Chapter 8

The evolution of health care in a data-rich environment

This chapter examines how large and diverse health data sets are being used to improve population health and support patient-centred care, health system management, and human health research. Among the aspects considered are electronic health records, smart models of care, the role of social media and crowdsourcing. The chapter also looks at barriers that will need to be overcome to pave the way for widespread data-driven innovation (DDI) in the health sector, examining issues raised by the use of personal health data not discussed in previous chapters. It concludes with a list of success factors that will enable governments to provide the leadership needed to progress further toward data-driven health research and care.

> *Big data is not just a quantitative change, it is a conceptual and methodological change. It will transform the way we do science and the way we deliver care. (Rossor in OECD, 2014a)*

The statistical data for Israel are supplied by and under the responsibility of the relevant Israeli authorities. The use of such data by the OECD is without prejudice to the status of the Golan Heights, East Jerusalem and Israeli settlements in the West Bank under the terms of international law.

The health sector is a knowledge-intensive industry: it depends on data and analytics to improve therapies and practices. There has been tremendous growth in the range of information that is being collected, including clinical, genetic, behavioural and environmental data. Every day, health care professionals, biomedical researchers and patients produce huge amounts of data from an array of devices, including electronic health records (EHRs), genome sequencing machines, high-resolution medical imaging, ubiquitous sensing devices (i.e. available anywhere), and smartphone applications that monitor patient health. The data generated are of great value to health care and research, and it is predicted that more medical information and health and wellness data will be generated in the next few years than ever before.

At the same time, the potential to process and analyse these emerging multiple streams and large volumes of data – big data – and to link and integrate them is growing. Such data-driven innovation (DDI) can yield many benefits, including new insights into the natural history of diseases and their diagnosis, prevention and treatment, and greater opportunity for further development of personalised therapies. Indeed, there is growing evidence that big data can be leveraged to transform health care.

Box 8.1 points to four basic categories of digital data use that can bring value to citizens, care providers and the system itself. The data can help: improve patient care; manage the health system; understand and manage population and public health; and facilitate health research. However, many challenges must be overcome before the benefits from DDI in the health sector can be reaped. One of these is that EHRs are being collected in health care systems that are often fragmented, with points of care functioning as silos. Questions of privacy also have to be addressed, and skill building will be needed to analyse voluminous health data sets.

This chapter reviews the evidence for big data's potential; equally, it considers the barriers that will need to be overcome to pave the way for widespread DDI in the health sector. The first section examines factors driving greater use of large-scale health data to generate knowledge and yield new intelligence for health system management and policy making. The second section reviews national health data sources and their capacity to be brought together, so as to understand health care pathways – that is to say, patients' progress through the health care system, from their earliest to last days – and raise the quality of clinical care. The third section provides examples of how large-scale data sources are already playing a role in DDI – helping create new and "smarter" models of care, care that is more centred on the patient, and a more efficient clinical research enterprise for improved prevention and better disease management. The fourth section provides an overview of the challenges of transforming health research with big data, including the need for infrastructure and analytical tools to analyse large health data sets.

The chapter concludes with a discussion of the factors that can enable governments to provide the leadership needed to progress further toward data-driven health research and care. The chapter thereby highlights specific opportunities for fostering Alzheimer's and dementia research in response to the direct mandate from the *G8 Dementia Summit Declaration* to the OECD to "take stock of our current national incentive structure for research […] and consider what changes could be made to promote and accelerate discovery and research and its transformation into innovative and efficient care and services".[1]

Box 8.1. **Uses of digital data in the health system**

Improving patient care – Secondary use of health data can improve quality initiatives in and the effectiveness of patient care, in both clinical and home care settings. For example, administrators and front-line clinicians can be alerted when measures related to quality and patient safety fall outside a normal range, and notified of factors that may be contributing to the deviations. Clinicians can aggregate and reuse data from their patients to evaluate their own performance against clinical practice guidelines. The data can also provide insights that lead to revised care protocols.

Managing the health system – Health data can be used to manage and improve the effectiveness and efficiency of the health system by informing decisions regarding programmes, policy and funding. For example, costs can be reduced by identifying ineffective interventions, missed opportunities, and duplication of services. Access to care can be increased and wait times reduced by understanding patient journeys across the continuum of care; by ensuring that patients receive the services most appropriate for their needs; by accurately projecting the future health care needs of the population; and by optimising the allocation of resources across the system.

Understanding and managing population and public health – Health data can be used to understand the burden of illness and quality of life of the population, and to manage and evaluate public health interventions, including for health promotion and prevention. For example, in addition to timelier public health surveillance of influenza and other viral outbreaks, data can be used to identify unanticipated side effects and contraindications of new drugs.

Facilitating health research – Health data can be used to support research in many fields that informs clinical programmes, health system management, and population and public health. For example, multiple sources of data can be integrated to find early (bio)markers of disease; the comparative cost-effectiveness of different interventions can be evaluated; and historical data can be used to simulate and model trends in long-term care needs, and evaluate different policy options to meet those needs.

Source: Adapted from Canadian Institute for Health Information, 2013.

8.1. Drivers of growth of digitised health data

The amount of digitised data available in the health sector is growing rapidly. There are five principal factors driving the increased collection and use of large-scale data in this sector. They are: i) demographic changes and the shifting of the global disease burden toward long-term non-communicable diseases; ii) fiscal pressures and the need for greater efficiencies; iii) the need for more responsive, patient-centred services; iv) increasing global co-operation to address common health problems; and v) the volume, velocity and variety of health data available (see Chapter 3).

Demographic changes and non-communicable diseases

The global burden of disease has shifted in the past 20 years, from infectious conditions to long-term non-communicable diseases (NCDs) brought on by lifestyle choices and environments: heart disease, stroke, diabetes, chronic neck and back pain, cancer, and depression (IHME, 2013). Further, populations are ageing and many people are living longer with multiple morbidities (concurrent diseases) and disabling conditions. Since 1970, average life expectancy has risen by 35 years worldwide, with gains in years achieved across the world's regions. OECD countries have witnessed extraordinary gains

in longevity, with average life expectancy at birth rising 10 years since 1970 to exceed 80 years in 2011 (OECD, 2013a). Globally, the proportion of people over 80 years of age in particular is anticipated to increase by 2.5 times between 2010 and 2050 (UN, 2013). The rapidly expanding cohorts of elderly and older elderly will include a significant proportion of persons with chronic diseases.

The rising burden of NCDs, multimorbidities and risk factors for NCDs has important implications, for how care is best organised and provided; where new treatment innovations and preventive approaches can be expected; and future cost pressures.

To address that burden, medicine must focus on preventing the onset of NCDs and controlling their progression, including through lifestyle changes. At the same time, health systems must focus on improvements in care co-ordination and delivery. The current pace of innovation in genomics, biological systems, and information and communication technologies (ICTs) has the potential to increase our ability to predict and prevent disease and promote healthy behaviours; develop cost-effective therapies; redesign health care systems to assure integrated and co-ordinated care; improve safety and quality for patients; and extend healthy lives. As will be explored further in this chapter, these advancements are closely linked to the generation and analysis of data that permit the study of full populations and their health care experiences and outcomes.

Fiscal pressures and the need for greater efficiencies

Fiscal pressures will continue to push governments to seek greater efficiency, accountability and quality in the health care sector. During the fifty years prior to 2009, health spending in OECD countries outpaced economic growth, resulting in an increasing share of GDP allocated to health. By 2009, 9.6% of GDP in OECD countries was allocated to health, up from under 4% in 1960 (OECD, 2013a). Average annual growth in health spending in real terms between 2000 and 2009 was 4.1%, compared to GDP growth of only 1.5% (OECD, 2013b). Since 2009, many countries have reduced budgets for health in response to the economic downturn. By 2012, health expenditures accounted for 9.3% of GDP (OECD, 2014b). Governments under pressure to protect funding for acute care have been cutting other expenditures, such as public health and prevention programmes. In 2012, on average across OECD countries, only 3% of health budgets were allocated to prevention and public health programmes in areas such as immunisation, smoking, alcohol, nutrition and physical activity. The long-term wisdom and sustainability of such budget reductions in spending on prevention is uncertain, as is the ability of governments to continue to contain rising costs.

Continuing pressure to find ways to make systems more productive has moved the focus from cost containment to performance-based governance. To evaluate health sector performance, managers and governments will need timely and accurate information about the prices and volumes of services provided and the health outcomes produced, at levels sufficiently detailed to take corrective policy action. The need to manage health system outcomes more actively will lead to greater use of clinical and administrative data to assess the comparative effectiveness of therapies and services. These data will also be needed for redesigning and evaluating new models of health care service delivery.

The need for more responsive, patient-centric services

The role of patients in the care process – managing their own health – has taken on much greater importance in recent years. In order to address patients' expectations for seamless care, it will be increasingly important to improve co-ordination and integration

of care provided by different parts of the health and social care systems. Patients' taking command of the management of their own health will support better outcomes and coordination of care, particularly for patients with chronic diseases who often require services from multiple health care providers.

The increasing use of electronic medical records promotes patients' participation in their care, self-management of health conditions, and informed decision making. Patients' interest in their diagnostic test results and medical records, in their options for care, in the quality of providers, and in scheduling visits online will keep growing. Over the past decade multiple studies have documented the value of electronic personal records (EPRs) in supporting greater patient-centred services. As an illustration, the US Department of Veterans Affairs offers patients access to an EPR that includes details from their medical history, such as clinical notes and laboratory test results. A patient survey to evaluate the service indicated that the majority of veterans viewed the personal record as helpful to them; as having made it easier to locate information they needed; as having improved their care; and as a tool they would recommend to others (Nazi et al., 2013).

Patients and practitioners are also increasingly interested in devices, tools and computer applications that assist in monitoring and improving health and well-being. They recognise that these can help patients live longer in their own homes rather than in considerably more expensive hospital or nursing home facilities; enable longer-term independent living; and encourage personal responsibility for healthier lifestyles (OECD, 2013c). Many such emerging information and communication systems have the potential to provide new streams of data for evaluating treatments and measuring and evaluating health care outcomes. However, in many countries challenges have yet to be met to unleash this potential.

Increasing global co-operation to address common health problems

The fourth driver is the need for co-operation to tackle global public health challenges such as infectious diseases, and improve early detection and warning of emerging health threats and events. Examples include the Program for Monitoring Emerging Diseases (ProMED), established in 1993, which has demonstrated the power of networks and the feasibility of designing effective, low-cost global reporting systems. ProMED has also encouraged the development of additional electronic surveillance data-sharing networks – such as the Global Public Health Information Network (GPHIN)[2] and HealthMap.[3]

Influenza surveillance is one of the most developed global surveillance and monitoring systems of the World Health Organization. It began in 1948 and has developed over the years into a highly successful global partnership. The network now involves 110 collaborating laboratories in 82 countries, constantly monitoring locally isolated influenza viruses and providing real-time streams of data on the emergence and spread of different strains.

Complementing these traditional case-based and syndromic surveillance systems, monitoring of unstructured events – through news and Internet media, web searches (e.g. Google Flu trends), etc. – has been a significant component of public health early warning and response over the past decade. More recently, with the increase of Web 2.0 platforms and social media, there is a new real-time source of intelligence provided by citizens that is immediately in the public domain and thus readily available. During the recent Ebola outbreak in West Africa, epidemiologists and telecommunication companies were exploring the potential of new data sources, such as mobile phones, to better model the spread of the disease (The Economist, 2014; Wall Street Journal, 2014).

In addition to monitoring, there is increasingly global interest in research to tackle the emergence of NCDs, through better preventive interventions and treatments. The OECD is actively engaged in a global project to improve data sharing and access internationally to accelerate innovation that addresses dementia (OECD, 2013d). The focus on dementia is the result of a direct mandate from the *G8 Dementia Summit Declaration* to the OECD.

Volume, velocity and variety of health data

The fifth and possibly most important driver of health data use is the sheer volume, velocity and variety of health data available. As will be discussed in the next sections of this chapter, many health care systems are rapidly digitising immense amounts of clinical, financial and operational data and using them for a wide range of activities, including:

- preventive care, e.g. early detection
- field data to support emergency and urgent care
- coaching, rehabilitation and maintenance
- intervention, e.g. reminders
- epidemiological assessments
- post-market surveillance and analysis
- health care quality and performance monitoring.

This will require real-time continuous archiving of multi-modal data sets and multi-domain collaborative annotations, as well as post-therapeutic visualisation of the archived data. The volume of this data is set to increase dramatically with advances in mHealth (mobile health, involving mobile devices), sensor and imaging technologies to support diagnosis and treatment. Further, these data are heterogeneous (structured, unstructured, text, etc.), reflecting the traditional silos across care settings, industry/research, and scientists/clinicians. The cost and complexity of linking data stored in these various formats are decreasing, enabling analysis of health care interventions and utilisation enhanced with additional information about personal behaviours, lifestyles and genetic profiles.

This remarkable expansion of digital health data is in turn largely driven by the confluence of important technological developments. These include notably the increasing ubiquity of broadband access and the proliferation of smart mobile devices and emerging smart ICT applications, empowered by sensor networks and machine-to-machine (M2M) communication. Cloud computing has also greatly increased data storage and processing capacity (see Chapter 3). Great reductions in storage costs over the past 20 years have also been a significant driver, as they have enabled the collection and use of large volumes of health-relevant data; electronic health records and genetic, neuroimaging and epidemiological data are just a few examples.

All of these drivers have greatly increased not only the availability of data in the health sector, but also – with developments in computing power – their use, creating new opportunities to obtain insights. The rise of chronic conditions and fiscal pressures will make it increasingly important to be able to follow health care pathways and determine which paths deliver better outcomes in an efficient manner. Patients will want the health sector to improve therapies, and will want health care experiences to be as modern as other business services in terms of service responsiveness, transparency and

communication. Opportunities for global co-operation in sharing data to find solutions to common challenges may continue to present themselves, particularly as the urgency of addressing NCDs and new infectious diseases rises. For all of these reasons, there will be continued interest in developing and using data to advance health care therapies, health care delivery, and health system governance. The next sections review the four basic categories of data use described in the introduction and then discuss the critical success factors and policy priorities for addressing challenges that may be limiting data sharing and use.

8.2. Data-driven innovation to improve health care quality and health system performance

Essential to the monitoring and evaluation of both health care quality and health system performance is the ability to track patients as they progress through the system, from primary health care to speciality care to hospitalisations, long-term care, home care, hospice care and death. Tracking data should also provide information about patient characteristics, illnesses, medications, therapies, laboratory tests and medical images. This type of follow-up permits a comprehensive view of health care services and evidence of what is effective and under what circumstances. It can also uncover, among other things, medical errors, adverse drug reactions, fraud, adherence to clinical guidelines or lack thereof, optimal care paths and patients with optimal treatment results.

Although the capacity to collect and analyse data related to health care pathways is increasing, only a handful of countries have health information systems organised to permit comprehensive views of patient care across the health care continuum. Key pieces of information about patients' care paths are instead often isolated in various databases, such as hospital admissions and discharges, primary care records, insurance claims, pharmaceutical databases, image banks and patient surveys.

This section begins with a review of current national health data assets and their capacity to be integrated to understand health care pathways. It goes on to discuss the development and use of electronic health record systems and their potential to improve measurement of health care pathways. Examples are provided of how some countries have developed data to follow the pathway of care in order to monitor health care quality and health system performance nationally, and to contribute to international comparisons. It then describes the emerging field of comparative effectiveness (or relative effectiveness) research, which uses health care pathway data to determine which therapies or processes of care are the most effective.

Key national health data sets

While countries are investing in data infrastructure, a 2011/12 OECD survey of national health data sets (OECD, 2013e) reveals significant cross-country differences in data availability and use. Some countries have seen significant progress in data use and its compatibility with robust privacy protection. Others have limited data and restrictions preventing access to its use, even by government (see Table 8.1). Most of the countries responding to the survey reported that they collect national data across the continuum of health care services and at the level of individual patients or persons. Users of national data included governments, insurers, research institutes and health care providers. About half of the countries reported that some of their key data sets were regularly linked together for research purposes or statistics to better understand health care pathways and outcomes. Similarly, about half of the countries reported that some of their key data sets

were routinely linked to monitor health care quality. Very few countries, however, link data routinely to monitor health care quality in several important areas of health care: prescription medicines (seven countries); mental hospital in-patients (five countries); primary health care (four countries); and long-term care (four countries).

Table 8.1. **Number of countries[1] reporting data and data linkages**

Data is about:	Hospital in-patients	Deaths	Cancers	Rx[2]	Mental hospital in-patients	Primary care	Long-term care	Health risks and behaviours	Socio-economics: income, education, employment, ethnicity
National data set available…	19	19	17	14	17	16	16	19	19
Contains records for patients or persons	16	17	16	12	14	13	13	16	16
Is linked to other data for health research or monitoring	14	15	13	12	8	10	11	10	11
A linkage study is usually[3] under way	12	15	11	10	7	8	6	7	11
A linkage study to monitor *health care quality* is usually under way	12	12	11	7	5	4	4	4	4

1. Nineteen countries responded to the survey. Australia, Belgium, Canada, Denmark, Finland, France, Germany, Israel, Japan, Korea, Malta, Norway, Poland, Portugal, Singapore, Sweden, Switzerland, the United Kingdom and the United States.

2. Pharmaceutical drug utilisation.

3. The data set is used to undertake record linkage projects on a regular basis, such that a data linkage project involving the data set is usually underway.

Source: OECD, 2013e.

Electronic health record systems

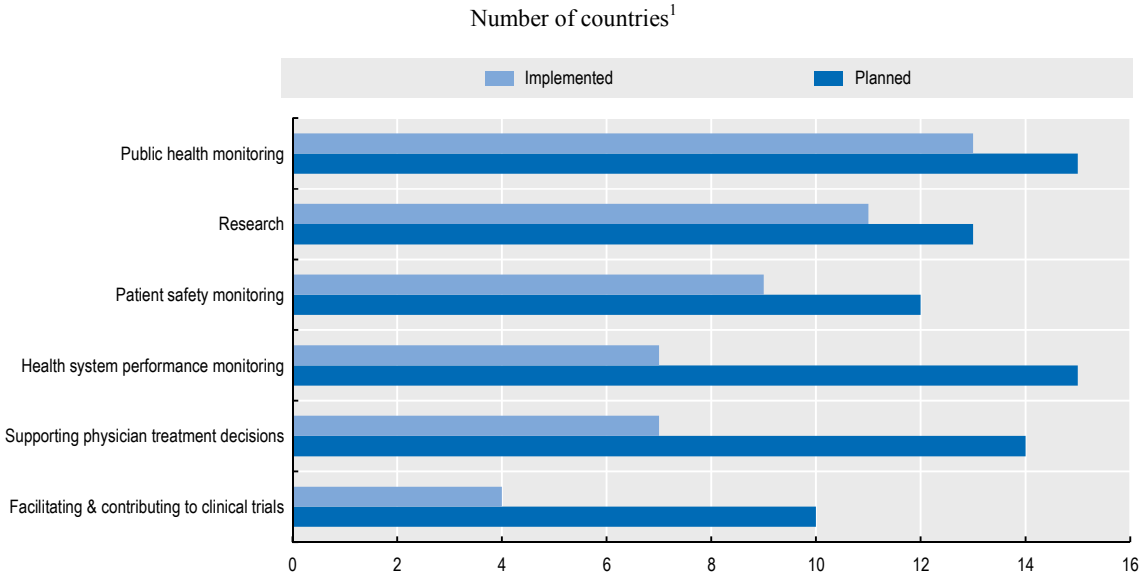
The development and use of data from electronic health records (EHRs) have the potential to support health care DDI and to improve the quality, safety and performance of health care systems. In its most mature form, an EHR system contains or virtually links together information about a patient's health care encounters, including diagnosis, radiology, laboratory tests and medications. Patient identifying information is necessary to bring the information together from various related data sets and then to retrieve the information when it is needed. As time passes, a comprehensive health care biography can emerge from the data available in the system to support the care of the individual patient; for population-level statistics and research, it can improve existing therapies, discover new ones, and improve the quality, safety and performance of health care systems.

In 2012 most of the countries studied (22 of 25) reported a national plan or policy to implement EHRs, and most had already begun to implement that plan (20 countries). EHR systems in some countries include data on patients' key characteristics and health problems, as well as their history of encounters with the health care system and the treatments they have received from a variety of health care providers. The greatest

contribution of these systems, as they develop, is the potential for secondary analysis of the data to monitor and conduct research to improve the health of the population and the quality, safety and efficiency of health care.

Of the 25 countries studied, 18 had included some form of secondary analysis of EHRs within their national plan (Figure 8.1). The most commonly included secondary uses reported were public health and health system performance monitoring. Fourteen countries also indicated that they intended for physicians to be able to query the data to support treatment decisions. The least commonly reported planned data use was for facilitating or contributing to clinical trials. This use was noted by ten countries.

Figure 8.1. **Planned and implemented uses of data from electronic health record systems**



1. Twenty-five countries responded to the survey. Austria, Belgium, Canada, Denmark, Estonia, Finland, France, Germany, Iceland, Indonesia, Israel, Japan, Korea, Mexico, Netherlands, Poland, Portugal, Singapore, Slovakia, Slovenia, Spain, Sweden, Switzerland, United Kingdom and United States.

Source: OECD, 2013e.

Many countries also reported that regular use of EHR data for public health monitoring (13 countries) and general research (11 countries) was already under way. There are currently several ongoing projects addressing the (re)use of EHR data for purposes of clinical research. In the United States, initiatives such as i2b2,[4] the eMERGE network,[5] the Kaiser Permanente Research Program on Genes, Environment and Health (RPGEH)[6] and the Million Veteran Program[7] are focusing on integrating EHRs and genomic data (Jensen, Jensen and Brunak, 2012). The Stanford Translational Research Integrated Database Environment (STRIDE) is an example of a US project that aims to create an informatics platform supporting clinical and translational research.[8]

In Europe, a number of research projects and initiatives such as the i4health network,[9] EMIF (European Medical Information Framework),[10] eTRIKS (Delivering European Translational Information & Knowledge Management Services),[11] INTEGRATE (Integrative Cancer Research through Innovative Biomedical Infrastructures),[12] Linked2Safety,[13] SALUS (Scalable, Standard based Interoperability Framework for Sustainable Proactive Post Market Safety Studies)[14] and TRANSFoRm (Translational

Research and Patient Safety in Europe)[15] are concerned with (re)using EHRs to facilitate clinical research by focusing on different disease domains and addressing different data use cases and scenarios.

There are several significant differences between countries whose national plans or policies called for at least four of the data uses outlined in Figure 8.1 (the engaged) and those who were planning on fewer or no secondary data uses (the cautious). Data privacy and security concerns have been one of the major barriers to the adoption of EHRs and implementation of a national health information exchange in a number of countries. These issues will be reviewed in some depth in the last section of this chapter. Other barriers include a lack of technology standards that could facilitate interoperability, and the cost of implementing such a system. Engaged countries were somewhat more likely than cautious countries to report having created national governing bodies responsible for clinical terminology and interoperability standards – 62% compared with 50%. Terminology standards ensure that the data is captured in a consistent manner through a structure that enables statistics and analysis. Interoperability standards ensure that records can be shared or exchanged. Nonetheless, where health care systems are fragmented, with points of care operating in silos, implementation of standardised EHR systems remains challenging.

Virtually all of the engaged countries (92%) have developed a national minimum data set that standardises the content of patient records that are intended to be shared among health care providers. In contrast, only one-half of the cautious countries have defined a minimum data set. Engaged countries (54%) are also somewhat more likely than cautious countries (42%) to report that their EHR system is already being used to create data sets for statistics and research. As a result, they are also more likely to have instituted processes for auditing the clinical content of electronic records for quality, although this is still relatively rare for both groups. Many engaged countries have also consulted with the public on privacy and security issues, and have developed data governance frameworks that permit privacy-protective data uses.

Harnessing value from data to improve health system performance

Countries that are actively monitoring health care quality and health system performance provide very interesting examples of how the data are being used and the benefits accrued. Examples of data use range from evaluation of the quality and cost-effectiveness of treatments to monitoring adverse events related to pharmaceuticals and medical devices; incorporating the results of care pathway analysis into evaluations of and revisions to clinical care guidelines; and building pathway data to promote world-class research.

Evaluation of treatment quality and cost-effectiveness

Finland monitors the content, quality and cost-effectiveness of a set of selected diseases and treatments (stroke, premature newborns, hip fracture, breast cancer, schizophrenia, heart attack, hip and knee replacement surgery, and invasive heart surgery) by linking patient data for the Finnish population across the whole cycle of care, from admission to hospital to care by their community doctor to the medications prescribed and deaths (OECD, 2013e). From both administrative data and data extracted from electronic health records, Finland has new indicators for each hospital to evaluate treatment quality and cost, including: mortality rates; emergency room visits and readmissions to hospital;

infections and complications; and stays in nursing homes and home care visits. Hospital quality is improving, as the results are publicly available.

Within the *United Kingdom, England* has a new initiative called care.data that aims to create data about episodes of care. Included are both health care and social care, with data pertaining to pathways between primary and secondary care and information about diagnosis, laboratory tests and prescription medications (NHS, 2013). The six aims of the care.data initiative are to support patient choice, advance customer services, promote greater transparency, improve outcomes, increase accountability, and drive economic growth by making England a centre for world-class health services research. Data for consenting patients within the entire population of England will be linked, with data extracts taking place monthly to ensure timely monitoring.

Japan has created a new medical insurance claims database to assist the Ministry of Health, Labour and Welfare in the preparation, implementation and evaluation of a plan to optimise medical care costs. The data were provided to researchers and to prefectures on a trial basis in 2011 and 2012. Several cost and quality studies were undertaken and published as a special issue of the *Journal of the National Institute of Public Health*. These studies included a linkage of insurance claim data with data on the provision of guidance to patients during periodic health check-ups regarding metabolic disease (Okamoto et al., 2013). The study found a reduction in the onset of metabolic disease and in health care expenditures among patients who received guidance about reducing disease risk during health check-ups.

Monitoring the underuse, overuse and misuse of therapies

Korea uses population-wide health insurance claim data to identify underuse, overuse and misuse of therapies and to reduce variation in care practices by regularly reporting quality indicators, including mortality and readmission after hospital procedures; inappropriate prescribing in primary care; and outcomes following discharge from mental health hospitals (OECD, 2013e). Korea links claims data for patients across the entire pathway of care, and is able to report timely results.

Quality and efficiency assessments of clinical care guidelines

Sweden is breaking new ground by using data to undertake both quality and efficiency assessments of clinical care guidelines (OECD, 2013e). These guidelines inform physicians and health care professionals about the most appropriate therapies for patients with different health profiles and problems. By following a patient's cycle of care, they are able to evaluate the extent to which guidelines are being followed and whether or not the health outcomes of the patient meet expectations. This evidence is then used to revise the guidelines, completing an ongoing cycle of improvement in care quality and efficiency.

Monitoring adverse events related to pharmaceuticals and medical devices

The *United States* Food and Drug Administration has implemented a sentinel project to transform how it monitors the safety of the medicines, medical devices and biologics that it regulates, by tapping directly into electronic health records, administrative data and insurance claim records. Building toward a nationwide rapid-response electronic safety surveillance system, the sentinel pilot study involves 17 data partners across the United States, and encompasses the data of nearly 100 million patients (FDA, 2013).

The *EU* Advanced Drug Reporting (EU-ADR) [nowhere do I find that name linked to the abbreviation. ADR instead would appear to stand for "Adverse Drug Reactions" – please confirm] initiative defined a proactive strategy for post-market drug assessment based on automating analysis of data stored in large electronic health record databases in four European countries (Denmark, Italy, the Netherlands and the United Kingdom) and covering 30 million patients (Coloma et al., 2012). EHR data are analysed to identify a ranked list of signals of potential adverse events and their significance in terms of health risks. Adverse events monitored include acute myocardial infarction, acute renal failure, anaphylactic shock and gastrointestinal bleeding. Results indicate that active surveillance for signal detection with health care database networks is feasible, but that it would be necessary to expand the data network coverage to a larger pool of patients – that is, to more participating countries – to monitor the effects of infrequently used drugs.

Generating clinical pathway data to promote health services research

In the *United States*, Kaiser Permanente, a health care maintenance organisation (HMO) with 8 sites and 9 million members, has 7 research centres conducting public domain research with patient-level data. Kaiser's experience with linking data across the health care pathway for research extends back 50 years. Kaiser is now at the forefront of this field with the data it can extract from its electronic medical system and the data it can link together with patient care pathways from its biobank. A new study that Kaiser described to the OECD involves examining whether certain prescription medicines for mental illness may be linked to the development of genetic mutations in humans (OECD, 2013e).

In *Canada*, the Institute for Clinical and Evaluative Sciences (ICES) is a research centre at the University of Toronto that provides population-based health services research for Canada's largest province, Ontario (OECD, 2013e). ICES collects personal health data from the Ontario Ministry of Health and Long-term Care and other entities. Findings in 2013 included that commonly co-prescribed statins and antibiotics are linked to muscle loss and kidney failure in seniors; that a recent colorectal cancer screening programme was not able to fully address inequities in access; and that implementation of North America's first stroke care facilities improved outcomes (ICES, 2013).

Cross-country comparisons of health system performance

Collaborative big data efforts to improve health system performance investments in the development of internationally comparable population-level health data are leading to new ways to benchmark and compare how health systems are performing to help countries to improve patient safety, health outcomes and system performance. Within Europe such efforts are funded by the European Union. Two examples from the EU Seventh Framework Programme are EuroHOPE and ECHO.

EuroHOPE – the European Health Care Outcomes, Performance and Efficiency Project – is evaluating the performance of European health care systems in seven countries, in terms of outcomes, quality, use of resources and costs (Häkkinen et al., 2013). Participating countries include Finland, Italy, Netherlands, Norway, Sweden and the United Kingdom (Scotland). Health care data for hospitalisations, pharmaceuticals, registered cancers and deaths are linked to follow patient pathways of care. The patient groups studied are those with acute myocardial infarction, stroke, hip fracture, breast cancer and low birth weight. EuroHOPE is developing indicators that it will recommend to the European Union for routine reporting; developing methods for international comparative health services research based on the linkage of person-level data; and

informing the public about the policy-relevant drivers of health care quality – including treatment practices, use of medicines and new medical technologies, waiting times, organisation of care, and costs.

ECHO, the European Collaboration for Healthcare Optimization project, has pooled hospital administrative and contextual data from seven countries (Austria, Denmark, England, Portugal, Slovenia, Spain and Sweden) to learn more about variation in care access and outcomes, and the relationship between this variation and the socio-economic status of the areas in which patients live.[16] ECHO intends to explore whether place of residence and access to particular health care providers have a bearing on whether or not care is safe and effective, by examining within-country and between-country variations. ECHO is the first international health system performance comparison to pool personal health data, into a data set of 200 million hospital discharges.

Assessing variability in health care treatment across countries

In terms of single disease areas, there is no doubt that the long history of databases to register cases of cancer has endowed research in that area with the evidence necessary to monitor and advance quality of care. Indeed, of the 19 countries responding to the OECD survey in 2011/12, 16 were maintaining national cancer registry data sets. There is also a long practice in most OECD countries of linking cancer registrations and death databases in order to estimate cancer survival rates. The CONCORD-2 study is a worldwide comparison of cancer survival from over 270 cancer registries in 61 countries for 10 cancer sites in adults and childhood leukaemia. This study examines underlying causes of differences in survival rates.[17] The International Cancer Benchmarking Project is advancing this research further. In this project, cancer registries with details about the cancer stage at diagnosis have been analysed in six countries (Australia, Canada, Denmark, Norway, Sweden and the United Kingdom), to compare differences in survival and to discover why differences occur. Thus far, the researchers have found that patients in Sweden are the most likely to survive at least one year after diagnosis of breast, bowel and lung cancers; those in the United Kingdom are the least likely (Cancer Research UK, 2013). The role of treatment in survival differences by cancer stage is the next stage of inquiry for the project.

The emerging field of comparative effectiveness research (CER)

Comparative effectiveness research is designed to inform health care decisions by providing evidence on the effectiveness, benefits and harms of different treatment options. The evidence is generated from research studies that compare drugs, medical devices, tests, surgeries, and ways to deliver health care (AHRQ, 2013). CER ultimately seeks to provide pragmatic knowledge that can be applied toward delivering "the right treatment to the right patient at the right time". Achieving this goal in an area as complex as health care, however, requires robust, accessible data sources capable of providing detailed patient-level information in a time- and cost-efficient fashion.

Patient registries have been used throughout Europe for CER as well as for patient-centred health outcomes research (PCOR). A recent review of disease and treatment registries in Europe (Larsson et al., 2012) describes how improvements in health outcomes, like fewer revisions after hip replacement surgery and gains in survival after acute myocardial infarction (AMI), followed implementation of public reporting of outcomes by providers and the engagement of the clinical community to address quality concerns.

Two incentive programmes in the United States have increased the volume and velocity of data generated for CER in recent years (*The Daily Briefing*, 2011). The first was an allocation of USD 1.1 billion for CER funding within the 2009 federal stimulus package. The second was the launch of the Patient-Centered Outcomes Research Institute in 2011, with the capacity to fund USD 550 million to conduct CER and establish priorities for national CER (PCORI, 2013).

Three representative examples of the CER approach have been published in the United States. The first is an evaluation of a comprehensive programme to control hypertension; the second is an evaluation of methods to improve colorectal cancer screening. The third identified an unexpected and dangerous interaction between two of the most widely prescribed medications: pravastatin, prescribed for hypertension control, and paroxetine, an anti-depressant (Jaffe et al., 2013; Mosen et al., 2010, 2013; Tatonetti et al., 2011).

The rollout of EHRs in many OECD countries will help health care systems reach the vision of CER as a valuable resource for informed health care decision making.

8.3. Data-driven innovation for smarter models of care

There is a broad and growing consensus that any systematic effort to address today's health and wellness challenges will also require data to support new and "smarter" models of care, that recognise the need to keep the elderly and the disabled in their own homes rather than in the considerably more expensive hospital or nursing home systems; that enable longer-term independent living; and that encourage personal responsibility for healthier lifestyle choices. The effort will require enhanced capacity for the sharing, processing and analysis of health and behavioural data to support patient-centred care, and a more efficient clinical research enterprise for improved prevention and better disease management.

Taking shape alongside these goals is a vision for a "learning health system". The Institute of Medicine, a long-time proponent of the concept, defines a learning health system as: "... one in which progress in science, informatics, and care culture align to generate new knowledge as an ongoing, natural by-product of the care experience, and seamlessly refine and deliver best practices for continuous improvement in health and health care" (Grossmann, Powers and McGinnis, 2011).

One example of a rapid learning health care system is the American Society of Clinical Oncology's Cancer Learning Intelligence Network for Quality (CancerLinQ) system,[18] CancerLinQ is designed to address the growing challenge of managing the deluge of data emerging from precision medicine for cancer care. The system incorporates data from researchers, providers and patients in order to continually improve comprehensive clinical algorithms reflecting preferred care at a series of decision nodes for clinical decision support.

These concepts, and the new models of care they represent, require a major shift from traditional practices. Today's care is reactive, episodic and focused on disease. The new health care will need to be proactive, preventive and focused on quality of life and well-being.

Current health care is usually provided within hospitals and clinics. New smart models of care could become more patient-centric, with greater opportunity for care to be

provided at home and include the broader social network (family and community) as a significant contributor to individual health and well-being.

Smarter models will be data-driven and promise to deliver greater safety and efficacy through evidence-based approaches and personalised care. This section examines the strategic directions that OECD countries are considering to realise this vision for health and wellness, from both the technological and policy viewpoints. It looks at the role of big data and ICTs, and discusses the research and policy options that could further the development of smarter models of care. It draws on the OECD – United States National Science Foundation workshop (and related report) entitled "Building a Smarter Health and Wellness Future", which was held in Washington, DC on 15-16 February 2011.

Personalised care

The power of health information processing is such that it is possible today to personalise therapy in wholly new ways. Culture, living style, belief systems and expressed choice comprise one dimension. A second dimension is the ability to search and process electronically recorded medical histories of individuals. This enables rapid identification of not just personal biological responses such as allergies, but also a much richer pattern of personal information such as results of diagnostic tests and outcomes of particular therapies. Thirdly, new genomic knowledge can help identify population group variations that influence care response, but also personal genetic profiles that can inform not just individual therapies but also selective targeted prevention.

Advances in DNA sequencing and whole genome analysis have made it possible to develop a greater understanding of response to treatment. In oncology, for example, pathologists measure whether the cancer is hormone sensitive to determine eligibility for tamoxifen therapy among those suffering from breast cancer. Effectiveness has been found to be contingent on an enzyme (cytochrome enzyme P450 2D6) needed to metabolise the drug, although the results have not always been consistent across studies (Roederer, 2009).

With the costs of whole genome sequencing declining, the expectation is for personalised medicine to be streamlined into medical practice. New data management and processing methods are needed in four areas to realise this potential: i) the processing of large-scale robust genomic data; ii) interpreting the functional effect and the impact of genomic variation; iii) integrating systems data to relate complex genetic interactions with phenotypes; and iv) making the data available at the point of care in such ways that the comprehensiveness of the information provided to the clinician supports the clinician's ability to accurately and rapidly prescribe drugs that are safe and effective for a specific patient (Fernald et al., 2011).

In addition, the full extent of patient data will need to be accessible so that questions spanning multiple data sources can be asked and answered. The consistency and completeness of patient EHRs will be increasingly important.

Ubiquitous and pervasive patient care

The ubiquitous care model is based on the utilisation of smart sensing and biometric devices for real-time monitoring, analysis and transmission of health data. The information can then be accessed by health care providers for informed diagnosis, clinical decisions regarding treatments, and evaluation of outcomes. It can also be viewed and acted upon by patients for both education and prevention.

The technology to support ubiquitous sensing already exists, and today an increasing amount of physiological monitoring data streams are displayed on medical devices. The key challenge is to combine these technologies with network infrastructure to create an integrated architecture that extends care outside the hospital to the home and to mobile patients – thus the term ubiquitous.

For example, in the case of managing patients with acute diabetes, the blood glucose level can be monitored continuously through an implant that controls the insulin delivery from a reservoir. In cardiology, there is increasing recognition of the value of implantable sensors for continuous monitoring of the most important physiological parameters for identifying the precursors of major adverse cardiac events, including sudden death. The data streams provide enormous potential for improved diagnostics, prevention, support of evidence-based practices, and remote health care. These data can yield answers to clinical questions, or raise new questions that influence care responses.

An area that is progressing rapidly is the Body Area Networks (BAN). Medical applications of BAN cover continuous real-time sampling of biomedical signals, monitoring of a person's vital signal information, and of low power medical devices. They can be broadly classified into two categories depending on their operating environments. One is the so-called wearable BAN, which is mainly operated on the surface or in the vicinity of body, such as medical monitoring.

Another is the so-called implantable BAN, which is operated inside the human body, e.g. the capsule endoscope and pacemaker. The former provide long-term health monitoring of patients in natural physiological states without constraining their normal activities. The latter allow communication between implanted devices and remote monitoring. One example of smart application is the "virtual ward", in which patients are monitored at home and visited by mobile medical teams when the data show that it is necessary. That is generally better for the patients and may be less expensive.

Ubiquitous computing can also be leveraged as a means to provide context and location-aware cues for health action. The greatest power of such techniques comes from the capacity to cross-link information drawn from multiple sensor systems and other information sources. For example, GPS data can cross-link with accelerometer-based physical activity estimates and geographic information systems (GIS). Local area communication standards, such as Bluetooth, can be used to determine the relative proximity of individuals to each other or fixed locations, relevant to the study of infectious disease.

Cross-linked sensor-based information can be used to persuade individuals to perform health behaviours – examples include encouraging people to take the stairs instead of the elevator in order to increase physical activity levels, and using text messages on a mobile phone to remind a person to measure their blood glucose. Ubiquitous sensors therefore have a particularly strong role to play in integrating health care by providing clinicians a novel and less-biased window into the habits and behaviours of their patients. This of course comes at a cost to individual privacy, and decisions must be based on voluntary and freely given consent.

The ever decreasing cost of sensor-based smart devices, together with the medical need for better information regarding a patient's habits outside the clinical environment, makes widespread adoption of these systems not only possible but indeed probable. Properly validated, these sensors have the potential to transform both personal and institutional care by providing reliable contextual information to individuals and practitioners.

Mobile health for greater patient engagement

Smarter health and wellness must address not only change in health care delivery, but also ways of engaging and informing the patient so they effectively achieve better health outcomes.

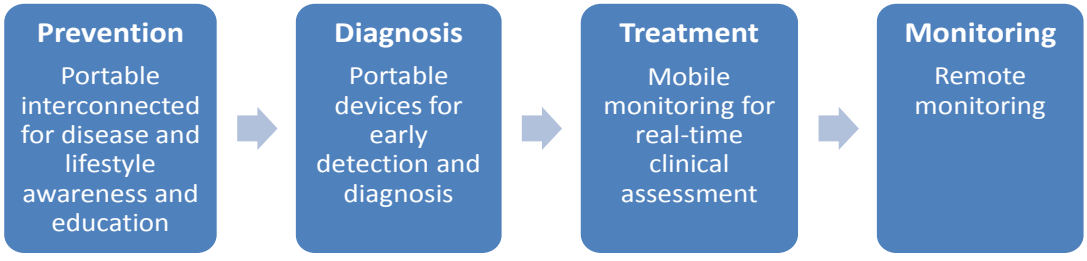
Advocates of patient-centred health have long argued for the citizen taking responsibility for their own health. This argument today applies to the prevention and management of chronic diseases such as diabetes and obesity, and health systems increasingly see their roles as agents of support. The chances of success for any prevention or care programme will depend on patient engagement and meaningful co-ownership and co-production of healthy behaviours. Indeed, a growing body of literature shows that when patients are engaged in their health care, that commitment can lead to measurable improvements in safety and quality (Dentzer, 2013; Laurance et al., 2014).

mHealth technologies can, for example, help to "nudge" people toward better decision making and remaining engaged in their care, although there are equity and safety issues that health professionals must bear in mind when recommending the use of these technologies.

Nonetheless, by putting the patient at the centre of health care transactions, health care providers can begin to overcome the silos within both specialty-based medical care and the various disciplines involved in alternative care. This requires a patient-centred data system, where every patient is a data point from which much can be learned.

mHealth offers a wide range of smart modalities by which patients can interact with health professionals, or with systems that can provide helpful real-time feedback along the care continuum, from prevention to diagnosis, treatment and monitoring (Figure 8.2). mHealth is of particular value in managing health conditions where continuous interaction is important, such as diabetes and cardiac disease. A wide range of devices is utilised for mHealth, including inter alia mobile phones, tablets, global positioning system (GPS) devices, mobile tele-care devices and mobile patient monitoring devices.

Figure 8.2. **Smart mHealth applications**



Source: OECD adapted from PricewaterhouseCoopers, 2012.

Among these devices, mobile phones in particular offer the potential to broadly and cheaply diffuse more intensive self-monitoring, feedback, self-management and clinical support than has been possible previously. This is especially true of smartphones, which support a diverse set of data streams and monitoring activities: automated traces of body movement, location, and other data that can infer physical activities, sleep, and the environment; automated and manually entered physiological measures (e.g. readings from a glucose meter); and prompted and user-initiated self-reports of the user's symptoms or behaviours. This information, appropriately managed, can be leveraged to trigger highly

personalised interventions, and thus significantly improve an individual's ability to understand and manage his or her own behaviours.

Five issues are, however, key to the successful widespread adoption of mHealth: i) establishing and sustaining engagement among participants; ii) increasing ICT knowledge across the society; iii) wide acceptance of privacy and security standards for personal data collection, analysis and use; iv) integration and interoperability – the new range of mobile devices have to function seamlessly and adapt to multiple user needs in the health sector; v) financing and new business models: there is a need to adapt regulatory structures and align incentives at different levels of the health delivery system to encourage investment in, and use of, mHealth.[19]

To achieve widespread use, mobile and health care industries will need to work toward interoperable solutions that enable economies of scale. Without agreed standards and connectivity for information exchange across the ecosystem of personal mobile devices and care services, there will be wide variation in the granularity and quality of the information collected and analysed and limited clinical utility, and payers will be reticent to invest. It is important to note, however, that individuals' engagement in mobile health requires a certain level of literacy and digital skills. Those with fewer of these skills, and who already experience poorer health conditions than those with higher levels of skills, could be excluded from using mobile health technologies, or could be at risk for not using them properly. Therefore, it is crucial for health professionals to ensure that people who are using these services either have the skills needed, or have access to opportunities for improving their skills to use the technologies effectively. Last but not least, with regard to business models, issues of cost to users of the applications as well as Internet access/bandwidth should also be considered. If not addressed properly, mobile health could create additional disadvantages for people of lower socio-economic status and/or those who live in regions with limited Internet access.

8.4. Transforming health research with big data

Big data is an integral part of the health research landscape, and indeed has even helped shape it (see Box 8.2). The next sections list a range of examples of how big data analytics offers new and more powerful opportunities to measure various aspects of disease progression and health for improved diagnosis and care delivery, as well as translational and clinical research.

Box 8.2. **Advances in genetic sequencing**

A remarkable example of the effect of big data is how, over the past two decades, the power of genetic sequencing has increased by one million-fold. No previous technology in history has increased in power that fast. DNA sequencing machines can now read about 26 billion characters of the human genetic code in less than a minute, and the sequencing cost per genome has dropped by 60% a year on average from USD 100 million in 2001 to less than USD 10 000 in 2012 (see Figure 3.10 in Chapter 3 of this volume). Whole genome sequencing programmes involve many terabytes of data. The Cancer Genome Atlas (TCGA) that uses next generation sequencing technology is expected, for example, to generate approximately 2.5 petabytes (PB) of data. The enormous growth of genomic and other biomedical data are at a scale that makes traditional centralised approaches to data management and analysis impractical. Centralised approaches have become untenable for individual laboratories and most small to medium research organisations due to the high cost of data storage, transmission, and analysis.

Box 8.2. **Advances in genetic sequencing (cont.)**

In this big data landscape, new models of global scientific collaborations are emerging that rely on shared e-infrastructures, cloud-based consortia and advanced computational capacities. The aim is to extract knowledge from these large streams of data in the most effective way possible, through global collaboration, open science, and bringing "computing to the data". An example is the International Cancer Genome Consortium (ICGC), a multidisciplinary, international collaborative effort on the part of nine countries to systematically and comprehensively characterise somatic mutations in over 24 000 tumour genomes from 50 different cancer types and subtypes, comparing tumour and normal tissues.

Since its launch in 2008, the ICGC has generated over 250 terabytes of data, adopting federated data architecture to address the data management needs. The scalability of the system is improved by having each member institution store and process data locally; the data federation software then presents these separate sources as a single access point for remote data access.

Source: NIH (2014), "DNA Sequencing Costs", National Human Genome Research Institute, National Institutes of Health, www.genome.gov/sequencingcosts/.

Systems biology to model complex molecular mechanisms

Network and systems biology strategies today offer a powerful means to explore the complex molecular mechanisms underlying many diseases (Chen, Shen and Sivachenko, 2006; Liu et al., 2006; APA, 2006). Research efforts are now increasingly directed to better understanding the interactions between cellular components (proteins, genes, metabolites and so on) (Vidal, Cusick and Barabasi, 2011; Barabasi, Gulbahce and Loscalzo, 2011).

In humans, the potential complexity of the resulting networks – the human interactome – is daunting, with the number of cellular components that serve as the nodes of the interactome easily exceeding 100 000. The number of functionally relevant interactions between the components of this network is expected to be much larger. With so much data available, the challenge is to integrate that information into a single meaningful interaction network.

The highly interconnected nature of the interactome also means that at the molecular level, it is difficult to view diseases as being consistently independent of one another. Indeed, different disease molecular mechanisms can overlap, so that perturbations caused by one disease can affect other diseases (Barabasi, Gulbahce and Loscalzo, 2011).

The systematic mapping of such networks has therefore culminated recently in the concept of the "diseasome": disease maps whose nodes are diseases and whose links represent various molecular relationships among the disease-associated cellular components.

Progress towards a reliable network-based approach – however promising – is currently limited by the incompleteness of the available interactome maps, and the need for powerful visualisation tools as well as statistical methods that are reliable in the context of interconnected environments (Barabasi and Oltvai, 2004). With continued advancements in data analytics, systems biology has the potential to yield a much more nuanced understanding of disease processes, and a greater personalisation of treatments.

Big data for early detection of neurodegenerative disease

Conventional structural neuroimaging, such as computed tomography (CT) or magnetic resonance (MR), has long played a supportive role in diagnosing memory disorders, and is today recommended for the routine evaluation of Alzheimer's disease (AD). However, because structural changes may not be detected at visual inspection until late in the course of the disease, more contemporary structural imaging techniques have emerged that aid in detecting subtle changes not readily apparent on routine images obtained at a single time point. These include positron emission tomography (PET), single photon emission CT (SPECT), and functional magnetic resonance imaging (fMRI).

Functional magnetic resonance imaging in particular offers the promise of revolutionary new approaches to studying human cognitive processes, provided we can develop appropriate data analysis methods to make sense of the huge volumes of data. fMRI measures brain activity by detecting changes in blood flow and blood oxygen levels (the ratio of oxygenated haemoglobin to deoxygenated haemoglobin in the blood with respect to a control baseline), over time and at many individual locations within the brain. It is widely believed that the blood oxygen level is influenced by local neural activity, and hence it is generally taken as an indicator of that activity.

A twenty-minute fMRI session with a single human subject produces a series of three dimensional brain images, each containing approximately 15 000 voxels, collected once per second, yielding tens of millions of data observations. Each voxel contains hundreds of thousands of neurons.

Accurate quantification of changes in regional brain volumes is time- and labour-intensive. If this limitation of fMRI-based methods can be solved via automation of scan analysis, such methods are almost certain to become useful tools for the early detection and monitoring of Alzheimer's and other neurodegenerative diseases in patients.

The Quantitative Imaging Network (QIN), driven by the US National Cancer Institute, grows out of this need to improve translational and clinical research in imaging sciences and technology. The network is designed to promote research in and development of quantitative imaging methods for the measurement of tumour response to therapies in clinical trial settings, with the overall goal of facilitating clinical decision making. Projects include the appropriate development and adaptation/implementation of quantitative imaging methods, imaging protocols, and software solutions/tools (using existing commercial imaging platforms and instrumentation), and application of these methods in current and planned clinical therapy trials.

Sensor-based systems to monitor behavioural changes

Sensor-based systems can also be leveraged to provide clues on emerging physical and mental health problems. Ubiquitous sensors have, for example, an increasingly important role to play in integrating a novel and less-biased window of cognitive and behavioural monitoring of older patients. These systems can also provide assistance in increasing the independence and security of people who have problems of memory, planning, and carrying out tasks in everyday life.

For individuals with chronic conditions, unobtrusive home-based monitoring can result in better patient outcomes by allowing the physician to verify compliance with pharmaceutical regimens and activity-level guidelines to better understand the range of variation of patient outcomes.

For older patients, sensor-based devices can also be utilised to monitor falls and near-falls, physical activity, socialisation, and even overall mobility. For example, wearable fall detectors that include accelerometers are a good example of information technology for assisted living at home (Brown, 2005). In most of these systems, a periodic report from the sensors is sent via wireless communication to a local base station.

The biggest hurdle to overcome to make these approaches useful is the development of efficient and user-friendly data-flow processing and effective conversion of the sensor events into clinically actionable knowledge. Context awareness imposes significant demands on the knowledge maintained by these systems.

Progress will thus depend on the development of robust algorithms and computational models that can fuse and derive meaning from the diverse sets of information. Key factors influencing scalability include: i) seamless integration and interoperability of the technology; ii) reliability of message capture, translation, and delivery to health care professionals and the amount of information transmitted per patient; iii) frequency of monitoring and transmission, and context awareness.

Using social media to research population and public health

Web- and mobile-based applications of social media are emerging as useful new approaches for the dissemination and collection of health and lifestyle information. They can reach a broad audience in a very short period of time; they are easy and affordable to access and use; and they cater to a wide variety of people.

Online social communities, for example, provide a vehicle for individuals with chronic diseases to share information on therapies and disease progression. Participants contribute personal stories that provide learning experiences for other participants who may be contending with a similar health problem. Some online communities are moderated by health care professionals who can offer expert advice via message board posts or synchronous chat sessions.

There is growing recognition that online communities not only provide a place for members to support each other, but also contain knowledge that can be mined for public health research, monitoring, and other health-related activities. By harnessing the power of global, widely disseminated user-generated content, social media is increasingly proving itself an important communication platform on health and disease, serving as an opportunity to collect data on patients' experiences to guide policy and communication planning. At the same time, analytical uses of social networking data must protect the privacy of data subjects. A lack of adequate methods to respect privacy in the use of this data can be a barrier to that use.

For example, the social network PatientsLikeMe developed a lithium-specific global data collection process to capture information about individuals suffering from amyotrophic lateral sclerosis (ALS) who were registered with the network and who began taking the drug off label via their physician (Wicks et al., 2011; CDC, 2009).

ALS is a chronic condition for which neither randomised trials nor nonrandomised clinical studies have yet provided an effective therapy. It is a rapidly fatal neurodegenerative disease causing progressive weakness and muscle atrophy; median survival from symptom onset is 2-5 years. In 2008, a small study suggested that lithium carbonate slowed ALS. Once that study was published, hundreds of ALS patients on PatientsLikeMe began taking the drug, and a few used freely available tools such as Google spread sheets to "crowdsource"[20] their own study. In response, PatientsLikeMe

upgraded its tools and developed new analytical techniques to evaluate whether lithium was effective (Wicks et al., 2011).

The social network also regularly imports the complete data set from ClinicalTrials.gov to let its membership know (free of charge) about the 30 000+ active trials for which they may be eligible. Government agencies are also using social networks to engage the public – for example, during product recalls and pandemic preparations (e.g. in the H1N1 flu pandemic) and as resource for investigating drug-related activity such as off-label use, side effects, product safety, and patient opinions.

Twitter (www.twitter.com) is also emerging as a suitable platform for this purpose. Twitter allows users to send and read short text-based messages limited to 140 characters, which contain a wealth of data. Mining these data provides an instantaneous snapshot of the public's opinions and health-related behavioural or other responses. Longitudinal tracking allows identification of changes in opinions or responses. In addition to quantitative analysis, twitter also permits qualitative exploration of likely reasons why sudden changes have occurred (e.g. a widely read news report), and may indicate what is holding the public's attention.

Twitter content has been studied recently to track flu epidemics (Chew and Eysenbach, 2010) to assess public misunderstandings surrounding antibiotic use (Scanfeld, Scanfeld and Larson, 2010; Signorini, Segre and Polgreen, 2011 ; Sadilek, Kautz and Silenzio, 2012), and more recently to gain insights on how online users share information about dementia and the type of information shared (Robillard et al., 2013).

There is also a large amount of literature proposing methods to extract useful information from online data-generated searches each day, including through Yahoo! and Google (Eysenbach, 2006). A Centers for Disease Control study conducted with Yahoo! in 2005 suggested that Internet searches for specific cancers correlated with their estimated incidence, estimated mortality, and volume of related news coverage. The authors concluded that "media coverage and prevalence appeared to play a powerful role in prompting online searches for cancer information" (Cooper et al., 2005).

Although current Internet search query data are no substitute for timely local clinical and laboratory surveillance, recent studies indicate that the intensity of certain web queries on influenza and influenza-like illness follows the same pattern as the laboratory and sentinel reports for influenza, and that they can be used as additional input data for estimation models (Hulth, Rydevik and Linde, 2009; Eysenbach, 2007).

In November 2008, Google Flu Trends was launched as an open tool for influenza surveillance in the United States. Engineered as a system for early detection and daily monitoring of the intensity of seasonal influenza epidemics, Google Flu Trends uses Internet search data and a proprietary algorithm to provide a surrogate measure of influenza-like illness in the population (Olson et al., 2013; Yin, 2012). The algorithms may still need to be refined however, as the journal *Nature* (2013a) reported in February 2013 that Google's Flu Trends data was significantly overestimating the number of influenza cases. Some researchers suggested that widespread news coverage led to spikes in influenza-related searches by people who were not ill (*Nature*, 2013a).

Crowdsourcing health care innovation

Crowdsourcing is emerging as a means to allow science to be conducted at scales of magnitude greater than before. It involves capitalising on the Internet and large groups of people, particularly via online Web 2.0 communities, to harvest "collective intelligence"

and accomplish tasks that might have traditionally been given to small research groups. Crowdsourcing is for example successfully being used by foundations and the public and private sectors for health research purposes such as to understand protein structure prediction and design.

Crowdsourcing can process data quickly and on unprecedented scales and with better quality control than any individual or small research group can attain, given the large number of participants. Crowdsourcing therefore has cost and speed benefits; it may allow science to be conducted at scales of magnitude greater than before (thousands of research participants recruited in months versus years) and huge numbers of data points, the potential for new discoveries in the patterns of large data sets, and the possibility of near real-time testing and application of new medical findings.

The success rate of crowdsourced innovation challenges is quite high, in some cases up to 40% – which is remarkable, especially since many "challenges" are generally put out on the web because they are, by definition, beyond the problem-solving ability of the organisation or the individuals posting them. Given its open, informal structure, crowdsourcing is inherently cross-disciplinary. In some cases, even gifted amateurs and people without direct experience with the problem provide valuable insights and solutions. However, at present there are several important unresolved ethical and legal issues that limit the use of crowdsourcing in health research (see also Chapter 5 of this volume):

1. Crowdsourcing may accelerate the sharing of information, but careful attention must be paid to policy regarding privacy, security, data stewardship and personal control. Rapid developments in this area have outpaced regulatory frameworks, raising a number of concerns that range from the potential of modest risks to the privacy of participating individuals and to the quality assurance of the large streams of data generated to severe safety risks.
2. Second, a range of new partnerships is emerging around these applications. There is a need to better understand this rapidly evolving ecosystem – the business models, the market potential and the related governance frameworks. This should be combined with the development of robust metrics for measurement and evaluation.

Despite these challenges crowdsourcing is increasingly being used by public and private sector to address complex and challenging problems. In 2011, 1 920 000 results were returned for a Google search of the terms crowdsourcing & health, linked with an ampersand; in 2010 and 2009 the figure was 669 000 and 318 000, respectively. In January 2012, the term "crowdsourcing" in a PubMed search yielded 16 publications, 13 of which were published in 2011 (Swan, 2012a).

InnoCentive, one of the first companies to crowdsource in the chemical and biological sciences, today has more than 300 000 registered "solvers", who stand to gain rewards of between USD 5 000 and USD 1 million if their solution works. Key to the success of InnoCentive's crowdsourcing has been: i) a governance structure carefully designed to protect intellectual property from both the Seeker and the Solver perspective; ii) reduced barriers to participation so that the challenge scales quickly; and iii) global reach, increasing the likelihood of solutions coming from very unexpected directions.

An illustration of public sector uses of crowdsourcing for health is the "Investing in Innovation (i2)" initiative of the US Office of the National Coordinator for Health Information Technology. The i2 program arranges challenge competitions to spur innovation in the developer community. These challenges, in which the winners are

awarded prizes, enable the sector to reach out to developers who have expertise in different fields: although they do not necessarily work in health IT, they can apply their knowledge from elsewhere to health IT issues.

An ongoing challenge, developed in partnership with the US National Cancer Institute, asks developers to come up with tools and applications for cancer survivors. Applications arising from the ONC challenge programme are already widely used – including Humetrix's iBlueButton, which enables patients to access and share their health records.

Citizen science initiatives are also growing in importance, internationally and in the bioscience field; often they are associated with crowdsourced challenges. Citizen science generally refers to a network of people, many of whom may have no specific scientific training, performing research-related tasks, such as recording specific observations over time to reveal patterns and trends. The approach leverages what Shirky (2010) called "cognitive surplus", referring to the vast amount of time that people collectively spend on activities such as watching TV.

Citizen science projects often involve non-professionals taking part in one or more of the following:

- crowdsourcing
- mass participation
- data collection
- data analysis.

Zooniverse, for example, works with researchers to design sites that present their data in a format that will permit the crowd help them achieve their objectives.[21] Zooniverse has a community of over 850 000 people, who have taken part in more than 20 citizen science projects over the years. These initiatives support a form of "scientific democracy", where data can be shared among and utilised by investigators in public and private sectors, policy makers, and the public.

Foldit is another popular online citizen science initiative, in which individuals are scored on noting changes in protein structure. The game records the structure and the moves that the players make; scientists can capture the data that are then used to improve the game in every aspect, from the quality of the scientific results that are returned to how long people play the introductory levels that teach the game. The whole game is like an ongoing, continuous experiment.[22] Foldit was successfully used to remodel the backbone of a computationally designed enzyme that catalyses the Diels-Alder reaction, bringing together two small molecules to form a particular kind of bond that the scientists were interested in making.[23] This catalysis can be useful in building other kinds of small molecules, such as drugs and chemicals. Scientists went back and forth with the players, and in the end designed an enzyme that was about 20 times more efficient in catalysing the reaction than the one the scientists had begun with. Tapping into the vast cognitive surplus online and incorporating crowdsourcing in research, to both enlist the public's help and engage public interest, holds tremendous promise for accelerating health innovation.

DIYgenomics, established in 2010, leverages crowdsourcing and citizen science in order to produce preventive medicine. At present the focus of DIYgenomics is on linking genetic mutation with phenotypic evidence and personalised intervention in a wide range of studies. DIYgenomics has vitamin deficiency studies under way to investigate two

possibilities.[24] The first is that one or more genetic polymorphisms (e.g. mutations) may lead to current blood marker levels that are already out of bounds per recommended levels; the second is that simple vitamin supplementation may be able to restore blood markers to recommended ranges. DIYgenomics also has a study examining how genetic variants may be related to dopamine processing in the brain, and how this may impact the processing of memories. DIYgenomics moreover hosts a longitudinal ageing study aiming to establish personal baseline norms for 50 blood markers and their potential correspondence to 1 000 gene variants associated with ageing, and experimenting with personalised intervention. The study provides an opportunity to apply the dozens of genome-wide association studies (GWAS) that relate to general and specific conditions of ageing in a comprehensive preventive medicine approach. Genomic data are linked with corresponding measures of phenotypic biomarkers and interventions. The participants' tasks are to complete an annual blood test (a comprehensive panel of approximately 50 markers available through DirectLabs [USD 79] or another source) and, if willing, share the data with the cohort and self-experiment with relevant interventions. One thousand genetic variants are reviewed that have been linked to a variety of ageing conditions (Swan, 2012a, 2012b).

Funding for crowdsourced and citizen science health studies comes from a variety of sources, including foundations, academia, the private sector and patient advocacy groups. "Next-generation research foundations" are finding crowdsourced cohorts suitable for their studies. This is especially true of foundations that are focused on genomics and personalised medicine – such as Sage Bionetworks, a non-profit research organisation based in Seattle. That organisation's Alzheimer's Big Data Challenges, the Global CEO Initiative on Alzheimer's Disease (CEOi), and IBM's Dream Team are examples of bio-informatics crowdsourcing aimed at identifying the best multivariate set of predictors of cognitive decline and of neuro-protective genotypes.[25]

The many concepts underlying crowdsourcing have been around for some time. But as the proportion of the world's population with access to the Internet climbs, the potential for crowdsourcing to both generate data and help interpret data will grow. At the same time there are challenges that must be overcome to allow the generalised use of crowdsourcing to improve health research and health care. These need to be addressed through supportive policies and environments. The next section outlines the factors that support data-driven research and care.

8.5. Critical success factors and policy priorities

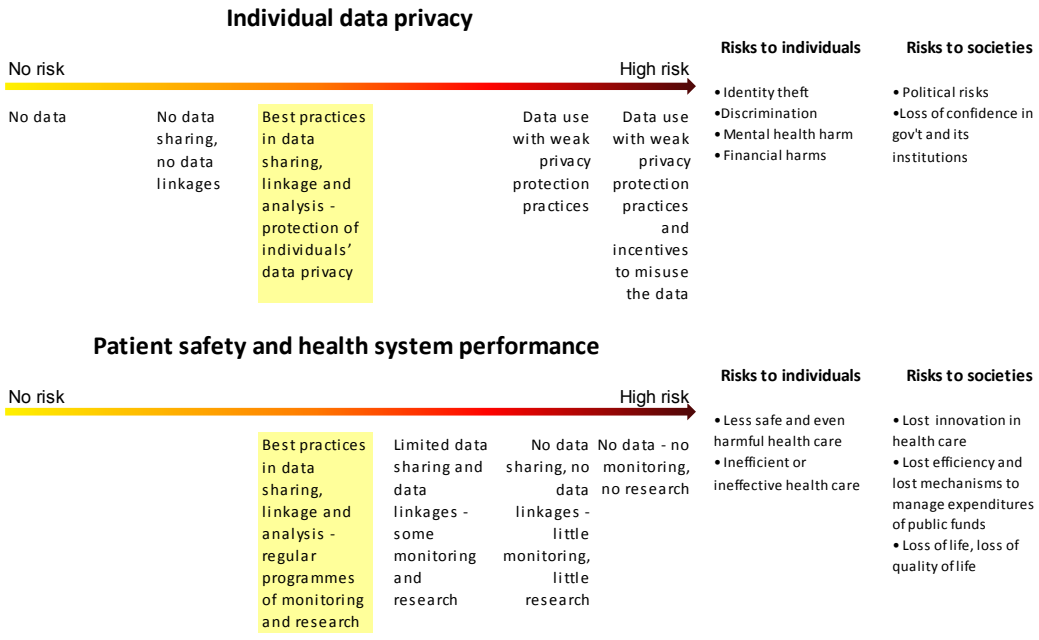
The most critical success factor for governments to realise value from investments in health data is data governance that is proactive, engaged and comprehensive. Such governance must be multi-sectoral, as the needs and conditions that would lead to a data-driven health sector reach beyond the control of health ministries. This section reviews priorities for policy action.

Minimising risks to the data subject's privacy

Collection and use of personal health data present a number of important risks to the privacy of individuals (Figure 8.3). These risks relate to the potential harm to individuals that could result from the misuse of their personal information. Losses to individuals can be severe and can include financial and psychosocial harm. Detriment to individuals can also produce a loss of public confidence in government and its institutions. Yet, there are equally significant risks to individuals and to societies when health information assets are

not developed, are unused, or are very difficult to use. These include a lack of evidence to detect and correct inefficient, ineffective and even harmful health care, and lost opportunities for research and innovation to improve health and health care outcomes.

Figure 8.3. **Risks associated with the collection and use of personal health data**



Source: OECD, 2013e.

Significant differences in approaches to the protection of data subjects' privacy among OECD countries have resulted in some countries advancing the generation of health data and its use for research and statistical purposes, and others restricting data collection, sharing and use. These differences are significant and can be attributed to differences in risk-benefit evaluations.

Many OECD countries report legislative barriers to the use of personal health data, including enabling data linkages and developing databases from electronic health records. Some of the countries with less developed information infrastructure have decentralised administration of health systems, and have not reached a consensus within the country of how the levels of government could work together. A principal challenge for some countries is the lack of clarity on how to translate into practice legislation concerning the protection of data privacy, including informed consent at the national and sub-national levels. This includes the legality of data sharing among public authorities and that of providing access to personal health data for research.

This complexity extends to multi-country data-sharing initiatives; the result is that such initiatives remain rare, challenged by concerns regarding differences in data privacy protection laws and whether shared data will be adequately protected in the receiving country. European countries have made the most progress among OECD members, having recognised in law that foreign entities can apply and be approved for access to data when the legal protection of information privacy in the foreign country adequately matches that of the home country. However, lack of resources to evaluate the adequacy of foreign laws continues to pose barriers to data sharing between European countries and other countries.

The need for new consent models

An important requirement prior to personal data collection for health research is to obtain patient consent. Explicit consent has become the pillar for protecting autonomy in research involving human subjects. The requirement for consent is underpinned by ethical principles of respect for persons. Consent is also the basis for data protection and privacy laws in most countries. Within the medical/scientific field, informed consent generally presumes the ability to indicate clearly to the participant the use and purpose of the particular research activity. While this is feasible for purpose-specific research, the new forms of biomedical research – and the ease with which multiple data from diverse sources can now be collected, stored, analysed and shared in greater volumes than ever before – renders provision of this type of information particularly difficult.

In the case of biobanks, for example, where there are multiple researchers and research projects, it is difficult to obtain explicit consent for all future research uses at the time of research recruitment, as is required in the original formulations of the Declaration of Helsinki (World Medical Association, 1964). The declaration states that use for research purposes different from the original would require re-contacting large population groups to obtain a new consent, which is often impossible or impracticable. Re-consenting is costly and time-consuming, and difficulty in locating people can result in high dropout rates. New approaches are clearly needed to meet ethical and legal requirements for consent and to accommodate the changes in data use and research practices.

A tiered or step-by-step consent approach has recently been adopted at Imperial College in the Chariot Register, a recently established cohort of over 20 000 healthy volunteers for the prevention of dementia and other age-related neurodegenerative diseases. Participants initially consented to be approached for individual observational or interventional studies, and were then offered a menu of options pertaining to such research uses, requests to re-consent, interest in returning results, etc. Other approaches recently proposed in the scientific literature include "adaptive" or "dynamic" models of consent forms, whereby (following the initial general consent) participants would be asked to re-consent for any "new" direction of travel/use of their data, potentially using web-based communication tools. This approach is dynamic because it allows interactions over time; it enables participants to consent to new projects or to alter their consent choices in real time as their circumstances change, and to have confidence that these changed choices will take effect (Kaye et al., 2014).

A key role for data custodians

Personal health data sets are in the custody of multiple organisations within countries, and legal frameworks and internal policies must be aligned to permit secure data sharing if big health data assets are to be brought together for research and statistics. Countries reported encountering difficulties in negotiating data-sharing arrangements among national organisations, with negotiations either unsuccessful or taking years to conclude (OECD, 2013e).

Countries that have centralised processing of personal health data within a single or a small number of organisations have the advantage of avoiding the need for complex data-sharing negotiations, as well as gaining efficiencies in data processing and data security protections.

These data custodians also play a central role in balancing data privacy protection and the use of data for monitoring and research, as they are responsible for the collection, processing, analysis and dissemination of personal health data. In many countries, data custodians are also responsible for vetting project proposals for the use of data from government and private entities; for maintaining the technical capacity to undertake data linkages; for maintaining a technical capacity for data de-identification; for providing data access modalities to internal and external researchers; and for ensuring that through all of their activities, the legal requirements for data security and data privacy protection are respected.

Development of privacy-enhancing technologies

Advances in techniques to ensure privacy through the design and development of privacy-enhancing technologies provide additional avenues to meet both health care data use and privacy protection needs (OECD, 2013e).

The practice of data de-identification and data pseudonymisation are widely used across countries, particularly before data are made available for research and analysis. While de-identification involves the removal of key patient identifying information, such as names, patient numbers, exact addresses and key dates, pseudonymisation replaces key patient identifiers with a meaningless code that can, for approved purposes, allow re-identification. An example of an approved purpose could be to conduct a new approved data linkage study with the same data set in the future. With this approach, a trusted party is usually employed to guard the key that enables data re-identification.

Data de-identification techniques, however, rarely remove all risk that a data set could be manipulated or combined with other data to rediscover the identity of data subjects. Importantly, some countries have developed data governance mechanisms that provide added security to protect de-identified data. These include independent review bodies that evaluate data use proposals for public benefits and adequacy of data security; contractual agreements that bind data receivers to required data security and disclosure practices; and security audits and follow-up mechanisms to ensure compliance with contractual obligations.

A few countries are pioneering alternatives to sharing de-identified data in order to further minimise data security risks. These include supervised research data centres, where authorised researchers analyse data within a physically secure location; and secure remote data access services, where authorised researchers enter a secure portal; there they can analyse data but cannot extract or otherwise remove data from the system (OECD, 2013e).

Engaging stakeholders

In an OECD 2012 survey of 25 countries, 13 reported having involved groups of stakeholders in their efforts to govern the development and implementation of their national electronic health record system, either through the groups' representation within the governing body or through consultation, or both. The groups included, for example, clinicians, pharmacists, professional associations, patients, insurers, IT professionals, lawyers and policy makers (OECD, 2013e). Engaging with all interested stakeholders would appear to be the best strategy for ensuring that all voices are heard and a consensus is reached on data use that respects privacy. Further, a public communication strategy that is open and transparent would go a long way toward demystifying data, opening data for monitoring and research, and generating positive public discourse about data risks and

utilities. Ideally, the strategy would enable all concerned stakeholders to know what data are being collected; how they are being used; how and with whom to apply for access to them; the conditions of approval; data security requirements; and details of the research projects that are approved.

Even with strong communication with the public; a high degree of transparency regarding data uses; and strong governance to safeguard patients' privacy in law and in practice, attention is needed to ensure that individuals who wish to restrict or withdraw their data from their contribution to research and statistics can reasonably do so. Strategies to enable individuals to exercise control over the use of their personal health data must be workable at the scale of population-level data collection, and in circumstances where there is a high volume of data and a high number of data use requests. Emerging techniques reported to the OECD in 2013 include the use of Internet patient portals, to request patient consent to data uses or to enable patients to opt out of data use. The portals provide information about data use that is broad enough to capture future uses for research and statistics that cannot be specified today, while being narrow enough to allow patients to fully understand the circumstances under which their data could be approved for use and how they would be protected throughout that use.

Promoting open data and data commons

Governments in several OECD countries have been engaged in initiatives to increase the openness and transparency of government data, including health data. Advocates for greater openness and transparency link the availability of government data and information to more socially inclusive service delivery; to participatory democracy; and to economic stimulation from the development of new products and services (Chapter 10 of this volume). In 2011, eight governments (Brazil, Indonesia, Mexico, Norway, the Philippines, South Africa, the United Kingdom, and the United States) founded the Open Government Partnership, and announced their country's action plan towards open government health data (UK Department of Health, 2012). This partnership has since grown dramatically, to 47 additional governments.

In 2013, of the 20 countries participating in the OECD Health Care Quality Indicators (HCQI) survey, 12 indicated that their country is planning a policy or programme to promote open government health data. For most countries, this effort is part of a whole-of-government initiative to provide citizens with a single entry point to government statistics, including health in the form of a web portal. In a few countries, this also includes developing mechanisms for citizens to more easily retrieve their own personal data.

The Health Data Initiative in the United States aims to increase data availability and transparency in order to improve community health. In particular, the initiative provides access to a broad range of health data at the local, state and national level, disaggregated by socio-economic characteristics, standardised and documented to enable ease of use.[26] The US initiative also involves working with clinicians, information technology professionals, policy makers and citizens to develop software applications and tools that turn data into actionable information, be they smartphone apps, interactive maps, indicators, social networking sites or games, etc. (DHHS, 2013).

In the United Kingdom, England launched a ten-year strategy to improve the National Health Service, public health and social care (UK Department of Health, 2012). An open government health data programme is an integral part of this strategy; it includes routinely releasing public service data sets in health to the public; providing health

service users with access to their own data; gathering and publicising health services user satisfaction and experiences data; engaging with data users to drive social and economic growth; and working to continuously improve data quality. The open government health data programme also connects with the overall strategy of the United Kingdom to promote open government data (United Kingdom, 2012).

The Australian government also has an open government data initiative with the inclusion of some government health data, as do Canada and Italy.[27] With its Digital Agenda for Europe, the European Commission is also promoting greater openness and reuse of government data, some of which is related to the health care system.[28] This project includes establishing a website to disseminate data held by the European Commission, and work toward a pan-European portal for all data from the EU, as well as from national governments and regional and local governments in Europe.

Building a new generation of health data scientists

Data scientists with good communications skills, and clinicians and other health professionals at ease with numbers and computing, can work together to produce remarkable results. Developed from a health data initiative of the Institutes of Medicine and the Department of Health and Social Services in the United States, a new event called Health Data Palooza was held in 2012 and 2013.[29] This event brings a diverse set of stakeholders together to discuss obtaining value from data. One of its key elements is a 48-hour competition where teams of clinicians and IT professionals compete for prizes by developing an app, tool or product from analysis of US Medicare databases. At the same event, start-up and established companies can showcase products leveraging information value from health data. In 2013, 80 companies showcased new products.[30] In England, events called NHS (National Health Service) Hack Days bring together clinicians, programmers and website designers. Some of their products that are now in use include Cell Countr, an app that provides haematology counts, and PatientList, an app that provides clinical task lists (Lewis et al., 2013)

England plans to engage clinicians directly in developing tools to exploit the power of big data to improve the NHS (Lewis et al., 2013) There are plans to involve medical professionals, patients and NHS managers in the development of applications and tools by training them how to write computer code; the initiative is called Code-4-Health. The thinking is that clinicians and managers with some knowledge of code will be better equipped to work with IT professionals toward developing useful tools, and able to unleash their creativity and innovative ideas. It will also avoid the old style of computer applications development, where a detailed specification is given to a programmer who works alone to try to develop code to meet it, often to the dissatisfaction of the end user.

Nonetheless, a scarcity of data scientists needs to be avoided through education and training initiatives, and such training needs to be responsive to the necessity of data scientists having the teamwork and collaboration skills to partner with health care professionals (Davenport and Patil, 2012). Adapting education and training programmes for other health care professions to ensure at least a minimum degree of skill development in statistics and programming is another worthy objective, both to build a generation of clinical data scientists and to increase appreciation for data and high-quality record keeping within the health care professional community (see further discussion on the skills implications of DDI in Chapter 6 of this volume).

Financial sustainability of big data projects

Big data is a costly activity. As an illustration, the cost of implementing the Canadian longitudinal study on ageing – which includes the first wave of data collection, follow-up on the initial cohorts, and management – was estimated at CAD 23.5 million. Generation of big data (e.g. imaging, microarray, phenotypi data etc.) can include costly processes, requiring expensive consumables as well as specialised equipment and personnel for their generation. If for financial reasons these networks or databases are unable to perform their tasks under conditions that meet the requirements of scientific research, scientists will see valuable information either lost or transferred into a strictly for-profit environment.

There is no magic bullet today with respect to the options or strategies required to achieve the long-term financial sustainability of big data projects. Financial sustainability is a critical issue for all big data initiatives – even those that are relatively more mature and directly funded by public sources.

As ongoing financial support is uncertain, large data networks very often must seek out multi-source financing – for example, by charging fees to those who want to gain access to a specific data set and associated database. Varying fee structures can be applied for access, depending on the nature of the data, its status and use. Another model that appears to have great potential for the prolonged financial sustainability of big data projects is the public-private partnership.

Increasing accessibility and sharing of existing data can be resisted due to mismatches in resources and incentives. Data collectors may not have sufficient resources to meet data access requests. Policies are needed that consider the bigger picture regarding the benefits of data use to improve innovation in treatment and care; they must provide incentive structures and resources that will enable key data holders to take part in making the necessary data available.

Setting standards to enhance interoperability

Standards and interoperability are central issues that must be addressed to advance big data in health care. While health care organisations have access to an ever-increasing number of information technology products, many of these systems cannot "talk" to each other and health information exchange remains a serious problem. If the systems cannot communicate, big data will not meet its potential in the health care system. Ensuring that electronic records can be transferred or shared among a patient's primary care physician and specialists is an issue that has yet to be addressed. This problem is common to all OECD countries, even those where deployment of EHRs has proved particularly successful.

In a networked environment, interoperability means common protocols defining the basic mechanisms by which users and resources negotiate, establish, manage and exploit data-sharing relationships. It means sharing not only data but anything that connects to the data's production and processing, including computing tools, applications, methods, software, metadata, workflows across different platforms and, as mentioned above, even communication.

There is, for example, a clear need to develop and promote international consensus standards for clinical information. The World Health Organisation has developed classification systems, which are essential for research and statistical reporting; however, these systems need to evolve to meet the needs of electronic clinical records and other

new forms of electronic data. Strategies to address a lack of standards for clinical terminology within electronic health records and to improve data quality and coverage include: laws or regulations requiring adoption and use of EHR systems that conform to standards; incentives and/or penalties relating to adoption and use; certification of software vendors; and quality auditing (OECD, 2013e; OECD, 2010). Many governments have set up specific bodies or agencies to co-ordinate standard-adoption activities, developing strategies at the national level.

Nonetheless, EHRs are expected to have important limitations compared with a clinical research data set designed to follow a strict research protocol with standardised definitions and rules to optimise statistical analysis. Simply put, the actors involved in inputting the data to EHR systems are primarily motivated by very different incentives – such as simplifying notes for their own use and saving time – from those of clinical scientists gathering data for their own research.

Technical solutions are on the horizon that could alleviate the burden of completing EHRs and address data quality problems. In particular, the advent of natural language processing (NLP) may enable health care providers to speak rather than to type; importantly, they may also require less personal knowledge of and competency with clinical terminology code sets (see Chapter 3 of this volume). This technology has not yet advanced sufficiently, however, for it to be widely adopted for use in EHR systems (Friedman, Rindflesch and Corn, 2013).

Analytic techniques are also emerging, to better cope with heterogeneity among genome experiment data, and to enable data to be pooled for systems-level research. These techniques include Bayesian integration, which enables prediction of the probability of an interaction between gene pairs and quantifies the contribution of each experiment to the prediction (Greene and Troyanskaya, 2012). Nonetheless, global standards for genomic data that would facilitate research are needed (*Nature*, 2013b).

The final area where standards are needed is in ensuring that observational health studies involving analysis of large databases follow scientifically sound methodologies. Certainly, the wealth of detail within electronic health records provides the ability to address challenges faced in the past due to an absence of information about important confounding factors. New methodological designs take this potential further and advance the science of comparative effectiveness research. These advances include sequential cohort studies, extensions of clinical trials with health record data, and modelling and trial simulation (Schneeweiss et al., 2011).

Providing incentives, investments, and grants

Options for stimulating the development and use of data abound. Eleven of twenty-five countries indicated in 2012 that they have introduced incentives, penalties or both to encourage health care providers to adopt EHR systems that conform to standards and use structured data. Seven countries have also introduced incentives or penalties to ensure that health care providers keep their electronic health records up to date. Eight countries have implemented or were implementing legislation or regulations requiring health care providers to adopt electronic health records and/or to conform to clinical terminology and interoperability standards.

The US National Institutes of Health recently launched the Big Data to Knowledge Initiative, providing grant funding to create centres of excellence in big data computing in the biomedical sciences. Objectives include developing new policies to encourage data

and software sharing; a catalogue of research data sets; data and metadata standards development; analytical software development; access to large-scale computing; engagement with users and software developers; increasing the number of computational and quantitative researchers; and strengthening the skills of existing biomedical professionals.[31] Further, as was previously noted, the USD 1.1 billion in federal stimulus funding in the United States has had a profound impact on the development of comparative effectiveness research. Over the next four years, Horizon 20/20 – the scientific granting programme of the European Union – intends to fund health research projects that involve the collection and processing of data for large populations, for long-term follow-up studies as well as studies to support the development of health data infrastructure (European Commission, 2015).

As was also noted earlier, the advent of prize competitions as an incentive is a new tool being leveraged in some countries to mobilise the business, clinical and research communities toward the use of health data to solve challenges – largely through the development of new computer tools and applications.

As decisions regarding investments in data infrastructure and analysis are taken, consideration should also be given to the balance between who is investing and who will receive the benefits from the investment. In situations where costs are borne by the state and benefits accrue to private sector businesses, compensatory mechanisms to share costs or to share profits should be considered.

In summary, the evidence discussed in this section strongly suggests that the most important prerequisite for advancement of DDI is proactive and comprehensive data governance. The next section summarises the key findings from this chapter and describes how work at an international level could help countries make further progress toward data-driven health research and care.

8.6. Key findings and policy conclusions

Data-driven innovation in the health sector is already taking place. This chapter has presented examples from many countries of how the huge volumes of clinical, genetic, behavioural and environmental data that can now be generated are being processed, analysed and integrated to support patient care, health system management and research. These advances have yielded extraordinary insights into the natural history of diseases and their diagnosis, prevention and treatment. There has also been a rapid increase in medical devices that produce streams of personal health data, as well as data provided by patients through personally controlled patient records and patient social networks – sources from which the research potential is beginning to emerge.

Data that encompass patients' care pathways – that is, that capture patient treatment flows through the medical system based on diagnoses, procedures, drug events and the outcomes and costs of those pathways – have tremendous potential to uncover medical errors, avoid adverse drug reactions, detect fraud, improve adherence to clinical guidelines and develop effective treatments. Care pathway data can support both scientific knowledge generation and intelligence for health care policy making and management. They also increase opportunities for synergy and feedback between R&D activities and policy and management activities. Electronic health records are an important resource for data on clinical care, and can contribute to examining patient pathways and their outcomes.

The most critical success factor for governments to realise value from investments in health data is proactive, engaged and comprehensive data governance. Elements of data governance proposed include strategic planning; ensuring legislative and regulatory requirements that support planning; introducing effective data privacy and security practices; engaging all stakeholders in planning and governance; developing a new generation of data scientists; promoting global co-operation; setting standards for data quality; and providing financial stimulus toward data development and use.

Countries that plan now for how they will harness the value of personal health data in a secure and regulatory-compliant fashion will have the opportunity to reap the considerable benefits of health care innovation, with the ensuing advantages of both high-performing health care systems and growth in health care innovation.

At an international level, the OECD will continue to support countries in strengthening their health information infrastructure and analytical capacity. Key opportunities for future international co-operation include:

- exploring technical and legal modalities that enable global sharing, linking and analysis of data for health research and health system performance improvement
- continuing support of countries in their quest to implement quality indicators to assess the performance of health services and systems through the OECD Health Care Quality Indicators (HCQI) programme
- ongoing monitoring of country progress in the development and use of health data, and exploring country skill development and technical infrastructure capacities to support data analytics
- reporting on emerging forms of health and wellness data, such as genetic data, medical device data and social media data
- reporting on government investments in data infrastructure and big data research in health care and on any cost-sharing or profit-sharing arrangements with the private sector that have been introduced to finance this work
- supporting the development of international coding standards for key elements of health data systems in collaboration with WHO
- exchanging information about how laws protecting privacy are implemented in this area and promoting approaches that effectively protect privacy while enabling the use of data for health research and health system performance improvement
- working with countries to identify good practices in data deposition, access, exchange and linkage to advance dementia and neurodegenerative disease research (OECD, 2013d).

Notes

1 The G8 Dementia Summit Declaration, released on 11 December 2013, is available at: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/265869/2901668_G8_DementiaSummitDeclaration_acc.pdf, accessed 7 May 2015. See also http://dementiachallenge.dh.gov.uk/category/g8-dementia-summit, accessed 7 May 2015.

2 The Global Public Health Intelligence Network (GPHIN), developed by Health Canada in collaboration with WHO, is a secure Internet-based multilingual early-warning tool that continuously searches global media sources such as news wires and websites to identify information about disease outbreaks and other events of potential international public health concern. See www.who.int/csr/alertresponse/epidemicintelligence/en/, accessed 7 May 2015.

3 HealthMap, developed at Boston Children's Hospital in 2006, uses online informal sources for disease outbreak monitoring and real-time surveillance of emerging public health threats. See www.healthmap.org/site/about, accessed 7 May 2015.

4 Information on the i2b2 initiative is available at: www.i2b2.org, accessed 7 May 2015.

5 Information on the eMERGE network is available at: https://emerge.mc.vanderbilt.edu, accessed 7 May 2015.

6 Information on the Kaiser RPGEH Program is available at: www.dor.kaiser.org/external/dorexternal/rpgeh, accessed 7 May 2015.

7 Information on the Million Veteran Program is available at: www.research.va.gov/mvp/default.cfm, accessed 7 May 2015.

8 Information on STRIDE is available at: https://clinicalinformatics.stanford.edu/research/stride.html, accessed 7 May 2015.

9 Information on the i4health network is available at: www.i4health.eu, accessed 7 May 2015.

10 Information on EMIF is available at: www.imi.europa.eu/content/emif, accessed 7 May 2015.

11 Information on the European Translational Information and Knowledge Management Services (eTRIKS) is available at: www.etriks.org, accessed 7 May 2015.

12 Information on the Integrative Cancer Research Through Innovative Biomedical Infrastructures (INTEGRATE) is available at: www.fp7-integrate.eu, accessed 7 May 2015.

13 A Next-Generation, Secure Linked Data Medical Information Space For Semantically-Interconnecting Electronic Health Records and Clinical Trials Systems Advancing Patients Safety In Clinical Research (Linked2Safety). Information is available at: www.linked2safety-project.eu, accessed 7 May 2015.

14 Information on the Scalable, Standard based Interoperability Framework for Sustainable Proactive Post Market Safety Studies (SALUS) is available at: www.salusproject.eu, accessed 7 May 2015.

15 Information on the Translational Research and Patient Safety in Europe (TRANSFoRm) is available at: www.transformproject.eu, accessed 7 May 2015.

16 Information on the European Collaboration for Healthcare Optimisation is available at: www.echo-health.eu, accessed 14 September 2013.

17 Information on the CONCORD Programme is available at: www.lshtm.ac.uk/eph/ncde/cancersurvival/research/concord, accessed 26 January 2015.

18 Information on CancerLinQ is available at: www.asco.org/institute-quality/cancerlinq, accessed 7 May 2015.

19 Similar conclusions were reached in the 2014 EC Green Paper on mHealth: http://ec.europa.eu/digital-agenda/en/news/green-paper-mobile-health-mhealth, accessed 7 May 2015.

20 Crowdsourcing is "the practice of obtaining needed services, ideas, or content by soliciting contributions from a large group of people and especially from the online community" (Merriam-Webster, 2014).

21 Information on Zooniverse is available at: www.zooniverse.org/ (accessed 26 January 2015).

22 In Foldit puzzles, for example, players are rewarded for solving clashes and voids, places where the protein is not consistent with known biochemical patterns. Players are able to build a hypothetical protein and see how it works in the game. The game's score is based on a proxy for how well the protein would work in a laboratory; whether it can catalyse some reaction that the scientists are interested in; or how well the protein sticks to some part of a virus – or even, in the case of the Symmetry puzzles, how well the protein sticks to itself. Solutions that are promising are then synthesised in the laboratory.

23 Information available at: www.nature.com/nbt/journal/v30/n2/full/nbt.2109.html, accessed 7 May 2015.

24 Information on these studies is available at: http://genomera.com/studies/vitamin-d-study, accessed 7 May 2015.

25 Information on Sage Bionetworks is available at: http://sagebase.org/about-2/, accessed 25 May 2015.

26 For information on HealthData.gov, see: www.healthdata.gov/ (accessed 30 September 2013).

27 For information on the initiative: in Australia, see http://data.gov.au/ (accessed 27 September 2013); in Canada, see http://data.gc.ca/eng; and in Italy, see www.dati.gov.it/ (accessed 27 September 2013).

28 For information on the Digital Agenda, see: http://ec.europa.eu/digital-agenda/en/news/commission-welcomes-member-states-endorsement-eu-open-data-rules, accessed 27 September 2013.

29 For information on the event's history, see: http://healthdatapalooza.org/history-of-the-health-datapalooza/, accessed 17 September 2013.

30 For details on that year's event, see: www.whitehouse.gov/blog/2013/06/07/health-datapalooza-iv-tops-huge-year-health-data-liberation-innovation, accessed 20 September 2013.

31 For information on the Big Data to Knowledge Initiative, see: http://bd2k.nih.gov/about_bd2k.html#areas, accessed 23 September 2013.

References

AHRQ (2013), "What is comparative effectiveness research?", Agency for Health Care Research and Quality, http://effectivehealthcare.ahrq.gov/index.cfm/what-is-comparative-effectiveness-research1/, accessed 16 September 2013.

APA (2006), "Funding outlook for NCI: News from the experts", American Psychological Association, May, www.apa.org/science/about/psa/2006/05/nci.aspx.

Barabasi, A.-L., N. Gulbahce and J. Loscalzo (2011), "Network medicine: A network-based approach to human disease", *Nature Review Genetics*, Vol. 12, pp. 56-68.

Barabasi, A.-L. and Z.N. Oltvai (2004), "Network biology: Understanding the cell's functional organization", *Nature Reviews Genetics,* 5, pp. 101-113.

Brown, G. (2005), "An accelerometer based fall detector: Development, experimentation, and analysis", University of California at Berkeley, July, http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.128.6988/

Canadian Institute for Health Information (2013), "Better information for improved health: A vision for health system use of data in Canada", in collaboration with Canada Health Infoway**,** www.cihi.ca/cihi-ext-portal/pdf/internet/hsu_vision_report_en, accessed 15 May 2015.

Cancer Research UK (2013), *International Cancer Benchmarking Partnership (ICBP),* ICBP Publications, www.cancerresearchuk.org/health-professional/early-diagnosis-activities/international-cancer-benchmarking-partnership-icbp, accessed 14 September 2013.

CDC (2009), "H1N1 web and social media metrics cumulative data report April 22, 2009 – December 31, 2009", Centers for Disease Control and Prevention, Division of eHealth Marketing, National Center for Health Marketing.

Chen, J.Y., C. Shen and A.Y. Sivachenko (2006), "Mining Alzheimer disease relevant proteins from integrated protein interactome data", Pacific Symposium on Biocomputing, November, pp. 367-78.

Chew, C. and G. Eysenbach (2010), "Pandemics in the age of Twitter: Content analysis of tweets during the 2009 H1N1 outbreak", PLoS ONE, http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0014118, accessed 15 March 2015.

Cooper, C.P. et al. (2005), "Cancer Internet search activity on a major search engine, United States 2001-2003". *Journal of Medical Internet Research*, www.jmir.org/2005/3/e36/, accessed 15 March 2015.

Dentzer, S. (2013), "Rx for the 'Blockbuster Drug' of patient engagement", *Health Affairs,* February, http://dx.doi.org/10.1377/hlthaff.2013.0037, accessed 15 March 2015.

DHHS (2013), "About the Health Data Initiative", Department of Health and Human Services, 30 September, www.hhs.gov/open/initiatives/hdi/about.html, accessed 30 September 2013.

European Commission (2015), Research & Innovation Participant Portal, https://ec.europa.eu/research/participants/portal/desktop/en/opportunities/index.html accessed 15 May 2015.

Eysenbach, G. (2007), "Five-year prospective study harvesting search and click data from Google 2004-2007", presented at AMIA Annual Fall Symposium, Chicago, www.jmir.org/2009/1/e11/.

Eysenbach, G. (2006), "Infodemiology: Tracking flu-related searches on the web for syndromic surveillance", AMIA Annual Symposium Proceedings, pp. 244-48.

FDA (2013), *Mini-Sentinel*, US Food and Drug Administration, http://mini-sentinel.org/ (accessed 13 September 2013).

Fernald, G.H. et al. (2011), "Bioinformatics challenges for personalized medicine", *Bioinformatics*, Vol. 27, No. 13, pp. 1741-48.

Friedman C., T.C. Rindflesch and M. Corn (2013), "Natural language processing: State of the art and prospects for significant progress, a workshop sponsored by the National Library of Medicine, *Journal of Biomedical Informatics*, Vol. 46, No. 5, pp. 765-73.

Greene, C.S. and O.G. Troyanskaya (2012), "Data-driven view of disease biology", *PLoS Computational Biology*, Chapter 2, Volume 8, No. 12, http://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.10028161002816.

Grossmann, C., B. Powers and J.M. McGinnis (eds.) (2011), "Digital infrastructure for the learning health system: The foundation for continuous improvement in health and health care – Workshop Series Summary", National Academies Press, Washington DC.

Häkkinen, U. et al. (2013), "Health care performance comparison using a disease-based approach: The EuroHOPE Project", *Health Policy*, 13 May, http://dx.doi.org/10.1016/j.healthpol.2013.04.013.

Hulth, A., G. Rydevik and A. Linde (2009), "Web queries as a source for syndromic surveillance", PLoS ONE, Vol. 4, Issue 2, www.plosone.org/article/info:doi/10.1371/journal.pone.0004378.

ICES (2013), "At a glance: Evidence guiding health care", Institute for Clinical and Evaluative Sciences, www.ices.on.ca/webpage.cfm?site_id=1&org_id=70, accessed 13 September 2013.

IHME (2013), "Global burden of disease", Institute for Health Metrics and Evaluation, www.healthmetricsandevaluation.org/sites/default/files/policy_report/2011/GBD_Generating%20Evidence_Guiding%20Policy_GBD2010_results.pdf, accessed 7 May 2015.

Jensen, P.B., L.J. Jensen and S. Brunak (2012), "Mining electronic health records: Towards better research applications and clinical care", *Nature Reviews Genetics,* 13, pp. 395-405.

Kaye, J. et alia (2014), "Dynamic consent: A patient interface for twenty-first century research networks", *European Journal of Human Genetics*, http://dx.doi.org/10.1038/ejhg.2014.71, accessed 15 March 2015.

Larsson, S. et alia (2012), "Use of 13 disease registries in 5 countries demonstrates the potential to use outcome data to improve health care's value", *Health Affairs*, Vol. 31, No. 1, pp. 220-27.

Laurence, J. et al (2014), "Patient engagement: Four case studies that highlight the potential for improved health outcomes and reduced costs", *Health Affairs,* September, Vol. 33, No. 9, pp. 1627-34; http://dx.doi.org/doi:10.1377/hlthaff.2014.0375, accessed 15 March 2015.

Liu, B. et al. (2006), "Exploring candidate genes for human brain diseases from a brain-specific gene network", Biochemical and Biophysical Research Communications, Vol. 349, pp. 1308-14.

Merriam-Webster (2014), "Crowdsourcing", Merriam-Webster.com, Merriam-Webster, www.merriam-webster.com/dictionary/crowdsourcing, accessed 24 September 2014.

Nature (2013a), "When Google got flu wrong", 13 February, www.nature.com/news/when-google-got-flu-wrong-1.12413, accessed 24 September 2014.

Nature (2013b), "Geneticists push for global data sharing", 4 June, www.nature.com/news/geneticists-push-for-global-data-sharing-1.13133, accessed 17 September 2013.

Nazi, K.M. et al. (2013), "Evaluating patient access to electronic health records: Results from a survey of veterans, *Medical Care*, Vol. 51, pp. S52-S56, March, http://journals.lww.com/lww-medicalcare/Fulltext/2013/03001/Evaluating_Patient_Access_to_Electronic_Health.11, accessed 6 March 2015.

NHS England (2013), "Care episodes statistics: Technical specifications of the GP extract", National Health Service, May, www.england.nhs.uk/wp-content/uploads/2013/08/cd-ces-tech-spec.pdf, accessed 13 September 2013.

NIH (2014), "DNA sequencing costs", National Human Genome Research Institute, National Institutes of Health, www.genome.gov/sequencingcosts/, accessed 3 May 2015.

OECD (2014a), "Unleashing the power of big data for alzheimer's disease and dementia research: Main points of the OECD expert consultation on unlocking global collaboration to accelerate innovation for Alzheimer's disease and dementia", *OECD Digital Economy Papers*, No. 233, OECD Publishing, Paris.
DOI: http://dx.doi.org/10.1787/5jz73kvmvbwb-en.

OECD (2014b), Health Statistics 2014., www.oecd.org/els/health-systems/health-data.htm accessed 15 May 2015.

OECD (2013a), Health Statistics 2013, www.oecd.org/els/health-systems/health-data.htm, accessed 15 May 2015.

OECD (2013b), *Health at a Glance 2013: OECD Indicators*, OECD Publishing, Paris. DOI: http://dx.doi.org/10.1787/health_glance-2013-en.

OECD (2013c), *ICTs and the Health Sector: Towards Smarter Health and Wellness Models*, OECD Publishing, Paris. http://dx.doi.org/10.1787/9789264202863-en.

OECD (2013d), "Unlocking global collaboration to accelerate innovation for Alzheimer's disease and dementia", OECD Publishing, Paris, www.oecd.org/sti/sci-tech/unlockingglobalcollaborationtoaccelerateinnovationforalzheimersdiseaseanddementia.htm.

OECD (2013e), "Strengthening health information infrastructure for health care quality governance: Good practices, new opportunities and data privacy protection challenges", *OECD Health Policy Studies*, OECD Publishing, Paris. http://dx.doi.org/10.1787/9789264193505-en.

OECD (2010), *Improving Health Sector Efficiency: The Role of Information and Communication Technologies*, OECD Publishing, Paris. http://dx.doi.org/10.1787/9789264084612-en.

Okamoto, E. et al. (2013), "Evaluation of the health check up and guidance program through linkage of health insurance claims", *Journal of the National Institute of Public Health,* Vol. 62, No. 1.

Olson, D.R. et al. (2013), Reassessing Google flu trends data for detection of seasonal and pandemic influenza: A comparative epidemiological study at three geographic scales", *PLoS Computational Biology*, Vol. 9, Issue 10, www.ploscompbiol.org/article/info:doi/10.1371/journal.pcbi.1003256.

PCORI (2013), "Important questions, meaningful answers", Patient-Centered Outcomes Research Institute, www.pcori.org/, accessed 17 October 2013.

PricewaterhouseCoopers (2012) Emerging MHealth: Paths for Growth

Robillard, J.M. et al. (2013), "Aging 2.0: Health information about dementia on Twitter", PLoS ONE, Vol. 8, Issue 7, www.plosone.org/article/info:doi/10.1371/journal.pone.0069861.

Roederer, M.W. (2009), "Cytochrome P450 enzymes and genotype-guided drug therapy". *Current Opinion in Molecular Therapeutics*, Vol. 11, No. 6, pp. 632-40.

Sadilek, A., H. Kautz and V. Silenzio (2012), "Predicting disease transmission from geo-tagged micro-blog data", Conference Proceedings from the 26th AAAI Conference on Artificial Intelligence, American Association for Artificial Intelligence, pp. 136-42.

Scanfeld, D., V. Scanfeld and E.L. Larson (2010) "Dissemination of health information through social networks: Twitter and antibiotics", *American Journal of Infection Control* Vol. 38, pp. 182-88.

Schneeweiss, S. et al. (2011), "Assessing the comparative effectiveness of newly marketed medications: Methodological challenges and implications for drug development", *Clinical Pharmacology and Therapeutics*, Vol. 90, No. 6, pp. 777-90.

Shirky, C. (2010) *Cognitive Surplus: How Technology Makes Consumers into Collaborators*, Penguin Group.

Signorini, A., A.M. Segre and P.M. Polgreen (2011), "The use of Twitter to track levels of disease activity and public concern in the U.S. during the influenza A H1N1 pandemic", PLoS ONE. http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0019467, accessed 15 March 2015.

Swan, M. (2012a), "Crowdsourced health research studies: An important emerging complement to clinical trials in the public health research ecosystem", *Journal of Medical Internet Research*, Vol. 14, No. 2.

Swan M. (2012b) "Scaling crowdsourced health studies: The emergence of a new form of contract research organization", *Personalized Medicine*, Vol. 9, No. 2, pp. 223-34.

The Daily Briefing (2011), "Will comparative effective research answer medicine's toughest questions?", 17 August, www.advisory.com/Daily-Breifing/2011/08/17/Will-comparative-effective-research-answer-medicines-toughest-questions, accessed 16 September 2013.

UK Department of Health (2012), "The power of information: Putting all of us in control of the health and care information we need", London, 21 May, www.gov.uk/government/uploads/system/uploads/attachment_data/file/213689/dh_134205.pdf, accessed 27 September 2013.

UN (2013), *World Population Prospects – The 2012 Revision*, United Nations, Department of Economics and Social Affairs, Population Division, DVD edition.

Vidal, M., M.E. Cusick and A.-L. Barabasi (2011), "Interactome networks and human disease", *Nature Reviews Genetics,* Vol. 12, pp. 56-68.

Wicks, P. et al. (2011), "Accelerated clinical discovery using self-reported patient data collected online and a patient matching algorithm", *Nature Biotechnology*, Vol. 29, No. 5, pp. 411-14.

Yin, J. et al. (2012), "ESA: Emergency situation awareness via microbloggers", paper presented at the Conference on Information and Knowledge Management (CIKM).

Further reading

Accenture (2012), "Patient access to electronic health records: What does the doctor order?", www.accenture.com/SiteCollectionDocuments/PDF/Accenture-Patient-Access-to-Electronic-Health-Records-What-Does-the-Doctor-Order.pdf, accessed 2 August 2013.

AliveCor, www.alivecor.com/, accessed 8 October 2013.

Anwar, M.N. and M.P. Oakes (2011), "Data mining of audiology patient records: Factors influencing the choice of hearing aid type, BMC Medical Informatics and Decision Making", in *Proceedings of the ACM Fifth International Workshop on Data and Text Mining in Biomedical Informatics*, Association for Computing Machinery, pp. 11-18.

Asfaw, B. et al. (2010), "Host-based anomaly detection for pervasive medical systems", *Risks and Security of Internet and Systems (CRiSIS) 2010 Fifth International Conference*, pp. 1-8.

Astolfi, R., L. Lorenzoni and J. Oderkirk (2012), "Informing policy makers about future health spending: A comparative analysis of forecasting methods in OECD countries", *Health Policy,* Vol. 107, pp. 1-10.

Avron, J. (2007), "In defense of pharmacoepidemiology – Embracing the yin and yang of drug research", *New England Journal of Medicine,* Vol. 357, pp. 2219-21.

Barth, J. et al. (2012), "Combined analysis of sensor data from hand and gait motor function improves automatic recognition of Parkinson's disease", *Conference Proceedings*, Institute of Electrical and Electronics Engineers, Engineering in Medicine and Biology Society, pp. 5122-25.

Beniwal, S. and J. Arora (2012), "Classification and feature selection techniques in data mining", *International Journal of Engineering Research & Technology*, Vol. 1, No. 6.

British Medical Journal (2013), "Turning doctors into coders", Editorial, Vol. 347, 24-31 August.

Brothers, K.B., D.R. Morrison and E.W. Clayton (2011), "Two large-scale surveys on community attitudes toward an opt-out biobank", *American Journal of Medical Genetics Part A,* Vol. 155, No. 12, pp. 2982-90.

Burgel, P.P. et al (2010), "Clinical COPD phenotypes: A novel approach using principal component and cluster analyses", *The European Respiratory Journal*, Vol. 36, No. 3, pp. 531-39.

Cavouklan, A. (2013), "Looking forward: De-identification developments – New tools, new challenges", Office of the Information and Privacy Commissioner, Ontario, Canada, May.

Cecchini M. et al. (2010), "Tackling of unhealthy diets, physical inactivity, and obesity: health effects and cost-effectiveness", *The Lancet*, Vol. 376, No. 9754, pp. 1775-84.

Chaturvedi, S.K., V. Richariya and N. Tiwari (2012), "Anomaly detection in network using data mining techniques", *International Journal of Emerging Technology and Advanced Engineering*, Vol. 2, No. 5, pp. 349-353.

Chee, B.W., R. Berlin and B. Schatz (2011), "Predicting adverse drug events from personal health messages", *AMIA Annual Symposium Proceedings*, pp. 217-26.

Chellaprabha, B. and M. Archana (2013), "Anomaly data leakage detection", *International Journal of Engineering and Innovative Technology*, Vol. 2, No. 10, pp. 210-15.

Coloma, P.M. et al. (2012), "Electronic healthcare databases for active drug safety surveillance: Is there enough leverage?", *Pharmacoepidemiology and Drug Safety*, Vol. 21, pp. 611-21.

Computer Weekly (2013), "What does a petabyte look like?", www.computerweekly.com/feature/What-does-a-petabyte-look-like, accessed 17 September 2013.

Copeland, L.A., J.E. Zeber and C.P. Wang (2009), "Patterns of primary care and mortality among patients with schizophrenia or diabetes: A cluster analysis approach to the retrospective study of healthcare utilization", *BMC Health Services Research*, Vol. 9, No. 127.

Data Protection Working Party (2013), "Opinion 03/2013 on purpose limitation", adopted on 2 April, 00569/13/EN WP 203, http://idpc.gov.mt/dbfile.aspx/Opinion3_2013.pdf, accessed 23 April 2014.

Davenport, T.H. and D.J. Patil (2012), "Data scientist: The sexiest job of the 21st century", *Harvard Business Review*, October.

Department of Health and Human Services Centres for Medicare and Medicaid Services (2012), *Report to Congress: Fraud Prevention System First Implementation Year*, www.stopmedicarefraud.gov/fraud-rtc12142012.pdf, accessed 16 September 2013.

Di Iorio, C.T., F. Carinci and J. Oderkirk (2013), "Health research and systems' governance are at risk: Should the right to data protection override health?", *Journal of Medical Ethics,* www.ncbi.nlm.nih.gov/pubmed/24310171, accessed 15 March 2015.

Eddy, D.M. et al. (2011), "Individualized guidelines: The potential for increasing quality and reducing costs", *Medicine and Public Policy*, 3 May.

Eder, J., H. Gottweis and K. Zatloukal (2012), "IT solutions for privacy protection in biobanking", *Public Health Genomics,* Vol. 15, pp. 254-62.

El Emam (2013), *Risky Business: Sharing Health Data While Protecting Privacy*, Trafford Publishing, United States.

Enterprise Irregulars (2011), "The enterprise opportunity of big data: Closing the 'clue gap'", www.enterpriseirregulars.com/40616/the-enterprise-opportunity-of-big-data-closing-the-clue-gap/, accessed 17 September 2013.

EUBIROD (2013), "European best information through regional outcomes in diabetes", www.eubirod.eu/index.htm, accessed 20 September 2013.

EUnetHTA, European Network for Health Technology Assessment, www.eunethta.eu/, accessed 16 September 2013.

Eurobarometer (2010), *Biotechnology*, European Commission.

EUROCARE (2013), www.eurocare.it, accessed 20 September 2013.

European Commission (2003), Pharmaceutical Forum http://ec.europa.eu/enterprise/sectors/healthcare/competitiveness/pricing-reimbursement/european-initiatives/index_en.htm#h2-the-high-level-pharmaceutical-forum-on-pricing-and-reimbursement, accessed 16 September 2013.

European Commission (2012), Proposal for a regulation of the European Parliament and of the Council on the protection of the individual with regard to the processing of personal data and on the free movement of such data (General Data Protection Regulation), COM(2012)0011 http://eurlex.europa.eu/smartapi/cgi/sga_doc?smartapi!celexplus!prod!CELEXnumdoc&lg=en&numdoc=52012PC0011, accessed 20 March 2013.

European Commission (2010), "Comparative study on different approaches to new privacy challenges, in particular in the light of technological developments", Working Paper No. 2: "Data protection laws in the EU: The difficulties in meeting the challenges posed by global social and technical developments", http://ec.europa.eu/justice/policies/privacy/docs/studies/new_privacy_challenges/final_report_working_paper_2_en.pdf, accessed 20 March 2013.

European Society of Cardiology, EURObservational Research Programme, www.escardio.org/eorp, accessed 16 September 2013.

Feldman, L. (2011), "Medical device integration – More than meets the eye", *For The Record*, Vol. 23, No. 13, 18 July, p. 20.

Foundation Centre, Number of Grantmaking Foundations 1975 to 2011, http://foundationcenter.org/findfunders/statistics/pdf/02_found_growth/2011/03_11.pdf, accessed 6 September 2013.

Global Fund to Fight Aids, "Tuberculosis and Malaria", www.theglobalfund.org/en/, accessed 6 September 2013.

Grand Challenges in Global Health, www.grandchallenges.org/Pages/Default.aspx, accessed 6 September 2013.

Health 2.0 Developer Challenge, www.health2con.com/devchallenge/blue-button-co-design-challenge/, accessed 20 September 2013.

Horgan, D., D. Byrne and A. Brand (2013), "EU directive on patients' rights to cross border healthcare: A largely theoretical achievement; so much more remains to be done, Editorials, *British Medical Journal*, Vol. 347, 7694f.

Hresko, A. and S.B. Haga (2012), "Insurance coverage policies for personalized medicine", *Journal of Personalized Medicine*, www.mdpi.com/2075-4426/2/4/201, accessed 8 October 2013.

Jaffe, M.G. et al. (2013), "Improved blood pressure control associated with a large-scale hypertension program, *Journal of the American Medical Association*, Vol. 310, No. 7, pp. 699-705.

Kaiser Permanente (2011), "Community benefit report: deeper and stronger, our commitment to total health continues to grow", http://share.kaiserpermanente.org/static/cb_annualreport/reports/docs/2011_CB_Annual%20_Report.pdf, accessed 13 September 2013.

Kimko, H.H.C. and C.C. Peck (eds.) (2011), "Clinical trial simulations: Applications and trends", *Advances in the Pharmaceutical Sciences Series*, American Association of Pharmaceutical Scientists, Vol. 1, Springer.

Lewis, C. et al (2013), "Public views on the donation and use of human biological samples in biomedical research: A mixed methods study", *British Medical Journal Open,* Vol. 3, No. 8, http://bmjopen.bmj.com/content/3/8/e003056.full, accessed 15 March 2015.

Lin, K.C. and C.L. Yeh (2012), "Use of data mining techniques to detect medical fraud in health insurance", *International Journal of Engineering and Technology Innovation*, Vol. 2, No. 2, pp. 126-37.

Liu, F., C. Weng and H. Yu (2012), "Natural language processing, electronic health records, and clinical research", in R.L. Richesson and J.E. Andrews (eds.), *Clinical Research Informatics (Health Informatics)*, Chapter 16, 293 Springer-Verlag London Limited.

Lusoli, W. et al. (2012) "Pan-European survey of practices, attitudes and policy preferences as regards personal identity data management", *JRC Scientific and Policy Reports*, European Commission.

Magoc, T. and D. Magoc (2011), "Neural network to identify individuals at health risk", *International Journal of Artificial Intelligence & Applications*, Vol. 2, No. 2, pp. 104-14.

Marx, V. (2013), "The big challenges of big data", *Nature*, Vol. 498, June.

Medcore Reveal Insertable Cardiac Monitors, www.medtronic.com/for-healthcare-professionals/products-therapies/cardiac-rhythm/cardiac-monitors-insert/reveal-dx-and-reveal-xt-insertable-cardiac-monitors-icms/index.htm, accessed 8 October 2013.

Microsoft Health Vault, www.healthvault.com/fr/en (accessed 8 October 2013).

Miksad, R.A. (2011), "When a decision must be made: Role of computer modelling in clinical cancer research*", Journal of Clinical Oncology,* 10 December, Vol. 29, No. 35, pp. 4602-4.

Mosen, D.M. et al. (2013), "More comprehensive discussion of CRC screening associated with higher screening", *American Journal of Managed Care*, Vol. 19, No. 4, pp. 265-71.

Mosen, D.M. et al. (2010), "Automated telephone calls improved completion of faecal occult blood testing", *Med Care*, Vol. 48, No. 7, pp. 604-10.

National Cancer Institute Fact Sheet, www.cancer.gov/cancertopics/factsheet/Risk/genetic-testing, accessed 8 October 2013.

National Centre for Health Statistics, National Health and Nutrition Examination Survey, www.cdc.gov/nchs/nhanes/response_rates_cps.htm, accessed 8 October 2013.

National Centre for Quality Assurance (2013), "NCQA to test pioneering way to measure quality, foster wider use of prevention strategies", press release, 9 April, www.ncqa.org/Newsroom/2013NewsArchives/NewsReleaseApril92013.aspx, accessed 27 September 2013.

National Institutes of Health, "Big Data to Knowledge Initiative", http://bd2k.nih.gov/about_bd2k.html#areas, accessed 23 September 2013.

NCBI Gene Expression Omnibus, www.ncbi.nlm.nih.gov/geo/, accessed 8 October 2013.

New York Times (2013), "Mining electronic health records for revealing health data", 14 January, www.nytimes.com/2013/01/15/health/mining-electronic-records-for-revealing-health-data.html?pagewanted=all, accessed 17 September 2013.

NIH Genetic Testing Registry, www.ncbi.nlm.nih.gov/gtr/, accessed 8 October 2013.

NIH National Human Genome Research Institute, www.genome.gov/10002335, accessed 8 October 2013.

OECD (2012), *Health at a Glance Europe*, www.oecd.org/els/health-systems/healthataglanceeurope.htm, accessed 7 May 2015.

OECD (2011), *Policy Issues for the Development and Use of Biomarkers in Health*, OECD Publishing, Paris.

OECD (2009a), OECD Guidelines on Human Biobanks and Genetic Research Databases, OECD Publishing, Paris.

OECD (2009b), *Pharmacogenetics: Opportunities and Challenges for Health Innovation*, OECD Publishing, Paris.

OECD (2007), *Genetic Testing: A Survey of Quality Assurance and Proficiency Standards*, OECD Publishing, Paris.

Open Government Partnership (2013), www.opengovpartnership.org/, accessed 27 September 2013.

Orphanet Report Series (2013), "Disease Registries" in *Europe*, www.orpha.net/orphacom/cahiers/docs/GB/Registries.pdf, accessed 18 October 2013.

O'Sullivan, B.P., D.M. Orenstein and C.E. Milla (2013), "Pricing for Orphan Drugs: Will the market bear what society cannot?", *Journal of the American Medical Association,* Vol. 310, No. 13, pp. 1343-44.

Patients Like Me, www.patientslikeme.com/, accessed 8 October 2013.

Pearson, J.F., C.A. Brownstein and J.S. Brownstein (2011), "Potential for electronic health records and online social networking to redefine medical research", *Clinical Chemistry* Vol. 57, No. 2, pp. 196-204.

Peptide Atlas, www.peptideatlas.org/, accessed 8 October 2013.

Rahm, A.K. et al. (2013), "Biobanking for research: A survey of patient population attitudes and understanding", *Journal of Community Genet*ics, October, Vol. 4, No. 4, pp. 445-50, http://dx.doi.org/10.1007/s12687-013-0146-0.

Rutter, C.M., D.L. Miglioretti and J. Savarino (2011), "Evaluating risk factor assumptions: A simulation-based approach", *BMC Medical Informatics Decision Making,* September, Vol. 7, No. 11, p. 55.

Simon, C.M., J.L. L'Heureux and B. Zimmerman (2011), "Active consent but not too active: Public perspectives on biobank consent models", *GENMED*, Vol. 13, No. 9, pp. 821-31.

Specimen Central (2013), www.specimencentral.com/biobank-directory.aspx, accessed 20 October 2013.

Statistics Canada, *Canadian Health Measures Survey Data Users Guide*, 2011.

Tatonetti N.P. et al (2011), "Detecting drug interactions from adverse-event reports: Interaction between paroxetine and pravastatin increases blood glucose levels", *Clinical Pharmacology & Therapeutics,* July, Vol. 90, No. 1, pp. 133-42, http://dx.doi.org/10.1038/clpt.2011.83 (accessed 15 March 2015).

Thompson Reuters (2013), Derwent World Patents Index, http://thomsonreuters.com/derwent-world-patents-index/, accessed 8 October 2013.

UCLA Centre for Networked Embedded Sensing, http://research.cens.ucla.edu/urbansensing/projects/find/, accessed 8 October 2013.

United Kingdom (2013a), Data.Gov.UK, http://data.gov.uk/, accessed 27 September 2013.

United Kingdom (2013b), www.gov.uk/government/news/health-secretary-to-strengthen-patient-privacy-on-confidential-data-use, press release, accessed 20 September 2013.

United States Environmental Protection Agency (2004), *Potential Implications of Genomics for Regulatory and Risk Assessment*, Washington, DC, December.

United States White House, Office of Science and Technology Policy, www.whitehouse.gov/blog/2013/06/07/health-datapalooza-iv-tops-huge-year-health-data-liberation-innovation, accessed 20 September 2013.

Ubaldi, B. (2013), "Open government data: Towards empirical analysis of open government data initiatives", OECD Working Papers on Public Governance, No. 22, OECD Publishing, Paris, http://dx.doi.org/10.1787/5k46bj4f03s7-en.

Wang, C. et al. (2011), "Health and economic burden of the projected obesity trends in the US and the UK", *The Lancet,* Vol. 378, No. 9793, pp. 815-25.

Wang, X. et al. (2009), "Active computerized pharmacovigilance using natural language processing, statistics, and electronic health records: A feasibility study", *Journal of the American Medical Informatics Association*, Vol. 16, No. 3, pp. 328-37.

World Medical Association (1964), *World Medical Association Declaration of Helsinki: Ethical Principles for Medical Research Involving Human Subjects*, archived at www.wma.net/en/30publications/10policies/b3/17c.pdf, accessed 7 May 2015.

Yeh, J.Y., T.S. Wu and C.W. Tsao (2010), "Using data mining techniques to predict hospitalization of hemodialysis patients", *Decision Support Systems*, Vol. 50, pp. 439-48.

Ylä-Herttuala, S. (2012), "Endgame: Glybera finally recommended for approval as the first gene therapy drug in the European Union, *Molecular Therapy*, Vol. 20, No. 10, pp. 1831-32.

Zhang, W. et al. (2011), "Role prediction using electronic medical record system audits", *AMIA Annual Symposium Proceedings*, American Medical Informatics Association, pp. 858-67.

Zubi, Z.S. and R.A. Saad (2011), "Using some data mining techniques for early diagnosis of lung cancer", *Proceedings of the 10th WSEAS International Conference on Artificial intelligence, Knowledge Engineering and Data Bases*, pp. 32-37.

Zucchelli, E., A.M. Jones and N. Rice (2010), "The evaluation of health policies through microsimulation methods", *HEDG Working Paper*, University of York, January.

From:

Data-Driven Innovation

Big Data for Growth and Well-Being

Access the complete publication at:
https://doi.org/10.1787/9789264229358-en

Please cite this chapter as:

OECD (2015), "The evolution of health care in a data-rich environment", in *Data-Driven Innovation: Big Data for Growth and Well-Being*, OECD Publishing, Paris.

DOI: https://doi.org/10.1787/9789264229358-12-en