities

There will be a great culture of innovation in Canada with technology solutions being used to improve preventative care as it stretches out into homes and communities, and tracks patients in real-time. Individuals will have holistic medical records that integrate with different health applications and include full genetic profiles. Cloud technologies will be used by applications and storage systems to ensure high scalability, low cost, and accessibility. Data mining, machine learning, and artificial intelligence will be applied to provide evidence-based decision making and predicted patient outcomes. Further, IoT will stretch across multiple areas of life to pull and analyze data relevant to the health and well-being of patients.

4. Suggestions

- Identify growth opportunities within organizations.
- Focus on scalable solutions.
- Invest in cloud technology.
- Do workflow evaluation to help with workflow redesign.
- Involve clinicians in implementation using bottom-up approach.
- Designate super users for new technologies
- Have a physician champion to voice concerns and involve staff
- Work with those most resistant to change
- Perform threat risk assessments (TRAs) and privacy impact assessments (PIAs)
- Use privacy by design approach when designing technology
- Leverage existing technology resources when possible
- Design analytics into all ICTs and information systems

3.1.3 Policy

1. Current State

From a legal and socio-ethical standpoint, there are policies and laws that exist to address consent, data collection and use, privacy, confidentiality and ethics (particularly in research). Often clinicians, researchers, and private players who require access to data need to jump through different hoops, including ethics review boards, privacy officers, and health information custodians (HICs).

2. Challenges

- Poorly structured best practices when it comes to data access and sharing
- Fragmented decision-making
- Lack of collective governance responsibility
- Accountability for implementation and use of technologies often unclear
- Liability can be in question too
- Disagreement about coverage for healthcare services (e.g. whether these services should be extended to community where predictive health tools could be used?)

3. Opportunities

There will be regulation and oversight surrounding the use of predictive analytics in healthcare. The governance structure for predictive health in Canada will focus on long-term organizational goals, setting up strategic plans, vision, and training healthcare leaders to ensure successful adoption for new technology. There will be legislation in place that addresses the proper use, collection, access, and sharing of health data, in line with privacy and personal information laws. Policies will also outline data content requirements and standards for interoperability.

4. Suggestions

- Push for policy coverage for home and community care.
- Ensure that new policy changes are in compliance with Canada Health Act.
- Work with stakeholders to bridge data policy development and healthcare industry.
- Expand regulatory and privacy legislation to field of mHealth and consumer health.
- Adhere to privacy and security guidelines for technology.

- Work with privacy officers to define policies in hospitals.
- Provide staff training on privacy, security, confidentiality, data sharing.
- Review HIAL for privacy and security services for personal health information protection.
- Use safeguards for the use, collection, and disclosure of personal health information.
- Actively promote transparency with use of patient data.
- Support government-sponsored big data initiatives.
- Support change of legislation to promote data accessibility.
- Policy-makers should use educational programs to attract students to learn about big data.
- Provide training to employees/students to increase the expertise in predictive health analytics.
- Work to increase the supply of data scientists.

3.1.4 Talent

1. Current State

Canada's Big Data Consortium estimates that there is a huge talent gap across all sectors for trained analytics professionals, around 10,500-19,000 workers [12]. The skills gap is being addressed by training programs and certificates that are available across the country to produce candidates for positions as data scientists, chief data officers, and data solutions architects [12]. For instance, Ryerson University offers certificates in Data Analytics, Big Data, and Predictive Analytics as well as in Health Informatics, and a Master in Data Science.

There are also program offerings such as the Master in Big Data program at Simon Fraser University and the Master of Management Analytics program at Queens University. Canada's Big Data Consortium is actively working towards strategic recommendations to advance data literacy from a young age, to encourage the development of analytics programs, to have employers define data scientist responsibilities, and for organizations to prioritize analytics in their business operations.

Additionally, a human resources outlook for health informatics positions from 2014-2019 projects that there is a high risk of a skills shortage, and that there will be a high replacement demand over these five years [15]. The skills gap is present and looming, helping to drive action by the government, industry, and academic partners.

2. Challenges

- Lack of skilled professionals in big data analytics and predictive analytics
- Few professional training programs exist
- Need greater advocacy for data scientist training
- Organizations are not identifying in-house talent
- External recruitment overlooks talent that can be found internally within organizations
- Academic institutions are not prioritizing changing the curriculum to include this new field
- Employer demand is high for people with analytics skills

3. Opportunities

There will be a strong analytics culture in every healthcare organization and/or facility across the country with EHRs demonstrating full analytics capabilities for real-time prediction. Students from kindergarten through grade 12 will be taught analytical skills and how to use different tools in cloud-based systems for learning hands-on skills [66]. Higher educational programs will have dedicated data science departments and course offerings specific to predictive analytics. The labour market forecasts, as well as increased demand from employers, will drive the need for change and increased numbers of skilled professionals; the healthcare industry will be thriving off data-driven solutions carried out by health data scientists and informaticians. Job growth will continue to boom for big-data analytics as more and more data will be exponentially produced. Open datasets for big-data competitions in the field of healthcare will be more prominent, and citizen data scientists will continue to help solve industry problems. The vast talent pool of people who are interested and willing to solve health issues using predictive analytics and other data related approaches is an increasingly source of innovation and economic growth.

All in all, the vision is that healthcare will no longer be lagging behind in the full-scale application of predictive analytics.

4. Suggestions

- Leverage existing talent.
- Increase training programs and educational certificates.
- Academic institutions must keep curriculum up-to-date.
- Identify talent internally within organization.
    - Government should strategize and increase funding, create policies for growth, and help open the door for new programs.
- Government, industry, and other public partners must provide input to structure curriculum.
- Labour market must be analyzed continuously.
- Use a shared-services model.

Open datasets for big-data competitions in the field of healthcare will be more prominent, and citizen data scientists will continue to help solve industry problems.

3.1.5 Financing

1. Current State

Healthcare funding is currently 70% and 30% from the public and private sector, respectively. The healthcare system is a $228 billion industry with expenditures reaching about 11% of Canada's gross domestic product GDP [67]. Physicians are generally paid using either fee-for-service, capitation, or salary-based models.

2. Challenges

- Willingness and ability of patient to pay for care
- Deciding which services will be insured
- How to maintain a sustainable healthcare system with growing health expenditures and an aging and sick population
- Choosing the right remuneration model for physicians
- Agreeing on the correct ratio between public and private sector funding

3. Opportunities

The funding model should include high investment by the government through public funding of the healthcare system. Specifically, this funding will be directed to address the most pressing necessities – primary care, community care, research and development - with the private sector being heavily involved in providing additional health services that cannot be covered by traditional means. Health spending will continue to grow as more investments are made into new health technologies and social determinants of health. Cost saving and operational efficiency is also an important part of the equation. Both improve funding and efficiency are needed for a well operated system.

4. Suggestions

- Invest in large-scale distributed data processing platforms.
- Use data to understand costs and impact on margins.
- Invest in cost-effective scalable solutions.
- Use analytics to inform decision-makers about business insights.
- Identify patient cohorts that are high-risk to healthcare system and provide appropriate treatment.
- Use analytics to identify fraud and abuse of healthcare system.
- Incentivize hospitals for technology use.
- Penalize hospitals for high readmission rates, poor use of resources, etc.

3.1.6 Entrepreneurship

1. Current State

The life-sciences sector is starting to capitalize on the application of big-data analytics, especially in the pharmaceutical industry, research community, and in biotech. This provides great opportunities for new players for start-ups and established companies to develop innovative ideas. There are currently many grants, competitions, and support, particularly from the government, to spur innovation in the healthcare sector.

The mHealth field is of particular interest, as new applications are being created and used for different chronic diseases, healthy lifestyle tracking, mental health, etc. Vendors are developing new EMR systems that can be open-source, customizable, secure, and hosted in Canada. These solutions are being tailored to meet the needs of care providers and facilities. The industry is focused on providing analytics services in many products such as IoT, mHealth services, and information systems (i.e., EMRs). The high demand for these technologies is driving the up-and-coming supply.

2. Challenges

- Deciding how to pay physicians for using different technologies
- Building solutions from top-down perspective
- Difficulty scaling out across wide regions (lack of transferability)
- Attaining physician buy-in

3. Opportunities

Increased funding by the private sector will go towards stimulating digital growth in the healthcare system. Incubators that capture all three domains of business, healthcare, and technology will be present to develop holistic solutions. Vendors will provide interoperable systems that integrate well with existing health solutions. In addition, a platform will exist to allow applications to be developed on top of it for easy accessibility and standardization. Entrepreneurs will have thought of creative ways to extend care into the home and community, offering a more decentralized delivery model for a sustainable healthcare system. As more data will become available, large-scale computing applications can help to deliver personalized medicine as well.

4. Suggestions

- Recognize successes and failures to deliver successful solutions.
- Collaborate closely with partners in academia, industry, and other public players.
- Always design technology with the end user (patient, clinicians, consumer) in mind.
- Introduce analytics capability in vendor solutions.
- Focus on bringing value to providers.
- Make sure that new solutions do minimal disruption to clinician workflow.
- Evaluate the existing technology to ensure successful integration.
- Establish shared governance principles to execute vision.

3.1.7 Leadership

1. Current State

There is leadership from the top government and non-profit organizations who are helping to guide Canada in investing in new health technologies. Scientific exploration continues to be encouraged and carried out by leading academic institutions and researchers. There is a national advisory council, an international liaison, and various consultation efforts to target the growth and strategy for Canada's digital infrastructure, data management, and research goals [64]. In addition, there are new and upcoming educational programs to address the skills gap and to better prepare healthcare workers for technology solutions and data analytics skills.

2. Challenges

- Not creating a clear vision
- Not incorporating short-term wins and letting momentum die early on
- Lacking proper governance structure and accountability
- Restrictions to use of data (accessibility, sharing)
- Lack of funding to take on new projects

3. Opportunities

There will be active engagement and contribution from all stakeholders, including patients, clinicians, vendors, healthcare leaders, healthcare facilities, etc. Forward-thinking leaders from different domains, including IT, clinical, government, and health informatics will represent and continue to build the digital health landscape in Canada. Academic institutions, R&D companies and others in the research community will lead wide-scale projects that lead to new discoveries, provide evidence in support of technology, and develop new solutions for healthcare.

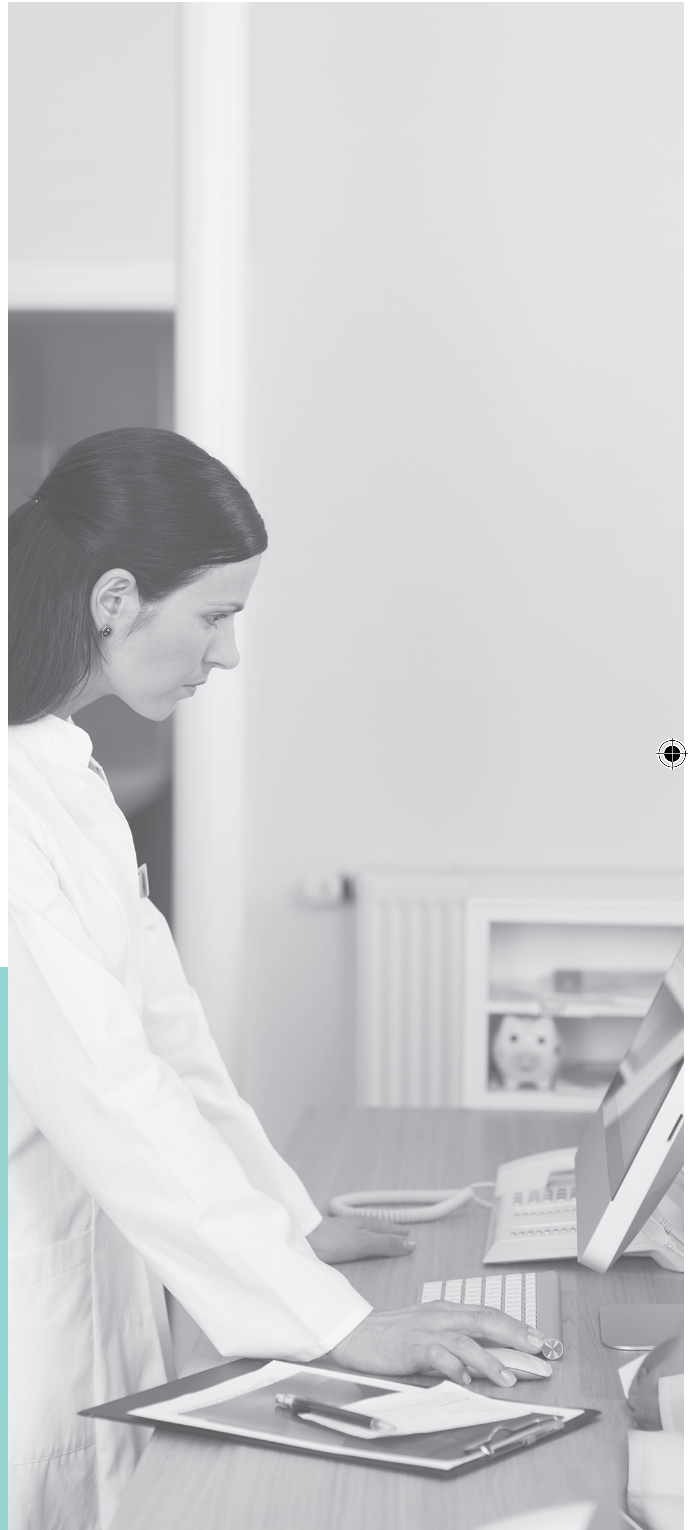
4. Suggestions

- Review clinical performance set against realistic targets.
- Build a guiding coalition and set a strategic vision.
- Gather feedback from end-users of technology and patients to ensure successful adoption and meaningful use.
- Clearly outline project scope and understand member roles within a team.
- Define change management strategies.
- Define a good governance framework and ensure governance throughout all phases of implementation.
- Appoint clinical leads, physician champions, and super users.
- Communicate successes to organization and learn from past failures.
- Join analytics networks and be active members to champion use.
- Explore other promising methods of healthcare delivery.

3.2 Examples of Predictive Tasks

Predictive Analytics as defined by Ji-Young An [68] utilizes tools and algorithms to analyze historical data to make an inference or prediction about future performance. Healthcare data that is expected to become much more readily available in the coming years can also be utilized to do this. There are patterns in health conditions (for example, coronary disease or diabetes) and associated outcomes (such as hospital admissions or death) can be studied to find predictor variables. These variables then can be applied to patient data to make predictions or even determine proper medication doses and treatment protocols with greater accuracy at the point of care [68]. This advancement in healthcare will, in turn, help reduce costs and time lag, thus making our healthcare system more robust and efficient.

As shown in Figure 3.2, predictive analytics in healthcare can be used to make a range of predictions including, but not limited to, real-time healthcare crisis management, prediction of the state of a disease as it progresses, locating the sources of outbreaks, diagnosis of diseases, lifestyle risk factors, and genetic and nutrition factors for and against good health. The sections below discuss how big data and relevant tools can be utilized in the healthcare domain to use predictive analytics to perform the essential tasks that will transform our healthcare system in the upcoming years.



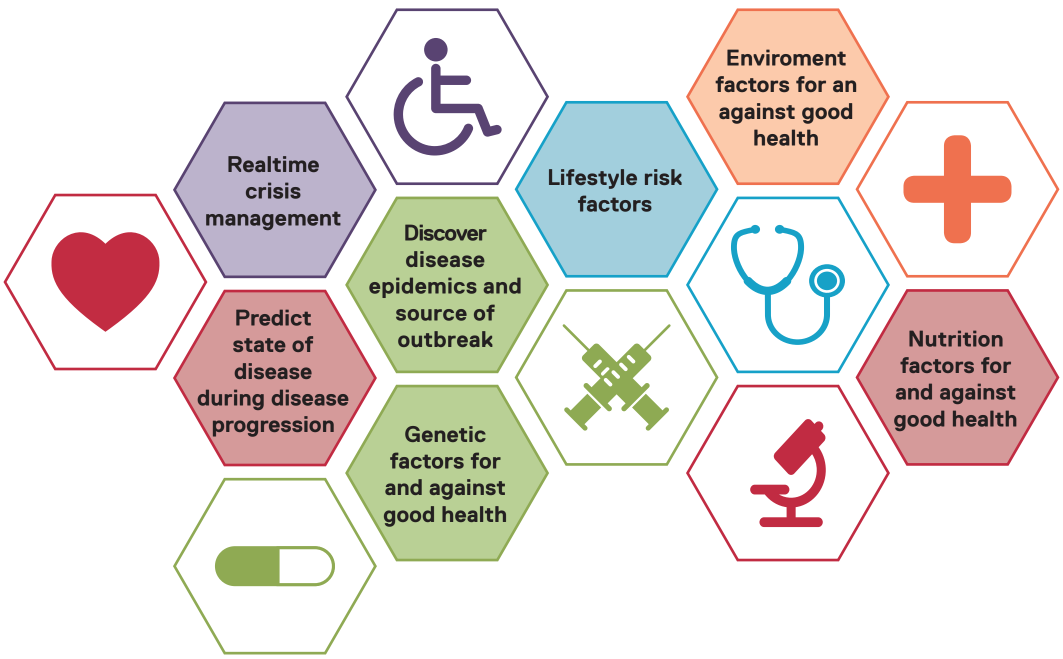
3.2.1 Real-time Intervention

Real-time analytics uses immediate information at the point of care [69] to make informed decisions. Instead of making decisions in hindsight through the batch method, real-time analytics allows choices to be made at the bedside so that immediate actions can be taken. Point-of-care decision-making holds the potential to truly revolutionize methods of patient diagnosis and treatment [70].

Real-time analytics with predicted consequences can suggest appropriate diagnosis and treatment for patients by generating real-time information regarding their medical history and current health status [71]. This methodology can be leveraged to treat critical healthcare issues such as heart attacks or strokes.

In a 2009 example, Parkland Health and Hospital System in Dallas, Texas, launched a predictive modeling system created in-house by a staff physician. The electronic system scans patients' information, identifies high-risk patients, and predicts outcomes for those patients. This allowed the hospital to save more than $500,000 since the system's implementation. In addition, the 30-day hospital readmission rate of patients suffering from heart failure has been reduced by 31%. Predictive analysis has also enabled the hospital to work towards future improvements [69]. As Selman writes in the article, *Paradigm of Prediction: Predictive Analytics to Prevent Congestive Heart Failure*, "predictive analytics can help healthcare organizations use information proactively, rather than reactively, to improve compliance and reduce the risk of chronic disease." [72]

Figure 3.2: List of Predictive Tasks in Healthcare



3.2.2 Predict Disease Progression

Another important task that big data can help with would be to predict the state of a disease during its progression.

When they know how severe the condition is, doctors can concentrate their efforts to help patients in the right direction. An example can be seen in the 2014 Bates, et al. article, *Big data in health care: using analytics to identify and manage high-risk and high-cost patients* [73]. The article discusses the way big-data analytics can be utilized to study neurodegenerative diseases in depth, and understand how these diseases progress. Combining the information learned about the disease with patients' symptoms and medical records can help practitioners make more accurate predictions on the state of an individual's disease so that the treatment plan can be tailored for the right stage for that person. This will allow patients to get the best treatment in time and also help healthcare professionals to find ways in which they can try to deter the disease from progressing [73].

3.2.3 Predicting Epidemics and Outbreak Locations

In the event of a disease epidemic, real-time availability of relevant data would make disease management far more effective and preventable. Healthcare personnel can use the information such as the location of the outbreak, nature of the disease, symptom and effects to issue timely alerts.

Internet and social media have become an important source in obtaining early warnings of disease outbreaks and "now casting" (predicting the current level of illness from recent clinical data and current social data). A 2016 article, *DEFENDER: Detecting and Forecasting Epidemics Using Novel Data-Analytics for Enhanced Response* [74], describes how we can use readily available Twitter data to explain disease outbreak, activity and progression.

The paper described a software system it called "DEFENDER" that capitalizes on live feeds from Twitter and other social media, and was designed to provide a combined disease outbreak detection and situational awareness capability. However, the focus of DEFENDER is primarily on the symptoms of the disease, versus the disease, itself. The authors believe that a "limited range of symptoms characterizes many common diseases, so a shift to symptoms adds flexibility without a great deal of additional complexity. Specific conditions can then be tracked by examining combinations of their symptoms." [74].

Development of the DEFENDER software is a promising example that shows we are not far from being able to predict disease epidemics and locating the source of outbreaks using big data in real-time.

3.2.4 Predict Diagnosis of Disease

Big data, its tools and the widespread use of electronic medical record systems have increased the importance of finding effective and scalable data-mining algorithms to develop useful models in healthcare to predict the appropriate diagnosis of disease. And there are many examples where it is already happening.

In his 2013 PhD thesis, *Healthcare Data Mining Using In-database Analytics to Predict Diagnosis of Inflammatory Bowel Disease*, the University of Washington's Eric Johnson describes the use of the Bismark in-database analytics framework in healthcare setting to apply logistic regression on a four-year set of patient demographics, encounters, and hospital records to produce predictive risk factors for a cohort of Inflammatory Bowel Disease patients [75].

In a 2013 article, *Comparative effectiveness for oral anti-diabetic treatments among newly diagnosed type 2 diabetics: data-driven predictive analytics in healthcare*, Jon Maguire and Vasant Dhar describe how machine learning methods were applied to large-scale medical and pharmacy claims for over 65,000 patients who were newly diagnosed with type-2 diabetes. The research was focused on the intersection of big data and health sciences, and posits that predictive analytics and pattern detection in real-world healthcare data can predict groups at risk, enabling proactive attention and prevention with the potential for significant cost savings [76]. They emphasized that the U.S. healthcare system could provide improved healthcare quality per unit of spend by analyzing the large volumes of healthcare claims data.

These examples are proof of the ways that big data and predictive analytics are shaping our healthcare system. With the help of big-data tools and predictive analytics techniques, we will be able to identify the onset of diseases, and also predict successful diagnoses.

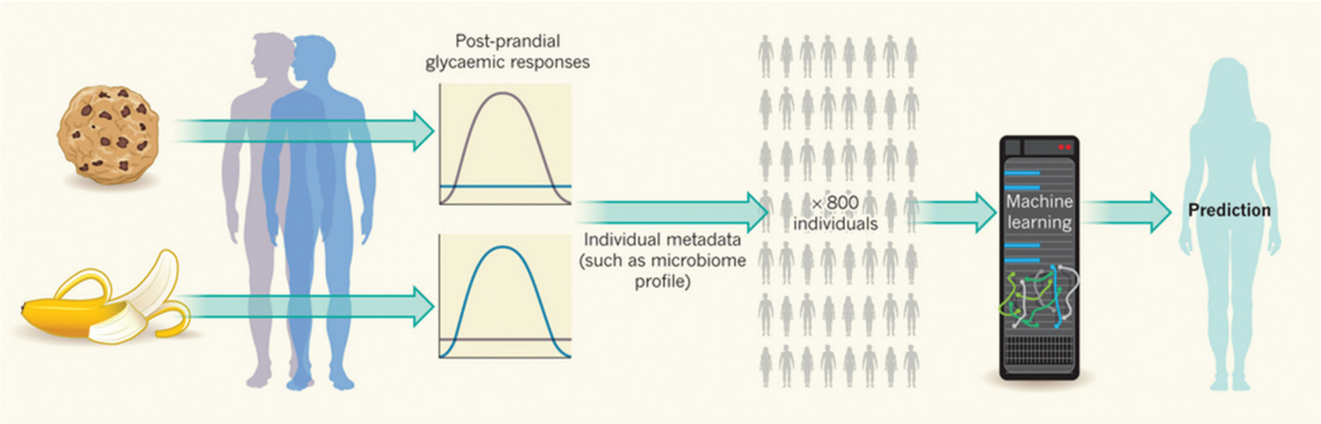
3.2.5 Predict Lifestyle Risk Factors

Modern cell phones come with sensors that can monitor vitals such as heart rate, physical activity, and peripheral capillary oxygen saturation in blood. By examining this information along with demographics, family history, healthcare records, ethnicity background and insurance claims, big-data tools can predict risks connected to lifestyles or environmental factors. Alerts can be sent out to people based on their activity, with recommendations on what they can do to minimize healthcare risks.

"Patient profile analytics", as this is called, uses devices or remote monitoring to capture fast-moving, real-time data to identify individuals who can benefit from proactive care or life changes [77].

This can improve health, but also save money in the long run as the number of individuals who become sick will be reduced due to the preventative measures. Figure 3.3 shows how data from various sources can be linked together to help individuals manage their health more effectively and proactively.

Figure 3.3: How Machine Learning can be Used for Nutrition Advice



3.2.6 Predict Environment and Nutrition Factors

Data from sources like Environment Canada can be analyzed along with hospital accounts, emergency medical records and demographics to find patterns to predict environmental factors linked to good health or poor health.

Similarly, nutrition information can be gathered from smart devices as shown in Figure 3.3, and machine learning can be used behind the scenes for nutrition advice. For example, machine learning can be applied to complex multidimensional data to provide personalized dietary recommendations for diabetics to control blood glucose levels [78].

3.2.7 Predict Genetic Factors

Genomic analysis is the identification, measurement or comparison of genomic features such as DNA sequences, structural variations, gene expression, or regulatory and functional element annotations at a genomic scale. As genomic studies look at high-dimensional data, big-data tools would be perfect to analyze these complex structures, allowing scientists to predict factors for and against good health.

With this availability of data, scientists and analysts can help researchers discover patterns and make inferences in genomic analysis. The EuResist Network[2] undertakes a clinical genomic analysis using IBM big-data tools to help HIV researchers optimize therapies for patients. The EuResist prediction engine can predict the response to combination drug therapies for a given patient with a given viral genotype, with an overall accuracy of around 77%. The EuResist project is an excellent example of big-data tools that help researchers discover treatments by better understanding clinical data from disparate sources.

An Ontario Genomics 2016 article, *Call for an Ontario Health Data Ecosystem*, (developed with support from the Data for Health Advisory Group and the Ontario Personalized Medicine Network)

[2] https://www.euresist.org

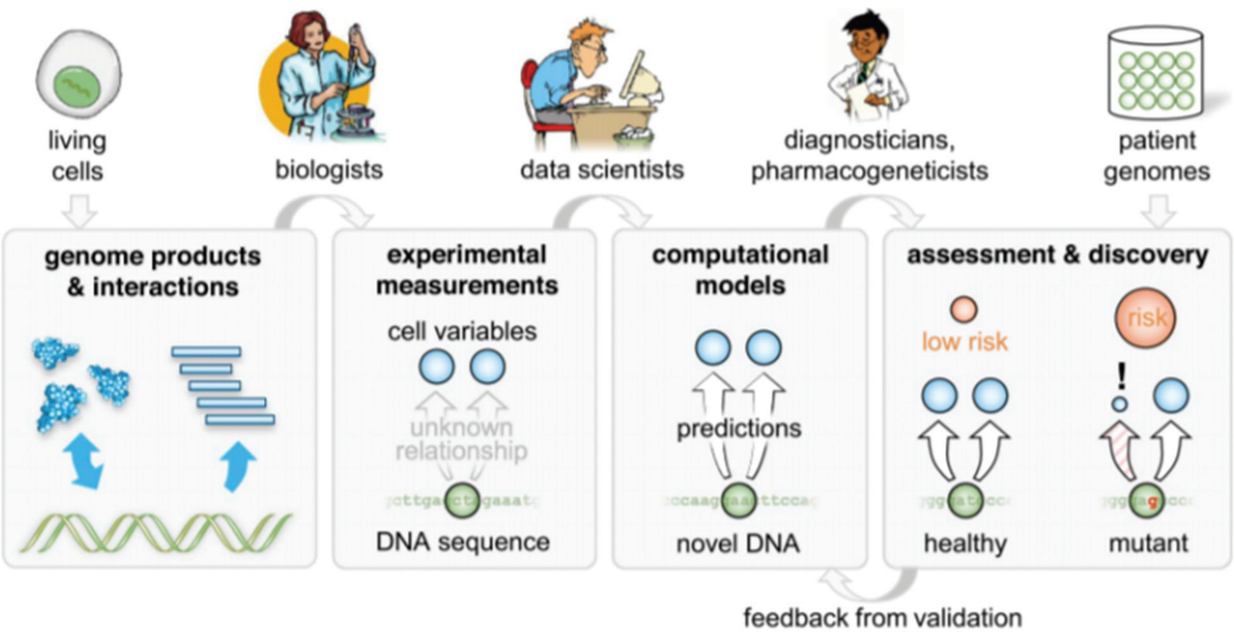
[29], describes new algorithms capitalizing on big data that allow for combined analysis of clinical and molecular data alongside social and environmental information. These algorithms can facilitate healthcare policy decision-making and enable evidence-based health system planning.

Another 2016 article, *Machine Learning in Genomic Medicine: A Review of Computational Problems and Data Sets*, examines the way machine learning can be used for important problems in genomic medicine. The article describes the ability to predict traits and disease risks from biomarkers as, "a supervised machine learning problem." Inputs and outputs are described as ". . . a stretch of DNA sequence (genotype) relevant to the underlying biology, and the outputs are the phenotypes." [79].

Figure 3.4 shows a "simplified view of how biologists, data scientists, and medical researchers can work toward genomic medicine." [79].

The value of machine learning is its ability to turn high-throughput measurements into specialized or general-purpose predictive models. With cell function, for example, by knowing how mutations affect disease, diagnosticians and pharmacogeneticists can more easily find direct correlates with the disease, and then develop treatments and targeted therapies for individual patients.

Figure 3.4: How Genomic Medicine can be Linked to Big Data and Predictive Analytics



3.3 Snapshots of Predictive Health

In the previous section[3], we discussed various predictive tasks that can be performed utilizing the data ecosystem proposed in this paper, including real-time intervention, prediction of the state of disease during disease progression, prediction of disease epidemic and diagnosis, prediction of lifestyle, environment, nutritional and genetic factors for and against good health. We also looked at some current studies that have been conducted in these areas for each of these sections. This section covers a series of case studies highlighting current and futuristic uses of predictive healthcare.

In Maguire and Dhar's 2013 paper, *Comparative effectiveness for oral anti-diabetic treatments among newly diagnosed type-2 diabetics: data-driven predictive analytics in healthcare*, it was found that the top 10% of newly diagnosed type-2 diabetes patients account for 68% of healthcare utilization in the United States. Knowing this, Maguire and Dhar sought to study how better healthcare could be provided by minimizing the cost and capitalizing on the power of predictive healthcare analytics. More specifically, the study explored the application of machine learning to large-scale medical and pharmacy claims data for over 65,000 patients who were newly diagnosed with type-2 diabetes [76]. This study is one of many done in search of cost savings and better healthcare by capitalizing on big data and predictive analytics.

A 2014 study by Mitra and Padman, *Engagement with social media platforms via mobile apps for improving quality of personal health management: a healthcare analytics case study*, looks at how social media platforms and mobile apps can be used for personal health and wellness management. In this study, the authors focused on how their findings can help health plans to develop and deploy targeted services and tools by integrating mobile apps and social media, another area for possible study cases [80] There are several studies illustrating the importance of community engagement as a potential key element in healthcare research and service, listed in this paper's bibliography at [81] [82][83] [84]. Here is a list of some of the findings:

- Community engagement serves as a cornerstone to numerous health improvement programs [82].
- The impact of community engagement on individuals is broad. From physical and psychological health, self-confidence and esteem, to personal empowerment and social relationships [82].
- "Community engagement is a process of trust and relationship-buiding rather than a one-off intervention and that, once trust is gained, it requires ongoing mutual commitment." [81]
- Community engagement allows for the identification of good practices from multiple stakeholders, and contributes to the collective documentation and case studies analysis [83].
- Implementing community engagement is not an easy task, as it includes maintaining the retention and recruitment of healthcare providers [84].

These examples and other case studies in this section are promising proof of how predictive analytics in healthcare can benefit everyone from the practitioners, to policy makers, insurance companies, drug manufacturers and most importantly, the individuals who might be facing healthcare risks.

[3] Both of Sections 3.2 and 3.3, are organized similarly for convenience, but present different content.

3.3.1 Real-time Intervention

One possible study for real-time intervention is the integration of data between the systems used by family doctors and those used at the hospitals via EHR (Electronic Health Record), with a health card as the unique identifier for each patient. When the patient visits their family doctor for a regular checkup, all their details would be recorded and tied to their card such that in an emergency, the doctors at the hospital can simply scan the card and view the full medical history in on a dashboard.

A dashboard would provide information regarding the patient's last visits to their family doctor, and any prescriptions, including a list and timing of all medication. With this information readily available at the click of a button, the doctors at the emergency center would be able to assess the patients without having to query the patient or caregiver to assess the severity of the situation. It can also show results from predictive analytics related to the patient.

The main tasks would be to consult with emergency department doctors and create an integrated system with an interactive dashboard offering all the necessary information and medical history of a patient.

Diseases like cancer and diabetes are diagnosed more frequently than ever before. However, advances in big-data tools and technologies can be used to our advantage to study diseases like these in depth, and assess the state of a disease during its progression.

Genomics is another area where scientists need to study highly complex multi-dimensional data to make inferences. Again, big-data tools and algorithms, along with availability of data can make this study practical.

More in-depth studies in genomics, cancer and other prevalent diseases can allow for precision medicine, and this is already possible with big data and predictive analytics tools and techniques available today.

Smart devices like wearables, smart homes and smart cities will only improve services, such as faster ambulance service and immediate bedside healthcare advice. Information and/or data obtained via diverse sources will improve predictive analytics, and will also allow for demographics and capacity planning, and facilitation of remote and tele-medicine where doctors could diagnose patients over a video chat. In the example of a video chat, smart wearables could provide the doctors with information on human vitals such as blood pressure, weight, blood sugar, pulse rate etc.

On the other hand, as the data ecosystem becomes more integrated, it is expected that the onset of serious illnesses such as heart attack or stroke could be detected earlier and alerts could be sent out to warn patients and physicians. In this example, alerts would allow stroke victims to manage their health symptoms and avoid putting themselves in further danger.

3.3.2 Predict Diagnosis of Disease

In Section 2, we discussed the need for links between information regarding interaction of drugs with food, and the study of profiles of people for whom a given drug worked, versus those for whom it did not work. Other links include environmental and lifestyle factors (smoker, non-smoker, etc.) as well as location data allowing researchers to group individual lives together to make inferences about the safety of drugs. In order to have meaningful studies in this realm, it would be necessary that the data can be linked appropriately.

3.3.3 Predict Nutritional, Lifestyle and Environment Factors

D. Grant Campbell, author of the 2014 paper, *Big data and the study of dementia: Epistemological promises and pitfalls* [86], describes how "big data could enhance our understanding of dementia, particularly in the context of behaviours that mark the later stages. Big data offers a shift in perspective and assumptions that provide additional discursive room for approaches and methods that could significantly alleviate some pressing problems in healthcare and elder care." The author emphasizes that although the promise of big data should be approached with caution, it should not stop us from exploring the potential of caring for people with dementia.

One idea is to use game-playing to improve, monitor, and alert patients and physicians about aspects of mental health such as dementia. As discussed in Section 2, data from various sources can be pooled together to make inferences and conduct predictive analytics. Home treadmills and other exercise equipment can also aid collection of health data such as balance, cardio, heart rate etc. Cameras can capture motion patterns that can be used to identify problems with gait, posture etc. during child development and accident recovery. All this information then can be integrated via links and analyzed using big-data tools and techniques.

Twenty years ago, mobile phones were far from what they are today. They were big, bulky, and only good for a phone call. Today smart phones are everything a person needs to carry; they allow us to talk, and to message others, record and make videos. And with the help of apps, we can also track our footsteps, manage the lighting and temperature of our houses remotely, check our pulse rates and track many other aspects of our lives; the sky is truly the limit today.

With the exponentially growing advancements in technology, there are other examples that we may see soon. We may have homes where smart toilets can test urine and stools automatically [87]. Clothes can have sensors to monitor activity; refrigerators know what one eats, and food items would have IDs to link them to nutritional information. Games could monitor mental health, and detection and monitoring of sleep disorders would be a breeze with all this information available.

Given this rate of technological advancement and the availability of big-data tools and techniques, predictive analytics have enormous potential in healthcare. It is expected that more research studies will be introduced in the coming years to evaluate how smart homes and predictive analytics will change the way we see healthcare today.

4 Vision Support

For the reasons outlined in the three prior sections, we feel that Canada is uniquely positioned to become a pioneer in health analytics and health-information system innovation in the near future. However, realizing this potential will require investment into a framework capable of making relevant health data accessible, in a privacy-sensitive manner.

Recommendations:

1. Create a new mechanism for the access of health data by individuals, research and healthcare facilities, and private businesses.
2. Engage the patient community more effectively by helping them make better-informed choices.
3. Attract top talent by offering the best raw materials in the form of comprehensive healthcare data.
4. Create healthcare innovation incubators to build a business, academic and healthcare ecosystem.

This data should include all health-related data points of an individual, as an integrated system from which data is easily accessible. This includes that owned by the individual, such as wearable outputs and nutrition data, and by related institutions such as hospitals, clinics, labs, diagnostic facilities, nursing homes, pharmacies, insurance companies, and research centres. Only then can we truly capitalize on other Provincial and Federal investments in analytic infrastructure and talent, realizing improvements in care value that benefit all Canadians.

When this system is available, and with a sufficient data supply, a strong talent pool of startup companies and educational programs can be developed to feed into the technology sector. This investment will also support the growth of innovative young companies into established enterprises, with the additional benefit of becoming targets for international investment into the Canadian market.

It is our position that advanced analytics of health data could ultimately become an important driver of the Canadian economy.

4.1 Create a New Mechanism for Health Data Access

We propose that a new mechanism be developed for the access of health data by individual, research and healthcare facilities, and private businesses to allow a broad range of personal data to be utilized that will improve healthcare and support the development of an innovative health industry. It will allow individuals to access their own information on an ad-hoc basis, while permitting startup companies, health service institutions, industrial enterprises and researchers to use anonymized data to develop new products and services, creating new value from the data. This idea can be realized through different approaches, three of which include:

1. Generation of a centralized data repository hosted and governed by a sanctioned institution (e.g. The Canadian Institute for Health Information (CIHI).
2. Creation of a multi-centered data repository coalition under a common governance.
3. Creation of a decentralized data ecosystem, e.g. blockchain of health data.

4.1.1 Purpose: Why are we Doing This?

Our reasons for proposing the framework are many:

- Data access remains the major bottleneck in the development of intellectual property in the health space. The inability to access data in a comprehensive way from multiple platforms (e.g. wearables, wellness and healthcare systems) means we are currently missing out on the real potential of big data to be used in improving the health and wellness of Canadians.
- For the greater good of society, centralizing our data and allowing access in a reasonable manner will fuel new discoveries to help people.
- It is good for business, whether not-for-profit or for-profit. With ready access to data, we can more rapidly measure the performance of our healthcare system and pinpoint where we need to do better and become more efficient; companies will be able to better understand healthcare issues to develop more effective products, etc.
- We can establish ourselves as world leaders in health data and data science. We are already leaders in data science but our data infrastructure is lagging behind the UK, Switzerland, Netherlands, Israel, etc. We need to remain competitive to retain talent and attract more investments, for example, hundreds of millions are being invested in Toronto and Montreal for initiatives like the Vector institute, to cement our position as global leaders in machine learning.

4.1.2 What are the Expected Outcomes?

We see this framework as benefiting individual Canadians, the healthcare industry, private sector, and government operating in the health space.

- Individuals will have secure access to the entirety of their health information, with the option to update their privacy settings as they see fit.
- With easier access to larger amounts of data, we can measure important associations we have not been able to in the past. This means researchers will be able to better understand important issues in healthcare, and individuals can better understand their own health.
- Healthcare institutions will have access to a higher quality, broader set of data. With that data, they will be able to determine the effectiveness of treatments and preventative health measures, and gain efficiencies through innovative automation of procedures and workflows.
- Companies and researchers can use this data to generate novel diagnostics and improve the health/wellness of the participants.
- Governments can reduce inefficiencies when they better understand the data. Governments at each level can plan better policies and assess the impact of their investments on population health.
- Canada's ability to attract top talent will grow. Leading scientists and companies will want to come here to leverage our data for its research potential. By offering massive health datasets for study, our universities will have a huge competitive advantage.

The bottom line will be the promise of improved efficiency through data insights and better health outcomes.

4.1.3 Target Audiences(s)

- There is a need for leadership at all levels to support the creation and maintenance of a comprehensive data collection platform, and to champion active participation from various stakeholders. Both the individual and community are essential for the success of this endeavor, and their support will be lost without sufficient public awareness and trust. As well, leadership needs to support incubation through collaboration between academic institutions and the private sector.
- It will be a massive undertaking to integrate, normalize and standardize data of this scale, and will require the hiring of skilled individuals as well as substantial computational infrastructure. Realizing this goal will require an investment from the Federal Government.

4.1.4 Types of Datasets and Tools

- Groups like Canadian Institute for Health Information (CIHI) and Institute for Clinical Evaluative Sciences (ICES) already have strengths in the area of administrative data, however, over time we need to broaden this base to include patient registries, mobile data, etc., with all data linked to the base.
- Advanced analytics tools, statistical packages and data visualization tools are needed.

4.1.5 The Need for Standardization and Definitions

As we acquire more "messy" (inconsistent) data, we will need a large team with expertise in data architecture, data mapping, data governance and stewardship, ETL (Extract, Transform and Load) developers, etc. to create standard definitions, map elements properly, clean data, etc. so that we have quality, useful data. This process needs to be led by a central group (e.g. CIHI) and will be an expensive undertaking, but can work within the business and financial models described in this paper.

4.1.6 Mandatory versus Voluntary Contribution

Data contribution from providers should be mandatory.

- If we rely solely on voluntary data contributions, the findings from analyses may not be valid representations of society at large, but only a subset chosen arbitrarily. In addition, voluntary submission may mean there are systemic differences affected by the cost or inconvenience to smaller, rural or less-affluent providers. (Dobra, Matthew et al., *Roadblocks, Regulation, and Red Tape: How American Health Policy Norms Threaten the Big Data Revolution* [88].)
- Some countries, such as Switzerland have implemented mandatory contribution. Public and nonprofit organizations should have no (or minimal) financial investment but require basic data literacy to access the data whereas for-profit organizations may need to pay.

4.1.7 Governance, Stewardship, and Operating Model

- Canada needs a new mechanism for owners to contribute data[1]. These topics have been discussed extensively but little sharing has occurred.
- All healthcare data will be open by default with privacy protection, however, for commercial usage, there will be different levels of access for different organizations. For instance, insurance companies will be allowed to access general data at a national, provincial or even municipal level, however, if they want further detail on the individual level, their request must be approved by a governing body that is capable of weighing the benefits of the analysis against possible privacy concerns of the individuals.
- A governing body would be established to oversee this new Canadian health data access mechanism, with the main responsibility to ensure proper usage of data by different organizations, and deal with privacy protection, and legal matters. This body will also maintain different access levels for the various organizations and ensure that they comply with the established rules and guidelines. In a nutshell, this body will be the gatekeeper for the pan-Canadian healthcare information system, including wellness data.
- As the individuals are the owners of their own data and the contributors to this data ecosystem, there would be no charge to access their personal health data (related to their own healthcare only).

[1] http://www.scienceadvice.ca/uploads/eng/assessments%20and%20publications%20and%20news%20releases/Health-data/HealthDataFullReportEn.pdf

- Universities and not-for-profit organizations using this information for research purposes will also be allowed to access the data ecosystem for free; however, secondary users such as insurance companies or drug manufacturers who want to use this data for profit will have to pay for use. This will initiate a data trade mechanism and pricing system for commercial organizations to pay, which can then circle back to the governing body to reinvest in further development and maintenance of this proposed system, creating a self-sustaining cycle. Start-up companies for innovative usage of health data should be supported with special consideration under this system.

4.1.8 Funding

This system could be funded in a number of ways. Government funding and private investment could be considered for different components, for example, seed funding for feasibility studies and early establishment of the organizational structures. Industrial investment and payment can be used to maintain the operations and new development.

4.1.9 Establishing Public Trust

Establishing public trust can be challenging. For example, in April 2014, the British National Health Service (NHS) had planned to launch an expanded centralized database. However, it was canceled in April 2014 because the NHS had failed to adequately educate the public and gain their trust, even though at the time the decision was made, £7.5 million had "already been spent constructing a database, printing leaflets, setting up a patient information helpline and researching public attitudes into data sharing"[2].

Gaining public trust in this area requires significant cultural change. Therefore, in addition to government leadership to enable such initiatives from a legal/policy standpoint (e.g. prescribed entity approach), we recommend that an independent body with significant resources to lead culture change be established.

[2] (http://www.telegraph.co.uk/science/2016/07/07/ how-the-nhs-got-it-so-wrong-with-caredata/

4.2 Engage the Patient Community with Better Information

The future of healthcare needs to include giving patients better information so they can make better informed choices regarding their health. The recent rise of the personal health record (PHR) has allowed patients direct access to their own health information, which is rightfully theirs. As these management platforms and portals evolve into lifelong data banks that cover their medical records as well as lifestyle and environment, individuals will be able to track, monitor, and trend data to have a comprehensive view of their health.

This level of individual awareness has the potential to improve prevention and early detection, as long as community attitudes and behaviours favour healthy lifestyles, regular check-ups and, where appropriate, a commitment to using technologies consistently (e.g. 24/7 monitoring using wearables). Building positive attitudes and behaviours will come by involving parents, education institutions, and community, as well as with support from business to promote work-life balance as part of a healthy lifestyle.

Through the sharing of personal health data, the general public will be better informed about health outcomes and feel empowered to take part in their own healthcare. The growth of mobile technologies and wearables in particular is already driving the generation of available personal health information and will continue to do so in the coming years. The full vision may not happen in the nearest future, as adoption of wearable technology and acceptance of integrated databanks of lifelong samples, medical records, lifestyle, and environmental data will take some time (and involve enlightened leadership).

However, without engaging the patient, the healthcare industry risks ineffective management of the increasing burden of disease, pressurized under a lack of resources.

4.3 Attract Top Talent with the Best Raw Materials

Universities and their affiliated research groups are important stakeholders, and talented researchers and students will be attracted by the best raw materials.

For Canada to establish itself as a leader in healthcare innovation, it is important that its academic researchers are well supported with data access. In addition, as the field continues to evolve due to the magnitude of health data that is collected every day, education providers need to ensure their programs remain current and relevant.

Access to high quality, comprehensive data for student projects will enable this, as students will experience real-world application of the concepts learned, and through this, it is hoped that students will become innovators when they enter the workforce. The end goal is to retain and grow talent as well as computational capabilities that will help Canada rise as a global centre for predictive analytics in healthcare.

4.4 Create Healthcare Innovation Incubators

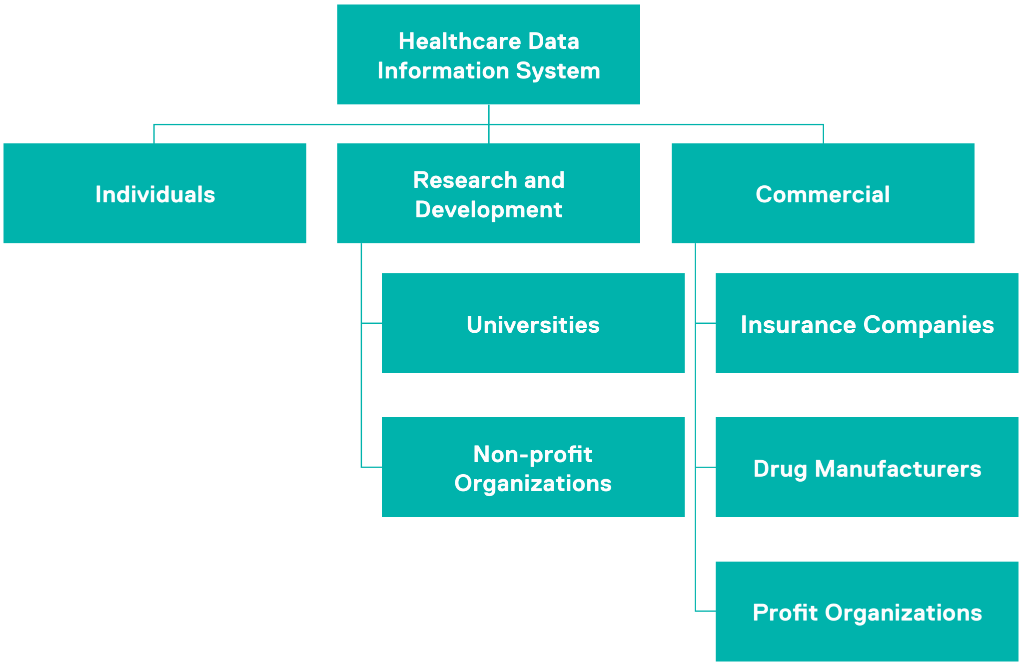
Construct big-data healthcare incubators to bring together business mentors, entrepreneurs, academics, students and healthcare professionals. These environments will enable the growth of a stronger interpersonal and inter-industry ecosystem for idea-generation and collaboration on new technology and opportunities.

As an example, the Vector Institute, a facility focused on artificial intelligence, was recently launched in Toronto with the goal of global leading-edge research and development. The Institute is funded by the provincial and federal governments, the University of Toronto, and private industry [89]. The launch has led to the creation of strong partnerships between industry, academia, and government, with promising potential for great economic growth and social value as a result.

This is exactly the type of collaboration that is necessary to propel the field of predictive analytics as it applies to healthcare. There is no time better than now to harness the brilliant talent in Canada, to engage all stakeholders from the public and private sectors, and to leverage existing IT infrastructure to realize the potential of what predictive health technology can bring to this country.

Figure 4.1 illustrates Canada as an innovative hub, and shows data access and community engagement. Individuals' own data will be available to them at no cost because they are the biggest contributors to the hub. Universities and non-profit organizations will be able to access aggregated data for research and development purposes free of charge as well. Although high level aggregated data will be open source, for-profit organizations such as insurance companies and drug manufacturers will have to pay to access detailed data.

Figure 4.1: Data Access and Community Engagement



5 Conclusion

Predictive Health is defined as applying predictive analytics to improve health and healthcare. This includes innovative data sourcing, advanced methodologies, multiple industrial collaborations, strategic investment, creative entrepreneurship and comprehensive governance and policies.

For predictive health to fuel innovation and become an economic driver, we recommend the following actions:

- Advance Canada's open access to data and analysis.
- Apply advanced predictive algorithms to improve efficiency within the healthcare system.
- Solidify an environment that is receptive to continuing innovation in health IT.

Health data sources have proliferated substantially in recent years. Data from wearables, IoT data from implanted instruments, social media sentiment from patients and many other types of data have become an important part of health data and influenced health related decisions of individuals and society. We recommend that Canada develops a new framework for health data to increase access in a privacy-sensitive manner and encourages innovation and entrepreneurship. This will help make predictive health, and the health industry overall, a major driver of economic growth.

It is our hope that, very soon, Canada will become the pioneer of a well-established, centralized health data system where individuals' health is better supported, and institutions have easier access and more powerful tools to analyze the data. We are confident that the Canadian health system has the potential to become a leader in creating a well-integrated health system, and we believe that the world will look to that system and follow suit.

Glossary

ASC X12 (EDI) The Accredited Standards Committee X12, Electronic Data Interchange. 23

CAP Certified Analytics Professional. 6

CARL Canadian Association of Research Libraries. 12

CASRAI Consortia Advancing Standards in Research Administration Information. 12

CCD Continuity of Care Document. 22

CDS Clinical Decision Support. 15

CEN's TC/251 Committee for European Normalisation (CEN) Technical Committee (TC) 251 Standardisation. 23

CFHI Canadian Foundation for Healthcare Improvement. 6

CHA Canada Health Act. 4, 5

CHIMA Canadian Health Information Management Association. 6

CIHI Canadian Institute for Health Information. 6, 12, 49

CNC/CODATA Canadian National Committee for CODATA. 12

COACH Canada's Health informatics Association. 6

Continuity of Care Record ASTM International Continuity of Care Record standard. 23

CONTSYS (EN 13940) Supports continuity of care record standardization.. 23

CPOE Computerized Physician Order Entry. 15

CPT Current Procedural Terminology. 14

DEFENDER Detecting and Forecasting Epidemics Using Novel Data-Analytics for Enhanced Response. 40

DICOM international communications protocol standard for representing and transmitting radiology (and other) image-based data. 23

DNA Deoxyribonucleic acid. 19

EHCR Electronic Health Care Record. 24

EHR Electronic Health Record, Contain information from all clinicians involved in the patient's care; designed to be shared with laboratories and specialists. 5, 9, 14, 15, 17, 20, 22, 23, 28, 29, 31, 34, 44

EMR Electronic Medical Record, Digital version of a clinician's notes. 22, 28, 31, 36

EN 13606 communication standards for EHR information. 23

FDA Food and Drug Administration. 14

FHIR Fast HealthCare Interoperability Resources. 23, 31

fMRI functional Magnetic Resonance Imaging. 20

GDP Gross Domestic Product. 35

GWA Genome-wide Association. 19

HDO Health Delivery Organization. 27

HIAL Health Information Access Layer. 33

HIC Health Information Custodians. 33

HIE Health Information Exchanges. 22

HIMSS Healthcare Information and Management Systems Society. 6

HISA (EN 12967) a services standard for inter-system communication in a clinical information environment. 23

HL7 Health Level Seven International. 22, 23, 31

IaaS Information as a Service. iv, 27

ICD International Classification of Diseases. 14

ICES Institute for Clinical Evaluative Sciences. 49

ICT Information and Communication Technology. 6, 9, 11

INFORMS Institute for Operations Research and Management Services. 6

IoT Internet of Things. 13, 32, 36

ISO/TC International Organization for Standardization's Technical Committee. 23

LCDI Leadership Council for Digital Infrastructure,. 12

LOINC Logical Observation Identifiers Names and Codes. 14

MRI Magnetic Resonance Imaging. 20

MU Meaningful Use Data. 22

NDC National Drug Code. 14

NEMA National Electrical Manufacturers Association. 23

NHS British National Health Service. 51

OCHIS Office of the Chief Health Innovation Strategist,. 5

OECD Organisation for Economic Co-operation and Development. 10

OLTP Online Transaction Processing. 20

OSI Operational Stress Injury. 31

P2P Peer-to-Peer. 24

PACS Picture Archiving and Information Systems. 29

PHI Personal Health Information. 27

PHIPA Personal Health Information Protection Act. 23

PHR Personal Health Record. 51

PIA Privacy Impact Assessment. 32

RDC Research Data Canada. 12

RDF Resource Description Framework. 25

RxNORM “normalized” notations for clinical drugs. 14

SaaS Software as a Service. iv, 26

SLA Service-Level Agreements. 21

SQL Structured Query Language. 21

TNA Trusted Notary Archive. 24

TRA Threat Risk Assessment. 32

xDT Family of data exchange formats for medical purposes that is used in the German public health system. 23

XML Extensible Markup Language. 22

Bibliography

[1] S. Allin, J. Veillard, and M. Grignon. Our health care system is inefficient. What can we do about it? 2014. url: http://healthydebate.ca/opinions/health-system-inefficient-can.

[2] N. Sahni, A. Chigurupati, B. Kocher, and D. Cutler. How the U.S. Can Reduce Waste in Health Care Spending by $1 Trillion. 2015. url: https://hbr.org/2015/10/how-the-u-s-can-reduce-waste-in-health-care-spending-by-1-trillion

[3] Basel Kayyali, David Knott, and Steve Van Kuiken. "The big-data revolution in US health care: Accelerating value and innovation". In: McKinsey Quarterly 2 (2013). url: http://www.mckinsey.com/industries/healthcare-systems-and-services/our-insights/the-big-data-revolution-inus- health-care.

[4] Canadian Institute for Health Information (CIHI). How Canada Compares: Results From The Commonwealth Fund 2015 International Health Policy Survey of Primary Care Physicians. Tech. rep. Ottawa, ON, 2016.

[5] Health Canada. Canada's Health Infostructure. 2016. url: http://www.hc-sc.gc.ca/hcs-sss/pubs/system-regime/2011-hcs-sss/index-eng.php.

[6] Canadian Institute for Health Information (CIHI). Spending. 2015. url: https://www.cihi.ca/en/spending-and-health-workforce/spending.

[7] Government of Canada. Canada's health care system. 2017. url: https://www.canada.ca/en/health-canada/services/canada-health-care-system.html.

[8] Justin Trudeau. Minister of Health Mandate Letter. Tech. rep. 2015. url: http://pm.gc.ca/eng/minister-health-mandate-letter.

[9] John N Lavis, Michael G Wilson, Jeremy M Grimshaw, R Brian Haynes, Steven Hanna, Parminder Raina, Russell Gruen, and Mathieu Ouimet. "Effects of an evidence service on health-system policy makers' use of research evidence: A protocol for a randomised controlled trial." In: Implementation science : IS 6.1 (2011), p. 51. issn: 1748-5908. doi: 10.1186/1748-5908-6-51. url: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3123565&tool=pmcentrez&rendertype=abstract.

[10] Jennifer Verma, Stephen Petersen, Stephen Samis, Nurdin Akunov, Improvement Analyst, Jeffrey Graham, and Former Intern. "Healthcare Priorities in Canada: A Backgrounder". In: Canadian Foundation for Healthcare Improvement April (2014), p. 26.

[11] CIHI. CIHI's Annual Report, 2015-2016 Charting a New Course. Tech. rep. 2015. url: http://www.cihiconferences.ca/annualreport2016/index_en.html

[12] Canada Big Data Consortium. "Closing Canada's Big Data Talent Gap". Tech. rep. Ryerson 2015. url: https://www.ryerson.ca/content/dam/provost/PDFs/Big_Data_Talent_Gap.pdf

[13] Ryerson University. Big Data and Advanced Analytics at The Chang School. 2016. url: https://ce-online.ryerson.ca/ce/default.aspx?id=3635

[14] OICR. Canadian Government Makes Big Investment in Big Data Research. 2016. url: https://news.oicr.on.ca/2016/09/canadian-government-makes-big-investment-in-big-data-research/

[15] Prism Economics and Analysis. Health Informatics and Health Information Management: Human Resources Outlook 2014-2019. 2014. url: https://books.google.ca/books?id=8Wm_oQEACAAJ.

[16] Komal Garewal. Wavefront Community Among the 20 Promising Canadian Digital Health Companies to Watch. 2015. url: https://hitconsultant.net/2015/07/08/canadas-top-20-digital-health-startups-to-watch/

[17] The Informatiion and Communications Technology Council. Digital Health in Canada. Tech. rep.2015. url: https://www.infoway-inforoute.ca/en/.

[18] E. MacIntosh, N. Rajakulendran, and Z. Khayat. MaRS Market Insights: Health innovation outlook for 2016. 2016. url: https://www.marsdd.com/news-and-insights/mars-market-insights-healthinnovation-outlook-2016/.

[19] Canadian Healthcare Technology. Ottawa funds 22 new eHealth research projects. 2016. url: http://www.canhealth.com/blog/ottawa-funds-22-new-ehealth-research-projects/

[20] Canadian Institute for Health Information. National Health Expenditure Trends , 1975 to 2014. Vol. 36. October. Canadian Institute for Health Information, 2014, pp. 423–5. isbn: 9780888068873. doi: 10.1007/s10916-010-9605-x. url: https://www.ncbi.nlm.nih.gov/pubmed/12132953

[21] Patty Kostkova, Helen Brewer, Simon de Lusignan, Edward Fottrell, Ben Goldacre, Graham Hart, Phil Koczan, Peter Knight, Corinne Marsolier, Rachel a McKendry, Emma Ross, Angela Sasse, Ralph Sullivan, Sarah Chaytor, Olivia Stevenson, Raquel Velho, and John Tooke. "Who Owns the Data? Open Data for Healthcare". In: Frontiers in public health 4.February (2016), p. 7. issn: 2296-2565. doi: 10 . 3389 / fpubh . 2016 . 00007.

[22] Daniel E Williams and Jay Leask. "People, Process, Technology Strategy for Enterprise 2.0". In: Booz Allen Hamilton (2011), p. 12.

[23] Canadian Institutes of Health Research. Health Research Data: National data stakeholders. 2016. url: http://www.cihr-irsc.gc.ca/e/49942.html.

[24] Vernon Turner, John F. Gantz, David Reinsel, and Stephen Minton. The Digital Universe of Opportunities: Rich Data and Increasing Value of the Internet of Things. Tech. rep. April. 2014, pp. 1–5.

[25] Internet Stats. Twitter Usage Statistics. 2016. url: http://www.internetlivestats.com/twitter-statistics/

[26] John Gantz and David Reinsel. THE DIGITAL UNIVERSE IN 2020: Big Data, Bigger Digital Shadows, and Biggest Growth in the Far East. Tech. rep. 2012. url: https://www.emc.com/collateral/analyst-reports/idc-digital-universe-united-states.pdf

[27] Andreas Holzinger, Christof Stocker, Bernhard Ofner, Gottfried Prohaska, Alberto Brabenetz, and Rainer Hofmann-Wellenhof. "Combining HCI, natural language processing, and knowledge discovery- Potential of IBM content analytics as an assistive technology in the biomedical field". In: Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). Vol. 7947 LNCS. Springer, 2013, pp. 13–24. isbn: 9783642391453. doi: 10.1007/978-3-642-39146-0{_}2.

[28] Basel Kayyali, David Knott, and Steve Van Kuiken. The big-data revolution in US health care : Accelerating value and innovation. 2013. url: https://digitalstrategy.nl/wp-content/uploads/E2 - 2013 . 04 - The - big - data - revolution - in - US - health - care - Accelerating - value - and - innovation.pdf.

[29] Ontario Genomics. Call for an Ontario Health Data Ecosystem. Tech. rep. 2016, pp. 1–8. url: http://www.ontariogenomics.ca/wp-content/uploads/sites/1/2016/10/Call-for-an-Ontario-Health-Data-Ecosystem.pdf

[30] National Human Genome Research Institute. The Human Genome Project Completion: Frequently Asked Questions. 2010. url: https ://www.genome.gov/11006943/human-genome-project-completion-frequently-asked-questions/.

[31] Michael Eisenstein. "Big data: The power of petabytes". In: Nature 527.7576 (2015), S2–S4.

[32] Jocelyn Kaiser. "Who has your DNA–or wants it". In: Science (New York, NY) 349.6255 (2015), p. 1475. issn: 0036-8075. doi: 10.1126/science.349.6255.1475. url: http://www.sciencemag.org/cgi/doi/10.1126/science.349.6255.1475%5Cnpapers3://publication/doi/10.1126/science.349.6255.1475.

[33] T Burdett, PN Hall, E Hasting, and LA Hindorff. The NHGRI-EBI Catalog of published genome-wide association studies. 2016. url: https://www.ebi.ac.uk/gwas/.

[34] The 1000 Genomes Project 1000 Genomes Project Consortium, Goncalo R Abecasis, Adam Auton, Lisa D Brooks, Mark A DePristo, Richard M Durbin, Robert E Handsaker, Hyun Min Kang, Gabor T Marth, and Gil A McVean. "An integrated map of genetic variation from 1,092 human genomes." In: Nature 491.7422 (2012), pp. 56–65. issn: 1476-4687. doi: 10.1038/nature11632. url: http://www.ncbi.nlm.nih.gov/pubmed/23128226%5Cnhttp://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC3498066.

[35] The 1000 Genomes Project Consortium. Statistics on the 1000 Genomes Project. 2013. url: http://www.internationalgenome.org/category/statistics/

[36] Vinod Menon. "Large-scale brain networks and psychopathology: a unifying triple network model". In: Trends in cognitive sciences 15.10 (2011), pp. 483–506.

[37] Michael D Greicius, Gaurav Srivastava, Allan L Reiss, and Vinod Menon. "Default-mode network activity distinguishes Alzheimer's disease from healthy aging: evidence from functional MRI". In: Proceedings of the National Academy of Sciences of the United States of America 101.13 (2004), pp. 4637–4642.

[38] Michael D Fox, Abraham Z Snyder, Justin L Vincent, Maurizio Corbetta, David C Van Essen, and Marcus E Raichle. "The human brain is intrinsically organized into dynamic, anticorrelated functional networks". In: Proceedings of the National Academy of Sciences of the United States of America 102.27 (2005), pp. 9673–9678.

[39] Eun Young Ko. S-Detect™ in Breast Ultrasound : Initial Experience. Seoul, Korea, 2014. url: http://www.danson.ro/images/Pdf/WhitePaper%20S-Detect.pdf

[40] Noel T Brewer, Talya Salz, and Sarah E Lillie. "Systematic review: the long-term effects of falsepositive mammograms". In: Annals of internal medicine 146.7 (2007), pp. 502–510.

[41] Benjamin Q Huynh, Hui Li, and Maryellen L Giger. "Digital mammographic tumor classification using transfer learning from deep convolutional neural networks". In: Journal of Medical Imaging 3.3 (2016), p. 34501.

[42] Shih-Neng Yang, Fang-Jing Li, Yen-Hsiu Liao, Yueh-Sheng Chen, Wu-Chung Shen, and Tzung-Chi Huang. "Identification of breast cancer using integrated information from MRI and Mammography". In: PloS one 10.6 (2015), e0128404.

[43] Sage Bionetworks. The Digital Mammography DREAM Challenge. 2016. url: https://www.synapse.org/#!Synapse:syn4224222/wiki/401743

[44] K D Mandl, P Szolovits, and I S Kohane. "Public Standards And Patients' Control: How To Keep Electronic Medical Records Accessible But Private". In: BMJ (Clinical research ed.) 322.7281 (2001), pp. 283–287. issn: 0959-8138. doi: 10.1136/bmj.322.7281.283.

[45] Pekka Ruotsalainen and Bryan Manning. "A notary archive model for secure preservation and distribution of electrically signed patient documents". In: International Journal of Medical Informatics 76.5-6 (2007), pp. 449–453. issn: 13865056. doi: 10.1016/j.ijmedinf.2006.09.011.

[46] P C Gioia. "Quality improvement in pediatric well care with an electronic record". In: Proceedings / AMIA Annual Symposium. American Medical Informatics Association. 2001, pp. 209–13. isbn: 1531-605X. doi: D010001619[pii].

[47] Vikki Papadouka, Paul Schaeffer, Amy Metroka, Andrew Borthwick, Parisa Tehranifar, Jessica Leighton, Angel Aponte, Ruron Liao, Alexandra Ternier, Stephen Friedman, and Noam Arzt. "Integrating the New York citywide immunization registry and the childhood blood lead registry." In: Journal of public health management and practice : JPHMP Suppl (2004), pp. 72–80. issn: 1078-4659.

[48] Scott D Williams and William Hollinshead. "Perspectives on Integrated Child Health Information Systems: Parents, Providers, and Public Health". In: Journal of Public Health Management and Practice 10 (2004), S57–S60.

[49] Rose Dieng-kuntz, David Minier, Olivier Corby, and Laurent Alamarguy. “Building and using a medical ontology for knowledge management and cooperative work in a health care network”. In: Computers in Biology and Medicine 36.7-8 (2006), pp. 871 –892. doi: 10.1016/j.compbiomed.2005. 04.015.

[50] Ora Lassila and Ralph R Swick. “Resource Description Framework (RDF) Model and Syntax Specification”. In: World Wide Web Internet And Web Information Systems 2004.October (1999), pp. 1–54. doi: citeulike-article-id:122376.

[51] Rachael Rettner. The Human Body: Anatomy, Facts & Functions. 2016. url: http://www.livescience.com/37009-human-body.html.

[52] Arshdeep Bahga and Vijay K. Madisetti. “A cloud-based approach for interoperable electronic health records (EHRs)”. In: IEEE Journal of Biomedical and Health Informatics 17.5 (2013), pp. 894–906. issn: 21682194. doi: 10.1109/JBHI.2013.2257818.

[53] Michelle Chibba and Ann Cavoukian. “Privacy, consumer trust and big data: Privacy by design and the 3 C’S”. In: ITU Kaleidoscope: Trust in the Information Society (K-2015), 2015. IEEE. 2015, pp. 1– 5.

[54] Adele Pauline Carty. “Child Health in the Province of Ontario: Establishing baselines for reproductive health and vaccine preventable diseases”. PhD thesis. University of Guelph, 2015. url: http://hdl.handle.net/10214/9182.

[55] Infoway Information and the Canadian Institute for Health. “Health Analytics for Informed Decision-Making: Health System Use Summit”. In: Health Analytics for Informed Decision-Making: Health System Use Summit. 2016. url: https://www.infoway-inforoute.ca/en/component/edocman/resources/reports/3156-health-analytics-for-informed-decision-making-health-systemuse-summit-executive-summary.

[56] N. C L Hess, D. J. Carlson, J. D. Inder, E. Jesulola, J. R. Mcfarlane, and N. A. Smart. Clinically meaningful blood pressure reductions with low intensity isometric handgrip exercise. A randomized trial. 2016. doi: 10.1017/CBO9781107415324.004.

[57] V Mayer-Schönberger and K Cukier. Big Data: A Revolution that Will Transform how We Live, Work, and Think. An Eamon Dolan book. Boston, USA: Houghton Mifflin Harcourt, 2013. isbn: 9780544002692. url: https://books.google.ca/books?id=uy4lh-WEhhIC.

[58] Susan M. Jack. “Qualitative Evidence in Health Policy Analysis”. In: Handbook of Qualitative Health Research for Evidence-Based Practice. Springer, 2016, pp. 527–551. doi: 10.1007/978-1-4939-2920-7{_}31. url: http://link.springer.com/10.1007/978-1-4939-2920-7_31.

[59] Canada Health Infoway. A Conversation about Digital Health ANNUAL REPORT 2015-2016. Tech. rep. 2016. url: https://www.cmdt.org.nz/file/file578ed39246e8b/open.

[60] Roger Collier. “National physician survey: EMR use at 75%”. In: Canadian Medical Association Journal 187.1 (2015), E17–E18. issn: 0820-3946. doi: 10.1503/cmaj.109-4957. url: http://www.cmaj.ca/content/187/1/E17%5Cnhttp://www.cmaj.ca/cgi/doi/10.1503/cmaj.109-4957.

[61] Feng Chang and Nishi Gupta. “Progress in electronic medical record adoption in Canada”. In: Canadian Family Physician 61.12 (2015), pp. 1076–1084. issn: 0008350X.

[62] Canada Health Infoway. Diagnostic Imaging Systems. 2016. url: https://www.infoway-inforoute.ca/en/solutions/electronic-health-records/diagnostic-imaging-systems.

[63] Hugh Watson. “Tutorial: Big Data Analytics: Concepts, Technologies, and Applications”. In: Communications of the Association for Information Systems 34.1 (2014). issn: 1529-3181.

[64] The Social Sciences and Humanities Research Council (SSHRC), The Canadian Institutes of Health Research (CIHR), The Natural Sciences and Engineering Research Council (NSERC), and The Canada Foundation for Innovation (CFI). Capitalizing on Big Data : Toward a Policy Framework for Advancing Digital Scholarship in Canada Consultation Document. Tech. rep. October. 2013, p. 30. url: http://www.sshrc-crsh.gc.ca/about-au_sujet/publications/digital_scholarship_consultation_e.pdf.

[65] Dennis Giokas. Data Analytics in Health Care. 2016. url: https://www.coachorg.com/en/newsandevents/resources/ONUP_2016/2016-ONUP-Presentations/Dennis-Giokas--final.pdf

[66] Canada’s Big Data Consortium. Closing Canada’s Big Data Talent Gap. Tech. rep. October. 2015.

[67] Canadian Institute for Health Information. Health spending in Canada 2016. 2016. url: https://www.cihi.ca/en/health-spending

[68] Ji-Young An. “Healthcare analytics for quality and performance improvement”. eng. In: Healthcare Informatics Research 19.4 (Dec. 2013), pp. 324–325. issn: 1098-6596. doi: 10.1017/CBO9781107415324.004.

[69] Mary JWills. “Decisions through data: analytics in healthcare”. In: Journal of Healthcare Management 59.4 (2014), p. 254.

[70] Jan Peter Ozga. How analytics can help employers measure and manage risks. 2013. url: https://www.benefitnews.com/news/how-analytics-can-help-employers-measure-and-manage-risks

[71] Kyle Murphy. What about healthcare must change for analytics to flourish? 2013. url: https://ehrintelligence.com/news/what-about-healthcare-must-change-for-analytics-to-flourish

[72] Deborah Helen Selman. “Paradigm of Prediction: Predictive Analytics to Prevent Congestive Heart Failure”. In: Online Journal of Nursing Informatics 18.2 (2014), p. 1. issn: 10899758.

[73] David W Bates, Suchi Saria, Lucila Ohno-Machado, Anand Shah, and Gabriel Escobar. “Big data in health care: using analytics to identify and manage high-risk and high-cost patients”. In: Health Affairs 33.7 (2014), pp. 1123–1131.

[74] Nicholas Thapen, Donal Simmie, Chris Hankin, and Joseph Gillard. DEFENDER: Detecting and forecasting epidemics using novel data-analytics for enhanced response. eng. Ed. by Christopher M Danforth. 2016. doi: 10.1371/journal.pone.0155417.

[75] E Johnson. "Healthcare Data Mining Using In-database Analytics to Predict Diagnosis of Inflammatory Bowel Disease". PhD thesis. 2013. url: https://digital.lib.washington.edu/researchworks/bitstream/handle/1773/22596

[76] Jon Maguire and Vasant Dhar. "Comparative effectiveness for oral anti-diabetic treatments among newly diagnosed type 2 diabetics: data-driven predictive analytics in healthcare". In: Health Systems 2.2 (2013), pp. 73–92.

[77] W Raghupathi and V Raghupathi. An Overview of Health Analytics. Vol. 4. 3. New York: Working paper, 20131. doi: 10.4172/2157-7420.1000132.

[78] Erica D. Sonnenburg and Justin L. Sonnenburg. "A personal forecast". In: Nature 528.7583 (Dec. 2015), pp. 484–486. issn: 14764687. doi: 10.1038/528484a.

[79] Michael K K Leung, Andrew Delong, Babak Alipanahi, and Brendan J Frey. "Machine Learning in Genomic Medicine: A Review of Computational Problems and Data Sets". In: Proceedings of the IEEE 104.1 (2016), pp. 176–197.

[80] Sinjini Mitra and Rema Padman. "Engagement with social media platforms via mobile apps for improving quality of personal health management: a healthcare analytics case study". English. In: Journal of Cases on Information Technology 16 (Nov. 2014), pp. 73+. issn: 15487717.

[81] A Nimegeer, J Farmer, C West, and M Currie. "Addressing the problem of rural community engagement in healthcare service design". In: Health & Place 17.4 (2011), pp. 1004–1006. issn: 1353- 8292.

[82] P Attree, B French, B Milton, S Povall, M Whitehead, and J Popay. "The experience of community engagement for individuals: a rapid review of evidence". In: Health & Social Care in the Community 19 (2011), pp. 250–260. doi: https://dx.doi.org/10.1111/j.1365-2524.2010.00976.x.

[83] S. Nakibinge, D. Maher, J. Katende, A. Kamali, H. Grosskurth, and J. Seeley. "Community engagement in health research: Two decades of experience from a research project on HIV in rural Uganda". In: Tropical Medicine and International Health 14.2 (2009), pp. 190–195. issn: 13602276. doi: 10.1111/j.1365-3156.2008.02207.x.

[84] Roger P. Strasser. "Community engagement: a key to successful rural clinical education." In: Rural and remote health 10.3 (2010), p. 1543. issn: 14456354. doi: 1543.

[85] Gabriel J Milinovich, Ricardo J Soares Magalhães, and Wenbiao Hu. "Role of big data in the early detection of Ebola and other emerging infectious diseases". In: The Lancet Global Health 3.1 (Nov. 2016), e20–e21. issn: 2214-109X. doi: 10.1016/S2214-109X(14)70356-0.

[86] D Grant Campbell. “Big data and the study of dementia: Epistemological promises and pitfalls”. In: Proceedings of the American Society for Information Science and Technology 51.1 (2014), pp. 1–6. issn: 1550-8390. doi: 10.1002/meet.2014.14505101033.

[87] Peter Paul Verbeek. “Ambient intelligence and persuasive technology: The blurring boundaries between human and technology”. In: NanoEthics 3.3 (2009), pp. 231–242. issn: 18714757. doi: 10. 1007/s11569-009-0077-8.

[88] Matthew Dobra, Dorothy Weinstein, and Christopher Broyles. “Roadblocks, Regulation, and Red Tape: How American Health Policy and Industry Norms Threaten the Big Data Revolution”. In: Big Data and Health Analytics (2014), pp. 201–222.

[89] Accenture. Accenture Becomes Founding Sponsor of Vector Institute for Artificial Intelligence in Toronto. 2017. url: https://www.newswire.ca/news-releases/accenture-becomes-founding-sponsor-of-vector-institute-for-artificial-intelligence-in-toronto-617621083.html

[90] Guomin Ren and Roman Krawetz. “Applying computation biology and “big data” to develop multiplex diagnostics for complex chronic diseases such as osteoarthritis”. In: Biomarkers 20.8 (2015), pp. 533–539. doi: 10.3109/1354750X.2015.1105499.

[91] Ivo D Dinov, Ben Heavner, Ming Tang, Gustavo Glusman, Kyle Chard, Mike Darcy, Ravi Madduri, Judy Pa, Cathie Spino, Carl Kesselman, and others. “Predictive Big Data Analytics: A Study of Parkinson’s Disease Using Large, Complex, Heterogeneous, Incongruent, Multi-Source and Incomplete Observations”. In: PloS one 11.8 (2016), e0157077.

[92] Hui-Ling Chen, Gang Wang, Chao Ma, Zhen-Nao Cai, Wen-Bin Liu, and Su-Jing Wang. “An efficient hybrid kernel extreme learning machine approach for early diagnosis of Parkinson’s disease”. In: Neurocomputing 184 (2016), pp. 131–144.

[93] Peter M Dawson, Marleen Werkman, Ellen Brooks-Pollock, and Michael J Tildesley. “Epidemic predictions in an imperfect world: modelling disease spread with partial data”. In: Proceedings of the Royal Society of London B: Biological Sciences 282.1808 (2015). issn: 0962-8452. doi: 10.1098/rspb.2015.0205. url: http://rspb.royalsocietypublishing.org/content/282/1808/20150205.

[94] Fahad Pervaiz, Mansoor Pervaiz, Nabeel Abdur Rehman, and Umar Saif. “FluBreaks: Early Epidemic Detection from Google Flu Trends”. In: J Med Internet Res 14.5 (Oct. 2012), e125. doi: 10.2196/jmir.2102. url: http://www.ncbi.nlm.nih.gov/pubmed/23037553.

[95] Gustavo Saposnik, Robert Cote, Muhammad Mamdani, Stavroula Raptis, Kevin E. Thorpe, Jiming Fang, Donald A. Redelmeier, and Larry B. Goldstein. “JURaSSiC: Accuracy of clinician vs risk score prediction of ischemic stroke outcomes”. In: Neurology 81.5 (2013), pp. 448–455. issn: 00283878. doi: 10.1212/WNL.0b013E31829d874e.

[96] Ravi B. Parikh, Meetali Kakad, and David W. Bates. "Integrating Predictive Analytics Into High-Value Care". In: Jama 315.7 (2016), p. 651. issn: 0098-7484. doi: 10.1001/jama.2015.19417. url: http://jama.jamanetwork.com/article.aspx?doi=10.1001/jama.2015.19417.

[97] Karin M Nelson, Christian Helfrich, Haili Sun, Paul L Hebert, Chuan-Fen Liu, Emily Dolan, Leslie Taylor, EdwinWong, Charles Maynard, Susan E Hernandez,William Sanders, Ian Randall, Idamay Curtis, Gordon Schectman, Richard Stark, and Stephan D Fihn. "Implementation of the patientcentered medical home in the Veterans Health Administration: associations with patient satisfaction, quality of care, staff burnout, and hospital and emergency department use." In: JAMA internal medicine 174.8 (2014), pp. 1350–8. issn: 2168-6114. doi: 10.1001/jamainternmed.2014.2488. url: http://www.ncbi.nlm.nih.gov/pubmed/25055197.

List of Contributors

David Hume
BC Government

Greg Butler
Concordia University

Young Lee
Deloitte

Mike Miller
Ernst & Young

Dan Zikovitz
GE Canada

Jas Klotia
GE Canada

Laura Morin
Industry Canada

Eugene Wen
Manulife

Greg Dashwood
Microsoft Canada

Paresh Yadav
Phemi

Gabriel Musso
BioSymetrics

Daniel Lewis
RightBlue Labs

James Schuback
RightBlue Labs

Ronen Benin
RightBlue Labs

Hira Fatima
Ryerson University

Linda Koechli
Ryerson University

Linda Kaleis
Ryerson University

Tamer Abdou
Ryerson University

Ayse Bener
Ryerson University

Greg Horne
SAS

Muhammad Mamdani
St Michael's Hospital

Meir Dick
VISR

Rob Reichmann
VISR

This Big-data Consortium project is supported by the Office of the Provost and Vice-President, Academic and led by Data Science Laboratory at Ryerson University.

Notes

Notes

Notes

Notes

We would like to thank the following organizations for their leadership and in-kind contributions to Canada's Big Data Consortium, Canada's Big Data Talent Gap Study, and to this paper, *"A Vision for Data Monetization in Canada."*