OECD publishing

BUILDING AND SUSTAINING COLLABORATIVE PLATFORMS IN GENOMICS AND BIOBANKS FOR HEALTH INNOVATION

OECD SCIENCE, TECHNOLOGY AND INDUSTRY
POLICY PAPERS
March 2021 **No. 102**

OECD
BETTER POLICIES FOR BETTER LIVES

This paper was approved and declassified by written procedure by the Committee for Scientific and Technological Policy (CSTP) on 17 February 2021 and prepared for publication by the OECD Secretariat.

Note to Delegations:
This document is also available on O.N.E under the reference code:
DSTI/STP/BNCT(2019)12/REV2/FINAL

Building and sustaining collaborative platforms in genomics and biobanks for health innovation

Hermann Garden*, Naomi Hawkins**, David E. Winickoff*

Genomic and biobank collaborative platforms hold significant promise for the development of new discoveries and therapies. This paper explores the complex technical, legal and business challenges arising from genomics and biobanks, and brings together ideas and best practices from major national and international platforms, and from a diverse range of experts. The global sharing of biological samples and genomic data has been critical for accelerating our understanding of the biology and spread of COVID-19, and for the development of vaccines and diagnostics. Although some of the policy challenges in the field are well known, they have been reconfigured by the digitalisation of health innovation combined with the increasing complexity and volume of data, the push for global collaboration, and the growing awareness of ethical, legal, and social implications.

Keywords: biobanks, collaborative platforms, genomics, governance, innovation policy, health, sustainability.

*OECD

**University of Exeter Law School, Exeter, United Kingdom

Foreword

This report discusses institutional arrangements and business models underlying collaborative platforms in genomics and biobanks for personalised health. It pools ideas and best practices from representatives of major national and international platforms in genomics and biobanks, and from a diverse range of experts at the triple interface of IP and data policy, business, and health systems.

In order to deepen the discussion a workshop "Collaborative platforms for personalised health: realising the potential of genomics and biobanks" (17-18 September 2019, Stockholm, Sweden) was organised under the auspices of the OECD Working Party on Biotechnology, Nanotechnology and Converging Technologies (BNCT) and hosted by Vinnova, Swedish Governmental Agency for Innovation Systems, Stockholm, Sweden.

This report was co-authored by Naomi Hawkins, Hermann Garden, and David Winickoff.

A special note of appreciation is extended to the Korean Legislation Research Institute (KLRI) and to Vinnova, Stockholm, Sweden, for supporting this project.

Table of contents

Executive Summary

Genomic and biobank collaborative platforms hold significant promise for the development of new discoveries and therapies. The global sharing of biological samples and genomic data has been critical for accelerating our understanding of the biology and spread of COVID-19 and for the development of vaccines and diagnostics. However, challenges remain: for instance, the nature of personal health and genomic data gives rise to complications in the use of this data. The productivity of these platforms will therefore depend on balancing a number of interests related to the control, access and linking of genomic data and samples. Building and sustaining the scientific, economic and social value of these platforms is a defining challenge. Platform governance must manage issues around privacy and data protection, fragmentation and interoperability, standards and federated learning.

There are a plethora of large-scale projects at the national and international levels, and increasingly genomics initiatives are linking up in larger health networks. These collaborative platforms reflect an array of models (governmental health system level, non-profit, private sector, and networked initiatives). There is growing integration between the public and private sector in this field, and attempts to find the right mix of public and private engagement are ongoing. Public-private partnerships (PPPs) are addressing issues around implementation, sustainability, and wider adoption but also raising difficult questions about the control of data and research.

The growth of genomic and biobank collaborative platforms has been driven by a need for access to vast quantities of high quality data, and has been enabled by increasingly powerful digital tools: indeed, advances in digital technologies are enabling changes in the field and reconfiguring the value of data. Most salient of these advances, cloud computing and artificial intelligence (AI) hold the promise of better sharing, mining and extraction of value from data. Accordingly, platforms that can collect and organise large amounts of data are able to take on greater scientific and economic value. At the same time, the use of these technologies has increased the complexity of the regulatory and governance frameworks against which collaborative platforms develop.

Business models in biomedical research and innovation tend to be closely tied to intellectual property (IP) rights, which are integral to the translation of research into clinically relevant outcomes in biomedicine. However, ethical concerns about the role of IP in genomics translational research remain, and failure to respond has the potential to impact public trust, and thus sustainability of this research endeavour. Restrictions or positive obligations as to the licensing of IP arising from the platform have the potential to advance public access to the technology in question. Attempts by collaborative platforms to use such conditions for IP rights to advance the purpose of the platform should be encouraged.

The sustainability of genomic and biobank collaborative platforms will depend on balancing competing interests in the control, access and linkage of genomic data and samples. Economic dimensions and financial considerations must align with social imperatives and public trust to ensure the maintenance of a "social contract" that entails mutual responsibilities across participants, publics, research institutions and researchers. Transparent and inclusive governance frameworks can help define terms of the social contract, resolve tensions therein, and promote the value of collaborative platforms in genomics.

Policy messages

Building and sustaining scientific, economic and social value around data is essential for genomic and biobank collaborative platforms. Main policy implications of this report relate to digital technologies, collaboration around data, economic and social aspects of sustainability, and governance frameworks:

Advances in digital technologies present important opportunities for genomic and biobank collaborative platforms

- Cloud computing provides flexible computational services which are hugely beneficial for genomic platforms. The inherent scalability of cloud resources enables genomic and biobank initiatives to adapt computation resources to their needs.
- Artificial intelligence (AI) and machine learning show great promise for genomics research. They are useful to enable both the generation and analysis of large volumes of data, as well as to facilitate new tools of genomic analysis. However, ethical, legal and social aspects of AI require attention. Due regard to the potential for bias, and the development of best practices for recognising and minimising the downstream effects of biased training data sets are necessary. Moreover, questions of explainability, responsibility and privacy in the AI context must be recognised and addressed.
- A collaborative platform's underlying digital system may become outdated within a short period of time. It is therefore important that digital systems are built on agile foundations that anticipate changes in the field.

Data is the primary currency of value for collaborative platforms in this area, carrying key implications for governance

- The need for access to vast quantities of high quality and often decentralised data has driven the establishment of genomic and biobank collaborative platforms. Genomic datasets are growing out of research and clinical practices worldwide, opening up important opportunities for collaborations holding clinical promise.
- Questions of privacy and data protection are central to the operation of collaborative platforms, and there are both ethical and legal obligations which flow as a result. One thing is clear: individuals, through rights of consent over data and material sharing, should be considered partners in the governance of genomic data bases and biobanks.
- Conflicting and divergent international regulations can represent important and time-consuming obstacles to smooth data flows, and compliance with data protection regulation is often time consuming for collaborative platforms. Partners and stakeholders will have to strike an appropriate balance between the level of protection of privacy and the liberal use of data.
- Interoperability of data is key to collaboration – it is imperative that different systems can speak to each other. Nevertheless, due to cultural differences, some variation in standards, laws and regulations across national systems and among platforms is inevitable and may be desirable: there is a need to recognise the importance of the context in designing effective governance. In these cases, meta-

standards, or higher-level standards might be useful in harmonising diverse approaches.

- Federated data analysis (i.e. analysing data stored in a network of local data collections without a centralised data repository) shows great promise for enabling effective and productive collaboration for platforms. This approach can leverage cloud computing and machine learning, and allows researchers to abstract analysis on top of secure multi-party computation systems. This functionality may address problems with data quantity, distributed and fragmented data sets and help ensure genetic data privacy and compliance. However, further work towards international consensus, standard setting and harmonisation is necessary to ensure that the promises of federation are realised.

The sustainability of collaborative platforms depends on a "social contract" that entails mutual responsibilities and trust across participants, publics, research institutions and researchers.

- Public funding agencies, research institutions, researchers and licensees are partners in a "social contract" that carries mutual responsibilities: each party has ethical claims that must be respected, but each owes each other certain obligations. The combination of reciprocal rights and obligations is likely to enable collaborative platforms to operate fairly and effectively for the benefit of all parties.
- Trust in genomics research is essential for the success of collaborative platforms in this field. The notion of the public good should help structure the interface between public and private actors. Participant engagement underpins trust in the research, and helps to maintain levels of participation in the research, as well as improves its relevance and utility. Relationships of trust do not occur spontaneously, but require active management.
- IP policies and practices need to be consistent with existing governance frameworks, and governance frameworks need to be open to commercialisation of downstream research.
- Collaborative platforms should be transparent in their communication, for instance employing the terminology of openness with caution, clarity and precision. Participants and the public must not be misled by the use of the rhetoric of openness, if in fact controlled access, fees for service, commercial secrecy and patenting will ultimately be employed, because to do so risks undermining public trust.
- Collaborative platforms are well placed to develop innovative and creative licensing policies. Attempts by collaborative platforms to use licensing and contractual conditions for IP rights to advance the public purpose of the platform should be encouraged. Moreover, collaborative platforms and policymakers should monitor the impact of licensing and IP policies on translation into clinical use and access.

Governance can be an effective means to address the challenges and realise the opportunities for genomic and biobank collaborative platforms

- Highest ethical standards, best practice privacy protections, benefit sharing arrangements, transparency, accountability and openness are the core elements of governance in population genomics and biobanking.

- Transparent and inclusive governance frameworks and public engagement can help define terms of the social contract, resolve tensions therein, and promote the value of collaborative platforms in genomics. There is for example, a need to support ongoing dialogue about the role of research partnerships between public and private sectors.
- Relationships of trust do not occur spontaneously, but require active management. Genomic research institutions should not seek to manufacture trust in individuals or communities, but rather to implement governance mechanisms such that the biobank deserves trust. Participatory governance, where participants have a say, is one potentially useful path forward.
- Ongoing ethical oversight is important to build trust and establish trustworthiness.
- Standardisation can help integrate multiple platforms and magnify the power of genomic analysis.
- Collaborative platforms could serve as a model for the development and testing of processes and standards in information technology networking, public deliberation, commercial strategies in biomedical research, and approaches to responsible innovation.
- The OECD Recommendation of the Council on Human Biobanks and Genetic Research Databases [OECD/LEGAL/0375], and the Recommendation of the Council concerning Access to Research Data from Public Funding [OECD/LEGAL/0347] provide useful assistance in designing good governance structures for this field.

1. Introduction

Health care systems in OECD countries and beyond are increasingly focused on patient centred and personalised medicine - a more precise approach to patient care that grows out of a better understanding of, and access to, an individual's genomic sequence and other omics data. Such an approach is enabling a more detailed understanding of health and disease risks (OECD, 2019[1]; Rehm, 2017[2]). Key rationales for government support of personalised medicines initiatives are, for example, to develop novel, more efficient diagnostics and therapies and to better monitor and improve existing therapies. Recent research has allowed advances in personalised medicine, and promises more efficient and individualised treatment (Burke and Psaty, 2007[3]; Rehm, 2017[2]).

It is an exciting and dynamic time for genomic databases and biobanks.[1] There have never been so many large-scale, national and international projects and attempts to sequence at the population scale, to integrate genomic data into healthcare systems, and to link genomics initiatives together in larger networks of health data and biological samples (Dubow and Marjanovic, 2016[4]). An analysis by Stark (Stark et al., 2019[5]) revealed that since 2013 governments of at least 14 countries have invested over USD 4 billion in establishing national genomic-medicine initiatives. The IQVIA Institute for Human Data Science has identified 187 genomic initiatives globally of which half are US based and close to one fifth in Europe (Aitken, 2020[6]). These collaborative platforms reflect an array of models, but, viewed together, demonstrate a trend towards finding the right kind of mix

[1] Biobanks can be defined as "a collection of biological material and the associated data and information stored in an organised system, for a population or a large subset of a population." Also referred to as Human genetic research database(s) (HGRD) or "population database(s)". https://stats.oecd.org/glossary/detail.asp?ID=7220

of public and private elements. Indeed, building and sustaining value – scientific, economic, and social – in this context is a defining challenge across many genomic and biobank collaborative platforms. Building sustainability is a useful lens for understanding policy goals in this arena, with data as the foundation for building and sustaining collaborative platforms, and the field as a whole. Collaboration within and across platforms on the sharing of data is central to building sustainability.

In 2001 the OECD led ground-breaking work on Biological Resource Centres (BRC) as a key component of the sustainable, international scientific and technological infrastructure of the life sciences and biotechnology (Müller et al., 2020[7]; OECD, 2001[8]). Biological Resource Centres (BRCs) are recognised as key to the realisation of the opportunities of biotechnology in health, industry, and other key sectors. Though outside the field of therapeutics and diagnostics, the OECD Best Practices Guidelines for Biological Resource Centres (2007[9]) provide useful references for quality management and for the development of national certification systems.

To support the development of Human Biobanks and Genetic Research Databases, and to address some of the governance challenges they entail, the OECD Council adopted the Recommendation of the Council on Human Biobanks and Genetic Research Databases in 2009 on the proposal of the Committee for Scientific and Technological Policy (CSTP). The Recommendation applies to human biobanks and genetic research databases, which are structured resources that can be used for the purpose of genetic research and which include: a) human biological materials and/or information generated from the analysis of the same; and b) extensive associated information (OECD, 2009[10]).

Population-based biobanks monitor the health status of participants over time to assess the natural occurrence and progression of common diseases. These biobanks combined with genomic and health data can enable a more personalised approach to medicine by locating the genetic component of human disease (Müller et al., 2020[7]; Kinkorová, 2016[11]; Zatloukal et al., 2018[12]). Moreover, growth of genomics markets and associated health sectors offers wider societal benefits including the potential to increase investment in innovation, generate new economic activity and create new jobs.

This document explores the complex technical, legal and business challenges arising from genomics – the study of all genes of an organism and their inter relationships in order to identify their combined functions – and biobanks. Although some of the policy challenges of the field are well known, the digitalisation of health innovation combined with the increasing complexity and volume of data, the push for global collaboration, and the growing awareness of the ethical, legal, and social implications (ELSI) are reconfiguring policy challenges (OECD, 2019[13]; Kinkorová and Topolčan, 2018[14]). As the recently enacted OECD Recommendation on Artificial Intelligence (2019[15]) and other similar statements attest, AI is one such factor reframing human rights and democratic values in the digital age.

The document highlights two key dynamics that are pushing a new array of policy issues to the fore: First, in terms of technology and data, a confluence of developments -- whole genome sequencing (WGS), the collection of identifiable, real-world health evidence and lifestyle indicators, combined with AI-driven innovation -- offer the potential to better understand genotype-phenotype associations for more effective therapies and diagnostics (Caulfield and Murdoch, 2017[16]; Robinson, 2012[17]; Tam et al., 2019[18]; Thorogood et al., 2019[19]). However, these personalised approaches demand the responsible management of ever greater and 'deeper' assemblages of –omics profiles, body and brain images, biomarkers, and clinical data (Bycroft et al., 2018[20]; Mallappallil et al., 2020[21]; Topol, 2019[22]). Policy issues around data integration, the return of individual genomic results to participants, workforce development, and cost effectiveness come further to the fore

(Minari, Brothers and Morrison, 2018[23]; Stark et al., 2019[5]; Yehia and Eng, 2019[24]). Despite the undisputed importance of personal health information and emerging digital technologies for the translation of genomic data into clinics, the risks of data ownership, potential misuse, genetic discrimination, and health inequality and disparities have to be addressed (Hindorff, Bonham and Ohno-Machado, 2018[25]; Joly et al., 2020[26]; West, Blacksher and Burke, 2017[27]).

Second, in terms of institutional change, many genomic and biobank collaborative platforms are seeking forms of sustainability and value creation that are less dependent on public funds, and instead are more self-sustaining, such as through public-private business models (Ciaburri, Napolitano and Bravo, 2017[28]; Livesey, 2019[29]; Rao et al., 2019[30]). At play are different aspects of sustainability that operate with different economies of value, from financial models of investment inputs and knowledge outputs to social models built on trust (Andry et al., 2017[31]). Together these economies constitute a kind of social contract that entails mutual responsibilities across participants, publics, research institutions, and the private sector. As research is translated into clinical use and markets, governance frameworks that seek to protect public interests and generate public benefit have had to engage more deeply with private sector incentives, commercialisation, and IP rights (Ballantyne and Schaefer, 2020[32]; Hofman et al., 2014[33]; Vaught et al., 2011[34]). However, the goals of generating private investment and promoting public value and social innovation are not always aligned, with implications for public trust (Livesey, 2019[29]; Stark et al., 2019[5]). Transparent and inclusive governance frameworks and business plans can help define the terms of the social contract, to strengthen social innovation, and to promote the value of genomic and biobank collaborative platforms (Hofman et al., 2014[33]; Matzke et al., 2016[35]).

2. Approaches to collaboration in genomics and biobanks

The goal of many genomic initiatives and biobanks is to maximise the use of data in order to optimise scientific, economic and social value. Genomic and biobank collaborative platforms enable access to scientific information, including meta-data and samples, provide important search functions and offer a variety of user services, for example online reservation of testing facilities (OECD, 2017[36]).

Collaboration has long been integral to genomics research and personalised medicine. The Human Genome Project (HGP) was a large, publicly-funded, collaborative effort, involving more than 2 000 researchers across the globe (Cook-Deegan, Ankeny and Jones, 2017[37]; Maxson Jones, Ankeny and Cook-Deegan, 2018[38]). Initiated in 1990 and successfully completed in 2003, the HGP provides an excellent example of how collaborative research can deliver fundamental information that would inspire many platform-based genomics projects.

Other large collaborations in genomics, such as the HapMap project (Belmont et al., 2003[39]), the International Cancer Genome Consortium[2] and recently, the ICGC/TCGA Pan-Cancer Analysis of Whole Genomes Consortium have given rise to important developments in the genetic underpinnings of human disease (Campbell et al., 2020[40]).

The fields of genomics and biobanking draws on this tradition in order to better realise the value of genomic and biobank resources. Data sharing, in a manner that appropriately protects the interests of the participants whose data is shared, is vital to enable the advancement of research (Rehm, 2017[2]), and sharing of data, strategies and standards has

[2] https://icgc.org

the potential to reduce duplication and speed progress in identifying and translating innovations to the clinic (Manolio et al., 2015[41]). Collaboration is also required by the need for vast quantities of high quality data, such that no individual researcher can generate sufficient data alone.

Key opportunities for genomic and biobank collaborative platforms are to:

- Deliver application-neutral resources (e.g. data, technologies, processes) for innovative solutions to complex challenges in health and society.
- Provide a means for multi-sector collaboration through shared investment and results, collaborative use of data and novel technologies, innovative approaches to organisational efficiency and sustainability.
- Enable technical and governance interoperability: foster standardisation, and data storage and data sharing through the implementation of novel technical solutions and innovative legal and ethical governance mechanisms.

Sharing in the genomics context is of course motivated further in the face of global crises. To give a contemporary example, the COVID-19 pandemic has presented researchers and clinicians with confounding variability of infected individuals' response to the disease, ranging from death, severe disease, mild symptoms, atypical symptoms, or no symptoms at all. Such variability amongst people leads to a high probability that there are strong genetic and epigenetic causes that affect the response, and a large genetic and additional omics study is necessary to investigate the genetic determinants of COVID-19 susceptibility, severity and outcomes (Murray et al., 2020[42]; Schäfer and Baric, 2020[43])

It is anticipated that such research will help to generate hypotheses for drug discovery and repurposing, identify individuals at unusually high or low risk, and contribute to global knowledge of the biology of SARS-CoV-2 infection and disease.

In this context, work in progress to develop collaborative research environments has been accelerated and enabled the rapid development of shared COVID-19 resources. Pre-existing platforms and long running studies in affected countries have given researchers a head-start on collecting high-quality data on the possible genetic contributions to the effect of the disease. Collaborative COVID-19 research based on rapid sharing of data, techniques and results has been transformative for the perception and impact of scientific research globally.

For example, the European COVID-19 Data Platform is a joint initiative by the European Commission, the European Bioinformatics Institute of the European Molecular Biology Laboratory (EMBL-EBI), the Elixir infrastructure and the COMPARE project, working with EU Member States and other partners. It is intended to support open science and open access, by enabling researchers to store and share datasets (such as DNA sequences, protein structures, data from pre-clinical research and clinical trials). This incentive is part of the effort to realise the objectives of the European Open Science Cloud. The entry point to the Platform is the COVID-19 Portal, which brings together relevant datasets submitted to EMBL-EBI and other major centres for biomedical data. Within the Portal, researchers should 'provide immediate and full open access and to share research outcomes (data, models, workflows, results) that are as fair as possible in real time'(European Union 2020).

2.1. Mission-oriented innovation policies and grand challenges

The field of genomics research has been built on extensive international public and charitable funding – from the early days, prior to the human genome project, the human genome project itself and subsequently. Moreover, many collaborative platforms

themselves were established through public funding, and continue to rely on public funding for their existence. Taking a "missions approach", i.e. what the OECD calls a mission-oriented innovation policy (MOIP) is a promising way to enhance collaboration in the field of genomic initiatives.

Strategic orientation and coordination of public and private activities lie at the heart of MOIPs. These bold initiatives are result-oriented, wide-spanning and coordinated sets of policy measures supporting the whole innovation chain to achieve common ambitious objectives. Facing mounting societal challenges such as climate change, population ageing, unmet medical needs, and increasing health-care costs, a number of governments in the OECD and beyond are experimenting different types of MOIPs.

Governments engage in MOIPs on the basis of their perceived strengths to make strides in fighting complex societal challenges such as climate change or emerging health threats, with the goals of:

- More open and exploratory challenge-oriented approaches, targeting selected problems, instead of supporting specific solutions.
- Holistic coordination, involving not only the authorities in charge of research and innovation policies but also the sectoral ministries (health, environment, energy, agriculture) that 'own' the challenges.
- Enhanced effectiveness through focused resources, policies and regulations.
- Reduction of unnecessary overlaps and greater coverage of various optional solutions.
- Increased political and public legitimacy, allowing bolder and longer-term interventions.

These measures can span different stages of the innovation cycle from research to demonstration and market deployment, mix supply-push and demand-pull instruments, and cut across various policy fields. They stand in sharp contrast with the traditional single science or technology-push policy instruments that have proved their limit to address complex and ambitious challenges. Such challenges require breaking away from established scientific and technologic trajectories, sectorial boundaries and ministerial silos, in order to devise novel solutions through joined up actions.

MOIPs are starting to be applied to the health area and more specifically in genomics and biobanks in order to address unmet public health needs and accelerate innovation processes. In the United Kingdom for instance a dedicated mission aims to expand the 100 000 Genomes Project to a total of 1 million whole genomes sequenced by the NHS and United Kingdom Biobank over five years. In Australia, the Genomics Health Futures Mission[3] aims to improve testing and diagnosis for many diseases, help personalise treatment options to better target and improve health outcomes, and reduce unnecessary interventions and health costs.

2.2. Typology of genomic and biobank collaborative platforms

There is no single model for genomic and biobank collaborative platforms. Instead, there are various approaches, which tend to share some key features that are explored in more detail below. As will be apparent, a collaborative platform could be constituted by a single collection of genomic and health data that seeks to bring partners to use and exploit it, but

[3] https://www.health.gov.au/initiatives-and-programs/genomics-health-futures-mission

it also refers to broader networks of such resources that share data and human biological samples (e.g. cells, tissues, bodily fluids or biomolecules) among themselves and with third parties.

Most of the collaborative platforms in this area have been initiated and are driven by governments, non-governmental organisation (NGOs)/ civil society organisations (CSOs)[4], and the private sector. Although platforms may span these categories, the typology is nonetheless useful for understanding the structures and operation of genomic and biobank collaborative platforms.

2.2.1. National, government led initiatives

Government led initiatives have been set up by countries as public or non-profit initiatives. Most genomic and biobank collaborative platforms focus on the development of diagnostics and individualised treatment through large-scale genomic sequencing techniques, fall in this category. These initiatives are set up with government funding as not-for-profit research infrastructures with differing levels of involvement and rules for collaborating with civil society, and private sector entities.

Government led initiatives often set the number of sequenced genomes or patients to collect and encourage public engagement and participation to achieve the goal, and establish infrastructure to collect, analyse, and share data. Initiatives differ in the type of data (e.g. genomic data, clinical data, electronic health care records) that they aim to collect, the way to store collected data (e.g. virtual storage, central repository), and the way to share curated data (e.g. differing levels of restriction on access). In addition, participant consent processes vary, as do the nature and extent of collaboration internationally, and with the private sector.

An important challenge for these types of initiatives arises due to the difficulties of long-term funding. Where the research infrastructure is funded through public research funding, funding cycles are often perilously short to produce outcomes, and without the renewal of funding, the outcomes of a project risk being lost. Moreover, the resources involved in applying for renewal of funding are significant and the focus of the project can be diverted away from the core research aims. Where collaborative platforms are supported predominantly through fixed term government and grant funding, they are vulnerable to changes in government policy, changes in funding structures or fragmented funding.

Examples of government let initiatives include:

2025 France Genomic Medicine Plan

An example of government led, integrated approaches to genomics and personalised health is the 2025 France Genomic Medicine Plan. It has been designed to advance science and innovation, introduce genome sequencing into public health care services, and to develop robust economic models that allow for collaboration with the private sector. As part of the initiative, a network of two sequencing platforms across France have been established.

[4] Civil society organisations (CSOs) can be defined as nonmarket and non-state organisations outside of the family in which people organise themselves to pursue shared interests in the public domain (OECD, 2010[155]). Non-governmental organisation (NGOs) can be defined as any non-profit entity organised on a local, national or international level to pursue shared objectives and ideals, without significant government controlled participation or representation. NGOs include foundations, co-operative societies, trade unions, and ad-hoc entities set up to collect funds for a specific purpose (OECD, 2019[156]). In this document the term non-governmental organisation (NGO) is used synonymously with the term civil society organisation (CSO).

The Reference Centre for Innovation, Assessment and Transfer (CRefIX, Centre de référence, d'innovation, d'expertise et de transfert) was established as a multi-institutional unit between Inserm (National Institute of Health and Medical Research), CEA (Alternative Energies and Atomic Energy Commissions), and Inria (National Research Institute for the Digital Sciences). The objective of CRefIX is to set reference standards, to manage implementation of those standards in PFMG 2025, and to promote innovation and collaboration with the industrial sector. CRefIX also works on standardization of conditions regarding new indications for genomic medicine, protocols for sequencing platforms, and collaborates with CAD (National Centre for Intensive Calculation) and COFRAC (French Accreditation Committee).

Maccabi Healthcare Services, The Israeli National Biobank for Research and Psifas, Israel

There are three collaborative platforms and biobanks being developed in Israel at present.

Maccabi Healthcare Services (MHS) has developoed the Tipa Biobank, a population-based biobank in Israel. MHS is the second largest healthcare provider in Israel, serving 2.3 million members, which constitutes a representative quarter of the Israeli population, and has electronic health records which form a longitudinal history of many patients throughout their entire lives. The Tipa Biobank, launced in December 2017, collects a wide variety of samples and links with electronic health records data in a de-identified manner for broad research use. The uniqueness of the Tipa Biobank is that the collection of samples is repeated from the same patients throughout their lives, therefore creating a "biological health record" which can be used for development of early detection tests and liquid biopsy development. Maccabi has already recruited over 135 000 members with more than 400 000 samples due to high trust levels in the provider.

The Israeli National Biobank for Research (MIDGAM) was established in order to promote academic research and biomedical industry in Israel. MIDGAM is funded by several governmental ministries and agencies and operates under the supervision of the Chief Scientist of the Israel Ministry of health. Since 2014, samples and annotated demographic and clinical data have been collected, processed and stored in several medical centres. The medical centres report to MIDGAM HQ, which handles investigator requests and operates a database, quality control measures, scientific and regulatory counselling services and price lists.

Psifas is Israel's National Precision Medicine Initiative, a research oriented project, designed to collect health data and biological samples from hundreds of thousands of Israeli Donors. Psifas will collect biosamples as well as clinical data from electronic medical records and questionnaires, genomic data and continuous physiological data utilising devices. The information obtained will create a national research framework and virtual research environment. Data will be anonymised and encrypted.

Initiative on Rare and Undiagnosed Diseases (IRUD), Japan

The Initiative on Rare and Undiagnosed Diseases (IRUD), Japan, was established in 2015, and is and coordinated by the Japan Agency for Medical Research and Development (AMED). The initiative has encouraged collaboration between paediatric and adult research consortia, with a focus on (1) reaching out to potential participating institutions and co-ordinating them as the 'All-Japan' clinical research program for those who are currently undiagnosed; (2) developing globally compatible databases and identifying data-sharing opportunities; and (3) accelerating research and development in the field of rare and undiagnosed diseases.

IRUD consists of three main pillars – IRUD Diagnosis Committees, IRUD Analysis Centres, and the IRUD Data Centre. IRUD Coordinating Center are in charge of overall coordination and discussion among experts and local communities, and IRUD Analysis Centres take patient samples and analyse them, while IRUD Data Centre manage the data. An important focus is on interoperability of the data, and the IRUD Exchange data platform is interoperable with diverse Japanese and foreign platforms.

BBMRI-ERIC, Europe

The Biobanking and BioMolecular resources Research Infrastructure–European Research Infrastructure Consortium (BBMRI-ERIC), an umbrella organization for biobanking in Europe, was founded in 2013 to provide a focal point for biobanking activities in Europe and to provide fair access to quality-controlled human biological samples and associated data for cross-biobanking research. BBMRI-ERIC currently includes 20 countries and one international organisation, making it one of the largest European research infrastructures.

BBMRI-ERIC provides a gateway for access to the collections, expertise and services of the European research community, ensuring coordination and efficiency, and new services and better access for users. It enables close collaboration between researchers, biobanks, patient advocacy groups and the biotech and pharma industry. It also seeks to improve the interoperability of the existing comprehensive population based or clinically oriented collections of biological samples and data. As existing biobanks have a strong national character and background, BBMRI-ERIC uses a distributed hub-and-spoke structure. This structure provides necessary flexibility so that new Member States and Observers can be connected at any time, and so that it is easy to respond to the emerging needs of biomedical research (Mayrhofer et al. 2016).

Korea National Bio Big Data Project, Korea

The Korea National Bio Big Data Project is carried out by a consortium of several institutes, with three playing a key role with respect to biosamples and data: the Korea Disease Control and Prevention Agency (KDCA), which is under the Ministry of Health and Welfare, the Korea Research Institute of Bioscience and Biotechnology (KRIBB) and the Korea Institute of Science and Technology Information (KISTI), both of which are under the Ministry of Science and ICT.

Three types of data are collected for rare disease patients – (1) blood and urine samples, (2) clinical data such as general information, family history, diagnosis results, family history, treatment history, and follow-up monitoring information, and (3) genomic data. Biosamples are collected by a total of sixteen hospitals consisting of regional base centres for rare diseases and tertiary hospitals. Clinical data are collected by the hospitals through a data management system of KDCA. Genomic data are produced by NGS service providers and transported to KRIBB. Biosamples are stored at the National Biobank of Korea, which is under KDCA. Clinical data are stored at KDCA. Genomic data are stored at the KRIBB. Both the clinical and the genomic data from this project are shared among the project consortium member institutes – KDCA, KRIBB, and KISTI. KISTI will construct a Clinical Interpretation Research Network (CIRN), which is based on a closed system among KDCA, KRIBB, and KISTI. External researchers access the virtual research environment provided by CIRN, which will accommodate access requests after consideration of research ethics and computing resource demand.

Genomic Medicine Sweden (GMS), Sweden

Exhibiting a nation-wide networked model of a collaborative platform, Genomic Medicine Sweden (GMS) is a publicly funded national initiative headed by 14 partners from regional healthcare and universities. In the field of precision medicine, GMS is currently

coordinating, or contributing to, a number of activities addressing various issues surrounding data sharing, including interoperability the ethical, legal, and social aspects of data sharing. The utilization of health care data for research and innovation is one of eight prioritised areas in Sweden's national strategy for life science.

Collaboration with the private sector is of key importance for GMS. The establishment of a national informatics infrastructure in Sweden, joint research projects as well as the accessing and utilizing of genomic data for biomarker discovery, development of new diagnostics and drugs, and for clinical studies all engage the private sector. Defining a sustainable business model is also critical to overcome the challenges imposed by the regional healthcare structure in Sweden. As a first step, GMS has formed a working group for innovation and industry collaboration. A focus on opportunities and challenges for collaborations between the public healthcare and industry in Sweden has led to the initiation of collaborative projects between GMS and specific private sector partners. An important precondition to industry collaboration is a mutual understanding that the ability to access and utilize health care data within GMS and its regional health authority partners is dependent on resolving the legal, ethical, and social aspects of data sharing. Involving industry partners at an early stage is important in order to secure that the national genomics database/informatics infrastructure meets the criteria to be an asset also for the private sector.

Genomics England, United Kingdom

In the United Kingdom, Genomics England Limited (Ltd), a company, was founded by, and is wholly owned by, the Department of Health & Social Care. Genomics England collaborates with stakeholders, including governmental initiatives, hospitals, and universities, and managed the 100 000 Genomes Project, which was launched in 2012 to catalyse the uptake of genomic medicine for the benefit of UK National Health Service (NHS) patients and research. The project was successfully completed in 2018 and involved over 130 NHS partner institutions across the UK, delivering a total of 125 000 whole genomes from approximately 85 000 patients and unaffected relatives. Genomics England is now leading the expansion of the project with further genomic sequencing and increased integration of genomic medicine into the NHS. Genomics England also collects samples of blood and tumours, clinical data, and Diagnostic Imaging Dataset (DID), and stores and shares curated data with own identifiers in the virtual data centre(s). Data is made available through a Trusted Research Environment for commercial and non-commercial research. Access to the data is governed by the Access Review Committee and requires users and their institutions to agree to the Participation Agreement and Rules. This includes expectations on the researcher and their institution around information governance, acceptable uses, and the approach to collaboration, publications and IP rights. Genomics England also is focused on developing capabilities in sequencing, data storage, ethics and public engagement.

Along with Genomics England, the Genomics England Clinical Interpretation Partnership (GeCIP), composed of 42 groups working on different diseases or topics, was established to facilitate interactions. NHS genomic medicine centres (GMCs) were established to deal with patient recruitment, informed consent management, and sample collection. The NHS Genomic Medicine Service was launched to transform outcomes and resources of the project into clinical use.

As a result of the COVID-19 pandemic, a new study led by Genomics England, the GenOMICC consortium and the UK National Health Service (NHS) is working to deliver whole genome sequencing of up to 20 000 individuals who have been severely affected by COVID-19 (requiring intensive care) and 15 000 additional individuals who have had mild symptoms. The aim of this study is to accelerate genomic research to better understand the

disease, provide better diagnosis, and to facilitate drug discovery in order to accelerate the response to the global crisis. Genomics England's first research priority for the new research environment is the COVID-19 participant cohort, quickly followed by leveraging the 100 000 Genomes Project data in the fight against cancer and rare diseases. The initiative will be delivered in partnership with Lifebit, United Kingdom, and the global cloud provider Amazon Web Services. Lifebit develops bioinformatics and cognitive software solutions that enable federated analysis in genomics and medical big data.

National Institutes of Health (NIH) initiatives, United States

ClinVar and ClinGen, two National Institutes of Health (NIH) genomic resources, have formed a partnership to improve our knowledge of clinically relevant genomic variation for use in research and precision medicine. This partnership includes efforts in data sharing, data archiving, and collaborative curation to characterise and disseminate the clinical relevance of genomic variation. To date, over 1 500 contributors from academic institutions, medical centres, and clinical testing laboratories have provided expert curation for 1 800 genes and submitted 1.3 million variant records to ClinVar. The FDA recognises ClinGen's variant curations as a valid source of evidence for test development and validation. ClinGen has established collaborations with other NIH-funded efforts such as CPIC and PharmGKB (two main resources for the pharmacogenomics research and implementation communities) along with the non-profit sector and CSOs (e.g the American College of Medical Genetics and Genomics (ACMG), the Association for Clinical Genomic Science (ACGS), and the American Society of Hematology (ASH)) and with companies (e.g. Genomenon, a genomic health IT company, and Concert Genetics, a technology company) to foster innovation, learn from each other, and share findings and recommendations.

The Estonian Biobank (EBB)

Estonia provides an example of government health system initiatives, representing a unique type of collaborative platforms that cuts across the health care and wider data environment in relation to citizens of that state. The Estonian Biobank (EBB), hosted in the Institute of Genomics, University of Tartu,[5] is one of the largest in Europe, with a cohort size of approximately 200 000 participants which closely reflects the age, sex and geographical distribution of the Estonian population. Participants have given broad consent, which includes consent to the linkage of the biobank to other Estonian databases and registries, which allows the collection of detailed information about participants, creating a powerful resource for research (Leitsalu et al., 2015[44]). Since 2002, Estonia has developed an e-government model, with extensive IT architecture and registries for a broad range of social functions, from homeownership and banking to healthcare.

Estonia, a country with 1.3 million inhabitants, has become a world leader in creating a more connected society, making public and private services more convenient, easy to use and personalised, all while protecting individual data privacy, and this is demonstrated in their biobank. For example, Estonia has implemented eHealth solutions, such as electronic health records, national image archiving, ePrescriptions, eReferrals, eAmbulance and eConsultations into the public health system (OECD, 2019[45]). However, the strongly integrated biobank model may prove difficult to implement in other countries with a larger population size or less well integrated IT systems. Moreover, many other countries would also face opposition to this level of governmental access and control of detailed information about citizens, which may represent an obstacle to adoption of a system offering this level of integration of medical and social data.

[5] https://genomics.ut.ee/en

2.2.2. Intergovernmental and civil society organisation initiatives

Intergovernmental and civil society organisation initiatives in genomics and biobanking often seek to leverage the power of networks to address particular challenges or issues. Such networks link platforms and key stakeholders with expertise to establish frameworks and standards. Two exemplars are the Global Alliance for Genomics & Health (GA4GH) and ELIXIR, which recently engaged in a strategic partnership for the development of technical standards and regulatory frameworks to facilitate responsible sharing of genomic data between countries and institutions.[6]

The Global Alliance for Genomics & Health (GA4GH) is an international, non-profit alliance formed in 2013 to accelerate the potential of research and medicine to advance human health.[7] GA4GH brings together more than 500 leading organisations from healthcare, research, patient advocacy, life science, and information technology to create frameworks and standards that enable responsible, voluntary, and secure sharing of genomic and health-related data. Twenty-three real world genomic data initiatives have signed on as GA4GH Driver Projects to help guide GA4GH's development efforts and pilot GA4GH tools. All GA4GH work builds upon the 'Framework for Responsible Sharing of Genomic and Health-Related Data', a guidance document founded on the human right to benefit from the advances of science (Knoppers, 2014[46]). GA4GH uses approved standards, including standard file formats for storing sequencing data, and standard Application Programming Interfaces (APIs) for discovering and accessing genomic data, and more than a dozen policies and frameworks to guide responsible, international data sharing.

ELIXIR is an intergovernmental organisation that brings together life science resources from across Europe, with the goal of coordinating these resources so that they form a single infrastructure.[8] ELIXIR is developing a local/ federated European Genome-phenome Archive (EGA), as a secure storage for sensitive human sequence and sequence-related data.[9] The EGA provides a service for the permanent archiving and distribution of personally identifiable genetic and phenotypic data resulting from biomedical research projects. EGA allows authorised users to search sequenced material, patient samples stored in biobanks, and the metadata around patients (their illnesses, treatments, outcomes). It also queries national search engines on behalf of the users. The Federated EGA extends and generalises the system of access authorisation and secure data transfer developed in the EGA. It aims to provide a framework for the secure submission, archiving, dissemination and analysis of human biomedical data across Europe.

The pan-European Biobanking and BioMolecular resources Research Infrastructure-European Research Infrastructure Consortium (BBMRI-ERIC)[10] is a research infrastructure that enables responsible health innovation between academic research, biobanks, industry, and patients (Van Ommen et al., 2015[47]). BBMRI-ERIC currently includes 20 countries and one international organisation, making it one of the largest European research infrastructures.

---

[6] https://elixir-europe.org/news/elixir-and-ga4gh-expand-collaboration

[7] https://www.ga4gh.org/

[8] https://elixir-europe.org/about-us

[9] https://ega-archive.org/

[10] https://www.bbmri-eric.eu/about/

2.2.3. Private sector research initiatives

The pharmaceutical industry and biotech companies have been driving the integration of genetic genomic information into the discovery and development process for novel therapies, vaccines, and diagnostics. Examples of private sector initiatives include The Geisinger MyCode Community Health Initiative[11], launched in 2007, offers an integrated biobank and electronic health record (EHR) infrastructure for research use by Geisinger and collaborators (Carey et al., 2016[48]). And, the Foundation Medicine, established in 2010, offers tissue-based genomic testing and more than 400 000 patient profiles to inform strategies in cancer therapy.

Collaborative platforms developed entirely by the private sector are rare in genomics and biobanking. Companies instead tend to build their own genomic and bio-sample repositories, or partner with public sector initiatives (Stark et al., 2019[5]). They may wish to partner with collaborative platforms for access to data, and in such a case may be treated similarly to other public sector entities accessing or contributing data, or may access the platform on differential terms (for example pricing). In addition, companies are developing tools and techniques to interface with and supplement analytical capacity of existing collaborative platforms. The interaction of the platform and private companies depends on the legal structure of the platform, and the contractual relationship between the parties, and these vary greatly.

Some platforms form a strong collaborative relationship with a number of commercial partners, and integrate those particular commercial partners more closely into the structure of the platform, for example:

- FinnGen,[12] a public-private partnership (PPP) between Finnish universities, biobanks, hospital districts, and several international pharmaceutical companies, aims to combine genome information from 500 000 blood samples collected by a nation-wide network of Finnish biobanks with digital health care data from national health registries. Partners have joined forces to drive research, implementation, and economic development in the field of personalised medicine.
- Illumina Inc.,[13] an innovative sequencing and array technologies company, and Genomic Medicine Sweden (GMS) pursue a collaborative project on the possible use of whole-genome and RNA sequencing for the diagnosis of acute leukaemia.[14] Illumina has also developed a partnership with Genomics England to deliver whole genome sequencing for National Health Service (NHS) Genomic Medicine Service.[15]
- Imagia,[16] a Montreal-based company seeking to leverage advances in AI in personalised medicine, provide both the infrastructure for the platform, as well as commercialization products and strategies for AI-derived solutions, with the intent to commercialise the innovations based on a unique business model.

[11] https://www.geisinger.org/mycode

[12] https://www.finngen.fi/en

[13] https://www.illumina.com/

[14] https://genomicmedicine.se/en/2020/05/28/gms-illumina-whole-genome-sequencing-acute-leukemia/

[15] https://www.genomicsengland.co.uk/genomics-england-illumina-partner-nhs-genomic-medicine-service/

[16] https://imagia.com/

3. Opportunities in novel digital technologies

Recent developments in digital technologies have the potential to advance innovation in personalised medicine. Cloud computing, AI, machine learning, and synthetic data are important and influential technological developments for genomics and biobanking, as they promise to allow sharing, mining and the bringing of value to data in new ways. Platforms that can collect and organise large amounts of data are able to take on greater scientific and economic value. Meanwhile, the push towards greater international linkage has resulted in new forms of institutional collaboration. At the same time, the use of these technologies has increased the complexity of the regulatory and governance frameworks against which collaborative platforms develop and data can be shared (Bombard and Hayeems, 2020[49]; OECD, 2017[36]; OECD, 2019[50]; OECD, 2020[51]).

3.1. Cloud computing

When genomic datasets were still manageable in terms of volume and analyses were less complex and computer-intensive, on-premises High Performance Computing (HPC) solutions were sensible. However, they no longer provide the scalability needed for the volume genomic data. Furthermore, HPC solutions require large amounts of capital upfront, incur significant maintenance overheads and require constant upgrades. Genomic and biobank research environments need to consider infrastructure solutions which provide flexibility of storage and computer resources needed for running analyses at scale.

The near-infinite scalability of cloud solutions and the reliability and security of the cloud make cloud computing a logical alternative to on-premises HPC. Cloud computing has advanced greatly in recent years (OECD, 2019[52]), and is "a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources ... that can be rapidly provisioned and released with minimal management effort or service provider interaction" (Mell and Grance, 2011[53]). It therefore allows the central aggregation of data from federated collaborative projects, and the accessing of that data from a single source by project collaborators and other investigators. If genomic and biobank research environments are designed in a cloud-native manner, they can take full advantage of the near-infinite resources of the cloud (Paul, Gade and Mallipeddi, 2017[54]; Yang, 2019[55]).

In genomics, cloud computing is significant in two areas. First, cloud computing enables the reanalysis of vast data sets available in existing data archives. Second, cloud computing has permitted collaborations on large amounts of shared data, with the distributed nature of the cloud facilitating collaboration through enabling collaborative and distributed computing efforts. Projects such as the International Cancer Genome Consortium (ICGC) and the related Pan-Cancer Analysis of Whole Genomes (PCAWG), the National Cancer Institute (NCI) Cancer Genomics Cloud (CGC) Pilots and the Encyclopedia of DNA Elements (ENCODE and Model Organism ENCODE (modENCODE)), all rely on cloud computing (Langmead and Nellore, 2018[56]; Yang, 2019[55]).

In the cloud computing model, computational resources are utilities to be rented from a provider. Many cloud providers are commercial services such as Amazon Web Services (AWS), Google Cloud Platform or Microsoft Azure. However, many other cloud services, including academic services, are available worldwide (Langmead and Nellore, 2018[56]).

Cloud computing provides elastic and flexible computational services which are hugely beneficial for genomic platforms. The inherent elasticity of cloud resources enables genomic and biobank initiatives to scale their computation resources according to the amount of genomic analyses researchers need to deploy. Unlike on-premises HPC

solutions, no money is spent on idle compute resources. Furthermore, as the cloud computing market becomes increasingly competitive, cloud-based computing resources and storage will become significantly cheaper than HPC solutions. Moreover, cloud computing allows both reproducibility and global access (Langmead and Nellore, 2018[56]).

3.2. Artificial intelligence and machine learning

Bioinformatics tasks have incorporated machine learning algorithms for some years (Raza, 2020[57]) and new tools and techniques in AI and machine learning algorithms show great promise for genomics research. In the context of the huge datasets which are needed for genomics research, researchers are increasingly using AI tools both to generate and analyse large volumes of data, as well as facilitate new tools of genomic analysis, including deep learning (Raza, 2020[57]; Williams et al., 2018[58]). Next-generation machine learning algorithms are able to search through large amounts of data, including medical databases and previously published medical literature, while collating information and looking for patterns. Many aspects of analysis in genomics analysis, from sequencing, phenotyping and variant identification, to downstream interpretation can therefore benefit from AI. Thus, AI is therefore becoming an indispensable tool for working with large amounts of data in health settings (OECD, 2019[15]; OECD, 2019[59]). In fact, and conversely, genomics science and AI are co-emerging, as an increasing number of AI applications have been developed in the genomics field.

While most current applications of AI in genomics are in the research phase, it is also increasingly beginning to have relevance in clinical settings (Koumakis, 2020[60]). Universities are often involved in the initial development, but from there on, companies and public-private partnerships have taken the lead in developing machine learning applications for the treatment of health. For example, DeepSEA, developed at Princeton, predicts chromatin states and evaluates variants associated to diseases (Zhou and Troyanskaya, 2015[61]). Face2Gene employs a facial image analysis framework, DeepGestalt, using computer vision and deep-learning algorithms trained on thousands of patient cases from a phenotype-genotype database, to suggest genetic syndromes a patient may have based on their facial features (Gurovich et al., 2019[62]).

For AI to be effective, it requires large amounts of high quality data in a useable form. Collaborative platforms are an ideal source of such data, and the potential for AI to be an important tool within the collaborative platform environment is clear. Moreover, the closer integration of AI specialists with genomics specialists will improve the power of AI in relation to genomics. New types of partnerships are key to the development of this new knowledge, and public and private sector collaboration will be important for developing expertise and high quality outputs (Paranjape, Schinkel and Nanayakkara, 2020[63]).

3.3. Synthetic data

Synthetic data is increasingly finding its way into economics, healthcare, and social sciences in a variety of applications (Sergey I. Nikolenko, 2019[64]). Synthetic data is artificially generated through techniques such as deep learning, to replicate the statistical components of real-world data. Synthetic data is generated with the aim of avoiding links to identifiable individuals, and has been used effectively for this purpose in other big data fields such as finance. It offers significant opportunities to better simulate complex molecular processes, such as gene expression, pharmacodynamics, and disease progression of 'synthetic patients' (Walonoski et al., 2018[65]). Also, synthetic data can help to augment sparse research data sets and health care records.

The role that synthetic data plays in collaborative platforms in genomics and biobanks is limited at present, and the potential for synthetic data to become useful in genomics and biobanking is not universally accepted. Lifebit, for example, does not believe that the synthetic data model will be useful in genomics and biobank initiatives in the near term, due to the importance of individual point mutations in genomics. Other solutions, such as a federated approach (see below) that brings computation and analysis to the data as opposed to moving data may be more appropriate and effective, in the near term. However, synthetic data approaches may offer an alternative in the future, if concerns about privacy and data protection can be adequately addressed.

3.4. Challenges to technology implementation and continuity

Although these emerging technologies offer important opportunities for genomic and biobank collaborative platforms, they also present significant challenges around stability, continuity and resilience, as well as ethical, legal and social issues.

The underlying system a collaborative platform implements may become outdated within a short period of time. It is therefore important that systems are built on agile foundations, which anticipate future challenges. Particularly in a field of emerging technologies, with technology disruptors, it can be difficult at an early stage to predict which systems and technologies will have long term staying power, and which will become the incumbent, and which will disappear. Such lack of stability has the potential to result in wasted effort and resources.

Other key challenges to the implementation of novel digital technologies are due to ethical, legal, and social implications (Tamminen, 2011[66]). There is some wariness of the part of both the public sector and some private sector actors about the adoption of AI based systems in the medical sphere. Important ethical concerns in this field must be considered and addressed as they have implications for public trust. Particular issues relate to questions of privacy, fairness and transparency in relation to the use of AI (Morley and Floridi, 2020[67]; OECD, 2019[15]; PHG Foundation, n.d.[68]; OECD, 2019[15]; Vayena, Blasimme and Cohen, 2018[69]). Questions of privacy in relation to medical data are well recognised in relation to the use of data for traditional genomics research, but the use of AI has the potential to raise different risks for participants in the future, which should be both recognised and addressed. Questions of fairness can arise where an AI algorithm is trained on poorly representative data sets, which can introduce bias. Regard to the potential for bias, and the development of best practices for recognizing and minimizing the downstream effects of biased training data sets are necessary (Vayena, Blasimme and Cohen, 2018[69]). AI also raises difficult ethical and legal questions in relation to transparency, whereby black-box algorithms are uninterpretable, with inner logic which remains hidden even to their developers. This lack of transparency is particularly problematic for the traditional applications of legal and regulatory principles in biomedicine. Further consideration of the problems of transparency in this field is warranted (Morley and Floridi, 2020[67])).

When using patient data or other sensitive data from citizens for the development of AI, there is a challenge to ensure that parties are both aware of the opportunities and operating in a safe, lawful and mutually beneficial manner. In response to concerns around AI tools for identifying high-risk patients, the UK has developed a Code of Conduct for Data-driven Health and Care Technology. It sets out principles and points to consider on user needs, context, legislation, transparency, evidence, security and commercial strategy (UK Department of Health and Social Care, 2019[70]). Others call for a more overarching analysis of the risks of AI, which address concerns beyond the level of risks for individual, and instead focus on the broader risks at the relationship, group, institutional, and societal levels (Morley and Floridi, 2020[67]).

As AI becomes applicable in the healthcare domain, additional challenges arise, around questions of explicability, liability and privacy (Emanuel and Wachter, 2019[71]; Paranjape, Schinkel and Nanayakkara, 2020[63]; Price, Gerke and Cohen, 2019[72]).

4. Collaboration on data: challenges and opportunities

In the field of population genomics, data is the primary currency of value. Large datasets are vital to realise the potential of genomics research, and linking datasets helps to avoid duplication and waste of resources. The need for access to vast quantities of high quality, often decentralised, data has driven the establishment of genomic and biobank collaborative platforms.

The optimisation of sequencing technologies in the past ten years has resulted in huge drops in the cost of sequencing (Schwarz, Gulilat and Kim, 2019[73]; Wetterstrand, n.d.[74]). In addition, the increasing proportion of genome sequencing occurring in the health care context presents both opportunities and challenges for research and for the advance of personalised medicine. In 2012, only 1% of genomic sequencing was paid for by healthcare, with the other 99% covered by research. In 2018 that number had jumped to 20% and by 2022 it is expected that more than 80% of all sequencing will be conducted and paid for in the healthcare context. If secondary use of these clinical genomic data for research is possible, for example by developing standards to support interoperability and federation, the emergence of a virtual cohort of more than 60 million samples by 2025 would be possible. Such a resource would allow the genomics community to deliver more statistical significance in analyses. Linkage could enable the match of similar patients at disparate ends of the globe leading to increased rare disease diagnoses, strong variant interpretations, and more informed clinical decisions support (Birney, Vamathevan and Goodhand, 2017[75]).

At the same time, the nature of personal health and genomic data gives rise to important challenges in the establishment and continued operation of collaborative platforms for genomics and biobanking. Their sustainability will therefore depend on balancing a number of interests in the control, access and linkage of data, with key issues around privacy and data protection, fragmentation and interoperability, standards and federated learning.

4.1. Privacy and data protection

Collaborative platforms in genomics and biobanking deal in sensitive personal data about identifiable individuals, with considerable privacy and data protection implications. The data in question are derived from and relates to individuals, who often remain identifiable at some level. This data is often detailed, containing many data points about individual health and physiology that can be connected to individuals and their relatives. Both full genomes, as well as much more limited genomic information can be identifying, so there are important privacy and data protection implications in the use and storage of all genomic information (Bonomi, Huang and Ohno-Machado, 2020[76]; Mitchell et al., 2020[77]). As a result, the governance around privacy and data protection is central to the operation of these collaborative platforms.

Data access and sharing are not binary: data need not be either open or closed. Instead, data access exists on a continuum of restricted access through to open to the public, with the location of the continuum determined by issues of sensitivity (Ballantyne and Schaefer, 2020[32]; Dyke, Dove and Knoppers, 2016[78]). Data is often treated a monolithic entity, when in fact considerations of access to data need to be nuanced and responsive to the nature of the data in question. Much important data which is not related to health may not

be sensitive. Data which is enriched through linkage may be more valuable, and at the same time, more sensitive and thus merit greater controls on access and sharing.

4.1.1. Interests of individuals in the control of data

Genetic privacy – and the control of personal data – is a complex topic, and the web of regulation is vast (Clayton et al., 2019[79]; Pormeister, 2018[80]). Privacy rights are closely linked to other central bioethical governance principles, including, most notably, consent. Attempts to address complexities around consent in longitudinal genetics research, broad vs narrow consent, consent to governance and dynamic consent all also have implications for privacy governance (Heeney et al., 2011[81]; Kaye and Hawkins, 2014[82]; Kaye et al., 2012[83]; Knoppers and Joly, 2018[84]; Nuffield Council on Bioethics, 2015[85]; Vayena and Gasser, 2016[86]). One thing is clear: individuals, through rights of consent over data and material sharing, are important partners in the governance of genomic data bases and biobanks.

Two particular aspects of genomic data complicate questions of privacy and data protection. First, genomic information relates not only to the person from whom the information was obtained, but also to their family members. The privacy implications of the sharing of genomic data are therefore complex (Heeney et al., 2011[81]). There is as yet no international consensus on the best way to address the familial nature of genomic information in governance frameworks. The second aspect of genomic information which gives rise to challenges is its detailed, inherently identifiable nature (Erlich et al., 2018[87]; Homer et al., 2008[88]; McGuire et al., 2011[89]). Anonymising genomic data is challenging. At its core, genomic data can never be truly anonymised because each person's genetic code is by definition, unique. It has been demonstrated that the re-identification of genomic sequence data is possible by linking to publicly available data (Erlich et al., 2018[87]; Gymrek et al., 2013[90]). Even if anonymization were technically feasible, in many cases is impossible in the context of the design of the collaborative platform, taking into account the need for ongoing linkage to medical records, gathering future data, or obligations to recontact or follow up participants. Although technical solutions are important to protect privacy (Bonomi, Huang and Ohno-Machado, 2020[76]), they must be used in combination with robust governance mechanisms, discussed further below, to address these important issues.

4.1.2. Privacy and Data Protection Regulation

Governance mechanisms are key to the protection of privacy, and important advances in the law relating to privacy and data protection, as well as in recommendations and guidance from bodies such as the OECD has been made in recent years (OECD, 2013[91]; OECD, 2019[13]). Respect for and protection of the privacy of participants is fundamental to the governance of collaborative platforms, and the governance and data sharing policies of platforms must set out the ways in which privacy is protected.

The General Data Protection Regulation (GDPR) is a particularly important development in the regulation of privacy, and although an EU instrument, has implications internationally (see Box 1).

Box 1. The General Data Protection Regulation (GDPR)

The GDPR governs the processing of "personal data" of natural persons. It applies to individuals, businesses and institutions established in the European Economic Area (EEA), and also extends to data controllers or processors based outside the EEA if they offer goods or services to data subjects in the EEA, or, monitor the behaviour of individuals on EEA territory (http://data.europa.eu/eli/reg/2016/679/oj). In the absence of European Data Protection Board (EDPB)[17] or other guidance on pathways for the legal transfer of research data from the EEA to researchers outside the EEA, the extra-territorial reach of the GDPR has caused significant challenges for researchers in collaborations where data from EEA-based participants would be transferred outside of the EEA for research purposes.

This extra-territorial reach of the GDPR has caused some concern among researchers based outside the EEA who use data from participants based in the EEA.

Determining when data are "personal data" under the GDPR is of fundamental importance for those using genomic information. There is lack of clarity and consensus about when genomic and associated health data are "personal data", particularly around questions of whether data that have undergone "pseudonymisation" always remain "personal data" (Mitchell et al., 2020[77]), and debate around these questions is ongoing in the field.

While the GDPR regime was intentionally written to allow flexibilities for research, one unintended consequence of GDPR has been the the curtailment of research collaboration across US and EEA researchers. Presently, there are limited options for long-term international transfers available for many third country governments. Standard contractual clauses, codes of conduct and certification mechanisms under Article 46, and binding corporate rules under Article 47, are often not feasible for governmental entities; and the EDPB's interpretations limit Article 49 derogations. Further, there is no current guidance on a consistent standard of anonymization under GDPR, nor an understanding of how anonymization might apply to genomics research (where anonymization may not be technically possible). It would be beneficial to adopt consistent approaches to risk-assessment of identifiability across Member States, including quantitative metrics and development of sector-specific codes or certification to establish harmonised standards for genomic data. In the absence of either guidance or recognised legal pathways for research data to be transferred outside the EEA, the burdens of compliance with the GDPR as well as associated provisions of national law are significant. Attention to these issues by collaborative platforms whose operations are within the scope of the GDPR is vital.

Other countries also have their own approaches to data protection regulation. As there is a perception that the GDPR imposes the most stringent requirements, to some extent the debate has been focused on its requirements, with less consideration of the differing but often lesser requirements of other jurisdictions. Moreover, as the GDPR has imposed a high level of protection, that seeks to have extra territorial application in certain respects, a number of countries have adopted similar regimes. For example, a number of African countries have developed national regulations on data protection that will impact on the use of health data, often in response to the GDPR. The national legislation is generally based on the GDPR but there are notable differences in each jurisdiction. Some, but not all have exceptions for research, but these are not uniform. For example, funder policies on open

[17] https://edpb.europa.eu/edpb_en

access and data sharing are likely to be contrary to these national frameworks in some instances. Given this trend towards extra-territorial applications of national privacy laws, it is vitally important that the OECD consider how to identify bases for international data transfers that are permissible under these laws so that we can fully realise the power and promise of international research collaboration.

A number of important responses at the institutional level serve to address the challenges outlined above. The GA4GH Framework for Responsible Sharing of Genomic and Health-Related Data provides guidance for the responsible sharing of human genomic and health-related data, including personal health data and other types of data that may have predictive power in relation to health, by reference to the right to privacy, and is supplemented by a specific Data Privacy and Security Policy. GA4GH has developed a useful three stage privacy test, which considers: the data's sensitivity, the potential harm resulting from possible re-identification of the data; and the expectations of individuals with respect to the sharing of that data (Dyke, Dove and Knoppers, 2016[78]).

4.2. Fragmentation and interoperability

Fragmented, dispersed or distributed data present major challenges for realising the scientific and commercial value of collaborative platforms. Data becomes distributed when it is generated and stored in isolated and inaccessible environments, and is siloed by type, disease, country, institution, and sector. Distributed data complicates data accessibility and collaboration amongst different entities. Furthermore, because of the size of the data in question, transferring data from one environment to another is no longer feasible as it leads to long transfer times, increases storage costs (as data is held in multiple places) and presents significant regulatory and privacy challenges. With the rise of population genetics initiatives, such as the UK Biobank in the United Kingdom, many now want to leverage these publicly available and rich datasets to integrate them with their own private data. The current workaround is to download datasets of interest and upload them into private environments, which is unsustainable for a number of reasons. This challenging issue is only set to escalate as many genomic medicine initiatives are still in their infancy.

Technical interoperability is also key to collaboration – it is imperative that different systems can speak to each other. It is important to share and access existing data, as well as to perform research and development across dispersed and fragmented data. But achieving interoperability of these datasets is hindered by differing technical approaches, regulatory regimes and approaches to governance worldwide. For data use to be optimised, data needs to be interoperable.

Fragmentation of the regulatory environment is also a concern. In terms of privacy and data protection, although there is some harmonisation as a result of the GDPR, there remains much fragmentation, overlapping and conflicting regulation. Even in Europe, although the GDPR has had great impact on harmonisation of data protection regulation on the whole, the regulation of research remains fragmented largely due to the discretion granted to Member States in this respect in the GDPR (Pormeister, 2018[80]).

Different policy approaches to data sharing in different projects also represent a practical obstacle to data sharing. Such approaches include a data commons approach (creating trusted, controlled repositories of multiple datasets); a hub and spoke model (where common data elements, structures, and access/ use rules facilitate interoperability); linking distributed datasets through a federated approach; or sharing only high level genomic knowledge rather than raw data. Networked genomics platforms, in particular, must negotiate these different structural approaches to data access.

4.2.1. Balancing pluralism and harmonisation

Some variation in standards, laws and regulations across national systems and among platforms is inevitable and may be desirable due to cultural differences. Legal regimes are specific to each jurisdiction, and local arrangements for ethical governance and the contractual bases of individual projects are necessary. There is a need to recognise the importance of the local context in designing effective governance (Gibbons, 2009[92]; Kaye et al., 2012[83]). Moreover, where there has been a history of exploitation, and where there is power imbalance, such as where research is conducted in resource limited settings, extra attention to inequalities is important, to guard against and provide redress for exploitative research practices (Staunton and de Vries, 2020[93]; Trust Equitable Research Partnerships, 2019[94]). In such cases, capacity building may require regional variation in policies such as, for example, reduced obligations for data sharing owing to resource constraints in low and middle income countries (de Vries et al., 2015[95]). Moreover, local and regional differences in approaches to ethical issues, including in areas such as consent, in light of histories of exploitative research, may need to be accommodated. The challenge is to recognise and accommodate local variation, whilst also designing a system which permits a level of standardisation sufficient to enable interoperability.

4.3. The role of standards

The development of standards to promote interoperability is key to enabling research through collaborative platforms. These may be formal standards, as in the case of the communications industry for example. Where formal standardisation does not exist, or is not necessary or desirable, a degree of industry consensus can serve to ensure that research and collaboration is not obstructed. Working towards developing the necessary consensus to build standards at a sufficiently early stage is important, as is building the consensus with the full range of stakeholders.

One of the most important factors in generating support for a voluntary code is its content, and in this respect, it is vital to involve the full range of stakeholders in its development. Stakeholder engagement is recommended as a principle of good governance (Kaye et al., 2012[83]; Staunton and de Vries, 2020[93]), and it is important to involve both the public and private sector, across the full life cycle of genomics research. Key players include policymakers, public and private sector researchers, industry (in the full range of areas involved, including pharma, diagnostics, AI and data science) and, importantly, the public and research participants. As genomics is a truly global endeavour, input from all regions of the world, not only the western world, is essential.

International standards have the potential to coordinate the technical specifications employed by different data sources, but too many standards can introduce potential conflict and further fragmentation. Standards in the field of genomic platforms proliferate, and the majority are voluntary standards, guidelines or codes of practice. In these cases, meta-standards, or higher-level standards might be useful in harmonising diverse approaches. For instance, as of February 2020, GA4GH has produced 15 standards, including application programming interfaces, data models, schemas, and ontologies. GA4GH Work Streams collaborate with Driver Projects to frame policy and develop standards needed by the international genomics community. Driver Projects put finished standards and frameworks to immediate use. The Partner Engagement initiative aligns with global genomics organizations to ensure uptake of GA4GH standards and frameworks, and harmonization of approaches across efforts. As GA4GH seeks to provide tools to enable cooperation and harmonisation, it makes its standards and software accessible to the genomics and health community. Its software is released under an Apache 2.0 licence, or

other Open Source Initiative licence, to enable open source compatibility. For Documents and Specifications, a bespoke license that is roughly equivalent to CC-BY in terms of the permissions it offers to users is used. Importantly, GA4GH has the rights in contributions to develop and disseminate standards.

4.4. Federated learning

Federated data analysis can be used with cloud computing and machine learning, and allows researchers to abstract analysis on top of secure multi-party computation systems, addressing problems with data quantity, distributed and fragmented data sets and helping to ensure genetic data privacy and compliance. Federated data analysis means that, essentially, data never moves. Instead of relying on certifications 'guaranteeing' safe handling of classified data, a federated approach allows two or more parties in a distributed system to perform secure analysis without exposing private data to risks. Federated learning introduces a model of distributed training that enables the training of a single model when the data is scattered nodes in a network, never accessible all at once or by a single actor. This allows federated learning to train a model that accounts for all the data, often exhibits comparable accuracy to the full-availability case, yet preserves privacy in individual nodes by never requiring raw data to be pooled, shared or otherwise aggregated. Federated learning works by sending the model to the data, instead of bringing the data to the model.

Increasingly, collaborative platforms aim to create digital environments that not only act as data portals but enable researchers to 'bring their own data'. However, an essential aspect of allowing researchers to bring their own data is to also allow them to bring their own tools to effectively mine and analyse this data. Currently, it is often impossible or difficult to enable researchers to utilise their own tools across widely distributed data without having to transfer or move data. However, federated systems will allow researchers to run analyses over genomic data in a more unified and seamless way - an approach that essentially eliminates concerns about data residency, and which infrastructure and computer environments are employed.

Both public and private sector entities are developing tools to enable federated data analysis. Imagia (Canada) and Lifebit (UK) are two companies developing federated learning tools which are integrating with public sector collaborative platforms in this area. Imagia's clinical network of partnering hospitals allows the exploration of associations between different types of data that are traditionally kept in silos, namely images, genetic data and physician reports, to discover insights early and to inform downstream patient and disease outcomes. Lifebit is another company which is developing tools to enable federated data analysis, and which has partnered with Genomics England. Lifebit's federated platform brings powerful computation to data and not the other way around, allowing researchers to combine and analyse disparate cohorts of large and sensitive datasets without ever copying or moving sensitive data. Users of Genomics England's new research environment can rapidly query, analyse and collaborate over large sets of disparate data, with the research environment interface, automated tools and collaborative functionalities allowing researchers to collaboratively access data, facilitating genomic research, diagnosis, and drug discovery.

Because federated learning offers a "private-by-default" schema of data access, whereby no party has a complete view of the entire dataset, there is reduced probability of leaks due to security and privacy failures. Because data analysis is possible without data moving from the secure environment of the collaborative platform, sensitive data is better protected. By keeping data in its original location, auditing and monitoring of activities taking place within the environment is possible, in contrast with the difficulties of audit and monitoring when data is transferred.

Federation can also save both time and money. It potentially reduces the burden of administration involved in approving multiple applications to access data in multiple jurisdictions – saving time and effort on both the part of investigators applying for access, as well those who generate and hold the data.

Federation of data in this way requires broad, reciprocal data access methods that respect the national processes and patient consents of each dataset. There are thus both technical and legal and policy obstacles to federation. Moreover, in order that maximum value can be extracted from analysis of the data in question, the data must be stored according to FAIR principles – it must be findable, accessible, interoperable and reusable (Corpas et al., 2018[96]). In this respect, standardised ontologies are necessary. However, storage of data in this way takes extra time and effort, and there must be consensus on the ontologies used. Metadata is also necessary. Further work towards international consensus, standard setting and harmonisation is necessary, to ensure that the promises of federation are realised.

5. Sustainability of collaborative platforms

Structures and business models for collaborative platforms are being transformed as they transition from publicly-funded research infrastructures to increasing integration with clinical healthcare and interface with industry. There is growing recognition, both in the public and private sector, of the importance of public-private collaboration in genomics and biobanking, with business models developing, but not yet settled.

The relationship between these changes and traditional patenting and commercialization strategies remains uncertain (van Overwalle, 2009[97]). According to traditional views of innovation, stronger engagement of the private sector in biomedical research typically involves increased use of IP rights, usually patents, with a view to ensuring a return for investment. However, translational outcomes from genomics research will include not only new diagnostics and therapies, but also changes to existing approaches to treatment, better risk stratification and preventive medicine (Khoury et al., 2007[98]; Zeggini et al., 2019[99]), and the role that IP plays in the commercialisation of these types of innovation is less clear.

At play are different aspects of sustainability that operate with different economies of value, from the financial aspects of investment inputs, business models, knowledge outputs (value generation) to social models built on trust (OECD, 2017[36]; OECD, 2017[100]). Together they depend on a kind of "social contract" surrounding genomic and biobank collaborative platforms that entails mutual responsibilities across participants, publics, research institutions and researchers. Amidst these diverse sources of value that must be built, trust and trustworthiness are key.

5.1. Economic dimensions

Collaborative platforms require financial sustainability. Financial sustainability of a collaborative platform itself may be addressed by ongoing core funding by the public or other funding institutions. Cost-recouping business models, e.g. through licensing fees, can help make platforms self-funding in the long term. Some projects and platforms employ differential pricing for commercial vs academic users, or for local vs international users.

Examples are illustrative as no biobank has exactly the same model. The Tipa Biobank in Israel, for instance, a population-based biobank, is available for the use of any academic or industry researcher who receives ethical approval (Beller[101]). As the biobank was created with internal resources and philanthropic donations, and no public funding, access to the biobank is charged in order to recoup costs and enable the continuation of the biobank.

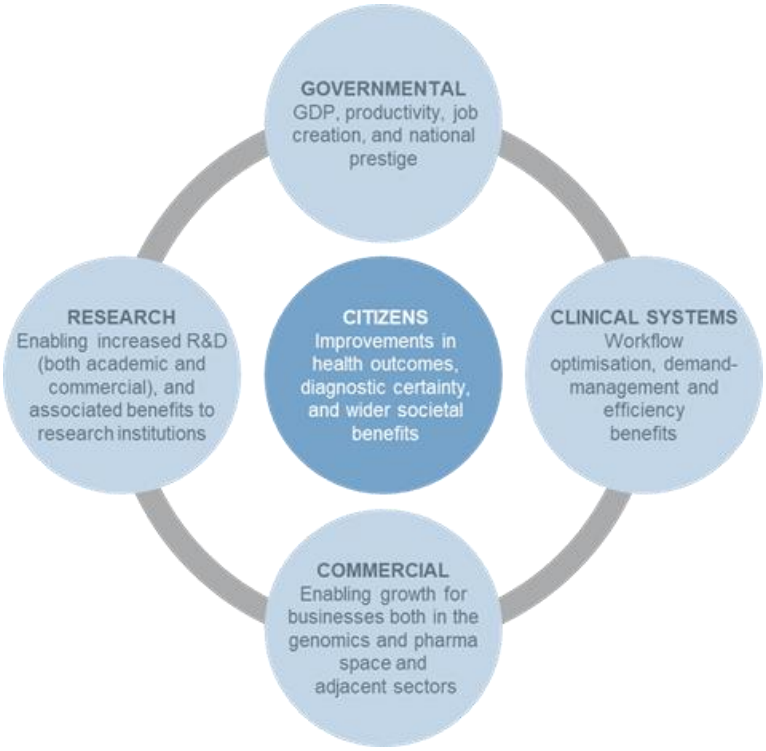
In Sweden, defining a sustainable business model is seen to be critical for overcoming the challenges imposed by the regional healthcare structure. An inclusive stakeholder forum between Genomic Medicine Sweden and industry, focused on discussing opportunities and challenges for collaborations between public healthcare and industry has led to the initiation of collaborative projects between Genomic Medicine Sweden and specific private sector partners. Involving industry partners at an early stage is important to ensure the national genomics database/informatics infrastructure is accessible to the private sector.

Finally, Israel's National Precision Medicine Initiative, Psifas, will only receive government funding for an initial five years, during which time most data infrastructure will be established, approximately 500 000 donors' samples will be collected and 100 000 genomes will be sequenced. It is expected that Psifas will be self-sustaining after the 5 years government funding, with long-term sustainability based on income from charging for services, such as usage of the platform and tools (cloud, data science, analytics tools), and partial payment for genome sequencing. All IP arising from research on Psifas resources will be owned by the researcher and/or the researcher organization.

The approach of industry in this field is very much driven by data. Commercial players seek to gain access to data in order to leverage commercial advantage. Much of this data is generated in the public sector, in research or clinical settings, and the interface between the public and private sector is also therefore focused around data access. However, to date, the focus of collaborative platforms tends to be on research, rather than on ongoing self-sufficient and sustainable business models.

Collaborative platforms involve varied stakeholders across the range of research - the public, participants and patients, clinical and research staff, industry, funders, government and policymakers (see Figure 1). The purpose of collaborations and business models is to create value, but the stakeholders involved have different conceptions of value. A company will be seeking to generate monetary profits. However, patients and governments value the outputs of research differently, and this value can be difficult to capture in traditional economic assessments. Awareness of the value to different stakeholders of the benefits of innovations in genomics is important as it can help produce mutually agreeable models (Ginsburg and Phillips, 2018[102]).

Figure 1. Stakeholders and agendas in population genomics



Source: Illumina, Inc.

5.1.1. Input/ output models

There are various types of innovative outputs of research utilising genomic and biobanking collaborative platforms, including therapeutic products, diagnostics and algorithms. Each type of output has a different manner of development, different issues in relation to IP protection, different regulatory regimes applicable, and different paths to market. Disruptive innovations may have great potential to improve healthcare for patients, but healthcare systems can be slow to adopt innovations which do not fit within existing care pathways and for which there is no existing funding stream. All these factors complicate the development of sustainability models, and there is no easy one-size-fits-all model.

Arguably, from an input/ output point of view, two key challenges have slowed the development of genomic and biobank collaborative platforms: 1) making the investment case, and 2) understanding how to develop the programme to realise the benefits (see Figure 2). There does not currently exist a cohesive view of potential benefits across different agendas (citizens, clinical systems, research, commercial, and governmental), and how they need to work together as part of a whole population genomics ecosystem.

Private sector actors have begun to create schematic models of inputs and outputs in order to better identify the challenges faced by genomic and biobank collaborative platforms. A promising collaboration between Illumina, Inc.[18] and PA Consulting[19], in collaboration with national genomics programmes, is developing a framework to support better economic

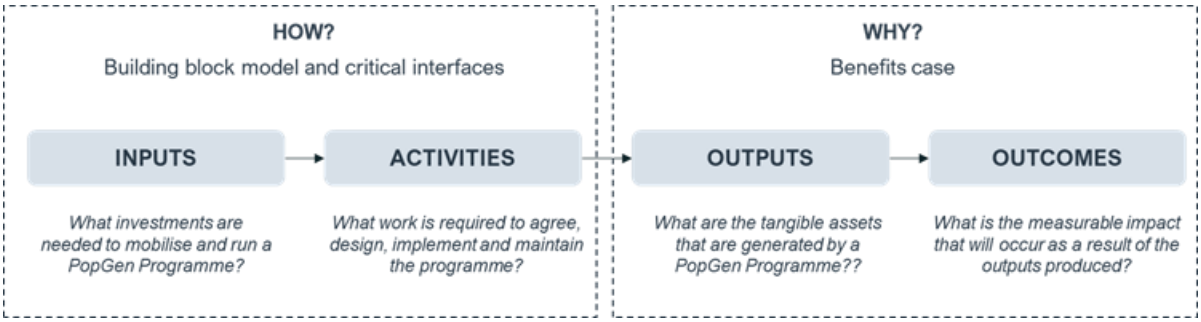
[18] https://www.illumina.com/

[19] https://www.paconsulting.com/

and ecosystem modelling on population genomics. The population genomics (PopGen) decision framework is designed to address two key questions (see Figure 2):

- How to identify the key inputs and activities required to deliver an effective population genomics programme?
- What are the outputs and potential benefits?

Figure 2. Decision-making framework for population genomics



Source: Illumina, Inc.

The first two components of the logic model might help address the question of 'how'. To identify the key inputs and activities required to deliver an effective population genomics programme, Illumina and PA Consulting are developing two views: firstly, a view of the foundational 'building blocks' required to develop any form of national population genomics programme; and secondly, a view of the overall ecosystem, the critical interfaces and building blocks required by different parts, and insights from emerging best practice on how to create these interfaces in a way which releases the desired benefits. The second two components of the logic model help address the question of 'why'. This sets out the potential benefits, the underlying causal relationships and how they relate to different agendas. Where feasible, the work will also seek to estimate the value of these benefits based on the literature currently available.

5.2. The social contract

Collaborative platforms in genomics and biobanking retain a strong public character and mandate due to the collective demands of this type of research. These collaborative platforms manage data and biosamples which are largely contributed by, and relates to, the public – patients and research participants. As a result, genomic and biobank collaborative platforms depend on the involvement of patients and the wider public as integral partners in the research. Moreover, public funding also underpins much of the research generating and analysing this data, and many of the collaborative platforms themselves. The role of the public in this research endeavour, the ways in which the public should be protected, and benefit from the research, and maintaining levels of public involvement and engagement in the research and are all important questions for the sustainability of collaborative platforms in this field. The ongoing project to construct genomic governance that acceptably orders the interface between public and private is unlikely to be successful unless we maintain the view that the enterprise is driven not by profit, but by collective political will.

These aspects of public involvement combine in the form of a social contract underpinning genomics research (Lucassen, Montgomery and Parker, 2016[103]), and indeed this language has explicitly been invoked in the case of Genomics England. This implies that for

stakeholders in the common enterprise, each party has important rights that must be respected, but also that each owes each other responsibilities, with the combination of reciprocal rights and obligations enabling the enterprise to operate fairly and effectively for mutual benefit.

The social contract in relation to genomics and biobanking can be conceived of as including three interrelated elements: reciprocity, altruism and solidarity, which feed into all levels of public involvement (IPSOS MORI, 2019[104]). However, this model is somewhat western-centric in its approach, and other cultures and societies have different value systems, with a more communitarian approach than a focus on individual autonomy. Solidarity is arguably an important basis for research globally, but solidarity can be manifested differently in different cultural contexts.

5.2.1. Trust and trustworthiness

It long been recognised that public trust is crucial if the promises of personalised, genomics-based medicine are to be realised. People will only actively take up the benefits that will come from data-driven innovation and adoption of AI systems, and will only participate in the research needed for these developments to progress if they have inherent trust in them (Caulfield, Borry and Gottweis, 2014[105]; Chalmers and Nicol, 2004[106]).

Trustworthiness, the quality of being deserving of trust, is an intrinsic ethical value, and is also instrumental in increasing research participation and improving the perception of research by the public (O'Doherty et al., 2011[107]; Kraft et al., 2018[108]). Given the ongoing nature of participant involvement in collaborative platforms, responsible institutions, with ongoing transparency and accountability, beyond the initial informed consent process, with robust institutional ethical oversight, are important to demonstrate trustworthiness (Kraft et al., 2018[108]; Samuel and Farsides, 2018[109]). Kraft and colleagues suggest that in order to cultivate long-term, trusting relationships with patients, institutions should address the role of history and experience on trust, engage with concerns about potential group harm, address cultural values and communication barriers, and integrate patient values and expectations into oversight and governance structures (Kraft et al., 2018[108]).

Issues of data security, privacy and commercial access to data also have important implications for public trust (Dheensa, Lucassen and Fenwick, 2019[110]; Nicol et al., 2016[111]). Protection of privacy, good data security and a clear and transparent governance framework are key in this respect (Critchley, Nicol and McWhirter, 2017[112]; Gaskell et al., 2013[113]; Kaufman et al., 2009[114]). Data sharing and research collaboration may not be a barrier to public trust provided that privacy is protected, there is sufficient ethical oversight and biobanks are shown to successfully contribute to health discoveries. However, collaborations with the private sector and with overseas researchers remain problematic for some participants, and attention must continue to be paid to these aspects.

Commercial involvement in research, and the idea that private organisations that may profit from using biobank resources is one of the most problematic aspects of cross-sector collaboration in terms of public trust (Caulfield, Borry and Gottweis, 2014[105]; Critchley, Nicol and Otlowski, 2015[115]; Garrison et al., 2016[116]; MORI for Wellcome Trust, 2016[117]; Milne et al., 2019[118]). Benefit sharing or returning financial or health benefits back to the biobank, donor or community could possibly help remedy any erosion in trust associated with these concerns (Haddow et al., 2007[119]; Nicol and Critchley, 2012[120]; World Economic Forum (WEF), 2020[121]), but benefit sharing is difficult to implement in practice. There is also however a recognition that the private sector can play a vital role in funding research and translating the research effort into clinical deliverables. There is therefore a need to support ongoing dialogue about the role of research partnerships between public and private sectors.

The role of the public is fundamental and frames much of the discussions of other aspects of this document. Patients, participants and the wider public are inextricably linked with collaborative genomics research and biobanks in three important ways: 1) policy and public funding, 2) data collection, 3) public benefit. Key questions include:

- As genomics moves from research towards commercialisation – how is the public interest accommodated?
- Does the public appropriately or adequately benefit from the research?
- How can we protect the public interest to ensure sustainability of collaborative platforms in the framework of private rights which traditionally govern commercialisation?

Participants who contribute to this research endeavour have traditionally often been viewed as 'donors' who merely allow their sample to be used. However, their role is increasingly becoming 'participants', 'partners', or even 'stakeholders', with recognition of their vital involvement in the research (Goisauf et al., 2019[122]; Gottweis, Gaskell and Starkbaum, 2011[123]; Winickoff, 2007[124]). Participant engagement underpins trust in the research, and helps to maintain levels of participation in the research, as well as improve its relevance and utility (Concannon et al., 2014[125]). However, the level of public and patient involvement (PPI) and community engagement in the governance of collaborative platforms is very variable. Some platforms involve patients and the public directly in their governance; others engage with the public but do not involve the public and patients in governance structures. As one group of social scientists have argued, relationships of trust do not occur spontaneously, but require active management. Genomic research institutions should not seek to manufacture trust in individuals or communities, but rather to implement governance mechanisms such that the biobank deserves trust (O'Doherty et al., 2011[107]). Community engagement may be key to establishing and maintaining the social contract for genomics research. Moreover, respect and community engagement is increasingly seen as an essential component in the ethical conduct of research (Staunton et al., 2018[126]; Tindana et al., 2017[127]).

5.3. Intellectual property considerations

The existence of intellectual property (IP) rights in relation to genomic inventions is controversial, a topic extensive considered in academic literature (Caulfield et al., 2006[128]; Caulfield, Bubela and Murdoch, 2007[129]; Dreyfuss, Nielsen and Nicol, 2018[130]). The nature and extent of the patentability of genomic inventions has been in a state of flux, and there has been considerable oscillation, legal uncertainty and controversy (Dreyfuss, Nielsen and Nicol, 2018[130]; Minn, 2019[131]).

Beyond the controversy around gene sequence patents in current and future genomics research IP is likely to arise on a much broader range of innovations. Creative approaches to the protection of inventions such as novel therapies, diagnostics, biomarkers, and algorithms and AI will likely arise. However, there is ongoing legal uncertainty in relation to the ownership and protection of genomics-related products (Aboy et al., 2019[132]; Cook-Deegan and Heaney, 2010[133]; Liddell, 2018[134]). Despite these uncertainties, business models in biomedical research and health innovation tend to rely on IP rights; indeed, no discussion of product development and commercialisation can ignore IP. Because of the length of the development and regulatory process, IP rights -- specifically patents -- are integral to the translation of research into clinically relevant outcomes in health innovation. IP frameworks become increasingly important as research projects develop and outputs mature.

The initial allocation of ownership depends to some extent on the terms of the IP legislation and case law of the jurisdiction in which the invention or creation first arises, the allocation of property rights is also highly dependent on the terms of the contracts and research agreements through which access to the collaborative platform is granted. Platforms therefore can exercise significant control in relation to IP arising from the platform, and the choice of terms can be controversial.

5.3.1. Ownership and exploitation of intellectual property rights

Key questions about IP in collaborative platforms arise in two areas with distinct ethical, legal, and social implications:

- The management of IP rights across a high diversity of actors, shared investment, and the development of broadly-applicable inventions or platform technologies.
- The responsible use of IP rights downstream of the innovation trajectory and in markets.

Issues around the allocation and licensing of IP rights on downstream innovations implicate questions of the social contract and trust discussed above. The existence of these IP rights, allocation of ownership, and the manner of their exploitation can unsettle expectations as ethical concerns about the role of IP in genomics translational research remain. Failures of transparency and the lack of public discussion have the potential to impact public trust, and ultimately the sustainability of the endeavour. Any IP policy should pay due regard to these debates.

Collaborative platforms normally have the power to limit or control the use of IP rights through contractual provisions in at least two important respects: (i) prohibitions on patenting of upstream technologies which have the potential to inappropriately restrict further research and development in the technological field, and (ii) requirements as to terms of licensing to advance the purpose of the collaborative platform

Restrictions or positive obligations as to the licensing of IP arising from the platform have the potential to advance public access to the technology in question. The imposition of terms and conditions about ownership of existing and future data, materials and IP rights are all common terms in existing access agreements for bioresources (Liddell, Liddicoat and Jordan, 2019[135]).

The extent to which platforms wish to exercise control over IP varies. In most cases, it may be practical and desirable that those accessing the platform would own their own IP, as they are in the best place to develop it appropriately. For example, The Israeli National Biobank for Research (MIDGAM) formally waives any IP rights, or ownership pertaining to inventions based on materials supplied by MIDGAM.

Some collaborative platforms, such as GA4GH, make their IP of this nature open source so that it can be freely used and adapted for the good of the community. As another example, the IP policy of Genomics England is shown in Box 2.

Box 2. Genomics England IP Policy

The Genomics England IP policy builds on the core aims of ensuring patient benefit, supporting research, enabling industrial collaborations and maintaining public trust and confidence. A set of broad IP principles address the need to recognise the level of public investment in the project and the need to address future access of the NHS to inventions and discoveries in any licensing agreements. The NHS institutions contributing samples agreed to assign ownership of the samples and genomic data to Genomics England to manage on behalf of the nation. If that had not been done then the data could have become difficult to manage with many potential rights-holders.

The Participation Agreement also sets out a default position that Genomics England owns the results of non-commercial research. This was intended to ensure that any collaborative research in the trusted research environment would be available to future users. Much of the research is collaborative, across different disease- or method-related 'domains' and often involves additional work to curate, tag and refine the complex genomic and phenotypic data. If users were to withdraw or otherwise restrict the use of their work it would affect the principles of collaboration. Genomics England has stated that it will not seek to own any pre-existing IP (such as outside data bought into the research environment for analysis). It will also assign IP to an institution or other user if they wish to commercialise but will retain an interest to help ensure that the invention is available for future NHS use (at an appropriate cost).

It has been noted that Genomics England's IP position is unusual, and it is possible that the status as a company has led to a perception that it is intending to maximise future value or profit. The future arrangements for IP and ownership of future genomic data is being considered as the NHS Genomic Medicine Service is established from late 2019. The aim will be to continue many of the arrangements pioneered in the 100 000 Genomes project to deliver the stated aim of 500 000 NHS-linked WGS by 2024.

Source: Genomics England

Much of the policy and academic literature in this field is at least notionally committed to the concept of openness (Levin and Leonelli, 2017[136]; Liddell, Liddicoat and Jordan, 2019[135]; OECD, 2015[137]). However, this concept is poorly defined and is used in many different senses. Collaborative platforms should be cautious about employing the terminology of openness without clear and precise definitions. Participants and the public must not be misled by the use of the rhetoric of openness, if in fact controlled access, fees for service, commercial secrecy and patenting will ultimately be employed, because to do so risks undermining public trust, as discussed above. Moreover, notions of openness are relatively unhelpful when platforms move beyond mere access to information to questions of ownership of the innovation developed from that information, where there is unlikely to be any intention or attempt to develop innovative products on an 'open' basis.

Some important attempts to formulate the principles for licensing genomic IP and to seek to harmonise high level policies and practices have been made by key bodies, including funders, academic and professional associations and policymakers (Aymé, Matthijs and Soini, 2008[138]; Department of Health and Human Services (HHS), 2010[139]; National Institutes of Health, 2004[140]; Organisation for Economic Co-operation and Development, 2006[141]; Stanford University, 2007[142]). Related to licensing, initiatives such as patent pools and clearinghouses may assist to make patented technologies more broadly available for research and development (van Overwalle, 2009[97]). Some high profile initiatives of

this nature exist in relation COVID-19, such as the COVID-19 Technology Access Pool (C-TAP) spearheaded by the World Health Organisation.

Beyond voluntary licensing approaches, patent law provisions such as research exceptions and compulsory licenses may also provide important flexibilities in ensuring public access to technologies (Bently, 2011[143]; Hawkins, 2012[144]). Much of the controversy around patents in biomedicine is associated with monopoly provision of healthcare (Hawkins, 2016[145]). Exclusive licences risk foreclosing further research and development of whole areas of technological development, and IP guidelines and policies therefore caution against exclusive licences unless unavoidable. However, it can be very difficult at an early stage in the development of a technology to know the optimal means of its development. Exclusive licences may give rise to problematic monopolies, or they may enable efficient and effective development of technologies. Moreover, the ultimate downstream impact of licensing policies and conditions often remains unclear, at least partly due to the long time lag between the imposition of the conditions, and the ultimate clinical use of the outcome of the research. Flexibility, and periodic reviews of licensing terms can help to ameliorate problematic effects. The complexities associated with licensing in this field are great, and further research to assess and determine the effects of such policies is vital.

Collaborative platforms have great potential to use governance and contractual access arrangements to seek innovative means to both advance innovation, and the public interest. Although IP rights are private rights, in this field, there is a significant public interest in access to innovation. Attempts by collaborative platforms to use such conditions for IP rights to advance the purpose of the platform should be encouraged. Collaborative platforms, serving as the interface between the public and private sector in genomic medicine, are well placed to advance the field of IP in this way, and they can employ innovative and creative policies. Public funders are increasingly recognising their important role in ensuring access to the fruits of their funded research (Wellcome Trust[146]), which may help to aid consensus building efforts. Additionally, and importantly, collaborative platforms are also well placed to assess and monitor the impact of their policies on downstream access and availability. Moreover, as the gatekeeper to information, they have sufficient bargaining power to implement these principles – and they have the imperative to do so, to safeguard the participants, and further develop and advance existing governance frameworks.

6. Strengthening governance frameworks

Collaborative platforms invest significant time and effort in the terms of their governance structures. Governance involves the decision making processes and procedures by which people organise themselves to achieve defined goals. Governance can have multiple layers, and there is both internal governance (for example of individual projects or companies) and external governance (legal frameworks, regulatory bodies). Governance is key to the successful operation of collaborative platforms in genomics and biobanking, and can be an effective means to address the challenges and realise the opportunities in novel therapies, vaccines and diagnostics.

Many of the key governance questions for collaborative platforms are similar to the questions which have been debated and addressed in the literature around genomics research and health innovation for the past ten years (Goisauf et al., 2019[122]; Phillips et al., 2020[147]). However, the nature of collaborative platforms in the current moment raises important and intractable governance issues in certain respects, which are deserving of particular consideration: the responsible implementation of novel digital technologies, the balancing of diverse interests of stakeholders, the need to promote equity and sustainability,

and coping with an increasing fragmentation of international governance. Governance structures – including accountable frameworks, principles, oversight boards, committees and participatory processes – can help mediate tensions, make decisions in accountable ways, and foster social innovation (OECD, 2010[148]). For example, the OECD Guidelines for Multinational Enterprises (2011[149]) states that "When granting licenses for the use of intellectual property rights or when otherwise transferring technology, do so on reasonable terms and conditions and in a manner that contributes to the long term sustainable development prospects of the host country."

6.1. Addressing novel digital technologies

As the novel digital technologies discussed above are applied to genomics data, and genomics data is linked to data available from other sources, such as through platforms such as Facebook, an extra layer of complexity arises in relation to the ethical issues (Nuffield Council on Bioethics, 2019[150]; Williams et al., 2018[58]), and new actors not traditionally included in existing governance frameworks begin to be important stakeholders.

Furthermore, with the greater use of AI and machine learning approaches in research, the existing focus on the accountability of data stewards for misuse of the data will be less effective to address harms. Instead, the difficulty will be to ensure that outcomes of data use are fair and ethical. Governance frameworks therefore need to continue to adapt in this respect. The harms and benefits of new technologies should be monitored, and governance frameworks adapted accordingly.

6.2. Balancing diverse stakeholder interests

Collaborative platforms feature interests that must be balanced. First, in the healthcare context the security and privacy of participant and patient data should be at the heart of governance frameworks but the privacy and security of data must be balanced against the need for data fluidity and secondary use (Ballantyne and Schaefer, 2020[32]). Second, local or national interests, e.g. those of healthcare or universities may conflict with global interests, e.g. those of international researchers or pharmaceutical companies. Third, the interests of sharing findings or benefits with the original data generator may conflict with downstream revenues and indeed there is evidence that the public is concerned about the need to prioritise public benefit over profit in biomedical research (MORI for Wellcome Trust, 2016[117]). Governance frameworks must face the challenge of balancing the potential harms to privacy with the potential benefits of openness.

Some governance structures are attempting to ensure that ethical and lawful research is supported through accountable decision making. The Genomics England case and others show how governance is starting to address these tensions in a transparent fashion, enabling opportunities arising from the changing nature of collaborative platforms. This both protects the integrity of the research community and also has the effect of promoting public confidence and trust. Good governance of the translational research process is especially important in relation to the involvement of commercial entities: indeed, research demonstrates that enhanced involvement of the public and data subjects in the governance of health data inspires trust and confidence in health data access (Bell, 2020[151]; Nuffield Council on Bioethics, 2015[85]).

6.3. Mitigating fragmentation and lack of interoperability

Conflicting and divergent international regulations can represent important and time-consuming obstacles to smooth data flows, and compliance with data protection regulation is often time consuming for collaborative platforms. In a large collaboration, involving public and private entities, perhaps across a number of countries, there is often great complexity in the governing frameworks themselves, with multiple regulatory regimes overlapping, each with their own different objectives (Kaye et al., 2012[83]). Attention to these questions, and whether regulation appropriately balances the importance of privacy protection against the risks and the drawbacks of compliance must continue.

Deeper integration could be achieved through standardisation and regulatory alignment. Here collaborative platforms could serve as a model for the development and testing of processes and standards in, for example, information technology networking, quality management, pubic deliberation, commercial strategies, education and training, and approaches to responsible innovation (Chalmers et al., 2016[152]). Increased interoperability of the governance frameworks themselves can help to facilitate collaboration. Important efforts to provide guidance and standards to encourage interoperability in international governance are ongoing. The policy framework and work programme developed by GA4GH signal a promising way forward in this regard (see Box 3).

Box 3. GA4GH's Framework for Responsible Sharing of Genomic and Health Related Data

GA4GH's Framework for Responsible Sharing of Genomic and Health Related Data, is founded on the human right to benefit from scientific advancement and its benefits and from the right to be recognised for one's contributions to science (Knoppers, 2014[46]). The Framework seeks to provide a principled and practical framework for responsible sharing of genomic and health-related data. Its primary goals are to:

- Protect and promote the welfare, rights, and interests of individuals from around the world in genomic and health-related data sharing, particularly those who contribute their data for biomedical research.
- Complement laws and regulations on privacy and personal data protection, as well as policies and codes of conduct for the ethical governance of research.
- Foster responsible data sharing and oversight of research data systems.
- Establish a framework for greater international data sharing, collaboration and good governance.
- Serve as a dynamic instrument that can respond to future developments in the science, technology, and practices of genomic and health-related data sharing.
- Serve as a tool for the evaluation of responsible research by research ethics committees and data access committees.
- Provide overarching principles to be respected in developing legally-binding tools such as data access agreements.

Source: (GA4GH, 2014[153])

6.4. Connecting up privacy and property

The approach to strengthen public interest in innovation governance and in commercialisation of IP rights law needs to be better reconciled in order to deliver sustainable public health systems and economic well-being. At this point, there is too often a disconnect between governance frameworks built on public interest, and the private rights of IP law, with narrower conception of the public interest.

The main focus of much of existing governance is to protect the participants who contribute data through privacy and data protection laws. Therefore, governance often focuses around the data held by and flowing through the platform, triggering data protection laws of the jurisdictions involved in the platform. Governance frameworks tend to focus on the protection of the individual, and bioethics scholarship has comparatively ignored property in favour of focusing on consent, ethical governance and privacy (Winickoff, 2007[124]). The ways in which the private monopoly rights of the patent holder clash with the public interest, in relation to, for example, access to medicine, is of ongoing concern (Pogge, 2011[154]).

Transparent and inclusive governance frameworks can help define terms of the social contract, resolve tensions therein, and promote the value of collaborative platforms in genomics. If handled well, data sharing could be viewed favourably, as serving the wider public interest. In sum, highest ethical standards, best practice privacy protections, benefit sharing arrangements, transparency, accountability and openness are the core elements of good governance in this space.

7. References

Aboy, M. et al. (2019), "Mayo's impact on patent applications related to biotechnology, diagnostics and personalized medicine", *Nature Biotechnology*, Vol. 37/5, pp. 513-518, http://dx.doi.org/10.1038/s41587-019-0111-5. [132]

Aitken, M. (2020), *Understanding the Global Landscape of Genomics Initiatives, Progress and Promise*, IQVIA Inc., https://www.iqvia.com/insights/the-iqvia-institute/reports/understanding-the-global-landscape-of-genomic-initiatives (accessed on 12 August 2020). [6]

Andry, C. et al. (2017), "Biobanking—Budgets and the Role of Pathology Biobanks in Precision Medicine", *Academic Pathology*, Vol. 4, http://dx.doi.org/10.1177/2374289517702924. [31]

Aymé, S., G. Matthijs and S. Soini (2008), "Patenting and licensing in genetic testing. Recommendations of the European Society of Human Genetics", *European Journal of Human Genetics*, Vol. 16/4, pp. 405-411, http://dx.doi.org/10.1038/sj.ejhg.5201929. [138]

Ballantyne, A. and G. Schaefer (2020), "Public interest in health data research: laying out the conceptual groundwork", *Journal of medical ethics*, Vol. 46/9, pp. 610-616, http://dx.doi.org/10.1136/medethics-2020-106152. [32]

Beller, D. (n.d.), *Biobank - Maccabitech*, https://www.maccabitech.com/biobank/ (accessed on 23 September 2020). [101]

Bell, J. (2020), "Governing Commercial Access to Health Data for Public Benefit: Charity Law Solutions", *Medical law review*, Vol. 28/2, pp. 247-269, http://dx.doi.org/10.1093/medlaw/fwz022. [151]

Belmont, J. et al. (2003), "The international HapMap project", *Nature*, Vol. 426/6968, pp. 789-796, http://dx.doi.org/10.1038/nature02168. [39]

Bently, L. (2011), "Exclusions from Patentability and Exceptions to Patentees' Rights: Taking Exceptions Seriously", *Current Legal Problems*, Vol. 64/1, pp. 315-347, http://dx.doi.org/10.1093/clp/cur011. [143]

Birney, E., J. Vamathevan and P. Goodhand (2017), "Genomics in healthcare: GA4GH looks to 2022", *bioRxiv 203554*, pp. http://dx.doi.org/10.1101/203554, http://dx.doi.org/10.1101/203554. [75]

Bombard, Y. and R. Hayeems (2020), "How digital tools can advance quality and equity in genomic medicine", *Nature Reviews Genetics*, Vol. 21/9, pp. 505-506, http://dx.doi.org/10.1038/s41576-020-0260-x. [49]

Bonomi, L., Y. Huang and L. Ohno-Machado (2020), *Privacy challenges and research opportunities for genomic data sharing*, Nature Research, http://dx.doi.org/10.1038/s41588-020-0651-0. [76]

Burke, W. and B. Psaty (2007), *Personalized medicine in the era of genomics*, American Medical Association, http://dx.doi.org/10.1001/jama.298.14.1682. [3]

Bycroft, C. et al. (2018), "The UK Biobank resource with deep phenotyping and genomic data", *Nature*, Vol. 562/7726, pp. 203-209, http://dx.doi.org/10.1038/s41586-018-0579-z. [20]

Campbell, P. et al. (2020), "Pan-cancer analysis of whole genomes", *Nature*, Vol. 578/7793, pp. 82-93, http://dx.doi.org/10.1038/s41586-020-1969-6. [40]

Carey, D. et al. (2016), "The Geisinger MyCode community health initiative: An electronic health record-linked biobank for precision medicine research", *Genetics in Medicine*, Vol. 18/9, pp. 906-913, http://dx.doi.org/10.1038/gim.2015.187. [48]

Caulfield, T., P. Borry and H. Gottweis (2014), "Industry involvement in publicly funded biobanks", *Nature Reviews Genetics*, Vol. 15/4, p. 220, http://dx.doi.org/10.1038/nrg3704. [105]

Caulfield, T., T. Bubela and C. Murdoch (2007), "Myriad and the mass media: The covering of a gene patent controversy", *Genetics in Medicine*, Vol. 9/12, pp. 850-855, http://dx.doi.org/10.1097/GIM.0b013e31815bf965. [129]

Caulfield, T. et al. (2006), *Evidence and anecdotes: An analysis of human gene patenting controversies*, NIH Public Access, http://dx.doi.org/10.1038/nbt0906-1091. [128]

Caulfield, T. and B. Murdoch (2017), "Genes, cells, and biobanks: Yes, there's still a consent problem", *PLoS Biology*, Vol. 15/7, http://dx.doi.org/10.1371/journal.pbio.2002654. [16]

Chalmers, D. and D. Nicol (2004), *Commercialisation of biotechnology: Public trust and research*, http://dx.doi.org/10.1504/IJBT.2004.004806. [106]

Chalmers, D. et al. (2016), *Has the biobank bubble burst? Withstanding the challenges for sustainable biobanking in the digital era Donna Dickenson, Sandra Soo-Jin Lee, and Michael Morrison*, BioMed Central Ltd., http://dx.doi.org/10.1186/s12910-016-0124-2. [152]

Ciaburri, M., M. Napolitano and E. Bravo (2017), "Business Planning in Biobanking: How to Implement a Tool for Sustainability", *Biopreservation and Biobanking*, Vol. 15/1, pp. 46-56, http://dx.doi.org/10.1089/bio.2016.0045. [28]

Clayton, E. et al. (2019), "The law of genetic privacy: Applications, implications, and limitations", *Journal of Law and the Biosciences*, Vol. 6/1, pp. 1-36, http://dx.doi.org/10.1093/jlb/lsz007. [79]

Concannon, T. et al. (2014), "A Systematic Review of Stakeholder Engagement in Comparative Effectiveness and Patient-Centered Outcomes Research", *Journal of General Internal Medicine*, Vol. 29/12, pp. 1692-1701, http://dx.doi.org/10.1007/s11606-014-2878-x. [125]

Cook-Deegan, R., R. Ankeny and K. Jones (2017), "Sharing Data to Build a Medical Information Commons: From Bermuda to the Global Alliance", *Annu. Rev. Genom. Hum. Genet.*, Vol. 18, pp. 389–415, https://doi.org/10.1146/annurev-genom-083115- 022515. [37]

Cook-Deegan, R. and C. Heaney (2010), "Patents in Genomics and Human Genetics", *Annual Review of Genomics and Human Genetics*, Vol. 11/1, pp. 383-425, http://dx.doi.org/10.1146/annurev-genom-082509-141811. [133]

Corpas, M. et al. (2018), "A FAIR guide for data providers to maximise sharing of human genomic data", *PLOS Computational Biology*, Vol. 14/3, p. e1005873, http://dx.doi.org/10.1371/journal.pcbi.1005873. [96]

Critchley, C., D. Nicol and R. McWhirter (2017), "Identifying public expectations of genetic biobanks.", *Public understanding of science (Bristol, England)*, Vol. 26/6, pp. 671-687, http://dx.doi.org/10.1177/0963662515623925. [112]

Critchley, C., D. Nicol and M. Otlowski (2015), "The impact of commercialisation and genetic data sharing arrangements on public trust and the intention to participate in biobank research", *Public Health Genomics*, Vol. 18/3, pp. 160-172, http://dx.doi.org/10.1159/000375441. [115]

de Vries, J. et al. (2015), "The H3Africa policy framework: Negotiating fairness in genomics", in *Getting to Good: Research Integrity in the Biomedical Sciences*, Springer International Publishing, http://dx.doi.org/10.1016/j.tig.2014.11.004. [95]

Department of Health and Human Services (HHS) (2010), *Gene Patents and Licensing Practices and Their Impact on Patient Access to Genetic Tests Report of the Secretary's Advisory Committee on Genetics, Health, and Society*, Department of Health and Human Services (HHS), http://oba.od.nih.gov/oba/SACGHS/reports/SACGHS_patents_report_2010.pdf (accessed on 11 September 2020). [139]

Dheensa, S., A. Lucassen and A. Fenwick (2019), "Fostering trust in healthcare: Participants' experiences, views, and concerns about the 100,000 genomes project", *European Journal of Medical Genetics*, Vol. 62/5, pp. 335-341, http://dx.doi.org/10.1016/j.ejmg.2018.11.024. [110]

Dreyfuss, R., J. Nielsen and D. Nicol (2018), *Patenting nature-a comparative perspective*, Oxford University Press, http://dx.doi.org/10.1093/jlb/lsy021. [130]

Dubow, T. and S. Marjanovic (2016), *Population-scale sequencing and the future of genomic medicine: Learning from past and present efforts*, RAND Corporation, http://dx.doi.org/10.7249/rr1520. [4]

Dyke, S., E. Dove and B. Knoppers (2016), "Sharing health-related data: A privacy test?", *npj Genomic Medicine*, Vol. 1/1, p. 1, http://dx.doi.org/10.1038/npjgenmed.2016.24. [78]

Emanuel, E. and R. Wachter (2019), *Artificial Intelligence in Health Care: Will the Value Match the Hype?*, American Medical Association, http://dx.doi.org/10.1001/jama.2019.4914. [71]

Erlich, Y. et al. (2018), "Identity inference of genomic data using long-range familial searches", *Science*, Vol. 362/6415, pp. 690-694, http://dx.doi.org/10.1126/science.aau4832. [87]

GA4GH (2014), *GA4GH Framework for Responsible Sharing of Genomic and Health Related Data*, https://www.ga4gh.org/genomic-data-toolkit/regulatory-ethics-toolkit/framework-for-responsible-sharing-of-genomic-and-health-related-data/ (accessed on 6 October 2020). [153]

Garrison, N. et al. (2016), *A systematic literature review of individuals' perspectives on broad consent and data sharing in the United States*, Nature Publishing Group, http://dx.doi.org/10.1038/gim.2015.138. [116]

Gaskell, G. et al. (2013), "Publics and biobanks: Pan-European diversity and the challenge of responsible innovation", *European Journal of Human Genetics*, Vol. 21/1, pp. 14-20, http://dx.doi.org/10.1038/ejhg.2012.104. [113]

Gibbons, S. (2009), "Regulating biobanks: a twelve-point typological tool.", *Medical law review*, Vol. 17/3, pp. 313-46, http://dx.doi.org/10.1093/medlaw/fwp016. [92]

Ginsburg, G. and K. Phillips (2018), "Precision medicine: From science to value", *Health Affairs*, Vol. 37/5, pp. 694-701, http://dx.doi.org/10.1377/hlthaff.2017.1624. [102]

Goisauf, M. et al. (2019), "Data in question: A survey of European biobank professionals on ethical, legal and societal challenges of biobank research", *PLoS ONE*, Vol. 14/9, p. e0221496, http://dx.doi.org/10.1371/journal.pone.0221496. [122]

Gottweis, H., G. Gaskell and J. Starkbaum (2011), *Connecting the public with biobank research: Reciprocity matters*, Nature Publishing Group, http://dx.doi.org/10.1038/nrg3083. [123]

Gurovich, Y. et al. (2019), "Identifying facial phenotypes of genetic disorders using deep learning", *Nature Medicine*, Vol. 25/1, pp. 60-64, http://dx.doi.org/10.1038/s41591-018-0279-0. [62]

Gymrek, M. et al. (2013), "Identifying personal genomes by surname inference", *Science*, Vol. 339/6117, pp. 321-324, http://dx.doi.org/10.1126/science.1229566. [90]

Haddow, G. et al. (2007), "Tackling community concerns about commercialisation and genetic research: A modest interdisciplinary proposal", *Social Science and Medicine*, Vol. 64/2, pp. 272-282, http://dx.doi.org/10.1016/j.socscimed.2006.08.028. [119]

Hawkins, N. (2016), "A red herring – Invalidity of human gene sequence patents", *European Intellectual Property Review*, https://ore.exeter.ac.uk/repository/handle/10871/19146 (accessed on 11 September 2020). [145]

Hawkins, N. (2012), "An Exception to Infringement for Genetic Testing - Addressing Patient Access and Divergence Between Law and Practice", *International Review of Intellectual Property and Competition Law*, Vol. 34/6, https://www.researchgate.net/publication/289195583_An_Exception_to_Infringement_for_Genetic_Testing_-_Addressing_Patient_Access_and_Divergence_Between_Law_and_Practice (accessed on 11 September 2020). [144]

Heeney, C. et al. (2011), "Assessing the privacy risks of data sharing in genomics.", *Public health genomics*, Vol. 14/1, pp. 17-25, http://dx.doi.org/10.1159/000294150. [81]

Hindorff, L., V. Bonham and L. Ohno-Machado (2018), "Enhancing diversity to reduce health information disparities and build an evidence base for genomic medicine", *Per. Med* 5, pp. 403-412, http://dx.doi.org/10.2217/pme-2018-0037. [25]

Hofman, P. et al. (2014), *Public-private relationships in biobanking: A still underestimated key component of open innovation*, Virchows Arch, http://dx.doi.org/10.1007/s00428-013-1524-z. [33]

Homer, N. et al. (2008), "Resolving Individuals Contributing Trace Amounts of DNA to Highly Complex Mixtures Using High-Density SNP Genotyping Microarrays", *PLoS Genetics*, Vol. 4/8, p. e1000167, http://dx.doi.org/10.1371/journal.pgen.1000167. [88]

IPSOS MORI (2019), *A public dialogue on genomic medicine : time for a new social contract ?*. [104]

Joly, Y. et al. (2020), "Establishing the International Genetic Discrimination Observatory", *Nature Genetics*, Vol. 52, http://dx.doi.org/10.1038/s41588-020-0606-5. [26]

Kaufman, D. et al. (2009), "Public Opinion about the Importance of Privacy in Biobank Research", *American Journal of Human Genetics*, Vol. 85/5, pp. 643-654, http://dx.doi.org/10.1016/j.ajhg.2009.10.002. [114]

Kaye, J. et al. (2012), *Governing Biobanks : Understanding the Interplay between Law and Practice*, Hart Publishing, http://dx.doi.org/10.5040/9781472565846. [83]

Kaye, J. and N. Hawkins (2014), "Data sharing policy design for consortia: challenges for sustainability.", *Genome medicine*, Vol. 6/1, p. 4, http://dx.doi.org/10.1186/gm523. [82]

Khoury, M. et al. (2007), "The continuum of translation research in genomic medicine: how can we accelerate the appropriate integration of human genome discoveries into health care and disease prevention?", *Genet Med*, Journal Article
Review
United States
official journal of the American College of Medical Genetics, pp. 665-674, http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=18073579. [98]

Kinkorová, J. (2016), "Biobanks in the era of personalized medicine: objectives, challenges, and innovation", http://dx.doi.org/10.1186/s13167-016-0053-7. [11]

Kinkorová, J. and O. Topolčan (2018), "Biobanks in Horizon 2020: sustainability and attractive perspectives", http://dx.doi.org/10.1007/s13167-018-0153-7. [14]

Knoppers, B. (2014), "Framework for responsible sharing of genomic and health-related data", *HUGO Journal*, Vol. 8/1, pp. 1-6, http://dx.doi.org/10.1186/s11568-014-0003-1. [46]

Knoppers, B. and Y. Joly (2018), "Introduction: the why and whither of genomic data sharing", *Human Genetics*, Vol. 137/8, pp. 569-574, http://dx.doi.org/10.1007/s00439-018-1923-y. [84]

Koumakis, L. (2020), *Deep learning models in genomics; are we there yet?*, Elsevier B.V., http://dx.doi.org/10.1016/j.csbj.2020.06.017. [60]

Kraft, S. et al. (2018), "Beyond Consent: Building Trusting Relationships With Diverse Populations in Precision Medicine Research", *American Journal of Bioethics*, Vol. 18/4, pp. 3-20, http://dx.doi.org/10.1080/15265161.2018.1431322. [108]

Langmead, B. and A. Nellore (2018), *Cloud computing for genomic data analysis and collaboration*, Nature Publishing Group, http://dx.doi.org/10.1038/nrg.2017.113. [56]

Leitsalu, L. et al. (2015), *Linking a population biobank with national health registries—The Estonian experience*, MDPI AG, http://dx.doi.org/10.3390/jpm5020096. [44]

Levin, N. and S. Leonelli (2017), "How Does One "Open" Science? Questions of Value in Biological Research", *Science, Technology, & Human Values*, Vol. 42/2, pp. 280-305, http://dx.doi.org/10.1177/0162243916672071. [136]

Liddell, K. (2018), "Realising Genomic Medicine: Intellectual Property Issues", *SSRN Electronic Journal*, http://dx.doi.org/10.2139/ssrn.3102224. [134]

Liddell, K., J. Liddicoat and M. Jordan (2019), "IP policies for large bioresources: the fiction, fantasy, and future of openness", in Minssen, T., J. Herrmann and J. Schovsbo (eds.), *Global Genes, Local Concerns*, Edward Elgar, http://dx.doi.org/10.4337/9781788116190.00023. [135]

Livesey, F. (2019), *Modelling the possible returns to the NHS from private sector use of the 100K genomes database*, Bennett Institute for Public Policy, University of Cambridge, http://www.bennettinstitute.cam.ac.uk (accessed on 14 August 2020). [29]

Lucassen, A., J. Montgomery and M. Parker (2016), "Chapter 16. Ethics and the social contract for genomics in the NHS", in *Generation Genome: Annual Report of the Chief Medical Officer 2016*, Department of Health. [103]

Mallappallil, M. et al. (2020), "A review of big data and medical research", *SAGE Open Medicine*, Vol. 8, p. 205031212093483, http://dx.doi.org/10.1177/2050312120934839. [21]

Manolio, T. et al. (2015), *Global implementation of genomic medicine: We are not alone*, American Association for the Advancement of Science, http://dx.doi.org/10.1126/scitranslmed.aab0194. [41]

Matzke, L. et al. (2016), "Fundamental Considerations for Biobank Legacy Planning", *Biopreservation and Biobanking*, Vol. 14/2, pp. 99-106, http://dx.doi.org/10.1089/bio.2015.0073. [35]

Maxson Jones, K., R. Ankeny and R. Cook-Deegan (2018), "The Bermuda Triangle: The Pragmatics, Policies, and Principles for Data Sharing in the History of the Human Genome Project", *Journal of the History of Biology*, pp. 1-113, http://dx.doi.org/10.1007/s10739-018-9538-7. [38]

McGuire, A. et al. (2011), "Ethical and practical challenges of sharing data from genome-wide association studies: The eMERGE Consortium experience", *Genome Research*, Vol. 21/7, pp. 1001-1007, http://dx.doi.org/10.1101/gr.120329.111. [89]

Mell, P. and T. Grance (2011), *The NIST Definition of Cloud Computing (Special Publication 800-145)*, National Institute of Standards and Technology, US Department of Commerce, http://dx.doi.org/10.6028/NIST.SP.800-145. [53]

Milne, R. et al. (2019), "Trust in genomic data sharing among members of the general public in the UK, USA, Canada and Australia", *Human Genetics*, Vol. 138/11-12, pp. 1237-1246, http://dx.doi.org/10.1007/s00439-019-02062-0. [118]

Minari, J., K. Brothers and M. Morrison (2018), "Tensions in ethics and policy created by national precision medicine programs", *Human Genomics*, Vol. 12/1, http://dx.doi.org/10.1186/s40246-018-0151-9. [23]

Minn, M. (2019), "Patenting of Genetic Research in Europe and the U.S.: A Questionable Future for Diagnostic Methods and Personalized Medicines", *Biotechnology Law Report*, Vol. 38/2, pp. 92-115, http://dx.doi.org/10.1089/blr.2019.29108.mm. [131]

Mitchell, C. et al. (2020), *The GDPR and genomic data*, PHG Foundation, http://www.phgfoundation.org (accessed on 9 September 2020). [77]

MORI for Wellcome Trust, I. (2016), *The One-Way Mirror: Public attitudes to commercial access to health data*, Ipsos MORI Social Research Institute, http://www.ipsos-mori.com/terms.http://www.ipsos-mori.com/terms.http://www.ipsos-mori.com/terms.http://www.ipsos-mori.com/terms. (accessed on 10 September 2020). [117]

Morley, J. and L. Floridi (2020), *An ethically mindful approach to AI for health care*, Lancet Publishing Group, http://dx.doi.org/10.1016/S0140-6736(19)32975-7. [67]

Müller, H. et al. (2020), *Biobanks for life sciences and personalized medicine: importance of standardization, biosafety, biosecurity, and data management*, Elsevier Ltd, http://dx.doi.org/10.1016/j.copbio.2019.12.004. [7]

Murray, M. et al. (2020), "COVID-19 outcomes and the human genome", *Genetics in Medicine*, Vol. 22/7, pp. 1175-1177, http://dx.doi.org/10.1038/s41436-020-0832-3. [42]

National Institutes of Health, P. (2004), *Best Practices for the Licensing of Genomic Inventions*, National Institutes of Health, Public Health Service, Health and Human Services (HHS), https://www.federalregister.gov/documents/2004/11/19/04-25671/best-practices-for-the-licensing-of-genomic-inventions (accessed on 11 September 2020). [140]

Nicol, D. and C. Critchley (2012), "Benefit sharing and biobanking in Australia.", *Public understanding of science (Bristol, England)*, Vol. 21/5, pp. 534-55, http://dx.doi.org/10.1177/0963662511402425. [120]

Nicol, D. et al. (2016), "Understanding public reactions to commercialization of biobanks and use of biobank resources", *Social Science and Medicine*, Vol. 162, pp. 79-87, http://dx.doi.org/10.1016/j.socscimed.2016.06.028. [111]

Nuffield Council on Bioethics (2019), *Artificial intelligence (AI) in healthcare and research*, Nuffield Council on Bioethics. [150]

Nuffield Council on Bioethics (2015), *The collection, linking and use of data in biomedical research and health care: ethical issues*, Nuffield Council on Bioethics, https://www.nuffieldbioethics.org/wp-content/uploads/Biodata-a-guide-to-the-report-PDF.pdf (accessed on 9 September 2020). [85]

O'Doherty, K. et al. (2011), "From consent to institutions: Designing adaptive governance for genomic biobanks", *Social Science and Medicine*, Vol. 73/3, pp. 367-374, http://dx.doi.org/10.1016/j.socscimed.2011.05.046. [107]

OECD (2020), *The Digitalisation of Science, Technology and Innovation: Key Developments and Policies*, OECD Publishing, Paris, https://dx.doi.org/10.1787/b9e4a2c0-en. [51]

OECD (2019), *Aid for Civil Society Organisations*, OECD, https://www.oecd.org/dac/financing-sustainable-development/development-finance-topics/Aid-for-CSOs-2019.pdf. [156]

OECD (2019), *Artificial Intelligence in Society*, OECD Publishing, Paris, https://dx.doi.org/10.1787/eedfee77-en. [59]

OECD (2019), *Enhancing Access to and Sharing of Data: Reconciling Risks and Benefits for Data Re-use across Societies*, OECD Publishing, Paris, https://dx.doi.org/10.1787/276aaca8-en. [50]

OECD (2019), *Estonia: Country Health Profile 2019*, State of Health in the EU, OECD Publishing, Paris/European Observatory on Health Systems and Policies, Brussels, https://dx.doi.org/10.1787/0b94102e-en. [45]

OECD (2019), *Health in the 21st Century: Putting Data to Work for Stronger Health Systems*, OECD Health Policy Studies, OECD Publishing, Paris, https://dx.doi.org/10.1787/e3b23f8e-en. [1]

OECD (2019), *Measuring the Digital Transformation* (Summary), OECD Publishing, Paris, https://dx.doi.org/10.1787/37044f2e-en. [52]

OECD (2019), *Recommendation of the Council on Artificial Intelligence, OECD/LEGAL/0449*, OECD. [15]

OECD (2019), *Recommendation of the Council on Health Data Governance, OECD/LEGAL/0433*, OECD, http://legalinstruments.oecd.org (accessed on 15 August 2020). [13]

OECD (2017), *Digital platforms for facilitating access to research infrastructures*, OECD Publishing, Paris, https://dx.doi.org/10.1787/8288d208-en. [36]

OECD (2017), "Strengthening the effectiveness and sustainability of international research infrastructures", *OECD Science, Technology and Industry Policy Papers*, No. 48, OECD Publishing, Paris, https://dx.doi.org/10.1787/fa11a0e0-en. [100]

OECD (2015), "Making Open Science a Reality", *OECD Science, Technology and Industry Policy Papers*, No. 25, OECD Publishing, Paris, https://dx.doi.org/10.1787/5jrs2f963zs1-en. [137]

OECD (2013), *The OECD Privacy Framework 2013*, http://www.oecd.org/sti/ieconomy/oecd_privacy_framework.pdf (accessed on 2 May 2019). [91]

OECD (2011), *OECD Guidelines for Multinational Enterprises, 2011 Edition*, OECD Publishing, Paris, https://dx.doi.org/10.1787/9789264115415-en. [149]

OECD (2010), "Addressing Global and Social Challenges through Innovation", in *The OECD Innovation Strategy: Getting a Head Start on Tomorrow*, OECD Publishing, Paris, https://dx.doi.org/10.1787/9789264083479-8-en. [148]

OECD (2010), *Civil Society and Aid Effectiveness: Findings, Recommendations and Good Practice*, Better Aid, OECD Publishing, Paris, https://dx.doi.org/10.1787/9789264056435-en. [155]

OECD (2009), "Recommendation of the Council on Human Biobanks and Genetic Research Databases", https://legalinstruments.oecd.org/en/. [10]

OECD (2007), *OECD Best Practice Guidelines for Biological Resource Centres*, OECD. [9]

OECD (2001), *Biological Resource Centres: Underpinning the Future of Life Sciences and Biotechnology*, OECD Publishing, Paris, https://dx.doi.org/10.1787/9789264193550-en. [8]

Organisation for Economic Co-operation and Development (2006), *Guidelines for the Licensing of Genetic Inventions*, Organisation for Economic Co-operation and Development Publications. [141]

Paranjape, K., M. Schinkel and P. Nanayakkara (2020), "Mainstreaming Personalized Healthcare-Transforming Healthcare through new era of Artificial Intelligence.", *IEEE journal of biomedical and health informatics*, http://dx.doi.org/10.1109/JBHI.2020.2970807. [63]

Paul, S., A. Gade and S. Mallipeddi (2017), "The State of Cloud-Based Biospecimen and Biobank Data Management Tools", *Biopreservation and Biobanking*, Vol. 15/2, pp. 169-172, http://dx.doi.org/10.1089/bio.2017.0019. [54]

PHG Foundation, U. (n.d.), *Black box medicine and transparency*, https://www.phgfoundation.org/research/black-box-medicine-and-transparency (accessed on 8 September 2020). [68]

Phillips, M. et al. (2020), *Genomics: data sharing needs an international code of conduct*, Nature Research, http://dx.doi.org/10.1038/d41586-020-00082-9. [147]

Pogge, T. (2011), "The health impact fund: How to make new medicines accessible to all", in *Global Health and Global Health Ethics*, Cambridge University Press, http://dx.doi.org/10.1017/CBO9780511984792.021. [154]

Pormeister, K. (2018), "Genetic research and applicable law: The intra-EU conflict of laws as a regulatory challenge to cross-border genetic research", *Journal of Law and the Biosciences*, Vol. 5/3, pp. 706-723, http://dx.doi.org/10.1093/jlb/lsy023. [80]

Price, W., S. Gerke and I. Cohen (2019), *Potential Liability for Physicians Using Artificial Intelligence*, American Medical Association, http://dx.doi.org/10.1001/jama.2019.15064. [72]

Rao, A. et al. (2019), "Critical Financial Challenges for Biobanking: Report of a National Cancer Institute Study", *Biopreservation and Biobanking*, Vol. 17/2, pp. 129-138, http://dx.doi.org/10.1089/bio.2018.0069. [30]

Raza, S. (2020), *Artificial intelligence for genomic medicine*, http://www.phgfoundation.org (accessed on 26 May 2020). [57]

Rehm, H. (2017), *Evolving health care through personal genomics*, Nature Publishing Group, http://dx.doi.org/10.1038/nrg.2016.162. [2]

Robinson, P. (2012), "Deep phenotyping for precision medicine", *Human Mutation*, Vol. 33/5, pp. 777-780, http://dx.doi.org/10.1002/humu.22080. [17]

Samuel, G. and B. Farsides (2018), "Public trust and 'ethics review' as a commodity: the case of Genomics England Limited and the UK's 100,000 genomes project", *Medicine, Health Care and Philosophy*, Vol. 21/2, pp. 159-168, http://dx.doi.org/10.1007/s11019-017-9810-1. [109]

Schäfer, A. and R. Baric (2020), "Epigenetic Landscape during Coronavirus Infection", *Pathogens*, Vol. 6/8, http://dx.doi.org/10.3390/pathogens6010008. [43]

Schwarz, U., M. Gulilat and R. Kim (2019), "The role of next-generation sequencing in pharmacogenetics and pharmacogenomics", *Cold Spring Harbor Perspectives in Medicine*, Vol. 9/2, p. a033027, http://dx.doi.org/10.1101/cshperspect.a033027. [73]

Sergey I. Nikolenko (2019), "Synthetic Data for Deep Learning", https://arxiv.org/pdf/1909.11512.pdf (accessed on 8 September 2020). [64]

Stanford University, U. (2007), "Nine Points to Consider in Licensing University Technology", http://news.stanford.edu/news/2007/march7/gifs/whitepaper.pdf (accessed on 11 September 2020). [142]

Stark, Z. et al. (2019), "Integrating Genomics into Healthcare: A Global Responsibility", *The American Journal of Human Genetics*, Vol. 104, pp. 13-20, http://dx.doi.org/10.1016/j.ajhg.2018.11.014. [5]

Staunton, C. and J. de Vries (2020), "The governance of genomic biobank research in Africa: reframing the regulatory tilt", *Journal of Law and the Biosciences*, http://dx.doi.org/10.1093/jlb/lsz018. [93]

Staunton, C. et al. (2018), "Rules of engagement: Perspectives on stakeholder engagement for genomic biobanking research in South Africa", *BMC Medical Ethics*, Vol. 19/1, p. 13, http://dx.doi.org/10.1186/s12910-018-0252-y. [126]

Tamminen, S. (2011), "Bio-objectifying European bodies: standardisation of biobanks in the Biobanking and Biomolecular Resources Research Infrastructure", http://dx.doi.org/10.1186/s40504-015-0031-1. [66]

Tam, V. et al. (2019), "Benefits and limitations of genome-wide association studies", *Nature Reviews Genetics*, Vol. 20, http://dx.doi.org/10.1038/s41576-019-0127-1. [18]

Thorogood, A. et al. (2019), "Return of individual genomic research results: are laws and policies keeping step?", *European Journal of Human Genetics*, Vol. 27, pp. 535-546, http://dx.doi.org/10.1038/s41431-018-0311-3. [19]

Tindana, P. et al. (2017), "Developing the science and methods of community engagement for genomic research and biobanking in Africa", *Global health, epidemiology and genomics*, Vol. 2, p. e13, http://dx.doi.org/10.1017/gheg.2017.9. [127]

Topol, E. (2019), *Deep Medicine: How Artificial Intelligence Can Make Healthcare Human Again*, Basic Books, New York. [22]

Trust Equitable Research Partnerships (2019), "Global Code of Conduct Resource-Poor Global Code of Conduct" January, pp. 61-64, http://www.globalcodeofconduct.org/. [94]

UK Department of Health and Social Care (2019), *Code of conduct for data-driven health and care technology*, https://www.gov.uk/government/publications/code-of-conduct-for-data-driven-health-and-care-technology/initial-code-of-conduct-for-data-driven-health-and-care-technology (accessed on 13 February 2020). [70]

Van Ommen, G. et al. (2015), "BBMRI-ERIC as a resource for pharmaceutical and life science industries: The development of biobank-based Expert Centres", *European Journal of Human Genetics*, Vol. 23/7, pp. 893-900, http://dx.doi.org/10.1038/ejhg.2014.235. [47]

van Overwalle, G. (2009), *Gene Patents and Collaborative Licensing Models: Patent Pools, Clearinghouses, Open Source Models, and Liability Regimes*, Cambridge University Press, Cambridge. [97]

Vaught, J. et al. (2011), "Biobankonomics: Developing a sustainable business model approach for the formation of a human tissue Biobank", *Journal of the National Cancer Institute - Monographs*, Vol. 2011/42, pp. 24-31, http://dx.doi.org/10.1093/jncimonographs/lgr009. [34]

Vayena, E., A. Blasimme and I. Cohen (2018), "Machine learning in medicine: Addressing ethical challenges", *PLoS Medicine*, Vol. 15/11, p. e1002689, http://dx.doi.org/10.1371/journal.pmed.1002689. [69]

Vayena, E. and U. Gasser (2016), "Between Openness and Privacy in Genomics", *PLoS Medicine*, Vol. 13/1, http://dx.doi.org/10.1371/journal.pmed.1001937. [86]

Walonoski, J. et al. (2018), "Synthea: An approach, method, and software mechanism for generating synthetic patients and the synthetic electronic health care record", *Journal of the American Medical Informatics Association*, Vol. 25/3, pp. 230-238, http://dx.doi.org/10.1093/jamia/ocx079. [65]

Wellcome Trust (n.d.), *Wellcome's approach to equitable access to healthcare interventions*, https://wellcome.ac.uk/what-we-do/our-work/access-healthcare-interventions/wellcomes-approach-equitable-access-healthcare-interventions (accessed on 11 September 2020). [146]

West, K., E. Blacksher and W. Burke (2017), *Genomics, health disparities, and missed opportunities for the nation's research agenda*, American Medical Association, http://dx.doi.org/10.1001/jama.2017.3096. [27]

Wetterstrand, K. (n.d.), *The Cost of Sequencing a Human Genome*, https://www.genome.gov/about-genomics/fact-sheets/Sequencing-Human-Genome-cost (accessed on 8 September 2020). [74]

Williams, A. et al. (2018), *Artificial intelligence, physiological genomics, and precision medicine*, American Physiological Society, http://dx.doi.org/10.1152/PHYSIOLGENOMICS.00119.2017. [58]

Winickoff, D. (2007), "Partnership in UK Biobank: A Third Way For Genomic Property?", *Journal of Law, Medicine and Ethics*, pp. 440-456. [124]

World Economic Forum (WEF) (2020), *Genomic Data Policy Framework and Ethical Tensions*, http://www.weforum.org (accessed on 10 September 2020). [121]

Yang, J. (2019), "Cloud computing for storing and analyzing petabytes of genomic data", *Journal of Industrial Information Integration*, Vol. 15, pp. 50-57, http://dx.doi.org/10.1016/j.jii.2019.04.005. [55]

Yehia, L. and C. Eng (2019), *Largescale population genomics versus deep phenotyping: Brute force or elegant pragmatism towards precision medicine*, Nature Publishing Group, http://dx.doi.org/10.1038/s41525-019-0080-0. [24]

Zatloukal, K. et al. (2018), *Biobanks in personalized medicine*, Taylor and Francis Ltd., http://dx.doi.org/10.1080/23808993.2018.1493921. [12]

Zeggini, E. et al. (2019), "Translational genomics and precision medicine: Moving from the lab to the clinic", *Science*, Vol. 365/6460, pp. 1409-1413, http://dx.doi.org/10.1126/science.aax4588. [99]

Zhou, J. and O. Troyanskaya (2015), "Predicting effects of noncoding variants with deep learning-based sequence model", *Nature Methods*, Vol. 12/10, pp. 931-934, http://dx.doi.org/10.1038/nmeth.3547. [61]