phg foundation
making science work for health

The GDPR and genomic data

The impact of the GDPR and DPA 2018 on genomic healthcare and research

A PHG Foundation report funded by the Information Commissioner's Office

ico. Information Commissioner's Office
Grant Recipient

Authors

Colin Mitchell, Johan Ordish, Emma Johnson, Tanya Brigden and Alison Hall

Acknowledgements

The GDPR and genomic data project was funded by the Information Commissioner's Office (ICO). We thank the ICO for their support

May 2020

URLs were correct as of April 2020

This report can be downloaded from www.phgfoundation.org

How to reference this report:

The GDPR and genomic data - the impact of the GDPR and DPA 2018 on genomic healthcare and research

Published by PHG Foundation

intelligence@phgfoundation.org

The PHG Foundation is a health policy think-tank and linked exempt charity of the University of Cambridge. We work to achieve better health through the responsible and evidence based application of biomedical science.

We are a registered company, no. 5823194

UNIVERSITY OF CAMBRIDGE

Contents

1. Introduction

This research assesses the current and likely near future impact of the GDPR and UK Data Protection Act 2018 on uses of genetic/genomic data in healthcare and health research. In particular, it addresses three linked research questions with significant practical importance for regulators, health services, clinical professionals, scientists, patients and publics:

- To what extent do genetic/genomic data used for healthcare and medical research in England and Wales count as 'personal data' under the GDPR?
- To the extent that they are 'personal data', what is the impact likely to be on the delivery of health and social care in the short-to-medium term (up to five years)?
- What can be done to mitigate or reduce any negative impacts?

The EU General Data Protection Regulation (GDPR) has replaced the previous Data Protection Directive (DPD) and updated the legal framework for the processing of personal data across the EU.[1] At the same time, rapidly developing genome sequencing technologies are driving ever increasing generation, collection and sharing of genetic data—and associated clinical information—for healthcare and medical research.

Together, these developments have generated significant uncertainty for healthcare professionals (HCPs), researchers and policymakers about how data protection law does, and should, apply to current and developing uses of genetic and genomic information, in particular, the extent of the types and applications of genetic and genomic data that are caught within the remit of the GDPR.

The UK Information Commissioner's Office (ICO) has funded research by the PHG Foundation on the impact of the GDPR and the UK Data Protection Act 2018 (DPA 2018) on the uses and regulation of genomic technologies in healthcare and biomedical research.[2] Through legal research, analysis of the literature, stakeholder interviews and a multidisciplinary workshop we have identified a number of ways in which the GDPR and DPA 2018 (or more frequently, uncertainty about interpretation of them) give rise to current and potential challenges for genomics initiatives, as well as some potential mitigations and ways forward.

In this report we present our key findings, discuss potential mitigations and make recommendations for genomics professionals, policymakers and regulators. We highlight five main areas of challenge for genomic healthcare and research under the GDPR/DPA 2018. These are:

- Uncertainty in determining when the GDPR applies to collaborators in genomics initiatives, in particular when professionals may become 'joint controllers' and when those outside the EU must comply with the GDPR (chapter 3);
- Uncertainty in determining when genetic, genomic and associated health data are 'personal data' governed by the GDPR and whether data that have been de-identified (e.g. through pseudonymisation) remain personal data (chapter 4);
- Challenges meeting the requirement for a lawful basis for processing personal data and specific conditions for processing 'special category' (e.g. health or genetic) data (chapter 5);
- Challenges fulfilling data subject rights and meeting obligations under the GDPR and DPA 2018 (chapter 6) and;
- Challenges making data accessible to others or data sharing both within the EU/EEA and to 'third countries' (chapter 7).

In chapter 8 we discuss technical, organisational and sector-led approaches to mitigate and address these challenges. We highlight, in particular, the potential for development of sector-specific codes of conduct and certification schemes for the genomics community to set appropriate standards for processing genomic and associated data in compliance with the GDPR and Member State regulation. Finally, in chapter 9 we make some overall conclusions about the current and short-to-medium term impact of the GDPR and UK DPA 2018 on uses of genomic data in healthcare and biomedical research. We make recommendations about how healthcare professionals, researchers, policymakers and regulators may address the challenges discussed in this report to ensure that the long-term impact of the GDPR is as positive as possible for genomic healthcare and research.

This introductory chapter first outlines the current and near future landscape of genomic technologies in healthcare and medical research and sets out our research aims and approach. Chapter 2 provides an overview of the changes brought about by the GDPR and DPA 2018, describing the principles and general approach to data protection under the GDPR which form the basis for our more detailed discussions in subsequent chapters.

1.1 The landscape of genomic technologies in healthcare and medical research

As a prerequisite to understanding the impact of the GDPR on the use of genomic technologies in healthcare and medical research, it is important to understand the nature and scope of these technologies. Clinical genetics is a comparatively new discipline: developments in biological understanding have, in turn, led to novel technologies which have transitioned from research to clinical settings. Over the last fifty years, this has enabled scientists, researchers and healthcare professionals to build a better understanding of the association between observed molecular and physiological differences and other clinical signs of disease.

Healthcare professionals and scientists seek to understand disease by analysing both a patient's genome and their phenotype. A phenotype is the observable manifestation of genetic and environmental influences on an organism, e.g. eye colour. The extent to which environment or genetics influences the phenotype will differ dependent upon the trait. As genomic technologies have evolved, the nature and scale of the data generated through these have increased exponentially, such that one of the most comprehensive technologies, whole genome sequencing, can generate billions of data points through the sequencing of a human genome.

The key interpretive challenge from a scientific and medical perspective, is that only a small proportion of those data points will be clinically useful. In the research context, genetic information about an individual may help to generate an understanding of the nature of disease processes, the variability in individual responses, and inform therapeutic interventions, including the drug development. In the healthcare context, the results of genetic or genomic tests may provide a disease diagnosis or provide an understanding of the progression and prognosis of current or future disease.

The increasing resolution and coverage of these technologies, either used alone or in combination, and their proliferation in medical research and in healthcare, have profound implications for data protection as they increasingly generate data which may be connected to individuals, thereby potentially falling within the definition of 'personal data' in the GDPR.

The sections that follow provide an introduction to the structure and function of the genome and to the breadth and scope of current technologies before turning to consider some of the implications for data protection.

The structure of the genome

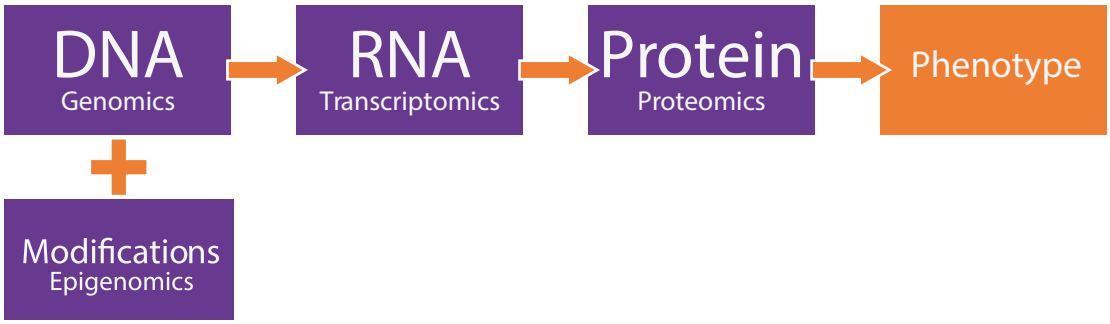
The genome is the complete set of DNA contained within an organism's cells. DNA is made up of four types of nucleotides, also referred to as bases, represented by the letters A, C, G and T. The order of these nucleotides is referred to as the DNA sequence or, more colloquially, the genetic code. This sequence determines much of an individual's biological make-up and has implications for health and disease amongst other things. DNA is arranged in long coiled strands, split into a few separate pieces, called chromosomes. All living things possess a genome; the genome is identical (with a few exceptions) in each cell of the same individual human being, but differs between all individuals, with the least variation between identical siblings.

Mutations or changes in the genetic code may occur in either germline or somatic cells, and have different implications. Germline cells are sex cells, eggs and sperm, which go on to produce the next generation - mutations in the DNA of these cells leads to heritable changes which will be present in all cells of the resulting offspring. All other human body cells are 'somatic' - changes in these cells will not be inherited by the next generation; cancer is a disease that may result from somatic mutations.

In healthcare and medical research, variants in patient or experimental genomes can be identified by comparing them with human reference genomes, which are built using information from the DNA of multiple individuals. In healthcare, references may also come from relatives of the patient, or from clinical databases e.g. Genome in a Bottle.[3] Even 'healthy' genomes differ - each individual's genome has around three billion pairs of bases and will differ from a reference at millions of these points. Some variants are more common than others, and some groups may show greater differences than others at various points in the reference. However, the vast majority of variation observed between genomes is not damaging, only a small proportion of variants result in disease.

Genomics is the study of the genome. It can be defined as the examination of genes and how they function, but it can also encompass structure, function, sequencing, mapping and evolution of DNA sequences and chromosomes. In healthcare and medical research, the term genomics generally indicates the examination of, part or all of, an individual's DNA sequence to gain information related to their current or future health, or the investigation of pathogens (and their genomes) that they may be hosting. On the pathway from DNA to phenotype there are additional molecules and processes that can be measured - since these are derived from, or related to the genome, they can provide some information about the genome itself and genetic disease.

Figure 1. Simplified genomic pathway, including terms used for the field of study associated with each type of molecule.



The following section describes how various approaches are currently used in healthcare and medical research and how they may develop in the future. As we shall explore in the rest of this report, these techniques have important implications for data protection as they increasingly generate data which may be connected to individuals, thereby potentially falling within the definition of 'personal data' in the GDPR. However, in order to make this assessment, it is necessary to understand the extent to which an individual's genome is unique to each person.

Discounting exceptional or atypical circumstances (e.g. identical twins, cancer, and mosaicism), an individual's entire genome is unique to each person and remains fixed throughout lifetimes - the sequence of our DNA remains mostly unchanged from birth through to death. Since the full genome is unique to an individual, it can be used as an identifier. However, the vast majority of that genome (around 99.9%) is shared in common with other humans. This means that almost all genetic and genomic information cannot, without additional information, be linked to a natural person and therefore be used as an identifier. The implications of this are discussed later in this chapter and in chapter 4.

Genomics and sequencing of DNA

Genomics is the study of the DNA of our cells - the genome, its structure and function. Although the terms 'genomics' and 'genetics' are sometimes used interchangeably, there are some differences. Clinically, the term 'genetics' relates to inheritance, and most prominently refers to the study or examination of hereditary (genetic) disease. Genomics is applied to health, especially within rare disease and cancer to provide diagnosis, prognosis and uncover links between genetic variation and disease, and the term implies a more holistic examination of DNA. A broad range of techniques exist for the examination of DNA, generating varying volumes of data at different resolutions and scale, for different purposes. Many techniques have existed for several decades, and are staples of clinical diagnostics. The National Genomic Test Directory aims to improve standardisation of the genetic and genomic tests which may be offered across the NHS in England.[4]

Low resolution DNA analysis

Techniques such as karyotyping examine chromosome structure, rearrangement or the presence or absence of chromosomes (e.g. in aneuploidies such as Down's syndrome) to provide low resolution information. Polymerase chain reaction (PCR) and microarray allow for the examination of known genetic variants, where the sequence to be detected is already known prior to investigation. These techniques range in scale from the detection of single base changes in a single section or small sections of DNA (e.g. use of PCR), to detecting the presence of variants in many thousands of different short DNA sequences in parallel (e.g. microarray). Large scale human genotyping projects such as the UK Biobank have utilised microarray technology to detect hundreds of thousands of variants in large numbers of participants. The extent to which these technologies can be linked to an individual (see chapter 4) will depend on the rarity of the structural change or the variant that is detected, and the availability of additional information. If a number of variants are tested for, or if technologies are used in combination, even low resolution DNA analysis could be identifying. However, if variants are common within a population, then testing for a number of variants will be relatively uninformative. The implications of such tests and testing strategies are heavily dependent upon context and are explored in more detail throughout this report.

Whole genome sequencing (WGS) and whole exome sequencing (WES)

Only DNA sequencing techniques are capable of reading the sequence of many nucleotides in DNA from scratch. Key approaches used for large scale investigation of the genome are whole genome sequencing, whole exome sequencing and targeted or panel sequencing (where selected subsections of the genome are sequenced). Whole genome sequencing (WGS) represents the most comprehensive approach to DNA sequence analysis. The whole genome (including non-coding regions), which consists of approximately 3 billion base pairs, is sequenced. Analysis may be subsequently computationally restricted to examine only parts of the data retrieved.

Whole exome sequencing (WES) consists of examination of all protein-coding (also called 'functional' or 'coding', parts of the genome), which is equal to about 2% of the whole human genome sequence. Exome sequencing is now used routinely in healthcare to aid the diagnosis of rare disease, through examining some or all of the exome. Trio exome sequencing, where the exomes of both parents and the patient are examined (a child requiring a diagnosis for a rare disease), can be used to determine whether a *de novo* (new) genetic variant rather than an inherited variant might be responsible for the condition in question. The recently announced expansion of rapid genetic testing for the diagnosis of disease in critically ill infants will utilise WES, and is likely to be available to around 700 newborns and children each year.[5]

Genomics and healthcare

Rare disease diagnostics using some form of genetic analysis are already well-established. The use of WGS and WES in healthcare will continue to expand over the next few years, with the aim of providing diagnosis, further disease and health information for patients, and building on established knowledge bases through linking genetic information to health and disease outcomes. This is a focus of the precision/ preventative medicine agenda. Both whole genome sequencing and whole exome sequencing create novel data protection challenges. As noted above, an entire whole genome sequence contains around three billion data points, each of which indicating the type of base present at a relative position within the genome sequence. Taken together the entire sequence can be used as an identifier, provided that sufficient information exists to link that genome with a natural person. Equally, less comprehensive genomic sequence data such as that produced by whole exome sequencing, could be similarly identifying depending on context (see further in chapters 4 and 8).

Additional applications of human DNA analysis for health

There are further fields of study which use forms of genomic investigation and DNA sequencing to retrieve health-related information. Broadly, these fields generate or rely upon different collections of individual-level and DNA sequence information, with consequent data protection implications. Examples are discussed below.

Pharmacogenomics (PGx) focuses on specific areas of the human or patient genome known or thought to be related to metabolism of drugs (pharmacodynamics and pharmacokinetics). Clinical implementation of PGx is currently limited to the targeted assessment of gene-drug pairs in some cancers, cystic fibrosis and epilepsy, and testing prior to administration of abacavir for HIV treatment. There is an aim to incorporate PGx testing into genetic services in the UK and between 40 and 50 gene-drug pairs could be of clinical relevance. Broader testing across the genome relating to drug metabolism is limited to research.

Broadly, **nutrigenomics** (or nutritional genomics) is the study of the relationship between genomic variants or gene expression, food intake, nutrients and nutrition, including its health implications. Genomic analysis focuses on particular regions of the genome that have been linked to outcomes related to nutrition, or to the examination of the genome for the purpose of identifying such regions. Scientific evidence and evidence to support its use in healthcare is currently limited. Studies have shown that the addition of nutrigenomic information for informing people's health and lifestyle decisions does not significantly impact people's behaviour and therefore does not necessarily lead to them making healthier lifestyle choices. Although there are significant exceptions, e.g. lactose intolerance, evidence of benefit from personalised diets based on genetic information and substantial outcomes for nutrition or vice versa is lacking.

Investigation of microbial genetic material

As discussed in chapter 4, 'genetic data' within the GDPR is defined in terms of personal data relating to a 'natural person'. A number of additional applications analyse genetic or genomic material from non-human organisms, but may inadvertently generate information about parts of the human genome as a consequence of sample preparation or sample contamination. Alternatively, these investigations may highlight relevant features of the inherited or acquired characteristics of a natural person, such as genetic factors which impact on an individual's susceptibility to disease such as their genotype or immune response. This section refers to those technologies.

Human beings are sometimes referred to as 'super-organisms', referring to the multitude of organisms that reside within and alongside us that help us function on a day-to-day basis. We also play host to pathogenic microbes which can cause disease and we act as vectors for their transfer to other humans and/or organisms. The study of these microbes can provide us with information about individual and population health, and as the investigation of human genomes becomes more common in healthcare, microbial genomics is also expanding as a field of interest. The microbiome is the composition of microbes in a particular environment and its investigation is not limited to genetic material. Pathogen genomics concerns the more detailed study of the genetic make-up of one or more pathogens for the purpose of understanding origins, evolution and informing treatment response.

Microbiome analysis (microbiomics) is the field of study associated with the examination of microbial communities in different environments and their interaction with that environment. The human microbiome is the complete microbial, primarily bacterial, complement present in/on a person. Examination of the microbiome includes several different approaches; in medical research, studies may include analysis of the genetic material (DNA or RNA) of microbes present within or on a patient. Changes in the microbiome are being linked to health outcomes ranging from improper gut function to depression, and have also been linked to the effectiveness of various cancer treatments and a range of other conditions, though evidence gathering in many areas is in the early stages.

Investigation of the microbiome is normally limited to a particular part of the body - researchers may choose to examine the skin, gut or mouth microbiome for example. Samples are collected in different ways depending on what is being examined; the use of mouth swabs or stool samples is common. Samples taken from patients will also contain human DNA, which it is necessary to remove or exclude for some forms of analysis. Sequencing used for investigation commonly focuses on short sections of microbial genetic information. Evidence to support the use of microbiome DNA analysis for health in several circumstances is lacking and it has seen only limited use in healthcare so far.

Pathogen genomics relies on the examination of DNA or RNA of pathogenic entities (e.g. viruses, bacteria) within a patient or in the environment for confirmatory diagnosis, pathogen sub-typing, discovery of pathogen origins, and tracking of infectious diseases. When deployed appropriately and in a timely manner, it can help to inform public health actions and influence the spread of disease.

Pathogen genomics does not aim to examine the human (patient) genome, however where samples are taken from patients there will likely be human DNA present – pre- or post-processing of samples can remove much human DNA contamination. Pathogen sequencing has a role in determining the speed and direction of transmission of infectious diseases. This is likely to be highly informative in understanding how diseases such as Bird Flu, Ebola or SARS can arise, mutate and spread. A better understanding of these principles can also inform how the effects of a novel disease might be mitigated through treatment and other interventions. However, the development of increasingly granular transmission chains based on pathogen genome sequencing raises the possibility that it might be feasible to generate transmission chains more routinely in the future to identify the source and route of an infection within a population. This has potential implications beyond hospitals and health providers, such as for local authorities and even individuals who are implicated as the original source or carrier of disease.

The recently announced COVID-19 Genomics UK Consortium (COG-UK)[6-7] aims to deliver rapid, large scale sequencing of pathogen genomes (Sars-CoV-2) from patients confirmed to have COVID-19 in the UK.

Other broadly applicable techniques

A further group of technologies selectively identify and analyse a sub-set of DNA from a particular sample type within an individual. Although the majority of these technologies are currently limited to research applications, they have significant data protection implications, because of the way they are likely to be adopted within health systems in the future. These technologies are likely to be used to indicate a change in health status over time relying on repeated observations in the same individual (e.g. ctDNA). Some of these technologies are likely to be cheaper and therefore more accessible than whole genome sequencing when used on a population basis. This data is likely to be collected and retained and form part of electronic health records in the future:

Cell-free DNA (cfDNA) is DNA that has been released from cells and is present in blood or in other fluids around the human body. This may be germline DNA, fetal DNA, cancer DNA, or potentially pathogen DNA, which is present in small broken sections and may contain changes (variants) not present in the main germline genome e.g. in cancer. One application of the analysis of cell-free fetal DNA (cffDNA) is to test a foetus for genetic conditions. Depending on how sensitive and specific they are, cffDNA tests can either be used for screening, which is known as non-invasive prenatal testing (NIPT) and requires a follow up diagnostic test, or upfront as a diagnostic test, which is known as non-invasive prenatal diagnostic testing (NIPD). Targeted sequencing (sequencing only a selected part of the genome) can be used for this purpose. This may involve collecting as much genome information from the sample as possible and using computational tools to limit the scope of the analysis to specific genomic regions.

Within the field of NIPT and NIPD, there is research ongoing to understand the feasibility and utility of performing whole exome and potentially whole genome sequencing using cffDNA, compared to current approaches which target specific genes or larger genomic regions. This could potentially be used to help screen for and/or diagnose a wider range of conditions than would otherwise be detected using current methods, and reduce the need for invasive procedures i.e. amniocentesis.

Another major application of this technique is for the examination of circulating tumour DNA (ctDNA) to identify if specific cancer mutations are present and which, if any, targeted therapies are appropriate. In contrast to standard DNA analysis for cancer, which relies on tumour biopsy, ctDNA can be obtained using a simple blood test, which is easily repeated if it fails, and can be repeated for the purposes of collecting temporal information about cancer recurrence in an individual and cancer prevalence within a population. At present, cell-free DNA analysis does not aim to examine the whole human genome.

Single cell sequencing (scSeq) aims to examine a single copy of the whole (or part) of the genetic material of a cell. This utilises one of several techniques for isolation and examination of the DNA and/or RNA inside one cell. Single cell 'omics may include examination of a cell's genome, transcriptome, proteome or epigenome. Projects like the Human Cell Atlas[8] aim to create a 3D 'Google map' of the 37 trillion cells of the human body which will allow scientists to zoom into organs, tissues and cells to reveal the location and gene activity patterns of each cell type. There are promising signs that single-cell sequencing can be used to identify genetic mosaicism - when cells in a tissue contain different genomic complements or copies of the genome - and assess the impact of intra-tumour genetic variability in cancer development or treatment response. However, there are technical challenges which must be addressed for single-cell 'omics to continue to advance[9]. Although single-cell sequencing can yield higher 'resolution' genomic data, a single test result is likely to give rise to similar data protection implications as genomic data derived from WGS or other techniques.

Examination of molecules derived from or associated with DNA

Although not directly equivalent to human DNA, other molecules derived from DNA including RNA and proteins can be examined to provide information not only about health, but about a person's genome. In addition, modifications to DNA can also provide information about the functional status of parts of the genome, and consequently about a person's health. The investigation of these molecules has implications for data protection.

Epigenomics

Epigenomics is the study of modifications to DNA (not including alterations to the genome sequence itself) that may or may not be inherited and can change over a person's lifetime. Types of DNA modification include methylation, histone modification and acetylation. These modifications can alter how genes are expressed, and change the structure of the genome. Epigenomic changes are implicated in rare genetic disorders such as Prader-Willi syndrome and Angelman syndrome, and some cancers. Altered methylation of genomic regions can result in altered expression of genes, failure to express proteins, or these processes being unreliable. Modifications to the genome accumulate over a person's life time, and modifications at some locations may correlate with a person's chronological age - referred to as 'epigenetic clocks'[10]. Epigenetic marks can also differ according to exposure to stress, and certain external influences. They can affect future generations through epigenetic inheritance. Methylation marks can be detected in several ways and form part of technologies used for diagnosis.

Transcriptomics

Transcriptomics is the study of the whole or part of the transcriptome, the collection of all RNA molecules present in a sample e.g. cell, tissue or organ, at a given time. The most commonly studied type of RNA is messenger, or 'coding', RNA (mRNA), which is produced from protein-coding genes and is used as a template for the production of proteins. The measurement of RNAs can be used to determine which genes are being expressed, and the level of that expression, in a sample at a given time, i.e. gene activity. The current use of transcriptomics in UK healthcare is limited to gene expression panel testing and single RNA biomarker assessment, much of which is conducted in cancer diagnostics, prognosis and for informing treatment selection. Transcriptome analysis which encompasses examination of the entire transcriptome, or large parts thereof, is being investigated for use in the diagnosis of rare disease alongside genome or exome sequencing data, but is currently restricted to research, for example as part of the 100,000 Genomes Project.

It is likely that the use of transcriptomics in rare disease diagnosis will expand in coming years, and more panels integrating multi-omic approaches (combining genomics, transcriptomics and/or proteomics) are being developed. Many NHS laboratories have the capacity to conduct RNA-sequencing for transcriptome analysis if it were required. The genomic test directory mentions the analysis of RNA as a future prospect, stating that in some instances, RNA samples should be stored for future use. Unlike DNA, the RNA content of cells differs substantially across cell type, time, and due to external influences. Different forms of RNA can be produced from the same section of DNA under certain circumstances. Even where the form of the RNA produced from a particular section of DNA is relatively consistent, the RNA may not be regularly produced.

Proteomics

Proteomics refers to examination of the proteome, the collection of all proteins present in a sample at a particular point in time. The proteome can reflect variation within the genome which may produce altered or non-functional proteins, or no protein at all. Unlike DNA, and much like RNA, the protein content of samples changes over time. The proteome can be examined using several different methods, including mass spectrometry, microarray and CHIPseq (for the examination of protein/DNA interactions).

Proteomics can aid in disease diagnosis and understanding, and help measure responses to treatment. For example, the detection of single or a low number of protein biomarkers is already used within healthcare to diagnose or influence treatment for several conditions. However, broader examination of the proteome is currently restricted to research and is unlikely to have significant applicability within healthcare in the next few years, as there are challenges associated with measuring individual proteomes and consistently drawing clinically relevant conclusions from data.

At present, since the amount of publicly available RNA and proteomics data is limited, and knowledge about the variability in RNA within and between individuals is partial, this suggests that RNA and other omic data are significantly less identifying than DNA data.

Summary: implications for data protection

The scale and nature of data that result from genomic and other 'omic techniques give rise to significant challenges for data protection. As outlined above, genomic technologies are being used to generate a greater level of data about individual health and physiology, for an ever-increasing number of people. These data are frequently high-dimensional, containing many data points that are capable of being connected to both individuals and their relatives. Because a full genome is unique to an individual it is potentially a powerful identifier. However, as discussed in this report, even much more limited genomic information can also be powerfully identifying so there are considerable privacy and data protection implications in the use and storage of all genomic information.

Although many of the data generated through the application of genomic techniques discussed above are highly complex and, in the case of transcriptomics and proteomics, will vary significantly across cell type and time even for the same individual, it is still possible that such data can be combined with other information to identify an individual. This is a challenge for genomics professionals, as data which may appear to be technical and unrelated to an individual may become increasingly identifying over time as technologies advance and are used in combination. For example, although pathogen genomics does not aim to examine the human (patient) genome, it will involve some processing of human DNA, and the tracking of transmission may in itself identify individuals. Because the majority of the genome (99.9%) is shared in common with others, the management of genomic data is complicated by the possibility of the same data relating to multiple individuals.

As an increasingly wide range of healthcare and research professionals are utilising genomic techniques and data, they need to consider whether the data they use gives rise to data protection implications and if so, how this impacts their work. In future, the routine use of these technologies, and retention of test results within electronic health records seems likely to increase the scale and nature of the information that will need to be taken into account by healthcare professionals and health providers. At the same time, regulators and policymakers require an understanding of the impact of the GDPR on this complicated and crucial area of healthcare and research in order to develop appropriate standards for data protection and genomic data.

1.2 Research aims and approach

This research assesses the current and likely near future impact of the GDPR and UK Data Protection Act 2018 on uses of genetic/genomic data in healthcare and health research. In particular, it addresses three linked research questions with significant practical importance for regulators, health services, clinical professionals, scientists, patients and publics:

- To what extent do genetic/genomic data used for healthcare and medical research in England and Wales count as 'personal data' under the GDPR?
- To the extent that they are 'personal data', what is the impact likely to be on the delivery of health and social care in the short-to-medium term (up to five years)?
- What can be done to mitigate or reduce any negative impacts?

To address these questions we conducted legal research, semi-structured interviews with key stakeholders and convened a multidisciplinary workshop with over thirty clinical/scientific professionals, policy makers, regulators and academic experts.

The legal research included primary research and analysis of the GDPR and DPA 2018 and reviews of the legal, scientific and medical literature for evidence of current or potential impact on genomic activities. As the GDPR and Member State legislation is at a relatively early stage of application, we also identified potential impacts based on analysis of the law and how it could apply to genomics health and research applications.

The stakeholder interviews and workshop provided evidence of current and likely impacts on healthcare and research, as well as range of expert insights on the correct interpretation of legal provisions, challenges and potential ways forward. Further details of the interviews and workshop are included in the Annexe to this report.

We report our findings and discuss key potential impacts of the GDPR and DPA 2018 throughout the rest of this report. We address the first research question—when are genetic/genomic data 'personal data— in particular in chapter 4, and the third question—what can be done to mitigate or reduce any negative impacts—in chapter 8. We make overall conclusions and set out recommendations for key audiences in chapter 9. In the next chapter, we provide an overview of the changes brought about by the GDPR and DPA 2018 and the principles and general approach to data protection under the GDPR, to provide a basis for more detailed discussions in the rest of the report.

2. The GDPR and Member State legislation

The General Data Protection Regulation entered into force on the 25th May 2018 and marks a considerable strengthening of data protection in Europe. Unlike the previous Data Protection Directive (DPD), the Regulation applies directly in all the Member States (MS) without the need for national implementing laws. In fact varying (or even duplicating) the Regulation in national law is unlawful unless it is in relation to one of the provisions that the GDPR explicitly allows MS to tailor.*[11] The fundamental basis for the GDPR is also strengthened: the previous Directive was based on the need to harmonise laws for the development of an internal market and it gave considerable discretion to MS to tailor the application of its rules. However, in 2007 the EU recognised the 'fundamental right to the protection of personal data'[12] within one of its primary treaties, providing a new independent legal basis for the development of a stronger Regulation (Art 16 TFEU).[13]

Against this backdrop, and in response to challenges brought by rapid technological developments and globalisation, the GDPR has a twin aim of facilitating the free flow of personal data within the Union while ensuring a high level of the protection of personal data. This has led to the introduction of new rights, such as the stand-alone right to erasure (Art 17), new obligations for data controllers and processors, new governance requirements (e.g. for a data protection officer reporting directly to the highest management level, Art 38) and greatly enhanced potential fines (up to 4% global annual turnover).

In this chapter we provide an overview of the GDPR and complementary Member State laws, particularly the UK's DPA 2018, to provide the context for discussion of specific data protection impacts and challenges for genomics in the following chapters. First, we briefly consider the influence of a human rights approach to data protection on the GDPR and its interpretation.

* The GDPR does not apply to areas outside EU competence, e.g. data processing for national security or intelligence purposes, so MS may implement their own laws governing these topics.

2.1 Overview of the GDPR

The GDPR aims to ensure both a consistent and high level of protection of 'natural persons' (living individuals) and to remove obstacles to flows of data within the EU. To do so, it replaces and updates the rules set in the previous Data Protection Directive which was finalised over two decades earlier, in 1995. Both the Directive and GDPR have their roots in human rights instruments, in particular those issued by the Council of Europe, the international organisation which is distinct from the European Union and responsible for the European Convention on Human Rights.

The Council of Europe adopted the first international instrument on data protection in 1981 to set standards and principles to ensure respect for the fundamental right of all individuals 'with regard to processing of personal data'. This was Convention 108, which aimed to promote the respect for privacy and protection of personal data at a global level, and which fifty-five countries have now either ratified or signed.[14] Convention 108 contains the seeds of the EU's data protection principles, including the core principles of lawful, fair, and purpose-limited processing,[15] and it was highly influential in the EU's development of its own Data Protection Directive in the 1990s. However, unlike the Council of Europe, the EU was not a fundamental rights organisation and its Directive was primarily designed to harmonise national rules on data protection and ensure the free movement of personal data within its economic area. In fact the treaty basis for the DPD was a provision that aimed to harmonise the 'internal market' of the European Community.* To remove obstacles to the free flow of data—the Directive also aimed to ensure that the level of protection of fundamental rights, in particular the right to privacy—were consistent across the EU (recital 7). However, as a Directive, the DPD left the means of implementation up to individual Member States and it also provided a considerable margin of appreciation as to the level of and nature of the protection of personal data in each State.

In the decades following the adoption of the Directive, technological and social changes coincided with a shift in the status of data protection in the EU. As the EU developed and expanded it adopted a Charter of Fundamental Rights, including a 'fundamental right to the protection of personal data'.[16] This was also incorporated in one of the EU's primary treaties alongside a new legal basis for the EU to legislate to establish rules relating to the processing of personal data, separate to the need to harmonise an internal market. This became the foundation for the GDPR, which was inspired by a need to 'put individuals in control of their personal data' and the strong statement that:

> '... individuals have the right to enjoy effective control over their personal information. Data protection is a fundamental right in Europe, enshrined in Article 8 of the Charter of Fundamental Rights of the European Union, as well as in Article 16(1) of the Treaty on the Functioning of the European Union (TFEU), and needs to be protected accordingly.'[17]

* The contemporaneous treaty basis for this being: Consolidated Version of the Treaty on Functioning of the European Union [2016] OJ C202/95, art 115.

This background is important because the recognition of a fundamental right to data protection, changes in treaty basis and the shift in emphasis between the DPD and GDPR may lead to different interpretations of concepts under the GDPR. As we discuss below, many of the concepts and principles of the Directive are carried over by the GDPR but their interpretation by the courts and authorities may change in light of these fundamental changes in EU data protection law.

A period of uncertainty

Although the GDPR entered into force on the 25th May 2018 there is inevitably some uncertainty about new provisions and the correct interpretation of more established concepts in light of the new context described above. National Supervisory Authorities (SAs) are in the process of updating their guidance and the new European Data Protection Board (EDPB)—which is charged with ensuring consistency across the EU—is also incrementally producing guidance and adopting some of the opinions of its predecessor (the Article 29 Working Party). Moreover, the GDPR does provide some room for the tailoring of certain provisions by MS, as the UK has through the Data Protection Act 2018, so the GDPR must be read together with relevant national laws, as and when they are introduced.

The scope of the GDPR

Although the GDPR is a significant updating of EU data protection law, it has retained many of the core concepts and principles of the previous EU data protection law. Unlike laws in other parts of the world (notably in the United States) the GDPR is not sector specific and it applies to most sectors and activities except those that fall outside the scope of EU law (e.g. national security). As we discuss in this report, this can create challenges in setting standards for unique areas of activity such as genomic healthcare and research.

The GDPR governs the processing (almost any imaginable activity, including storage) of the 'personal data' of living individuals (this is known as the 'material scope'). The concept of 'personal data' is discussed in detail in chapter 4 and it is based on the identifiability of a unique individual (as opposed to a group) from information that 'relates' in some way to that person. Purely personal or household activities are, however, excluded.[18]

The primary 'territorial scope' of the Regulation is the EU (and EEA) Member States. However, as we discuss in detail in chapter 3, the GDPR also applies to processing of the personal data of individuals outside the EU, if sufficiently connected to the activities of an EU-based individual or organisation, and to some processing of the personal data of individuals on EU territory by data controllers or processors based elsewhere in the world.

Principles (Article 5)

The GDPR sets out seven key principles for data processing (Art 5). Six of these are broadly similar to the principles contained in the Directive but the seventh, the principle of 'accountability' is a new addition. These are all relevant to uses of genomic data but some, as we discuss below, present particular challenges in this context.

Lawfulness, fairness and transparency

The first principle requires processing to be lawful, both on the terms of the GDPR itself, but also in terms of the wider legal framework. In the UK this means data must be processed within statutory and common law obligations. In particular, processing must also be in line with duties of confidentiality under the common law and a right to privacy under the Human Rights Act 1998. This means that although those areas are governed by separate laws, they are relevant to the assessment of lawful processing and may be considered by SAs in their investigations.*

A core element of fairness relates to whether individuals would expect their data to be used in a particular way and whether processing can be justified. As we discuss at several points in this report, this may be a powerful support for processing in the genomics context if controllers and processors can show that individuals would expect their genomic data to be used in a certain way (for example to support the healthcare of another or in clinical research). Transparency requires processing to be clear and open and understandable. This principle is reinforced by many further rights and obligations in the Regulation, in particular the rights to information in Arts 13 and 14 (see chapter 6). However, as we discuss further in chapter 5, changes to consent under the GDPR have challenged the ability of some genomic data controllers to be clear and transparent because they are having to explain the differences between consent under the GDPR and consent to research more broadly. This may not be easy to communicate.

Purpose limitation

Another principle that is potentially challenging in the genomics context is that data must be collected for specific purposes and not further processed in a way that is incompatible with those purposes. As we discuss in chapter 5, this means that genomic data controllers should be very specific about their purposes and choose an appropriate lawful basis for processing accordingly. This can be challenging when new purposes are considered, such as further forms of genomic research. However, as we discuss below, there is a qualification that allows further processing for scientific research purposes.

Data minimisation

This principle requires data to be adequate, relevant and limited to what is necessary in relation to the purposes for which they are processed. Part of the challenge of the GDPR in the genomic context is that data minimisation runs counter to the fact that genomic data are of increasing utility for healthcare and research over time. This is because our level of understanding of the interaction between genetics, physiology and disease is constantly improving. From this point of view, the genomics community and data protection law have different starting points so it is important that genomics professionals are able to justify the retention and use of more data, over a longer period, if this is necessary for healthcare and research.

* For example, the Information Commissioner explicitly considered compatibility with the common law duty of confidentiality (taking advice from the National Data Guardian on this issue) in her investigation of the arrangement between Royal Free NHS Foundation Trust and Deepmind. This investigation is also discussed in chapter 5. See further: https://ico.org.uk/media/action-weve-taken/undertakings/2014353/undertaking-cover-letter-revised-04072017-to-first-person.pdf

Accuracy

The fourth principle is that every reasonable step must be taken to ensure that personal data which are inaccurate are erased or rectified without delay and, where necessary, personal data are kept accurate and up-to-date. This is supported by an allied right to rectification (Art 16) and we discuss this in more detail in chapter 6 because the question of when genomic data can be considered 'inaccurate' could have significant implications for the updating of genetics results and curation of health records.

Storage limitation

Personal data must not be kept for longer than necessary, and the retention of personal data must be justified. However, this does not apply if data have been anonymised so that they are no longer identifying, nor does it apply as strictly to data if they are solely stored for scientific research purposes.

Integrity and confidentiality

Data must be stored and processed in a secure way and protected against unauthorised processing, loss, destruction or damage, using technical or organisational measures. We discuss some of these obligations, and the technical and organisational safeguards they require in relation to genomic data, in chapters 6 and 8.

Accountability

The final principle is a new one to the GDPR. It requires data controllers to be responsible for, and be able to demonstrate, compliance with the Regulation. This requires genomic data controllers to be proactive in justifying their approach under the Regulation and demonstrate they have done all they should to comply with the GDPR. There are some mechanisms such as a Data Protection Impact Assessment (Art 35) that can help with this process.

Although these principles may seem general and broad compared to the more specific provisions in the rest of the GDPR, it is important to recognise that non-compliance with them can be the subject of fines and enforcement action on their own, without the need to find a breach of other rights or obligations. As we consider at several points in this report, understanding these principles can also help genomic data controllers to assess what is appropriate in terms of complying with more specific obligations.

Data subjects, controllers and processors

The GDPR refers to 'data subjects' (a natural person who may be identified from personal data), data controllers and processors. 'Data controllers' are the individuals or institutions who determine the purposes and means of the processing of personal data. 'Processors' are those who perform an action or service with those data on the controller's behalf. Data controllers are responsible for most of the rights and obligations under the GDPR. However, the GDPR also imposes obligations on data processors so they are also responsible to individuals and the authorities. We discuss the interpretation of when purposes and means are determined in the context of genomic data in chapter 3. This is very important because it is potentially broad enough to capture some genomics professionals who may not expect to be considered data controllers, particularly in collaborative genomics projects.

Rights, obligations, liability and sanctions.

The GDPR sets out a range of rights and obligations that may apply to data processing. We discuss those which our research has identified as most pressing for genomics throughout this report but particularly in chapters 5, 6 and 7. Some new rights and obligations have captured public attention but it is the potential sanctions that have perhaps driven a much greater awareness of the GDPR and a concern about its impact in all sectors, including the genomics sector. Fines may be levied by Supervisory Authorities up to 4% global annual turnover and some early high-profile examples have demonstrated that the penalty for non-compliance may be significant. However, any person who suffers material or non-material damage as a result of the infringement of the Regulation is also entitled to receive compensation (Art 82) so it is open to data subjects to bring court proceedings against data controllers or processors as well as make a complaint to their Supervisory Authority.

Member State legislation & Data Protection Act 2018

The GDPR aims to harmonise data protection law across the EU/EEA and, unlike the previous Directive, it applies directly in all Member States without the need for further implementing legislation. However, there is in fact significant scope for MS to tailor how rights and obligations apply (even disapplying them in certain contexts) and some forms of processing are only justified (such as processing of health and genetic data for research purposes) without consent if MS have enabled them. We discuss these issues throughout this report and highlight the scope for national variation as a particular challenge for cross-border genomics projects. The UK has applied, supplemented and varied aspects of the law that MS have been expressly allowed to tailor, in the Data Protection Act 2018. We discuss these provisions where they are relevant throughout the report but it is important to note that unlike the previous Directive-based regime, the GDPR is the direct source of most data protection principles, rights and obligations in the UK, not the Data Protection Act.

Supervisory authorities and the EDPB

Like other forms of EU law (such as medical device regulation) there are national Supervisory Authorities (SA) whose primary role with respect to the GDPR is to monitor and enforce its application.* [19] The UK's SA is the Information Commissioner's Office (ICO). UK's ICO wields more power than that granted by the GDPR, the ICO upholding information rights according to several other pieces of legislation, for instance, the Freedom of Information Act 2000 and the Privacy and Electronic Communications Regulations.[20] Notably, the ICO must balance the interpretation of legislation under its remit, and take account of potential conflict between different elements, for instance, between the Freedom of Information Act 2000 and the GDPR.†

The European Data Protection Board (EDPB) is the European body that contributes to the consistent application of data protection rules throughout the EU. The EDPB is the successor to the Article 29 Working Party 29 (WP29) which had a similar function under the Data Protection Directive and the EDPB has deliberately adopted a significant number of WP29 Guidelines and Opinions.[21] The EDPB also has a complementary body, the European Data Protection Supervisor (EDPS) that seeks to ensure the consistent application of data protection rules as they relate to EU institutions and bodies.[22] How do SAs and the EDPB ensure consistent application of the GDPR across the EU?

Consistency of interpretation

To meet its secondary treaty basis of harmonisation, the GDPR must be interpreted and applied in a consistent manner across MS. There are two broad mechanisms that work to provide such consistency: the mechanisms anchored in Article 63 and the interpretation of the Court of Justice of the European Union (CJEU).

Article 63 requires SAs to contribute to the consistent application of the GDPR and cooperate with the European Commission through the consistency mechanisms set out in Articles 64-67. There are two notable mechanisms laid out in these Articles:

1. Article 64 lays out the various circumstances under which the EDPB can issue an opinion. These opinions can be given upon request from various bodies (Article 64(2)) or be automatically issued when a supervisory authority intends to adopt certain measures (Article 64(1)). Broadly, their opinions will be issued to provide interpretative clarity to GDPR provisions.
2. Article 65 provides for a dispute resolution process and for the EDPB to issue binding decisions in relation to such disputes.

With respect to the CJEU, it is also important to note that the Court offers the most authoritative interpretation of EU law. Accordingly, the CJEU's capacity to respond to preliminary questions referred by national courts and its ability to enforce the GDPR's provisions via state liability are both important consistency mechanisms.

* Under the Medical Devices Regulation and In Vitro Diagnostic Medical Devices Regulation, these national authorities are called 'competent authorities.'

† For instance, see: *University of Bristol v John Peters* [2018] EA/2018/0142.

Notably, in view of Brexit the UK post-transition period will lack the above consistency mechanisms, the 'UK GDPR' being beyond the CJEU's jurisdiction and the UK ICO at best retaining observer status within the EDPB and European Commission.[23] Consequently, it is unclear how consistent the interpretation of the UK GDPR will be to the referent GDPR. More worryingly, there is also the possibility of two (perhaps divergent) versions of the GDPR – the GDPR and the UK GDPR which may be applicable to the same data.

Genomic data and scientific research under the GDPR

'Genetic data' under Article 4(13), 'biometric data' under Article 4(14), and 'data concerning health' under Article 4(15) all count as special category data and so are subject to further MS variation. Article 9(4) provides for such variation, noting that MS 'may maintain or introduce further conditions, including limitations, with regard to the processing of genetic data, biometric data or data concerning health.' In this way, MS have significant leeway to restrict the processing of special category data through further legislation. Interestingly, the UK's DPA 2018 does not add restrictions, but further flexibility for the processing of data concerning health in Schedule 3, Part 2.

The GDPR provides further flexibility for 'scientific research.' For instance, as discussed in chapter 5, consent for scientific research is allowed to be broader, allowing controllers to state the purposes for processing with less specificity.[24] Further, Article 89(2) allows MS to legislate derogations from and provide for restrictions from certain data subject rights where personal data is processed for scientific research. Notably, the GDPR requires that these derogations and restrictions only occur where the rights would otherwise 'render impossible or seriously impair' the purposes for processing and must be accompanied with 'appropriate safeguards.'[25]

In addition, the UK's DPA 2018 also adds further definition to what 'appropriate safeguards' must be in place if the scientific research is to benefit from Article 89. For instance, Section 19 DPA 2018 specifies that research that will likely cause 'substantial damage or distress' to data subjects cannot benefit from Article 89.[26]

Moreover, research that is 'carried out for the purposes of measures or decisions with respect to a particular data subject' also cannot benefit from Article 89 unless this research constitutes 'approved medical research'[27] namely research carried out with the approval of a research ethics committee or another similar body.[28] Importantly not all Member States have the same derogations or restrictions for scientific research.

2.2 Conclusions

This overview has introduced aspects of data protection law that our research has identified are currently having, or are likely to have, an impact on genomic data processing. In the chapters that follow, we discuss parts of the law in greater detail and their application to the genomics sector. The results of our analyses are also informed by the key impacts identified through our interviews, and policy workshop (described in more detail in the Annexe to this Report), and through evaluation of relevant literature and the media.

Determining how the GDPR applies to the generation and sharing of genomic data involves the interpretation of key provisions (and relevant national laws) and consideration of how they apply in context. There are a range of sources that can help. First, the recitals (background text) to the GDPR set out further detail about the intention of the lawmakers and, although not legally binding text itself, aids the interpretation of the Articles that follow. Second, the opinions and guidance published by national SAs like the Information Commissioner's Office (ICO) and by the EDPB (and its predecessor WP29) are very influential.

Such guidance may also require some interpretation and, at the current time, much of it is out of date since it related to the Data Protection Directive [Directive 95/46/EC], and should be read with caution. Third, the decisions of both the European Court of Justice and national courts or tribunals are important, and may interpret or reject the opinions and guidance of the regulators. These must also be used with caution as they may only be directly binding in relation to specific circumstances and currently the vast majority of judicial decisions are based on the previous legal framework. We draw on these sources, expert insights from our research and the academic literature to discuss some of the main impacts of the GDPR on genomic data processing, beginning with the impact of its territorial scope and approach to data 'controllership'.

3. When and where does the GDPR apply?

An initial but crucial set of challenges for genomic medicine and research are knowing when the GDPR applies.

We will address the 'material scope' of the Regulation: 'personal data', in the next chapter. In this chapter we consider how the GDPR may apply to individuals and organisations within the EU/EEA who are involved in the processing of genomic or health data elsewhere in the world (the territorial scope of the GDPR) and the potential responsibility of those who do not process data themselves but who work with others to determine the purposes and means of processing (data controllership) in genomics collaborations.

3.1 The 'territorial scope' of the GDPR

The GDPR governs the processing of 'personal data' of all natural persons (living individuals), wherever they are in the world, by individuals, businesses and institutions established in the European Economic Area (EEA).* The GDPR also extends its jurisdictional reach in certain circumstances to data controllers or processes 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.

This means that the GDPR will frequently govern the activities of data controllers or processors who are established in the EU/EEA even if the individual data subjects are elsewhere in the world. To a lesser degree the Regulation may impact on the activities of controllers or processors who are based in third countries (for example in North America) if they 'target' data subjects on EEA territory with an offer of goods or services, or, target them by monitoring individual behaviour.

Key Impact
Our research has found that there is concern about the potential scope of the GDPR applying, for example, to govern EEA-based university researchers who are involved in the processing of data from participants elsewhere in the world.

In the first part of this chapter we analyse the breadth and limits of the GDPR's territorial scope.

The territorial scope of the GDPR is laid out in Article 3:

> **Article 3 Territorial scope**
>
> 1. This Regulation applies to the processing of personal data in the context of the activities of an establishment of a controller or a processor in the Union, regardless of whether the processing takes place in the Union or not.
>
> 2. This Regulation applies to the processing of personal data of data subjects who are in the Union by a controller or processor not established in the Union, where the processing activities are related to:
>
> a. the offering of goods or services, irrespective of whether a payment of the data subject is required, to such data subjects in the Union; or
>
> b. the monitoring of their behaviour as far as their behaviour takes place within the Union.

Art 3 splits the 'territorial scope' of the GDPR in two ways; Art 3(1) applies the GDPR according to the territory and jurisdiction in which processing takes place (the establishment criterion), whereas Art 3(2) applies the GDPR to processing which targets data subjects within the EEA (the targeting criterion).[29]

* The EEA includes the EU and Norway, Lichtenstein and Iceland.

'Establishment'

According to Art 3(1), processing is governed by the GDPR when it takes place 'in the context of the activities of an establishment of a controller or processor in the Union'. It is clear that the processing does not have to take place on the territory of the EU/EEA itself, all that is necessary is that it takes place in the context of the activities of an 'establishment' in the EU/EEA.

'an establishment … in the Union'

The CJEU has adopted a flexible definition of 'establishment' rather than a more formalistic approach, for example, based on the place of legal registration of a business or institution. Recital 22 clarifies that an establishment implies the 'effective and real exercise of activities through stable arrangements' and the 'legal form of such arrangements' is not the determining factor.

What matters here is that there is some form of 'stable arrangement'. The CJEU have confirmed that this must be interpreted in light of the specific nature of the economic activities or services concerned.[30] As the EDPB explained in their recent guidelines, the threshold for 'stable arrangement' can be quite low; the presence of a single employee or agent can constitute 'stable arrangements' in certain contexts, for example in the provision of online services.[31-32] There is certainly no need for a legally registered branch or subsidiary in the Union for there to be an 'establishment' in the Union.

In the genomics and health contexts, a broad approach to establishment could perhaps apply the GDPR to processing conducted by non-EU service-providers or institutions, if, for example, they have a subsidiary or some employees in the EU/EEA. For example, if data are being processed for research purposes in North America but the research institute also employs some researchers in the EU, it is possible that the GDPR will apply to the North American institution's processing if it can be shown that the processing takes place in the context of the activities of the establishment (the agents) in the Union.

This could also be the case if a non-EU processor (for example a cloud service provider) has an establishment in the Union and processing can be shown to take place in the context of its activities. Another example might be a direct-to-consumer genetic testing service which has agents in the Union, however, they would be likely to fall under the targeting aspect of the territorial scope in Art 3, as we discuss further below.

'In the context of the activities of an establishment … in the Union'

As with the interpretation of 'establishment' itself, the CJEU has adopted a 'teleological' approach to the interpretation of determining when activities take place in the context of an establishment in the EU: this means that legal provisions are not read literally but are 'understood in the light of the purpose, values, legal, social, and economic goals these provisions aim to achieve.'[33] Drawing on the objectives of EU data protection law in ensuring the effective and complete protection of the fundamental rights and freedoms of natural persons, including their right to privacy, even under the Directive the CJEU interpreted 'in the context of the activities of an establishment' broadly to avoid a circumvention of rules.[34]

As discussed in chapter 2, this will now require the provisions to be considered in light of the similar objectives of the GDPR, which include the protection of the fundamental right to protection of personal data (Art 8 of the Charter and Art 16 (1) of the Treaty on the Functioning of the European Union). However, the EDPB caution that there will need to be a balance so that the territorial scope is not interpreted too broadly and that the 'existence of any presence in the EU with even the remotest links' to the activities of a non-EU entity is sufficient to bring this processing within the scope of the EU law.[35]

EU law attempts to walk a tightrope between ensuring effective protection of fundamental rights on one hand, and avoiding EU legal overreach on the other. Unfortunately, as with many aspects of the GDPR, the proper balance can only be properly considered in context but there are some indications from the case law and EDPB guidance of the factors that that are important and which could be useful to consider in the context of genomic healthcare and research.

For example, it could be that a case analysis shows that the activities of the establishment in the Union are 'inextricably linked' to those of the non-EU entity. The CJEU found that the activities of Google Inc.'s Spanish subsidiary, which was established to market advertising services and provide other commercial functions, were inextricably linked to the operation of Google's search engine outside the EU because those commercial services rendered the operation of the search engine economically profitable and because the operation of the search engine enabled advertising space to be sold.[36] If this is the case, then there is no need for the processing to be carried out by the 'establishment' itself. By analogy, it would seem likely that the marketing/revenue raising link could apply to capture the processing of data by, for example, a direct-to-consumer genomics service provider outside the EEA if some of the revenue raising activity is carried out by their agents in the EEA.

Another example is given by the EDPB in their guidance:

> 'Example 5: A pharmaceutical company with headquarters in Stockholm has located all its personal data processing activities with regards to its clinical trial data in its branch based in Singapore. In this case, while the processing activities are taking place in Singapore, that processing is carried out in the context of the activities of the pharmaceutical company in Stockholm i.e. of a data controller established in the Union. The provisions of the GDPR therefore apply to such processing, as per Article 3(1).'

In this example, it is made clear that the EEA based pharmaceutical company has control over the purposes and means of data processing and that therefore this processing is firmly within the 'context of the activities' of the EEA based organisation. What is less clear is what the case would be if the EEA based entity was not the headquarters of a group but perhaps just one part of an international genomics collaboration. If the EEA entity's activities are inextricably linked with the overseas processing, for example because the data are being used in research by researchers at an EEA institution, then the GDPR would be likely to apply. However, it is much less clear if the GDPR would apply to govern overseas processing where an EEA-based collaborator is merely advising other collaborators, albeit having some influence over the means and purposes of processing. To what extent would that processing take place in the context of their activities, even if this is not a core part of their business/operating model?

The 'teleological' approach to interpreting Art 3(1) would draw on the fact that the GDPR aims to protect the fundamental rights and freedoms of all natural persons 'whatever their nationality or place of residence' (the wording used in recitals 2 & 14). Against this backdrop it may more often be the case that some involvement from an EEA establishment in the processing of individuals outside the EEA will mean that the processing is governed by the GDPR. As we discuss in the second part of this chapter, the wide approach the CJEU has taken to determining data controllership could lead to the scope of the GDPR applying very broadly, across international genomics collaborations. This is something that requires urgent attention from authorities and those engaged in international genomics initiatives.

Another way that Art 3(1) may apply is to govern the processing in the context of the activities of an EEA-based processor (whether this is actually carried out by the EEA based entity or its overseas but inextricably linked establishment). The EDPB discuss this and their advice is that the processor obligations will apply to govern the activities of the EEA-based processor but not a controller who has commissioned the processing if they are not established in the EU.[37] This suggests that, for example, an EEA based genetics service which is contracted to process and analyse genetic material by a healthcare provider outside the EEA would have to comply with the GDPR's processor obligations but the healthcare provider is not governed by the EU Regulation.

'Targeting'

The second part of Art 3 applies the GDPR, in certain circumstances, to processing by a controller or processor outside the EEA, even where they do not have a branch or connected entity within the EEA/EU. The circumstances are referred to as 'targeting' of data subjects in the EEA/EU, either by offering goods or services to those data subjects, or, by monitoring the behaviour of data subjects within the Union. This means that, for example, direct-to-consumer (DTC) genomics services based outside the EEA will be required to comply fully with the GDPR when processing EEA/EU data subjects' data and any health app providers who draw on the live data of users while in the EU/EEA (regardless where those users are from) will be within the scope of the GDPR. One of the ambiguities in this part of Art 3 is that it will apply where processing 'relates to' targeting activities, as well as where it directly involves those targeting activities.

'Offering of goods and services'

The GDPR applies to the offering of goods and services 'irrespective of whether a payment of the data subject is required' (this is frequently the case in a world of data as payment).[38] What is not as straightforward, is determining when goods or services are offered to data subjects in the EU/EEA or whether they are simply and incidentally available to users in Europe. As the EDPB note, recital 23 of the GDPR clarifies that the intention of the provider is a relevant factor and that this can also be deduced from other factors such as the ability to pay in European currencies, use of European languages and references to customers in Europe. As with so many aspects of the GDPR, this can ultimately only be assessed on a case by case basis.

In the EDPB's opinion, where goods or services are offered to the world at large, without specific targeting of data subjects in the EEA, this will not fall within the scope of the GDPR*. They give the example of a Swiss University making an online application platform openly available but taking payment only in local currency and not specifically advertising to students in the EU. What is not clear is how much an intention to provide services to the whole world (not specifically the EEA) would take the processing outside the scope of the GDPR. It seems that any form of marketing to data subjects in the EEA would bring the processing within the scope but if a service provider, e.g. a North American DTC genomics analysis provider, is agnostic about the location of their users, arguably this would also fall outside the scope. Perhaps European Courts and authorities would be tempted to ensure the protection of EEA based data subjects by adopting a broad interpretation of 'offering goods and services … to data subjects in the Union' in these circumstances.

'Monitoring of behaviour'

Recital 24 of the GDPR makes clear that this aspect of territorial scope is particularly relevant to internet tracking and profiling and the EDPB add that this is also clearly relevant to other forms of tracking, including through wearable and other smart devices. It seems clear that this would also extend to apps, such as health apps which gather live behavioural information from the data subject.

The EDPB does not consider any 'online' collection or analysis of personal data to automatically count as monitoring, again something more is needed in the form of a specific purpose such as profiling or carrying out behavioural analysis. There are not likely to be considerations in this part of the territorial scope of the GDPR that are specific to the genomic context. However, if personalised health websites, devices or apps draw on genetic information and continuous monitoring of an individual in the EU, then the GDPR will apply.

As mentioned already, the scope extends to processing activities that relate (directly or indirectly, according to the EDPB) to the offering of goods and services but there is minimal guidance on how this should be interpreted. Again, this means that there will be a case by case analysis of whether the processing in question is sufficiently related to the targeting of individuals in the EEA/EU.

Although the full impact is not yet clear, the territorial scope provisions are likely to have a significant impact on genomic healthcare and research in the coming years. Our research has already identified that the potential for genomics researchers and others involved in international genomics collaborations to fall within the scope of the GDPR even when they are not directly processing or benefitting from data themselves, is a potentially challenging impact of the Regulation. Part of this challenge is due to the broad view of controllership taken in several CJEU cases which could lead to researchers and other collaborators being viewed as data controllers with relatively minimal involvement in such international collaborations.

* See example 14 in: European Data Protection Board. Guidelines 3/2018 on the territorial scope of the GDPR (Article 3). 2018, 17.

3.2 The scope of 'controllership'

We briefly introduced the categories of processor and controller in chapter 2. Although these are relatively straightforward categories at first glance, the CJEU's interpretation of controllership, in particular 'joint controllership' threatens to widen the application of the GDPR to a surprisingly large range of actors. For example, in the genomic context it could be the case that researchers or healthcare professionals who are part of genomic initiatives are considered data controllers, even when they have relatively minimal involvement in setting policies or steering the actual processing involved. This could come as a surprise to those professionals and institutions, particularly, as discussed above, where the processing of data relates entirely to data from individuals outside the EU/EEA.

Key impact
Our legal analysis and policy workshop found that the broad approach to 'controllership' in recent CJEU case law is causing concern that collaborators in genomic initiatives could be found to be legally responsible for compliance with the GDPR where they only have a limited influence on, or involvement with, genomic data processing.

A broad concept of controllership

In their 2010 Opinion, WP29 explained that the 'first and foremost role of the concept of controller is to determine who shall be responsible for compliance with data protection rules, and how data subjects can exercise their rights in practice. In other words: to allocate responsibility.'[39] Over the years, the CJEU has built on this to emphasise that 'controller' is a deliberately broad concept because its objective is to ensure 'effective and complete protection of data subjects'.[40] This has included going beyond an analysis of the influence of an individual or organisation on the processing at hand, by considering the effect of an activity on the privacy and data protection interests of data subjects.

As Bygrave and Tosoni highlight, in *Google Spain* (discussed above) the CJEU emphasised that Google Inc.'s role as a search engine operator was decisive in facilitating profiling of data subjects and the dissemination of those data, and accordingly is 'liable to affect significantly … the fundamental rights to privacy and the protection of personal data.'[41] In the genomics context, if the processing is of sensitive health or genetic data, by analogy, it is likely that activity which is strongly influential in shaping how those data are processed or disseminated will similarly colour the determination of controllership. Although the concept and the definition of controller has hardly altered significantly since the Directive, there may be scope for a different approach to develop under the GDPR given that data processors now have extended and direct obligations in a way that they did not under the Directive. However, for the time being we can only be guided by existing case law and authoritative guidance.

'Controller' is defined in Art 4 (7) of the GDPR:

> 'controller' means the natural or legal person, public authority, agency or other body which, alone or jointly with others, determines the purposes and means of the processing of personal data (Art 4(7))[42]

The key part of this definition is whether an individual or organisation in some way determines the purposes and means of processing of personal data. There is no need for a controller to actually take part in any processing or have any direct view of personal data. As always, the determination of controller status can only take place on a case by case basis, in context.

However, the approach that the CJEU has taken in recent years suggests that the level of involvement in determining the means and purposes of processing may only need to be relatively minimal to confer controller status. In other words, where there are multiple collaborators helping to determine the means and purposes of processing, even relatively limited influence from one party may be sufficient to make them a joint controller of personal data.

Joint controllership

One of the important impacts of the GDPR we have identified for those involved in genomic initiatives is an uncertainty about whether they have sufficient control over the processing of data by another party to be considered a 'joint controller'. This could be crucial where genetic or clinical data are never held or used by one party in an identifiable form but are part of a broader enterprise (e.g. a large-scale sequencing project with separated clinical and research elements) where identifiable data are processed. As under the Directive, the GDPR makes clear that there can be more than one data controller. A controller may act alone or 'jointly with others' (Art 4(7)).

What constitutes 'determining the purposes and means of the processing' of personal data? Guidance and case law demonstrates that a 'substantive and functional approach should be taken' in assessing joint control,[43] and CJEU case law (although under the previous regime) suggests that even very limited involvement in the collection or processing of personal data can suffice. In the *Wirtschaftsakademie* case (C-210/16) the ECJ concluded that administrators of a Facebook fan page are jointly responsible for processing of personal data by Facebook, even though they did not have access to any personal data themselves. And in *Tietosuojavaltuutettu* (C-25/17) the ECJ determined that the Jehovah's Witnesses religious community was jointly responsible for notes made by individual members in door-to-door preaching, even though it neither required or obtained those notes.

The essence of these decisions is that if you 'influence', create an 'opportunity' for, 'encourage' or 'coordinate' the processing of personal data by another, there is a significant chance that you are jointly determining the purposes and means of the processing of personal data. Formal arrangements are not necessary to establish joint control. One consequence of joint control is that the controllers are obliged to determine their respective responsibilities for compliance with the GDPR (Art 26) and communicate them to the data subject. Importantly, it is clear that joint controllers will be jointly and severally liable for all the obligations owed under the GDPR.[44]

Applied to the genomics context, where there are frequently multiple collaborators dealing with data in different forms as part of a shared enterprise, it is likely that many professionals and institutions will cross the threshold of 'joint control.'

However, some participants in our research were hopeful that this expansive approach to joint controllership could be distinguished to the precise facts of the cases and context in which they occur. For example, it may be possible to interpret the approach taken in *Wirtschaftsakademie* as relating specifically to online processing. If so, it is possible that the genomics community could make the case that the appropriate application of the Regulation in this context should be very sensitive to the distinction between determining the means and purposes of processing genomic data and that both must be present to find that an advisor or remote collaborator is in fact a data controller.

> I agree that cases are problematic and several unclear but I think there is scope to emphasise the fact specific nature of those cases and in other contexts you could see things differently.
>
> **Workshop participant**

Further, as another workshop participant noted, focusing on the 'granularity' of control and that various operators may be involved at different stages and to varying degrees, could diminish the risk that an operator is found to be a data controller for the purposes of a whole collaborative project. It would assist the genomics community to have further guidance about the appropriate application of the Regulation in this context and to obtain clarity about what activities are likely to meet the threshold of 'data controller'.

3.3 Conclusions

Our research has identified that one of the impacts of the GDPR on genomic medicine and research is an uncertainty about how broadly the law may apply to individuals and organisations within the EU/EEA who are involved in the processing of genomic or health data elsewhere in the world (the territorial scope of the GDPR) and the potential responsibility of those who do not process data themselves but who work with others to determine the purposes and means of processing (data controllership) in genomics collaborations. In both these aspects, the CJEU and authorities have taken a broad view to ensure protection of data subjects' fundamental rights and freedoms, regardless where they are in the world. Unfortunately, this has the potential to surprise many scientists, researchers and others who have relatively minimal involvement in international genomics collaborations but who may bear responsibility for agreeing how to comply with the GDPR and national data protection law with project partners.

4. When are genetic or genomic data 'personal data'?

The GDPR governs the processing of 'personal data'. Determining when data are 'personal data' is of fundamental importance for those using genomic information.

Our research has identified that professionals are experiencing challenges reaching consensus about when genomic and associated health data are 'personal data', in particular, whether data that have undergone 'pseudonymisation' always remain 'personal data'.

Our legal analysis has identified further potential impacts, including the danger that the reasoning of UK courts in recent caselaw could be applied to determine that genomic data are inherently identifying 'personal data' without any other link to, or impact on, actual individuals.

Another interpretive challenge is that the GDPR explicitly incorporates a category of 'genetic data' for the first time in EU data protection law but again, there are ambiguities in how this category is defined. The consequences for genetic information falling within the scope of 'genetic data' may be significant, both in terms of existing data protection requirements and in terms of the potential for specific regulation or guidance on the processing of genetic data across the EU Member States.

In this chapter we analyse and evaluate the definition and approach the law adopts to determining when data are 'personal data' (section 4.1). We highlight the challenges and factors that regulators, policymakers and data controllers should take into account when assessing the identifiability of genomic data. We evaluate the position of 'data which have undergone pseudonymisation' and argue that this category of data should be capable of being sufficiently de-identified to fall outside the Regulation (section 4.2).

Finally, we analyse the definition of 'genetic data' in the GDPR, highlight the consequences for data within that category and identify some specific challenges that genomic data give rise to under the Regulation (section 4.3).

4.1 When are genetic and health data 'personal data'?

The GDPR only applies to 'personal data' (Art 2(1)) of living individuals.* This means that the first and fundamental assessment when considering if genetic, genomic or health data are governed by the GDPR is whether the data fulfils the definition of 'personal data'.

Key impact
Our interviews and policy workshop found that determining whether genetic/genomic data and associated health data are personal data is a significant challenge for some healthcare and research professionals. Disagreement about whether certain data are 'personal data' is challenging local, national and international flows of genomic and health data.

In this section we consider in turn the requirements of 'personal data', evaluating what this means in the genomic context and highlight some of the particular challenges faced in determining whether genetic/genomic data and associated health data are personal data under the GDPR. To understand 'personal data' we begin with the definition provided in Art 4(1).

'Personal data' are defined as:

> any information relating to an identified or identifiable natural person ('data subject'); an identifiable natural person is one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person (Art 4(1))

This clarifies that a natural person or individual does not have to actually be identified by the data but that that they need only be 'identifiable', which means capable of being identified directly or indirectly based on the data. As well as some more obvious identifiers, names and online identifiers (e.g. cookies), individual-specific genetic 'factors' are also listed as a potential reference point for identification of an individual. However, this does not mean that genetic 'factors' (as Art 4(1) refers to them) will always lead to identification, only that they are one of the range of factors that could be used to identify a person. What Art 4(1) makes clear is that to understand when genomic data are identifiable, a controller is expected to assess whether they 'relate to' an 'identified or identifiable natural person' who can be identified 'directly or indirectly'. We consider these concepts in turn, emphasising particular challenges they raise in the genomics context.

Information must 'relate to' an individual

An important requirement for 'personal data' is that information must 'relate to' an individual. In their guidance on the concept of personal data, the Art 29 Working Party emphasised that this is important as a means of understanding if data are sufficiently linked to an individual to constitute personal data.[45] In many cases the content of the information will be clearly 'about' an individual, for example medical genetic test results are clearly about the person tested.

* Member States may provide for rules governing the processing of deceased persons, see: GDPR, recital 27.

However, other information may not obviously be about a person but they could be used to, or result in an, impact on their rights and interests. According to WP29 and CJEU, this requires a focus on the 'content', 'purpose' and 'effect' of the data,[46-47] to determine whether it is linked to a particular person. This means although the definition of 'personal data' is at first glance agnostic to the intentions of processors and controllers, the purpose or effect of the processing are important considerations in establishing whether the data relates to an individual.

Taking the 'content' of the data first, the ICO discuss this as being a question of whether data are obviously 'about' an individual, focus on an individual or have 'biographical significance'.[48] Some cases under the previous data protection directive and DPA 1998 took a narrow approach to personal data and required that data which was not obviously about an individual should have some focus on the individual, or that it was biographical in some significant sense.[49]

Interpreted in this way, the requirement for data to 'relate' to an individual was a significant qualification of the scope of personal data, so that, for example, a name in a database, with no other information would not necessarily be personal data, even if an individual were identifiable from the data. However, the narrow approach was confined to limited circumstances over time as the European Courts and Art 29 Working Party made clear that the concept of 'personal data' is a broad one.

Now, the position is that the content of the data can either obviously be 'about' an individual (the ICO give the example of medical records being obviously about an individual), or, data can relate to an individual because they have some biographical significance or the individual as their focus.

Moreover, as the CJEU confirmed in the case of *Nowak*, the scope of the concept of 'personal data' is very wide. In that case, examination answers were found to be 'personal data' related to an individual by their content, purpose and effect.[50] The content related to the individual because they reflected the candidate's knowledge, the purpose related to them because the answers were used to evaluate their abilities and the effect related to them because they would impact their rights or interests by influencing their chance of entering a profession or obtaining a post. Moreover, the CJEU found that opinions could constitute personal data if they relate to the data subject.[51]

The interpretation of when data relate to an individual may be important in the genomic context. Just as it is uncontroversial that medical records and healthcare data are obviously 'about' an individual, in many uses of genetic, genomic or health data, it will be obvious that data relate to a person. However, with more technical or minimal forms of data, for example human cell level data or some biological or physiological measurements, it may not be obvious that they relate to an individual. In such cases, the question of whether the purposes or effects of processing such data impact individuals' rights and interests may be important in determining if data are 'personal data'. This is a separate requirement to the requirement we will now discuss, that data relate to an 'identified or identifiable' natural person. However, the purpose or impact of processing on an individual's rights and interests may also be relevant to determining when an individual should be considered to be identified by data.

An identified or identifiable person

As with the previous law, data must relate to an identified or identifiable person in order to constitute 'personal data'. The Art 29 Working Party stated that in general terms, a 'natural person' can be "identified" when, within a group of persons, he or she is "distinguished" from all other members of the group.[52]

As we discuss below, we believe that distinguishing individual-level data from others in a dataset is necessary but not sufficient, without something further, to connect that data to an actual individual. 'Identifiable' refers to the possibility that an individual can be identified, so it is important to appreciate that for most assessments of 'personal data' there must be a risk assessment of the probability (or reasonable likelihood- as we discuss below) that an individual could be identified. Article 4(1) specifies that an identifiable person is one who can be identified either directly or indirectly on the basis of the information. Direct identification refers to the ability to identify an individual based on identifiers that are present in the data (such as a name, or, a combination of other identifiers such as date of birth and address). Indirect identification refers to the potential identification of an individual by combining or comparing the data in question with other information, for example, comparing a genome sequence with publicly available genetic datasets containing names or other identifiers.

Because 'personal data' clearly includes indirectly identifiable information and therefore there must be an assessment of the probability that data are identifiable using wider methods or available data sources, the distinction between direct and indirect identification is not particularly helpful. Indeed, even common identifiers, such as a name, must be assessed in context: 'John Smith' may not identify an individual in a population of millions. It is good practice to start with a consideration of 'identifiers' within data but an assessment of 'personal data' must always consider the prospect of indirect identification. However, as we now discuss, the approach that the English courts have taken to direct identification poses potential difficulties for genomic data.

Direct identification

As we have noted, direct identification generally refers to the ability to identify individuals solely based on identifiers that are present in the data. For this reason, it is common practice across healthcare and research to remove all obvious identifiers such as names and individual patient numbers when de-identifying data. However, a challenge for genetic information is that it is sometimes referred to as an 'identifier' in itself, alongside biometric information like fingerprints or retinal patterns. The Art 29 Working Party referred to DNA data in this way as 'providing information about the human body and allowing unambiguous and unique identification of a person'.[53] They noted that biometric information and DNA could have a dual quality whereby it may be both the content of the information about an individual as well as an element to establish a link between one piece of information and the individual. The 'linkability' of genetic, and particularly genomic information is—as we concluded in our previous research—one of the reasons that genomic data are strongly identifying.[54]

This is because genomic data can contain many markers which can be cross-referenced, for example with publicly available genetic data or because it can be linked to clinical signs and symptoms (phenotypic information) to identify an individual. However, it is important to acknowledge that this does not necessarily make genetic information directly or inherently identifying in and of itself. As the PHG Foundation concluded in our previous report on Identification and genomic data, it is crucial to recognise that although such data allow for identification, something further is needed to link them to an individual.

Despite this, direct identification has often been equated to 'singling-out' or 'distinguishing' an individual from a group or a dataset. In the GDPR, 'singling out' is referred to as a means of directly or indirectly identifying an individual (recital 26) and nothing in the Regulation supports the conclusion that singling out, with nothing to connect information to an actual individual, necessarily constitutes identification.

For example, if a dataset of thousands has a separate row for each individual, it is possible to single an individual out from the dataset. But if there is no identifying information in a row and there is no way to connect that row with any natural person, the data do not identify any person and should not be 'personal data'. Of course, it is true that distinguishing or singling out an individual record will very often mean that identification is reasonably likely (and therefore will constitute indirect identification- see below). This close connection between singling-out or 'individuating' and identification has been emphasised in guidance which warns that individuation may often lead to identification.

For example, in their 2007 opinion on the concept of personal data, the Art 29 Working Party noted that it will not be necessary to have a name in all cases to identify an individual and that 'this may happen when other "identifiers" are used to single someone out'. They warned that is particularly the case online where online identifiers do not require an actual name but can distinguish an individual from others.[55] The ICO similarly state that '[a]n individual is also identifiable if you are able to distinguish that individual from other members of a group.'[56]

Unfortunately, this guidance has been taken to an extreme by the courts in England and Wales who have elided the concepts of individuation and identification and removed the necessary distinction between the two. In recent cases the courts have concluded that 'individuation' is a form of identification under the previous Data Protection Act. Because there is not thought to be any significant change in the concept of 'personal data' under the GDPR, there is a very real risk this may be applied to data under the new Regulation.

In *Vidal-Hall v Google Inc* the Court of Appeal considered the identifiability of 'browser-generated information' and whether 'identification for the purposes of data protection is about data that 'individuates' the individual, in the sense that they are singled out and distinguished from all others.'[57]

Both these criteria were met in this case, and in a 2019 challenge to use of automated facial recognition technology by South Wales Police (*R (Bridges) v Chief Constable of South Wales Police*) the High Court drew on Vidal-Hall as authority for 'individuation' as a sufficient test for determining whether data are directly identifiable 'personal data'. This meant that the processing of biometric facial data was processing of personal data under the DPA 2018 even though such data were immediately deleted and never connected to an individual's name or other form of record.

In our view, it is possible that the Court could have reached the same conclusion without eliding individuation and identification by finding the biometric facial data in question reasonably likely to indirectly identify an individual (discussed further below). Although this case concerns law enforcement processing and is not directly governed by the GDPR, much of the court's reasoning is relevant not only to biometric data but also to interpretations of 'directly identifying' under the GDPR.

The judges drew on some of the guidance mentioned above as well as CJEU case law to conclude that biometric facial data 'permits immediate identification of a person' and is therefore (directly) identifiable personal information.[58-59] This was despite the evidence that such data permitted no tangible or intangible identification of an individual in a crowd unless they were matched with other information held on a 'watchlist'. The judges were more cautious about other forms of information and they expressly distinguished facial data from dynamic IP addresses or online identifiers, which they found to be qualitatively different.[60]

It is our concern that genomic or 'genetic data', which is one of the special categories of data alongside biometric data and which is frequently referred to in guidance as highly identifying alongside biometric data, could be treated in a similar way by the courts. This is because following these decisions, English courts might conclude that individual genetic or genomic information which allow the singling-out or 'individuation' of an individual are 'personal data' even though there may be a very low likelihood that such information would ever be connected to an actual individual or their other personal information. This could lead to some genetic or genomic information being treated as inherently identifiable.

Key impact

Our analysis has highlighted the potential that the courts in England and Wales could determine that genomic data are directly, inherently identifying, if they follow their recent approach to biometric facial data.

Are genomic data directly identifying?

For example, if an individual's whole-exome or whole-genome sequence forms part of a database but all other records and identifiers have been fully deleted so that all that remains is sequence data, this data could easily be 'individuated' or singled out but—we argue—it cannot be connected to a natural person without something further which can relate it to them.

However, the latest ICO guidance on genetic data states that:

> 'in practice, genetic analysis which includes enough genetic markers to be unique to an individual is personal data and special category genetic data, even if you have removed other names or identifiers.'

In our view, although there is an increased risk of identification from such data, it would be a mistake to treat it as inherently identifying simply because it is unique, without a further assessment of the likelihood that it could lead to identification in reality. This is particularly important given that it could have significant consequences for important uses of genetic data in health research if all genomic data are treated as personal data.

We believe that there should be careful scrutiny of the position of 'personal data' and identifiability under the GDPR and that, although some information is highly identifying and must be protected, the need for a further factor to connect individuated information to an individual must not be overlooked. In our view, as Jeff Skopek, a leading scholar of privacy and anonymity emphasises, 'identification requires more than individuation'.[61]

Skopek refers to 'individuation' as a matter of the 'uniqueness' of a 'trait' and adds that for identification a second factor is required. This 'second relevant factor is the extent to which a unique trait is connected to other relevant information'.[62]

This does not have to be a connection to a person by name. As Skopek suggests, ' "identifying" a person can have both tangible and intangible dimensions' and the 'type of identification' turns on the 'type of access to the person' that is sought. In other words, for many purposes, identification may be nameless. For example, the use of online identifiers to identify an individual does not necessarily require a linkage to the physical individual but certainty that this is a specific individual who can be categorised, creating the ability to 'attribute certain decisions to him or her', in the words of the Art 29 Working Party.[63]

In practice, this is the tenor of much of the guidance which, although emphasising the likelihood of identification via individuation, is concerned by the impact of singling-out, distinguishing or individuating, an individual. For example, although the authors of the explanatory report to the amending protocol to Convention 108 introduce yet another term, 'individualise' or 'individualisation' which they equate with 'single out' as a form of identification that is not of the individual's civil or legal identity, their focus is on how such individualisation allows an individual to be treated differently.[64]

And as noted above, the ability to 'attribute certain decisions' to an individual is a reason that the singling out of an individual using online identifiers is considered by WP29 to be a form of processing of personal data.

At the more fundamental level, data protection law is concerned with the impact of the use of personal information on the rights and freedoms of individuals and this is partly the function of the requirement that data 'relate' to an individual. As we discuss above, this requires some consideration of the purpose or effect of processing for example, if data are used to 'evaluate, treat in a certain way or influence the status or behaviour of an individual' (WP29).[65]

We believe that it is logically coherent for a similar assessment to be made about the identifiability of an individual: if there is no possible connection with, or impact on the rights and interest of, a natural person resulting from data being individuated or singled-out from other data in a larger pool, this should not be treated as 'personal data'.

We suggest that the way of resolving some of the confusion caused by these terms and on the part of UK courts about the identification and individuation is twofold.

First, the position which Skopek lays out, whereby singling-out or distinguishing information as unique (individuation) is insufficient for identification unless this unique information somehow relates to an individual, should be adopted. Is there some meaningful connection, either tangibly or intangibly to an individual or do they remain anonymous in a crowd with no possible decisions to be made about them?

Second, data should not be viewed in isolation and the distinction between direct and indirect identification is an unhelpful analytical frame for assessing 'personal data' which should be avoided. Even common 'direct' identifiers such as a name may not actually identify an individual if the pool is large enough and there is no further information to distinguish one individual from several. Moreover, it is rarely helpful to focus on the identifiability of the information held by the controller in isolation from other sources that could be available to indirectly identify an individual.

An assessment of the identifiability of information should always consider it in the round, including the function and potential impact of processing on the individual.

This is not to say that there should not be a close analysis of individuated data and a cautious approach to determining that there is no reasonable likelihood of it leading to identification, but the two concepts of individuation and identification should not be elided.

> It is our strong recommendation that genetic, and particularly genomic information should not be viewed in itself as inherently or directly identifying without some further link to or impact on an individual

We note that there is an argument, as Hallinan and colleagues discuss that genetic data 'remain a biological representation of one single individual even if the link to civil identity has been cut.'[66] With his colleagues, Hallinan challenges the assumption that a link can be cut between genetic data and an individual, and concludes that a range of interests related to genetic data are not fully protected by data protection law which generally treats anonymity as the severance between data and 'civil identity'.

We agree that there are some communal, individual and on-going interests in genetic data that data protection law is plainly ill-equipped to protect but, as Hallinan and his colleagues conclude, to extend the protection of genetic data to non 'personal data' is clearly not the intention of the Regulation.

It is generally true that data protection law is concerned with information that is linked in some way with an individual's civil identity and not—as the genetic data may remain—their biological identity. However, the scope of 'personal data' is broad enough to incorporate information that relates to individuals through the purposes or effect of processing, so genomic data are potentially safeguarded more broadly than at first sight.

An example of individuation v identification:

An individual's whole genome has been sequenced and the data is part of a dataset of genome sequences where it is possible to distinguish one genome from another.

If …

a) The data are stored without any identifiers (such as names or medical record numbers) or other information associated with the data

then…

an individual's genome sequence should not be treated as directly identifying information or 'personal data' unless there is an assessment of the chance that the genome data could be combined or compared with other information (including the background knowledge of individuals who may have access to the data) to identify an individual.

But if…

b) Decisions may be made using the individuated genome sequences which will impact whether individuals will be offered experimental treatment or be informed about certain pathogenic results

then…

this information is likely to be 'personal data' because the distinguishing of the individual's data allows it to be connected to them in a way which has an impact on their interests or fundamental rights.

Unfortunately, ambiguity about wording abounds. For example, Shabani and Marelli refer to 'individuation' 'as the connection between a record and an individual, such as a unique code assigned to a tissue sample', whereas 'distinguishability' is the 'ability to distinguish records from one another'. [67]

We adopt Skopek's definition of individuation as the unique distinction of information from others rather than the connection of that information to an individual but it is this sort of ambiguity that may have led the English Courts to their position.

Indirect identification

Leaving aside the issue of whether genetic information may be inherently identifying without further information, most of the challenges for the identifiability of genetic/genomic information relate to the possibility that they may be indirectly identifiable, in combination with further additional information, including the background knowledge of family members or professionals.

As we mentioned above, one of the challenges of assessing identifiability is that it requires a risk assessment of the probability that an individual could be identified through combination with other information or using other means. The GDPR sets the level of probability of identification for data to constitute 'personal data' as at least 'reasonably likely' but as we now discuss, this is to be assessed in context and according to the means that may be available now and, in the future, to identify an individual.

Key impact

Interview and workshop participants demonstrated that there is a challenging diversity of views and approaches regarding when genomic data are reasonably likely to lead to identification of an individual.

A reasonable likelihood of identification?

The approach that should be taken to assessing identifiability is explained in recital 26 of the GDPR:

> To determine whether a natural person is identifiable, *account should be taken of all the means reasonably likely to be used*, such as singling out, either by the controller or by another person to identify the natural person directly or indirectly. To ascertain whether means are reasonably likely to be used to identify the natural person, account should be taken of all objective factors, such as the costs of and the amount of time required for identification, taking into consideration the available technology at the time of the processing and technological developments. [Emphasis added]

As this explains, there must be a determination of whether there are 'means reasonably likely to be used', either by the controller or by another person. As part of this, 'account should be taken of all objective factors' required for identification. These include the technology that may be available to connect data or make inferences and a range of contextual factors that have been highlighted as relevant to an assessment of whether data are identifiable. Although the reasonable likelihood relates specifically to the 'means' in this text, in practice the same general standard of probability is applied to the overall assessment of the likelihood of identification by 'any'[68] party who could access the data.

Determining when identification is 'reasonably likely' is not straightforward. Case law and guidance help to set the interpretative limits to this concept and demonstrate when identification is not reasonably likely. For example, WP29 have referred to 'mere hypothetical possibility',[69] and the UK courts to a 'remote' chance[70] of identification, which would be insufficient.

However, there have been signs from WP29 and the CJEU that a relatively low level of risk of identification could result in data being classified as 'personal data'. In 2014 WP29 issued an Opinion on Anonymisation Techniques which emphasised that anonymisation should be 'irreversible' and the European Court of Justice seemed to adopt an expansive interpretation of 'reasonably likely' in the case of *Breyer*. In this case the Court made clear that not all information enabling identification need be in the hands of one person for it to be 'personal data.'

Furthermore, the existence of legal channels allowing online media service providers to request that German State authorities obtain information from an internet service provider to identify a person from their dynamic IP address, constituted means that may be reasonably likely to be used to identify a data subject. This was despite the evidence that the legal channels would only be used in exceptional circumstances, such as in the event of cyber-attacks, and that otherwise German law did not allow an internet service provider to transmit identifying data to an online media service provider. Approaching the question in the negative, the court said it would not be reasonably likely if identification was practically impossible or if 'the risk of identification appears in reality to be insignificant.'[71]

We discuss this case further below in the context of 'pseudonymisation'. In relation to the standard of likelihood of identification that is required for data to be considered 'personal data', *Breyer* could be read as setting a very low level of risk as the threshold for personal data.[72] If 'realistically insignificant' is the standard of risk to be applied then a great deal of information will be captured as personal data, not least in the genomic field. It could also be argued that the express inclusion in the GDPR of pseudonymisation and data which have undergone pseudonymisation as a form of 'personal data' (discussed further below) sets a benchmark for 'personal data' more generally, if such data, which have been coded, and heavily safeguarded against individual identification, normally constitute personal data

Key impact

Following the CJEU's judgement in the *Breyer* case, some of our research participants argued that there is a low threshold for risk of identification required for data to constitute 'personal data': any risk greater than practically or realistically 'insignificant' would lead to information being (or remaining) 'personal data'.

Part of the problem is that differences in the specific contexts for processing discussed in the guidance and addressed in case law inevitably lead to uncertainty when determining an overall standard.

For example, the decision in *Breyer* has to be viewed in the specific context of processing dynamic IP addresses. Indeed, in a recent English case, Recorder Douglas was asked to consider how *Breyer* should be applied to an analogous situation where copyright owners and partners sought access to identifying information held by an ISP which would enable them to identify individuals from their IP addresses, in order to pursue allegations of copyright infringement.[73] Although for the sake of argument he proceeded on the basis that the 'mere fact' that a party is able to obtain a court order to identify a natural persona under the English civil system would lead to IP address data being 'personal data', Recorder Douglas was cautious about importing the result of *Breyer* into domestic law. His view was that the result of *Breyer* turned on the 'specific factual aspects of the German legal system.' In his view it would be surprising if the ability to obtain a court order for disclosure of identifying information from an ISP led to otherwise unidentifiable IP addresses being classed as 'personal data'.

Similarly, the WP29 Opinion discusses the standard of identifiability in the context of specific techniques (such as aggregation, differential privacy and hashing). Stalla-Bourdillon reports that national supervisory authorities themselves are viewing the Opinion as relevant to three types of re-identification risks—singling-out, linkability and inference—but that another 'broader approach' could also be taken to assessing risk.[74] It may be that context specific descriptions of the risk threshold for identifiability, according to the form of data or the technical nature of the processing, will be developed over time as the GDPR beds in. For now, all we can do is refer to the wording of the GDPR, the guidance of the EDPB and case law of the CJEU for advice.

What is clear is that an assessment of the risk of identification can only be made in context.* In this light we think there are several factors, drawn from the guidance, legal decisions and the literature, which are particularly relevant to the assessment of whether genetic/genomic and phenotypic data are reasonably likely to identify an individual. These are the 'richness' of the genetic and any associated phenotypic data, restrictions on who may access the data and what additional information may be used with the genetic data to identify an individual.

The 'richness' of genetic data

As discussed in chapter 1, complete genome sequences are unique and capable of being connected to many and diverse data. This doesn't make them inherently identifying but the richness of this data can make them 'strongly identifying', depending on the data with which they can be connected.† Even more limited genetic data may be highly identifying, for example very rare variants, and even a relatively limited level of more common data may enable identification; for example, researchers demonstrated that under 100 single nucleotide polymorphisms (SNPs) are required to distinguish an individual's genetic data as long ago as 2004.[75]

* The ICO provide some useful guidance as a starting point for data controllers:

The Information Commissioner's Office. What is personal data Available from: https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/what-is-personal-data/can-we-identify-an-individual-indirectly/ [Accessed 11th May 2020].

† However, it is noted that limited genetic data may also be highly identifying if, for example, they are relate to particularly rare genetic variants, see: Finnegan T, Hall A. Identification and genomic data. PHG Foundation. 2017, 5.

A particular challenge for genetic data is the general truth that the richer the data, the greater the utility for healthcare and research. This also applies to the level of clinical data that may be associated with genetic or genomic data. However, the richer the individual level genetic and clinical data, the greater the potential for linkages, inferences and singling-out of an individual.

As we discuss below, this places a significant burden on the safeguards that are used, for example pseudonymisation, in complying with data protection law and the contextual analysis required to determine if such data are sufficiently de-identified to fall outside the category of 'personal data'.

Who can access the data?

One factor that has been important for the courts and in the guidance is consideration of who may be likely to identify an individual. Recital 26 refers to identification by the 'data controller' or 'another person'.

The 'data controller'?

The previous UK law only referred to identification by the data controller,[76] causing some confusion where data were disclosed in a de-identified form – by removing any links to identifying information. In these circumstances, it may sometimes be that case that re-identification is possible by the controller because they have retained identifying information.

The courts and authorities have sometimes distinguished between scenarios where this is a reasonably likely possibility and those where it is not according to the context or purpose of the processing. This frequently relates to an assessment of data that have been shared by a controller in a coded form, but where the controller has retained the key (see further on pseudonymisation below).

For example, the Article 29 Working Party distinguished between the disclosure of data without a 'key' (linking coded data to the individual) for research, where there are organisational, professional or contractual safeguards to prevent the controller re-identifying an individual, from disclosure where re-identification by the data controller is specifically contemplated in the research protocol (e.g. in case medicines turn out to pose a danger in a clinical trial).[77] In the latter case, WP29 specifically advocated that data remain personal data.

In practice, courts in the UK have grappled with some confusion over whether the potential for re-identification of data by the disclosing data controller is relevant to the assessment of identifiability in the context of the case before them. On one hand, in his leading judgment in *Common Services Agency v Scottish Information Commissioner*,[78] Lord Hope suggested that in the disclosure of statistical information by a Government agency in compliance with a freedom of information request, the assessment is of whether the data are identifiable by the data controller as well as third parties. On the other hand, subsequent courts and tribunals in similar cases have taken a different approach, distinguishing between the data remaining 'personal' in the hands of the controller but being sufficiently de-identified to fall outside this category in the hands of recipients.[79]

Perhaps these cases can be reconciled with a contextual approach to identifiability that asks whether it is reasonably likely that the data controller will attempt to re-identify data? In many cases this would be unlikely. For example, where data are disclosed in response to a freedom of information request, the controller already possesses an identifiable dataset and will have no need to attempt re-identification of the published data. Also, where data are disclosed for research there will often be technical, organisational and legal safeguards to prevent the controller attempting to re-identify data which also makes identification unlikely. In other circumstances, the potential desirability of clinical re-contact may require re-identification by a data controller (although it may not if they are simply acting as an intermediary between clinicians and/or researchers).

'another person'

Most often, the focus of an assessment of identifiability will be on the potential for identification by recipients of data or third parties – 'another person' in the text of the recital. The range of people who may be in a position to attempt identification using data will depend on the circumstances. In some cases this could include recipients of data for research, in others it may extend to anyone who can access data that have been published openly. The law requires an objective assessment of the means that such parties might be reasonably likely to have available to identify an individual, taking into account any technological developments that could assist them.

The UK Information Commissioner's Office adopted the concept of a 'motivated intruder' as one way of approaching this assessment in cases where many people may be able to access the data (for example on publication of information following a freedom of information request).[80] They referred to such a person as someone who has no prior knowledge but wishes to identify an individual from the data; the motivated intruder test assumes that this person is reasonably competent, has access to information resources (such as the internet/public documents) and would employ investigative techniques, such as enquiring after people with additional knowledge of the identity of the data subject. They are not assumed to have any specialist knowledge, such as computer hacking skills, specialist equipment or to resort to criminality.

As both ICO and the Art 29 Working Party[81] noted, some data will be more attractive to a motivated intruder, so to a degree there should be consideration of potential motivations and how this may alter the level of skill and knowledge involved (for example, journalists or campaigners). The courts and tribunals have often applied the motivated intruder test but there is some uncertainty, for example, whether it should be assumed that there is a motivated intruder in a case where it is not obvious there would be.[82]

There are further limits to relying solely on a 'motivated intruder' approach. As the WP29 noted, it may be a speculative test[83] and could either lead to assessments that are not grounded in reality or false assurances that data are not at risk of identification if an intruder is not foreseen by the controller. This could be a particular problem with data that are available online and potentially accessible to many people with relevant skills.

There are other reasons to question whether the 'motivated intruder' is fit for purpose, or potentially too narrow in certain cases; to what extent is it appropriate to rely on the expertise (or lack thereof) of a data controller to contemplate all possible motivations for attempting to access and identify data? Is it sensible to exclude the possibility of criminal actions in attempting to re-identify data if data are particularly attractive or valuable?

The ICO is in the process of updating guidance under the GDPR and their latest online guidance does not refer to a 'motivated intruder'. Instead they refer to a 'determined person with a particular reason to want to identify individuals' and that 'it is likely that there will be some who are willing to use extreme measures to identify that individual.'[84] This is a welcome updating of the concept and a useful tool for data controllers in assessing risk of identification.

The 'data subject'?

Another person who may be able to identify an individual from data is the data subject themselves; their detailed prior knowledge can make it easier for them to self-identify where others could not. Could this lead to information being categorised as 'personal data' if accessible by the data subject in cases where it would otherwise be anonymous? As under the previous law, the GDPR does not exclude this possibility but the guidance and judicial decisions are not completely clear about when an assessment of identifiability should include the potential for self-identification or not.

The issue was raised in *Queen Mary University of London v The Information Commissioner and Alem Matthees*[85] a case concerning a freedom of information request for clinical trial data. The University argued that there was the potential for participants to self-identify from the data but the Information Commissioner (on the basis of the previous law) disagreed and added that 'in any event self-identification … is insufficient for the data to be deemed personal, as identification for the purposes of the Act must be able to be made by a third party.' The tribunal did not directly address this point but it is not clear that self-identification would always be irrelevant.

In some cases, self-identification could lead to harm if the individual would learn new information, e.g. about their position on a scale of risk or their place in a cohort. Indeed, genomic information is particularly likely to reveal information about individuals that they do not already know and the unexpected discovery of, for example, disease predisposition could lead to worry and harm.

For controllers of genomic information, it would be sensible to consider the potential for self-identification by data subjects, particularly in circumstances where the information would be unexpected or unwanted and there is a risk of harm.

The availability of 'additional information'

A key component of the assessment of whether there are 'means reasonably likely to be used to identify an individual' is whether there is additional information available that could be combined with the data to identify one or more individuals. This is intimately connected with analysis of data that have been 'pseudonymised' (see below). However, it may require a much broader assessment in some contexts, for example, it could require a consideration of whether genetic data can be combined with publicly available genealogy data that could allow identification by inference.[86]

We suggest that a two-part assessment is required. First, is there additional information which could be used to identify an individual? Second, is it reasonably likely that such information could be obtained and used to identify an individual?

The existence of safeguards

One factor that can be influential in limiting the circumstances where data are considered identifiable and personal, is the existence of safeguards that restrict the ability of recipients of data, or other persons, to identify an individual. This was emphasised by the ICO in their anonymisation code of practice, as being particularly useful in 'borderline cases' where it is very difficult to determine if it is likely that re-identification will take place. There are a range of possible safeguards: for example, obligations of confidentiality, restricted access and restrictions on re-use or re-identification may all be imposed by contract or required by law and can help to reduce the likelihood of a misuse of information and re-identification.[87] It will also be highly relevant if there are legal barriers to prevent a person obtaining any additional information that could be used to identify an individual.[88]

Though research data are frequently safeguarded by such restricted access procedures, where data have been requested on the basis of freedom of information law, both the Information Commissioner and tribunals have sometimes found that such data may be published without those safeguards in place, as non-personal data. As the public authorities in those cases have frequently complained, this would seem to minimise the importance of safeguards in reducing the identifiability of individuals,[89] although this approach could perhaps be confined to the specific context of freedom of information requests where there are competing legislative and public interests. We discuss these and further safeguards that could help reduce risks of re-identification in chapter 8.

One form of data safeguard that is promoted by the GDPR is pseudonymisation. Our research has shown that the status of pseudonymised data is one of the uncertainties challenging uses of genomic data. There has been significant debate whether its inclusion in the GDPR means that pseudonymised data are always 'personal data'[90] and participants in our research identified this is a key challenge for them.

4.2 Pseudonymisation

Pseudonymisation is generally understood to involve the removal and replacement of real-world identifiers with a key, cipher or code so that an individual cannot be easily identified from the data without that key or code. One of the significant changes under the GDPR is the inclusion of 'pseudonymisation' as a measure to safeguard data and reduce risks[91] to data subjects rights and interests. However, this has in itself created some uncertainty in the literature and on the part of our research participants about whether pseudonymised data should be treated the same as other forms of de-identified data (such as aggregated data) and assessed according to the reasonable likelihood standard of identification, or, if a different standard should be applied to pseudonymised data so that they almost always remain 'personal data'.

Key impact
Our interviews, workshop and review of the literature have identified that there is uncertainty about data that have been pseudonymised and whether they always remain 'personal data'.

Pseudonymisation is defined in Article 4:

> 'pseudonymisation' means the processing of personal data in such a manner that the personal data can no longer be attributed to a specific data subject without the use of additional information, provided that such information is kept separately and is subject to technical and organisational measures to ensure that the personal data are not attributed to an identified or identifiable natural person. (Art 4(5))[92]

Pseudonymisation is encouraged as a safeguard for data (especially for scientific research) throughout the GDPR and it is emphasised in recital 29 that it should be possible for one controller to separate personal data and 'additional information' internally to a satisfactory standard, providing they have taken technical and organisational measures.[93] As well as explicit inclusion of the process of pseudonymisation in the Regulation, recital 26 also refers to a new category of data, 'personal data which have undergone pseudonymisation'. It is discussion of this category which has caused some uncertainty about whether pseudonymised data should always be treated as personal data or, if it is possible to separate the 'coded data' sufficiently well from its 'key' for it to become anonymised and fall outside the Regulation.[94]

In guidance on the previous data protection law, the ICO suggested pseudonymised data could be sufficiently well separated—using technical and organisational safeguards—from the 'key' ('additional information' in the GDPR definition) needed to link these to particular individuals, so that it would lead to effective anonymisation.[95] This is the position under the US Common Rule (which governs research in the U.S.).[96] It is also the logical conclusion of applying the test for 'personal data' to data which have undergone pseudonymisation, namely that it must be reasonably likely that an individual could be identified from those data.

However, in their 2014 Opinion on Anonymisation Techniques, WP29 referred to pseudonymisation more in line with the approach taken in the GDPR; as a measure to reduce the linkability of data but not, in itself, sufficient to lead to anonymisation.[97] Part of recital 26 of the GDPR could also be read as suggesting that a mere possibility of re-combining data with additional information means it will be 'personal data':

> Personal data which have undergone pseudonymisation, *which could* be attributed to a natural person by the use of additional information should be considered to be information on an identifiable natural person. [emphasis added]

And, in their latest online guidance, the ICO state that:

> pseudonymisation is effectively only a security measure. It does not change the status of the data as personal data. Recital 26 makes it clear that pseudonymised personal data remains personal data and within the scope of the GDPR.[98]

If pseudonymised data are to be treated as personal data in almost all cases, there would be a significant difference between the risk assessment required of data more generally—that it must be at least reasonably likely that an individual can be identified—and the risk assessment for pseudonymised data and its identifying key. With the consequence that those working with pseudonymised genetic or clinical data would be likely to be processing 'personal data' if there is a key in existence which can be used to re-identify an individual.

The word 'could' in recital 26 raises the prospect that a mere possibility of re-combining data with additional information would render it 'personal data'. However, this would contrast with the approach to 'personal data' described above, which requires that (re)identification of data is reasonably likely and more than a mere hypothetical possibility to be governed by the Regulation. Is 'pseudonymised' data meant to be governed by a different standard to other forms of de-identified data, e.g. data which have been aggregated to a point of relative anonymity?

No rationale is given for such a difference and, as Mourby and colleagues argue,[99] the approach taken by the ECJ in the case of *Breyer v Bundesrepublik Deutschland*[100] strongly suggests that the conventional test of identifiability will be used where data are separated from additional identifying information.

Breyer concerned dynamic IP addresses and whether they may constitute personal data in the hands of an online media service provider even though they could only be attributed to an individual in combination with information held by an entirely independent internet service provider. Although this did not directly concern pseudonymised data – as opposed to potentially identifiable partial information – and it was governed by the previous legal regime, the European Court of Justice was clear that the question of whether data separated from additional information constitute 'personal data' was governed by the same test of identifiability discussed above; is it reasonably likely – not merely possible – that the data could be combined with the additional information and an individual identified?*

* There remains scope for some ambiguity under the GDPR because the wording of recital 26 suggests a different standard of likelihood is required, merely that the data 'could' be attributed to a natural person by the use of additional information.

Unfortunately, because the data in that case had not actually undergone pseudonymisation and because the case was governed by the previous Directive, not the GDPR we cannot yet be certain that this approach will be followed.

Although the express inclusion of pseudonymisation in the GDPR has created some doubt, we agree with those who argue that data which have been so well separated from additional identifying information that there is no reasonably likelihood of them being re-combined will fall outside the scope of 'personal data' and the Regulation.[101]

In our view, there are good reasons to doubt that a different standard of risk is applied to pseudonymised data: pseudonymisation is primarily a safeguard for data in the Regulation, it is referred to as a process not a category of personal data and no rationale is provided to justify an exceptional standard for this form of de-identified data versus another (e.g. aggregated data). However, in the absence of certainty around the status of pseudonymised data, professionals who wish to adopt a precautionary approach could treat the data as 'personal data' unless technical and organisational safeguards make it highly unlikely for the key and the data to be recombined.

4.3 'Genetic data' and genomic data under the GDPR

Although we have mentioned that the GDPR has expressly incorporated a category of 'genetic data' for the first time in EU data protection law we have deliberately not discussed what this definition involves or its implications until now. This is because the fundamentally important point is that data generated by genome sequencing, and similar analyses, will only be governed by the GDPR if they first meet the definition of 'personal data'. If that is the case, then it is a second order question whether those data also fall within the specific category of 'genetic data', or another related special category of data.

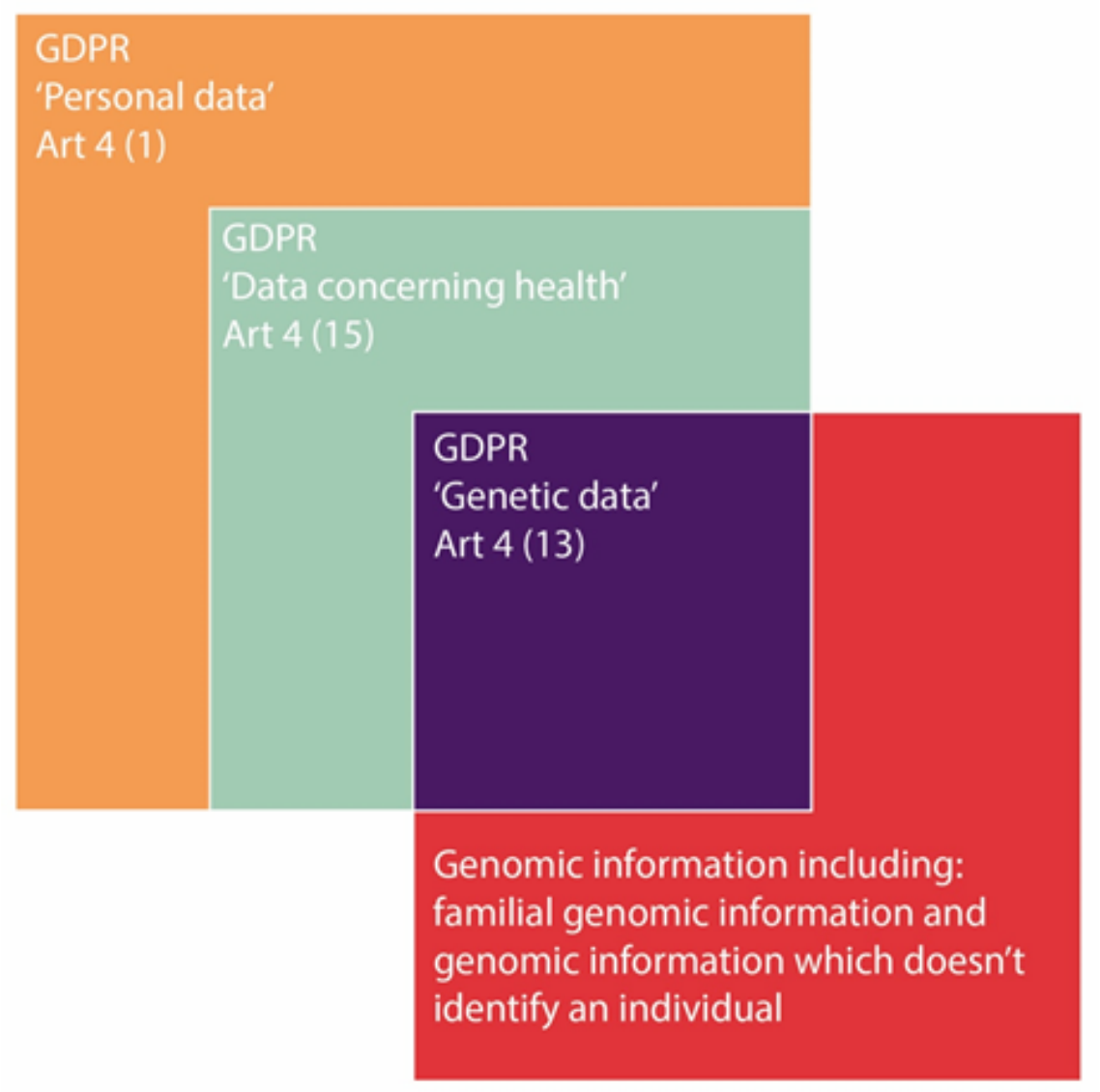
The GDPR introduces this definition of 'genetic data':

> 'genetic data' means personal data relating to the inherited or acquired genetic characteristics of a natural person which give unique information about the physiology or the health of that natural person and which result, in particular, from an analysis of a biological sample from the natural person in question (Art 4(13))[102]

Through the drafting process, this definition went through multiple iterations, including an initial reference to 'characteristics of an individual that are inherited or acquired during early prenatal development'.[103] This echoes the Council of Europe's description of a 'genetic test' in its additional protocol to the Convention on Human Rights and Biomedicine, concerning Genetic Testing for Health Purposes.[104] By the final draft the reference to 'prenatal development' had been removed so that genetic data now includes 'genetic characteristics' that are acquired over a lifetime, not only during prenatal development. For example, this would clearly include somatic mutations that occur later in person's life course.

In some ways this is a broad definition of genetic data: it extends to genetic characteristics that provide information about the physiology—not just health—of a person (See Figure 2) which are inherited or acquired over the course of their life. However, aspects of the definition are ambiguous and potentially problematic.

Figure 2: The relationship between GDPR 'personal data', 'data concerning health', 'genetic data', and genomic information



Unique information?

There is a requirement that genetic data 'give unique information about the physiology or the health of that natural person.' Does this require that the information about the physiology or health of a specific person is not available from other sources of information e.g. a family history or pedigree? To the extent that a genetic test result confirms something that is strongly suspected from other information, for example, the inheritance of an autosomal dominant variant, this could clearly fall under this interpretation as valuable new, 'unique' information.

Would this also apply if a genetic test result confirms the presence of a 'class 3' variant, or variant of uncertain significance? In this scenario, the evidence for the influence of that variant may be insufficient to contribute any extra information about the health or physiology of an individual above what is already known or suspected from other phenotypic information or test results. On this interpretation there is no new or 'unique information about the physiology or health of that natural person' and the test results would fall outside the definition of 'genetic data'. This would be a surprisingly narrow definition of genetic data and turns it into a narrow and exceptional category of 'extra' health information.

An alternative interpretation of the word 'unique' in this definition is that it requires information to be unique to a single specific individual. If so, by definition it must exclude data that give information about the physiology or health of more than one person, such as a familial shared variant or condition* As others have noted, such an individualistic view of genetic data ignores the shared nature of genetic information and that although a whole genome may be unique to an individual, 'this uniqueness is relative and each persona may expect to share much of their genetic code with others.'[105] Indeed, human beings share 99.9% of their genetic makeup with other humans by virtue of their being human, so it is only differences in the remaining 0.1% that can be considered unique to that individual.

As Dove notes, this interpretation aligns with the 'law's general fixation with the individual ('data subject') rather than familiar or group protections'.[106] For example, the only other definition in the GDPR to use the word 'unique' is the neighbouring definition of 'biometric data' and here (as in the wording of Art 9(1)) the word is used to emphasise that the data should enable the singling out or 'unique identification' of a person. If the word unique has the same meaning within the definition of 'genetic data,' then genetic information that do not allow the individual identification of one person—as opposed to identifying a family or number of related individuals—fall outside the GDPR's particular definition of 'genetic data'.

It could also be argued that this must obviously be so because it accords with the fundamental approach to 'personal data', and as we emphasised earlier, genetic data must first and foremost be 'personal data'. As we discussed earlier in this chapter, 'personal data' requires that an individual must be capable of being individually singled-out or distinguished from others,[107] even close family members or identical twins. If it is not possible to distinguish between two people, the data are technically not 'personal data'.

* This tension between the concept of 'genetic data' as on one hand 'unique', and on the other, also 'likely to reveal information on several people' has been present in EU Data Protection law since the Art 29 Working Party 'Working Document on Genetic Data', published in 2004. Although 'family' is used as a proxy for biological relatedness, it may be the case that the social relationships within a family do not equate to biological relatedness in cases of adoption or misattributed paternity for example.

In reality, it is hard to imagine circumstances where there is no additional information, for example in a patient or research record, which could be used to single-out the individual who was the subject of a genetic test. If this is possible, then the standard test of the risk of it occurring must be applied.

However, we suggest that there should be extreme caution when deciding that health and genetic data are familial as opposed to personal because, as we have outlined, if the data are reasonably likely to be combined with other available information or background knowledge (e.g. of family members) to identify an individual data subject, the data will fall within the scope of the GDPR. It is also likely that knowledge of who a data subject is in a database will constitute 'personal data' most of the time, even if no other information from the database is available. This is because simply knowing that an individual is part of a database tells you something about that person and so falls within the requirement that the information 'relates' to an identifiable individual (see discussion above).

Given the requirements for personal data to distinguish between individuals, it is not clear why this would need to be expressly incorporated in the definition of genetic data itself rather than treated as part of the pre-requisite for all 'personal data'. It remains to be seen what the correct interpretation of 'unique' in the definition of genetic data is. The danger of a narrow definition of 'genetic data' in the GDPR which excludes familial or shared variants, is that it may lead to a view that such familial information should not be treated as sensitively, although in many cases it is likely that other information will allow the identification of one person in combination with the 'shared' genetic information.

Analysis of a biological sample?

A final component of the definition of genetic data is that they result, 'in particular', from the analysis of a 'biological sample'. As Mark Taylor discussed in relation to earlier drafts and proposals for the Regulation, this may be read as recognising that analysis of a biological sample is a sufficient, but not a necessary element of genetic data.[108] This would broaden the scope of the definition to that originally proposed by the Commission and include information about genetic characteristics which result from other forms of analysis. Hallinan gives the example of a diagnosis of Reiger Syndrome, which may be made by looking at the features of the eyes of the sufferer.[109]

However, the explanatory text (recital 34) is less equivocal:

> (34) Genetic data should be defined as personal data relating to the inherited or acquired genetic characteristics of a natural person which result from the analysis of a biological sample from the natural person in question, in particular chromosomal, deoxyribonucleic acid (DNA) or ribonucleic acid (RNA) analysis, or from the analysis of another element enabling equivalent information to be obtained

This suggests that 'genetic data' must result from analysis of a biological sample, and that this is a necessary condition of the category. The recital is a non-binding part of the Regulation but if it is being used to help interpret the definition in Art 4(13) it would seem to tip the balance in favour of the view that 'genetic data' must result from the analysis of a biological sample. That there is ambiguity in this important aspect of the definition is very unfortunate.

Analysis of DNA, RNA or another element?

If genetic data must result from analysis of a biological sample, the recital emphasises that this will include chromosomal, DNA or RNA analysis. However, recital 34 adds that 'equivalent information', obtained from the analysis of 'another element' which 'relate' to the 'genetic characteristics' of a natural person would fall within the category of genetic data. The scope of this definition is not clear but it could mean that information derived from the analysis of a biological sample, such as a blood test, which reveal 'genetic characteristics' also constitute 'genetic data' even if it has not been derived from analysis of DNA or RNA (for example, diagnosis of a genetic disorder such as sickle cell disease by inspection of the blood cells).

The potential for analysis of 'another element', as opposed to DNA, RNA or chromosomal analysis, to generate genetic data could become more important over time, as other 'omics technologies may lead to individually identifying information about the genetic, health and physiological characteristics of an individual.

For example, advanced protein analysis methods have been shown to identify an individual from as little as 1cm of a single strand of hair.[110] Could the results of this analysis be classified as 'genetic data'? Arguably it could. Providing there is a means of cross-referencing results, such as a variant database, the data could be argued to provide identifiable information relating to the inherited or acquired genetic characteristics of a natural person which give unique information about their physiology (e.g. their hair type and/or colour) and which result from analysis of a biological sample. The breadth of this description of genetic data means that it is capable of adapting to future scientific developments.

What are the consequences of falling within or outside the definition of 'genetic data'?

The GDPR is clear that genetic factors may lead to the identification of an individual,[111] and that 'genetic data' are one of the 'special categories' of personal data that 'merit higher protection'.[112]

However, another special category, 'data concerning health', is arguably broad enough to encompass most identifiable data being used for healthcare and medical research (for example both genotype and phenotype data):

> 'data concerning health' means personal data related to the physical or mental health of a natural person, including the provision of health care services, which reveal information about his or her health status[113]

Indeed, recital 35 describes data concerning health as including 'information derived from … genetic data' and frequently it will be relatively unimportant whether the data derived from genetic testing fall within the specific definition of 'genetic data' or not, because they will fall under the broader category and be subject to the same, increased level of protection. This is because there are no rights or obligations in the GDPR that apply only to the processing of 'genetic data'; three of the special categories of 'genetic data', 'biometric data' and 'data concerning health' are always taken together where more specific rules are mentioned.

Of course, genetic data are to be treated sensitively and guidance, for example on Data Protection Impact Assessments (DPIAs), emphasises that the processing of genetic data could represent a 'high risk' to individuals' rights and freedoms.[114] However, as the EDPB clarified 'the processing of genetic data on its own is not necessarily likely to represent a high risk';[115] something else is also required, such as processing data on a large scale or processing data concerning vulnerable data subjects (e.g. patients). In spite of explicitly incorporating 'genetic data' for the first time, the GDPR does not mark the introduction of genetics exceptionalism in the EU's data protection law.

Despite this, some scope for regulatory divergence remains depending on whether data fall within the definition of genetic data or not. It may be argued that some data derived from genetic testing do not reveal information about a person's 'health status' and therefore would not fall under the category of 'data concerning health'. However, if it could be argued that such data provide 'unique information about [a person's] *physiology*' (emphasis added) they could meet the definition of 'genetic data' and remain subject to the GDPRs higher level of protection.

There is also the prospect for divergence across Europe as Member States are granted the power to introduce further conditions or limitations to the processing of genetic data under Art 9(4). The UK has yet to implement specific conditions but some nations have gone further including Finland and Italy, by introducing specific requirements for the processing of genetic data.[116]

Another possibility is that Supervisory Authorities or the European Data Protection Board (EDPB—the body charged with ensuring consistency across Europe) will develop further guidance on 'genetic data'.

Given some of these definitional challenges, at present, the pragmatic approach for those using or storing data in genetic or genomic research or healthcare might be to consider all genotype or phenotype data as sensitive and part of the specific categories of data given higher protection under the GDPR, provided they are sufficiently individually identifying to constitute 'personal data'.

What about 'genomic data'?

Although the GDPR refers to 'genetic data', both this special category of data, and the category of data concerning health, are clearly capable of incorporating data derived from the sequencing of a person's whole-genome, whole-exome or large parts thereof (commonly referred to as genomic data). Moreover, as we discuss above, the reference to analysis of DNA, RNA or another element would incorporate the data resulting from the techniques outlined in chapter 1, including pathogen sequencing and metabolomics. In this sense, there is no difference between 'genetic' or 'genomic' data.

However, where 'genomic' data are perhaps different is that they provide more scope for identification of an individual; they are potentially 'strongly identifying'* and are more likely to fall within the scope of the GDPR and require a high level of data protection. This will still depend on the context, availability of other sources of information and identification techniques.

One of the ways that genomic data are special is in their potential to yield information that relates to more than one individual. This means that genome sequence data should be handled carefully even if, for example, it is derived from a deceased individual, and otherwise out of scope. It could be argued that aspects of the genome that are highly likely to be shared with living family members will constitute 'personal data' if they can be connected with those data. For example, if a tissue block or sample from a deceased relative is analysed to inform the treatment of a living individual, the data should be treated as 'personal data' relating to the living individual because the GDPR does not regulate based on the source of the information but focuses on its content. This also has implications for the data subject's right of access (discussed in the following chapter) which is qualified if the data also contain another individual's personal data.

Finally, genomic data are special in that our understanding improves with time. This means that there are significant advantages to the long-term storage or preservation of genomic data (and related medical data) to both enable research that advances scientific knowledge and to allow new insights to be made about an individual's health state or predisposition to disease. This characteristic of genomic information potentially conflicts with the obligations and principles discussed in chapter 2 to minimise, anonymise or even delete data as far as possible to protect the rights and interests of individuals.

4.4 Conclusions

Some of the most significant challenges our research has identified relate to determining when genetic, genomic and associated health data are 'personal data'. In particular, uncertainty and disagreement about whether data are 'personal data' is challenging local, national and international flows of genomic and health data (see further in chapters 7 and 8). This is especially the case where there are disagreements about when data may have been sufficiently de-identified to fall outside the GDPR and whether data which have undergone pseudonymisation are capable of falling outside the category of 'personal data'.

Because the GDPR takes a risk-based approach to identification, requiring a probabilistic approach – namely that it must be reasonably likely to give rise to 'personal data', assessing whether data are sufficiently identifiable or whether there is a very limited likelihood of identification can only take place in context. There are some factors that genomic data controllers in particular should take into account. These are the 'richness' of the genetic and any associated phenotypic data, consideration of who may access the data, and whether other information may be used in combination with the genetic data to identify an individual.

* However, it is noted that limited genetic data may also be highly identifying if, for example, they relate to particularly rare genetic variants. Finnegan T, Hall A. Identification and genomic data. PHG Foundation. 2017, 22-29.

Unfortunately, it is not currently clear if the same approach should be taken to assessing the identifiability of data that have undergone pseudonymisation. However, our view is that the same approach should be adopted and that there is no rationale to justify applying a different standard to this form of de-identified data versus another (e.g. aggregated data).

Genomic data raise particular challenges under the GDPR, including some that relate specifically to ascertaining the status of genetic or genomic data. For example, in our legal analysis, we highlight the approach the Courts in England and Wales have recently taken to assessing whether data are 'directly identifiable' because they single-out or 'individuate' an individual from a group.

We are concerned that if the same reasoning is applied to genomic data as to the forms of data involved in those cases (most recently, biometric facial data), it could be concluded that individual level genomic information is inherently identifiable. We contest the logic involved in these cases and it is our strong recommendation that genetic, and particularly genomic information should not be viewed in itself as inherently or directly identifying without some further link to or impact on an individual.

Indeed, our analysis of the inclusion and treatment of the category of 'genetic data' in the GDPR is that the Regulation does not mark the introduction of genetic exceptionalism in the EU's data protection law. 'Genetic data' are included alongside health and biometric data (amongst others) as a special category of data deserving higher protection but the EDPB has clarified that the processing of genetic data on its own is not necessarily likely to represent a high risk. However, there are implications of the inclusion of 'genetic data' in the GDPR which present some challenges for data controllers and genomics professionals.

One challenge we have discussed is that the scope of the definition of 'genetic data' is not completely clear. In some ways it is a broad category that extends to information about the physiology—as well as health—of a person and can include information derived from analysis of blood test or other biological sample, not only data derived from sequencing. Moreover, 'genetic data' can include information from analysis of other molecules, including RNA, so it is capable of adapting to future scientific developments.

In other ways, the category is more narrow. The requirement that personal data relate to a unique individual who is distinguishable from others also applies to genetic data. This means that shared biological or familial genetic information is not governed by the GDPR unless it can be related to one specific individual from a group. This does not reflect the medical or scientific understanding of genetic data but it does reflect the focus of data protection law on the individual, as opposed to communal interests.

Despite this, we recommend that data controllers are very cautious in concluding that genetic information is purely familial or shared information unrelated to an individual because it is frequently likely to be the case that other information (in particular, the background knowledge of family members or healthcare professionals in the healthcare context) will be available to link genetic information to a specific individual. This also holds true for genetic information derived from a deceased person, if it is possible to link that same information to a living relative.

The consequences of genetic and associated information constituting 'personal data' are that all the requirements of the GDPR will apply, including the need for a higher level of protection in accordance with Art 9. The processing of 'genetic data' may be subject to a greater level of regulatory variation as Member States are granted the power to introduce further conditions or limitations to the processing of genetic data under Art 9(4). Supervisory Authorities or the European Data Protection Board could also develop further guidance on 'genetic data'. Although this may complicate cross-border processing and data sharing (see further in chapter 7), it also may provide the opportunity for the genomics community to advocate for appropriate standards to be applied to the processing of genetic and associated health data in healthcare and research.

This possibility is discussed further in chapter 8, following analysis of other impacts for genomic data processing in chapters 5-7. We turn next to the challenge of fulfilling the requirements for a lawful basis for processing genomic data under Art 5, and meeting a condition for processing special category data under Art 9.

5. Lawful processing of genomic data for healthcare and research

Our research has identified a range of challenges for genomic healthcare and research which relate to establishing a lawful basis for forms of processing personal data and meeting the conditions for processing health and genetic data under Art 9 and relevant national law.

These challenges have been raised in the literature, by interviewees and participants in our research, and through our legal analysis. In this chapter we consider how these requirements apply to uses of genomic data in healthcare and research and some of the challenges involved in meeting them.

5.1 Establishing a lawful basis for processing

Although the GDPR carries over many of the requirements of the previous Data Protection Directive, it still gives rise to some challenges when establishing a legal basis for processing of personal data, as required by Art 6. In particular, the GDPR has introduced new, more stringent standards for consent leading some data controllers to consider alternative legal bases for processing of genetic or health data. The choice of legal basis is also important because different rights, or exceptions to them, flow from each choice of legal basis. Under the GDPR, personal data shall only be processed if one of six legal bases can be satisfied (Art 6(1)). These are:

> (a) the data subject has given consent to the processing of his or her personal data for one or more specific purposes;
>
> (b) processing is necessary for the performance of a contract to which the data subject is party or in order to take steps at the request of the data subject prior to entering into a contract;
>
> (c) processing is necessary for compliance with a legal obligation to which the controller is subject;
>
> (d) processing is necessary in order to protect the vital interests of the data subject or of another natural person;
>
> (e) processing is necessary for the performance of a task carried out in the public interest or in the exercise of official authority vested in the controller;
>
> (f) processing is necessary for the purposes of the legitimate interests pursued by the controller or by a third party, except where such interests are overridden by the interests or fundamental rights and freedoms of the data subject which require protection of personal data, in particular where the data subject is a child.

One of the significant impacts of the GDPR identified in the literature and in the course of our research, is the strengthened requirements for consent and the challenges relying on this lawful basis, in particular for genomics research. We discuss the challenges of consent below followed by consideration of the alternative legal bases and how they apply in the genomics context.

Key impact

Our research has identified a significant range of challenges associated with consent under the GDPR. In particular, research participants raised challenges meeting the GDPR's standards for consent, the implications of withdrawal of consent in the research context, determining whether broad consent is lawful, and, transparently explaining the differences between GDPR consent and ethical or common law consent.

The challenge of consent

Article 4(11) defines the 'consent' of the data subject as:

> any freely given, specific, informed and unambiguous indication of the data subject's wishes by which he or she, by a statement or by a clear affirmative action signifies agreement to the processing of personal data relating to him or her

Art 7 sets out further conditions, including that the data subject must be informed that they have the right to withdraw their consent at any time (Art 7(3)). The background text (recitals) provide further guidance. Recital 42 explains that at a minimum, a data subject should be informed of the identity of the controller and the purposes of the processing, and that there must be a genuine or free choice to refuse or withdraw consent without detriment. Recital 32 explains that consent should be given by a clear, unambiguous affirmative act (excluding silent or opt-out consent) and that consent should cover all processing activities carried out for the same purpose or purposes.

A clear imbalance?

One of the challenges under the GDPR is that consent is presumed not to be freely given and is therefore not valid where there is a 'clear imbalance' between the data subject and the controller. Recital 43 further explains that this may particularly be the case where the controller is a public authority, which is frequently the case in healthcare and research. The European Data Protection Board (EDPB) – the independent body which ensures consistency of data protection rules across the EU – has cautioned that where a participant is not in good health, or when they belong to economically or socially disadvantaged groups, there is likely to be an imbalance of power.[117] This could be read as meaning that consent is an inappropriate legal basis for a great deal of health research (as the NHS Health Research Authority concluded and we discuss below).

Others push back on this interpretation and point out that recital 43 refers to a need to consider 'all the circumstances of the specific situation' which could mean there is no power imbalance in the particular circumstances of certain genomics research projects.[118] Moreover, the EDPB discussion of consent does not rule out consent as a lawful basis for health research but rather requires a 'particularly thorough assessment of the circumstances' before relying on consent as a legal basis.[119]

Broad consent for research?

A particular challenge for genetic/genomic research is the ambiguity around how specific the consent must be. In general, it is clear that consent is lawful only if given for one or more 'specific purposes' (Art. 6(1)(a)) and it should allow for separate choices to be given to different 'personal data processing operations' (recital 43). This is explained by the Article 29 Working Party (the predecessor to the EDPB) as requiring 'granularity', so where appropriate, separate consent should be obtained for different purposes.

An important question for genetic/genomic research, particularly long-term big data research, is how specifically defined the research purposes must be at the outset. The Article 29 Working Party previously explained that enough information should be provided about specific purposes for the data subject to understand the implications of their choice,[120] and to assess whether the law and safeguards have been complied with.[121] They concluded that a purpose which is 'vague or general, such as … "future research" … will – without more detail – usually not meet the criteria of being "specific"'.[122]

However, in recital 33 of the GDPR it is acknowledged that:

> it is often not possible to fully identify the purposes of personal data processing for scientific research purposes at the time of collection. Therefore, data subjects should be allowed to give their consent to *certain areas* of scientific research when in keeping with recognised ethical standards for scientific research. [Emphasis added]

This suggests that a broader consent may be possible for processing of genetic/genomic data in 'areas' of scientific research purposes, subject to ethical oversight. Such flexibility would align with other aspects of the GPDR, such as the explicit exemption from the purpose limitation principle for scientific research.

However, the Article 29 Working Party cautions that the requirement for specific consent will apply unless it is not possible to sufficiently specify the purposes for data processing at the outset, providing an 'exception that the purpose may be described at a more general level.'[123] The Article 29 Working Party state that the 'flexible approach' of recital 33 will be subject to a higher degree of scrutiny when special categories of data, such as genetic and health data, are processed on the basis of explicit consent (see below).[124]

Further, a controller 'must seek other ways to ensure the essence of the consent requirements are served best', such as seeking consent to each defined stage of research as it progresses, or providing regular updates on the development of research purposes.[125] This could be read as an encouragement for dynamic consent[126] if purposes cannot be sufficiently specified at the outset and constitutes a restrictive interpretation of recital 33.

The WP29's approach has already been followed by the Association of German Supervisory Authorities in their 2019 paper on broad consent for research.[127] This has caused some concern for the genomics and scientific research community[128] who fought to change earlier drafts of the Regulation so that it did not hamper research.[129] It is argued that broad consent is a legitimate approach in the field of genomic research and that the omnibus nature of the GDPR and WP29 guidance means that these are not appropriate standards for the more specific sector of genomics research.[130]

Hallinan mounts a strong defence of broad consent in genomics research under the GDPR, even in light of the WP29 guidance. He argues that because the WP29 guidance does not differentiate between different types of scientific research, it can be argued that genomic research is a special category of health research which should not be subject to the same omnibus guidance.[131] Ultimately, Hallinan argues that neither the WP29 nor the EDPB have power to interpret provisions against the express wishes of the legislator and, because it is clear that the legislator inserted recital 33 to provide for more flexible consent in scientific research, the WP29 guidance should be interpreted in this light.

However, the potential power imbalance between researchers and participants, and the challenge of ensuring sufficiently specific consent, have contributed to recommendations against consent as a legal basis for scientific research processing. The NHS Health Research Authority have stated that 'the legal basis for processing data for health and social care research should not be consent',[132] for both public authorities and commercial companies or charitable research organisations. Participants in our research identified further challenges with consent, including a view that individual consent may not be an appropriate approach to take when processing genomic data because those data are inherently 'familial' or 'relational'. Others highlighted the further challenge of managing consent to uses of genomic data from a child.

Parental or child consent

A significant challenge may arise in large-scale genomics research where consent to sequencing and processing of a child's genomic information has been obtained from a person with parental responsibility.

Key impact

Several interviewees highlighted the challenge of managing a transition from parental consent for processing of child's data, to obtaining the consent of the mature and competent minor as they develop in maturity. They were concerned that it was unclear when and how new consent should be sought from the mature data subject.

Over time, the child will develop in competence and maturity and this raises the question of whether there should be efforts to obtain consent from the child, and if so, the point at which this should take place. This was considered by the Art 29 Working Party in their 2009 Opinion on the protection of children's personal data:[133]

> If the processing of a child's data began with the consent of their legal representative, the child concerned may, on attaining majority, revoke the consent. But if he wishes the processing to continue, it seems that the data subject need give explicit consent wherever this is required.

In the UK the age of full majority is 18 but for certain purposes, including providing consent to medical treatment, the law recognises that a minor aged 16 has capacity to consent for themselves.[134] The implication of the WP29 advice is that fresh consent is required from the data subject at least when they reach that age in the UK. However, the WP29 also recognised that data controllers should adapt to the degree of maturity of the child, in line with the way that Member State laws deal with developing maturity and competence. In the UK this is governed in healthcare by the standard of *Gillick* competence which enables a child below 16 to consent to healthcare if they are competent to make that specific decision (a context-specific assessment). Generally, this also applies to medical research.[135] The ICO makes clear that this also applies to consent under the GDPR:

> For other types of processing, the general rule in the UK is that you should consider whether the individual child has the competence to understand and consent for themselves (the '*Gillick* competence test').[136]

Genomic data controllers should therefore seek to refresh consent as the child develops competence. However, because the primary consideration in UK law is that of the best interests of the child until the age of 18, as Taylor and his colleagues emphasise,[137] it is arguable that a parental consent to processing will remain valid as long as it is also in the best interests of the child for processing to continue. The ICO also acknowledge this possibility:

> Parental consent won't automatically expire when the child reaches the age at which they can consent for themselves, but you need to bear in mind that you may need to refresh consent more regularly.[138]

Taylor and colleagues advise that 'data controllers would be wise to develop strategies for when and how to seek fresh consent from children as soon as is practicable after the data controller has reason to be aware they are mature enough to consent on their own behalf.'[139] They argue that the original consent may remain valid (if in the best interests of the child) while these efforts are being made to seek consent even if a child has reached 16.

Such strategies will be labour intensive and it is clear that the controller must make considerable efforts to ensure that consent is refreshed more regularly as children grow up. Indeed, the ICO recommends that controllers seek to refresh even adult consent every two years to ensure good levels of trust and engagement unless a longer period may be justified.[140]

A further challenge for genomic data controllers may be how to respond to a refusal of consent from a competent minor. Advice is unanimous that it is not simply possible to switch the legal basis for processing.

The ICO cautions:

> Even if a different basis could have applied from the start, retrospectively switching lawful basis is likely to be inherently unfair to the individual and lead to breaches of accountability and transparency requirements*

The WP29 guidelines on consent (adopted by the EDPB) state that 'under the GDPR, it is not possible to swap between one lawful basis and another',[141] although they make clear in the next sentence that it may be possible to rely on another lawful basis 'as a one off situation'. The legal position is likely to be, as Taylor and colleagues suggest, one where it will only be possible to rely on an alternative legal basis if 'that alternative legal basis as a ground for processing has been openly and transparently communicated to both the parent (or other legal representative) originally providing consent and the data subject.'

As well as these considerable challenges for consent, as discussed in the following chapter in greater detail, if processing is based on consent, a number of rights and obligations follow, including the right to withdraw consent at any time. This is a potential challenge for uses of genomic data, for example in research, because it may require the removal of data from databases and thereby weaken research power. Moreover, it may be incompatible with the ethical or legal imperative to maintain some links, for example for adverse drug reporting in medical research, or, for the increasingly advocated recontact of genomic research participants, if significant, particularly clinically actionable information is generated through research.[142]

In the following chapter we discuss whether the right to erasure may be satisfied by anonymisation or whether it requires full removal and deletion of data. Finally, even if consent is not chosen as a legal basis under the GDPR for processing personal data, consent may still be required in order to comply with other legal and ethical standards, including the common law of confidentiality.

Participants in our research highlighted that this can make explaining to individuals how their data will be used and managing their expectations difficult, especially if that withdrawal of consent will not prevent further processing for research purposes (see below). This is a challenge for those who are seeking to be clear and transparent about processing.

> [One challenge is] definitely consent. It's confusing to people who have an idea about what it already means and it is different to the common law concept of consent … it's a difficult message to convey that you're not using consent as the legal basis.
>
> **Interview participant**

*The Information Commissioner's Office. Lawful basis for processing.
Available from: https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/lawful-basis-for-processing/

Alternative legal bases

There are five other legal bases available for data controllers under Art 6. However, distinctions between public authorities and private organisations influence which are most appropriate. In this section we will briefly discuss the requirements for each of these, highlighting when they may be of particular relevance to the genomic context and highlighting challenges identified in our legal analysis. Many of these considerations are also relevant to the evaluation of the appropriate legal basis for international transfers of data (chapter 7) where this aspect is discussed in greater detail.

Performance of a contract (Art 6(1)(b))

Processing is lawful if it is necessary for the 'performance of a contract' (Art 6(1)(b)). This could be the case in private healthcare or direct-to-consumer genetic services. However, processing must be necessary for performance of (or entry into) the contract.[143] 'Necessity' has its own independent meaning in EU law, being interpreted in a way that reflects the objective of the law in question.[144] In the data protection context, the fundamental right to privacy and protection of personal data, and the principles in Art 5 are key to this interpretation. Considering the 'necessity' of any given term includes a 'combined, fact-based assessment' of the purposes of processing and the term in question.

The necessity of any given contractual term often turns critically upon the specific purpose of a contract.[145] The EDPB, the CJEU, and the European Data Protection Supervisor (EDPS) all stress that this assessment must also consider whether there are other less intrusive means to achieve the same end. If there are 'realistic, less intrusive alternatives, the processing is not 'necessary.'

The mere reference to a term being included in a contract is insufficient to establish that the term is 'objectively necessary' to perform or conclude that contract. The EDPB again endorse WP29 comments on how to interpret 'necessary for the performance of a contract' and make clear that:

> 'Even if these processing activities are specifically mentioned in the small print of the contract, this fact alone does not make them 'necessary' for the performance of the contract.'[146]

This means that where several services are offered, consideration of whether each service can be performed separately may indicate whether it is objectively necessary to deliver the individual services requested by the data subject. To repeat the EDPB's words, you cannot bundle together services to create a 'take it or leave it situation for data subjects who may only be interested in one of the services.'[147] If, for example in relation to direct-to-consumer testing (DTC), the collection of some personal data is unnecessary to carry out the genetic testing, another legal basis will be required. Alternatively, in private healthcare contracts, the inclusion of a wide range of processing activities in the small print of a contract would not necessarily make them necessary for performance of the service and thereby provide a legal basis for processing.

Compliance with a legal obligation (Art 6(1)(c))

Some forms of processing of genetic/health data may be necessary for 'compliance with a legal obligation' (Art 6(1)(c)). This applies if processing of personal data is a reasonable and proportionate way of complying with a clear statutory or common law obligation (Art 6(1)(c)). For example, sharing healthcare data via a statutory gateway with bodies such as NHS Digital. Of particular note for genomic information is the possibility that this could provide a lawful basis for disclosure of genetic information to relatives of the data subject without consent, in accordance with the recent duty of care established in UK law.[148]

Protecting the vital interests of the data subject or of another natural person (Art 6(1)(d))

As a matter of last resort, processing will be lawful if it is necessary to protect the vital interests (generally interpreted as matters of life and death)[149] of the data subject or another person (Art 6(1)(d)). This is unlikely to be met for most genetic tests unless the threat of death is significant and proximate, for instance, where using whole genome sequencing to inform the diagnosis of rare genetic diseases in severely ill babies and young children. It will not apply if processing can be undertaken using an alternative basis (recital 46).

Performance of a task carried out in the public interest (Art 6(1)(e))

For public authorities, the most appropriate legal basis is often likely to be 'performance of a task carried out in the public interest'. This is likely to be the most useful legal basis for public sector healthcare or research organisations processing genetic data. It will apply if the processing is necessary to perform a task, or in the exercise of official authority, function or power, which has a clear basis in law.[150] This basis can only be relied upon by public authorities in the UK subject to freedom of information legislation or otherwise specified in Regulations,[151] so it will most clearly apply to provision of healthcare by the NHS* but not to some of the major research institutions which are not public authorities. The explanatory notes to the UK Data Protection Act 2018 clarify that this basis should be available to a university undertaking processing of personal data necessary for medical research purposes.[152]

However, this basis is not available to private or charitable organisations. For them, the legal basis of 'legitimate interests' (Art 6(1)(d)) is likely to be the best option in many cases.

Legitimate interests (Art 6(1)(d))

Legitimate interests are the most flexible basis for processing personal data (although they are no longer available as a legal basis for processing by public authorities performing their tasks or exercising their official authority). The concept of a legitimate interest is very broad and under the GDPR has been extended to include the legitimate interests of any third party, not only recipients of data. The Information Commissioner's Office (ICO) guidance is that it may extend to even trivial or controversial interests but that vague, unethical and unlawful interests would not count as legitimate.[153]

* For the range of statutory sources of official authority, see: Information Governance Alliance. The General Data Protection Regulation: Guidance on Lawful Processing. 2018. Available from: https://www.swft.nhs.uk/application/files/8615/3744/8007/Lawful_Processing.pdf

However, data processing must be necessary for the purposes of those legitimate interests, which means it must be a reasonable and proportionate way to achieve those ends. Crucially, legitimate interests may be outweighed by the impact of processing on the interests or fundamental rights and freedoms of the data subject. The ICO advises that these are to be interpreted broadly e.g. including loss of control over data and social or economic disadvantage,[154] and recital 47 makes clear that this must include an assessment of what the data subject may reasonably expect (placing importance on the information that they are given by the controller).

The ICO recommend that controllers perform a 'legitimate interests assessment' (LIA) to assess the balance of the controller's interest and the impact on the data subject, and to record the justification for processing.* The reasonable expectations of the data subject are an important factor in this assessment and it could be the case in the genetic/genomics context, that evidence of patient or participant expectations could provide support for processing. However, it is also the case that the sensitivity of genetic or health data and the need for heightened protection of such data (especially of childrens' data) are significant but not insurmountable barriers to processing on the basis of legitimate interests. There is no specific requirement to provide the reasoning or a LIA to data subjects in the GDPR, but providing information on the decision helps to fulfil the responsibility to process data fairly, transparently and accountably.

Overall, the challenges with consent mean that for many uses of health and genetic/genomic data other legal bases will often be more appropriate, particularly for research purposes. There are obvious alternatives to consent: public task for public authorities such as university medical researchers, or, legitimate interests for private or charitable organisations. However, in the context of genetic/genomic data, the legitimate interests of the processor need to be balanced sensitively with the rights and interests of the data subject and even if consent is not the legal basis for processing it will often be required for other legal and ethical reasons. Communicating these distinctions to data subjects may be challenging.

5.2 Conditions for processing genetic and health data under Art 9(2)

Even if a legal basis has been established for processing personal data, the processing of genetic data, data concerning health and biometric data (amongst other 'special categories') is actually prohibited by Art 9(1) GDPR unless one of the ten conditions in Art 9(2) apply.

* The ICO provides more detailed guidance and an optional template to assist data controllers. See:
The Information Commissioner's Office. How do we apply legitimate interests in practice? Available from: https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/legitimate-interests/how-do-we-apply-legitimate-interests-in-practice/#LIA_process [Accessed 11th May 2020].

We discussed some of these special categories in chapter 4 and it should be noted that they are potentially very broad: 'data concerning health' includes all data which reveal information about the health status of an identifiable or identified individual and 'genetic data' extends to data which give unique information about the physiology of an individual. The inclusion of 'biometric data' is also potentially relevant for genomic contexts because it includes facial images and other images which allow the unique identification of an individual, which potentially includes images of dysmorphology and other phenotypes that frequently accompany some genetic records. Taken together, most of the individual-level data that result from genetic analysis and frequently accompany genetic results as phenotypic data will fall within these special categories of data, provided—as we discussed in chapter 4—that they are reasonably likely to identify an individual.

Our analysis of Art 9 conditions has highlighted some challenges that may arise in the genomics context, in particular a challenge in divergent implementation of Art 9 options across the EU/EEA.

Special category data may only be processed if one of the conditions in Art 9(2) applies (see below).

Article 9(2):

a. the data subject has given explicit consent to the processing of those personal data for one or more specified purposes, except where Union or Member State law provide that the prohibition referred to in paragraph 1 may not be lifted by the data subject;
b. processing is necessary for the purposes of carrying out the obligations and exercising specific rights of the controller or of the data subject in the field of employment and social security and social protection law in so far as it is authorised by Union or Member State law or a collective agreement pursuant to Member State law providing for appropriate safeguards for the fundamental rights and the interests of the data subject;
c. processing is necessary to protect the vital interests of the data subject or of another natural person where the data subject is physically or legally incapable of giving consent;
d. processing is carried out in the course of its legitimate activities with appropriate safeguards by a foundation, association or any other not-for-profit body with a political, philosophical, religious or trade union aim and on condition that the processing relates solely to the members or to former members of the body or to persons who have regular contact with it in connection with its purposes and that the personal data are not disclosed outside that body without the consent of the data subjects;
e. processing relates to personal data which are manifestly made public by the data subject
f. processing is necessary for the establishment, exercise or defence of legal claims or whenever courts are acting in their judicial capacity
g. processing is necessary for reasons of substantial public interest, on the basis of Union or Member State law which shall be proportionate to the aim pursued, respect the essence of the right to data protection and provide for suitable and specific measures to safeguard the fundamental rights and the interests of the data subject
h. processing is necessary for the purposes of preventive or occupational medicine, for the assessment of the working capacity of the employee, medical diagnosis, the provision of health or social care or treatment or the management of health or social care systems and services on the basis of Union or Member State law

i. processing is necessary for reasons of public interest in the area of public health, such as protecting against serious cross-border threats to health or ensuring high standards of quality and safety of health care and of medicinal products or medical devices, on the basis of Union or Member State law which provides for suitable and specific measures to safeguard the rights and freedoms of the data subject, in particular professional secrecy;

j. processing is necessary for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes in accordance with Article 89(1) based on Union or Member State law which shall be proportionate to the aim pursued, respect the essence of the right to data protection and provide for suitable and specific measures to safeguard the fundamental rights and the interests of the data subject.

In the genetic or genomic context some of the Art 9(2) conditions are particularly relevant. These include that processing is based on (a) explicit consent, is necessary for (h) medical purposes, (i) public health purposes or (j) for scientific research (j). The potential for special category data to be processed when (e) they have been manifestly made public by the data subject and for (g) reasons of substantial public interest may also become relevant in the genomics context. In this section we discuss these conditions and their requirements, highlighting how they may apply in the genomics context and particular impacts or challenges associated with them.

Explicit consent (Art 9(2)(a))

As well as meeting the requirements for consent in Articles 6 and 7 (discussed above), Art 9(2)(a) requires that specific informed consent to processing of special category data is 'explicit'. This requires an express statement of consent, for example in a written statement signed by the data subject or a similar positive action using an electronic form.[155] The data controller has to demonstrate that the data subject has consented to the processing of their data (Art 7(1)). The Article 29 Working Party recommended that as best practice, consent 'should be refreshed at appropriate intervals' to ensure data subjects remain well informed about how their data is used and how to exercise their rights.[156]

The same challenges of consent and facilitating withdrawal discussed in relation to Article 6 apply to processing based on explicit consent. This has led to calls for researchers, in particular, to move away from consent where alternatives, such as scientific research (j) are available. However, as we discuss below, such alternatives may not be available in all EU/EEA Member States, so where organisations are subject to multiple overlapping Member State laws, consent may still be the best available option, despite its challenges. This could mean that the genomics community should, and will, continue to advocate for appropriate consent standards for the specific processing of genomic data in research.

Medical (h) or public health (i) purposes

The GDPR makes clear that genetic and health data may be processed for medical or public health purposes: Art 9(2)(h) allows processing for 'the purposes of preventive or occupational medicine … medical diagnosis, the provision of health or social care or treatment or the management of health or social care services on the basis of Union or Member State law or pursuant to contract with a health professional'. The DPA 2018 provides a basis in UK law[157] and implements the 'secrecy' safeguards required by the GDPR (Art 9(3)) so that processing must be by or under the responsibility of a health or social work professional or another person who in the circumstances owes a duty of confidentiality.[158]

If there is a public health objective to processing data – such as separating human DNA from virus DNA as part of pathogen sequencing during an outbreak – the Art 9(2)(i) condition will be more suitable, which applies if 'processing is necessary for reasons of public interest in the area of public health'.[159] This includes ensuring high standards of quality and safety of health care, medicines or medical devices, and must be supervised by a professional under a duty of confidentiality.[160]

One of the challenges in the genomics context may be determining which Art 9 condition is most appropriate when there is a blurring between health care, research or technology development and clinical safety testing purposes. As artificial intelligence methods such as machine learning are increasingly being deployed in health or public health settings, health and even genomic data are likely to be processed in order to test decision support tools and other advanced technologies. This means that the deployment of a tool could involve overlapping forms of processing: processing for direct patient care or diagnosis; processing to test and further develop the accuracy of the tool, and/or; processing for research purposes.

The example of the Royal Free London NHS Foundation Trust partnership with Deepmind/Google Health's Streams application provides an example of the complexity of such scenarios. Streams is an application for clinical detection, diagnosis and prevention of Acute Kidney Injury ('AKI'). In 2015 and 2016 Streams was fed with data from ~1.6 million patients for clinical safety testing of the app as part of its development.[161]

A subsequent ICO investigation found that the data controller, the Royal Free, had failed to establish that this processing was fair, lawful, transparent and necessary or proportionate. In particular, the ICO found that such processing differed significantly from what data subjects may have reasonably expected to happen to their data when presenting at the Royal Free for treatment.

In response, the Royal Free has remedied its data protection in relation to Streams and now clearly sets out to patients and the public its legal basis and Art 9 condition for processing patient data in Streams. This has now satisfied the ICO,[162] so the data protection decisions made by Royal Free are a useful indication for other data controllers. What is clear is that the ICO has been satisfied that the processing on such a scale of patient data was necessary and proportionate for the clinical testing of an application and that the Royal Free is entitled to rely on the Art 9(2)(h) condition for provision of healthcare services and management of healthcare systems for clinical safety testing purposes.

This makes it likely that advanced applications (for example clinical decision support tools) that could be used in genomic medicine may also use this condition for clinical testing purposes.[163]

However, there is a potentially fine margin between clinical activities, non-clinical/non-research activities - such as audit or service evaluation, public health activities, and research. Where an activity is better described as research, the regime for scientific research we described in chapter 2 becomes relevant, including a specific condition for processing special category data under Art 9(2)(j).

Scientific research purposes

An important feature of the GDPR's regime for scientific research (see chapter 2) is that special category data may be processed if necessary, for scientific research purposes in accordance with Art 89(1) based on Union or Member State law (Art 9(2)(j)). This requires implementation in Member State law which must provide 'suitable and specific measures to safeguard the rights and freedoms of the data subject, in particular professional secrecy'. As we discuss further below, this gives rise to a challenge in itself because it leads to fragmentation of rules for scientific research (if actually implemented) across the EU/EEA.

The UK has enabled this condition through the DPA 2018 which establishes further safeguards. One is that research processing must be in the public interest.[164] However, the DPA does not define or explain what this means. As we noted in chapter 2, the concept of scientific research in the GDPR is very broad, so the requirement that research is in the public interest is a potentially important limit to this breadth in the UK's data protection regime for research. The concept of the public interest is itself potentially very broad and it can be argued to mean different things in different contexts.

An inherent challenge of the UK's implementation of the public interest test is uncertainty about when it may be met. It could be that it will be met by reference to a general goal or aim of the research itself or it could be, as Taylor and Whitton propose, that it introduces a much more substantive restraint to research use without consent. Their proposal, based on considered analysis of the concept of public interest in this context, is that 'it is only in the public interest to allow research processing without consent when individuals can be provided with reasons to accept this use without consent.'[165] On their account, the public interest test in the 2018 Act could not be satisfied if it were practicable to conduct the research and meet the high GDPR threshold of valid consent.

The UK has also implemented further safeguards which are less open to interpretation. One is that scientific research under Art 9(2)(j) must not be 'likely to cause substantial damage or substantial distress to a data subject'.[166] More significantly, if processing is carried out for the purposes of 'measures or decisions with respect to a particular data subject' it must have been approved by a research ethics committee.* This means that any genomic research which anticipates the return of genomic results or a need for recontact with a research participant in order to provide individual information which could influence their care or life choices, must have valid ethics approval. However, research which is purely observational and does not involve return of results or in any way influences decisions in relation to the data subject, does not require ethics approval under this provision. It is, however, highly likely that research ethics approval will have been required and obtained for such research as part of the broader regulatory framework for research in the UK.

Finally, Art 9(2)(j) requires researchers to conduct research in accordance with Art 89(1). As noted in chapter 2, this requires that researchers must also put in place 'appropriate safeguards' for the rights and freedoms of the data subject (Art 89(1)). These could include measures such as pseudonymisation and technical and organisational safeguards (as we discuss in chapter 8). The European Data Protection Supervisor has also suggested that 'even where consent is not appropriate as a legal basis under GDPR, informed consent as a human research participant could still serve as an 'appropriate safeguard' of the rights of the data subject.'[167] In line with the broader principle of data minimisation, Art 89 also requires that if it is possible to achieve the research purposes by further processing which 'does not permit or no longer permits the identification of data subjects', this should be done.

Overall, to meet this condition genetic or genomic data research must be as limited as possible and include technical and organisational safeguards such as pseudonymisation. If individual measures, for example recontact with results, are anticipated, research must have REC approval. It may be challenging to rely on this basis for cross-border research if Member States implement different requirements (or even if some do not enact this derogation at all). However, research processing has the benefit of a range of exemptions from some of the rights and obligations in the GDPR (see below) which could considerably reduce the burden on researchers.

Data which are manifestly made public by the data subject (Art 9(2)(e))

Some other Art 9(2) conditions may apply to particular uses of genetic/genomic data. For example, special category data may be processed if the processing relates to personal data which are manifestly made public by the data subject (Art 9(2)(e)). It is possible that this could apply if individuals have intentionally and consciously published their genetic or health information on an open website for anyone to access, for example on an ancestry database. However, this derogation should be interpreted in line with the higher protection required for special category data and the data subject herself must have positively and knowingly made data public, as opposed to making it accessible within a limited group of people, even if the data have become publicly accessible thereafter.[168]

* This includes research ethics committees and other bodies approved by the HRA, NHS organisations, the Secretary of State and some other sources of authority. DPA 2018 s 19(3), (4).

One of our research participants raised the prospect that courts and regulators may focus on the need for an on-going intention on the part of the data subject to make their data public, based on the present tense wording of the condition: 'processing relates to personal data which *are* manifestly made public by the data subject' [emphasis added].

> What needs to be looked at here - is it rescindable? … A court will say 'is being made public' is important. Need present tense. If the individual removes it or doesn't want it [to continue in the public domain] then it is not a legal basis.
>
> **Workshop participant**

Vital interests of the data subject or of another natural person (Art 9(2)(c))

Special category data may also be processed if it is necessary to protect the vital interests of the data subject or of another natural person where the data subject is physically or legally incapable of giving consent (Art 9(2)(c)). As previously discussed in relation to legal bases, this is intended as a last resort where it is a matter of life or death or there is a need for an essential diagnosis and where alternatives are unavailable, so it will not apply in many circumstances. Most importantly, under Art 9 this condition will only be met if the data subject is physically or legally incapable of giving consent. This could frequently occur where the data subject is incapacitated but it may also occur in emergencies. In their guidance on the similar provisions for international transfer, the EDPB refer to situations such as natural disasters where the data subject is 'considered to be unable to provide their consent.'[169]

This condition therefore, may be used for urgent genome sequencing, for example, where a patient is incapable of providing consent and the sequencing is necessary to track a life-threatening disease. In such cases there would be overlap with potential alternatives in Art 9(2)(h) or (i).

Substantial public interest (Art 9(2)(g))

A further option that could become relevant to the genetic/genomic context, is if processing is necessary for reasons of substantial public interest based in EU or UK law (Art 9(2)(g)). The DPA 2018 sets out specific circumstances where this condition may be met in the UK law.[170] None are currently directly relevant to healthcare, public health or research purposes but there is scope for this list to be amended over time.

National variations

One challenge for cross-border genomic medicine and research initiatives is that though the GDPR harmonises many aspects of data protection law, it also provides considerable scope for national variations for processing special category data. Article 9(4) explicitly allows Member States to introduce (or maintain) further restrictions, conditions or limitations to the processing of genetic, health and biometric data. The UK has yet to implement specific conditions but other countries such as France have.[171] As mentioned earlier, some nations have gone further including Finland and Italy, by introducing specific requirements for the processing of genetic data.[172] Most of the Art 9 conditions also require implementation or authorisation in EU or Member State law, so some, e.g. for scientific research, may not be available in all jurisdictions.This could mean explicit consent becomes the only realistic cross-border option.[173-174]

Finally, when Member States provide a legal basis for derogations like scientific research purposes, they must provide 'suitable and specific measures to safeguard the fundamental rights and interests of the data subject' (Art 9(2)(j)). The choice of these measures is left to Member States so significant differences arise. For example, Ireland has implemented measures which require explicit consent in health research unless a committee convened for this purpose can be satisfied that the public interest in the research 'significantly outweighs' the public interest in requiring the explicit consent of the data subject.[175]

This means that cross-border genetic or genomics projects are likely to face a complicated regulatory environment and will have to ensure that their approvals, policies, processes and patient/participant information meet the requirements of each relevant jurisdiction. This could mean that despite the challenges already discussed (and noting, in particular, different national conditions and safeguards relating to genetic data), explicit consent for the processing of special category data may be the preferred option for those seeking a single streamlined cross-border approach. Some of these challenges are discussed as part of chapter 7, focusing on impacts on international data sharing. This variation could mean, as Hallinan suggests, that it will be important for the genomics community (and in particular the genomics research community) to continue to advocate for the legitimacy of consent as a lawful basis for processing personal and special category data.[176] The genomics community can also highlight the legitimacy of even broader consent within genomic and biobanking research.

For example, the World Medical Association Declaration of Taipei supports consent for multiple or indefinite uses[177] and even in UK guidance, the Human Tissue Authority's code of practice on consent allows for broader or 'generic' consent.[178] Although there is a caution that there may be a power imbalance between researchers and participants where the controller is a public authority, this is not a hard and fast rule. It is open to the genomics community to resist the impression/misconception that there is always a clear power imbalance between researchers and participants, by providing a more nuanced and context based account, informed by the specific circumstances of individual genomics research projects. Fundamentally, it is open to the genomics community to argue that the GDPR, as a cross-sector Regulation, must be interpreted flexibly and in a way that takes account of the specific context of processing.

5.3 Conclusions

Some of the most high profile changes brought about by the GDPR are the enhanced standards for consent if consent is chosen as a lawful basis for processing under Art 6, or as a condition to enable the processing of 'special category' data under Art 9 (so long as consent is 'explicit').

The standards are undoubtedly high, but as discussed in this chapter there may be scope for the genomics community to argue for consent standards that are appropriate to genomics activities, in particular the need for broader consent to certain areas of genomic research. There are alternatives available, notably the provisions for scientific research, but these may not be available in other Member States, and may involve uncertain and inconsistent safeguards and additional requirements in cross-border processing collaborations. If alternatives are chosen, there is a potential challenge in ensuring they are appropriate for the specific genomics processing they apply to.

Even if consent is not chosen as a legal basis for processing, it may be difficult to communicate this to data subjects in ways that are both transparent and accessible. In complicated uses of genomic data that span healthcare, testing, technology development and research purposes, data controllers may need to identify multiple lawful bases and Art 9 conditions for different purposes. One resultant challenge is that these bases and conditions may have different consequences for data subject rights under the GDPR, a topic we now consider in detail.

6. Fulfilling data subject rights and meeting obligations under the GDPR

There is a direct link between the legal bases and Art 9 conditions discussed in the previous chapter and the data subject rights set out between Articles 13-22. The choice of legal basis for processing or Art 9 condition for processing special category data will make a significant difference to whether certain rights apply in the context of that processing. Because the application of data subject rights varies so significantly according to context, it can be complex for data subjects and data controllers to determine when rights and consequent obligations arise. This is one of the challenges that our analysis highlights in this chapter, and it is important because genomic data controllers will need to determine upfront and be able to explain clearly to patients or research participants which rights apply and how. Perhaps a greater challenge is determining what the fulfilment of certain rights require in the genomics context. For example, determining when genomic data are 'inaccurate' and thereby require rectification if requested by data subjects.

Our research suggests that we are at relatively early stage in terms of substantiating the requirements of some data subject rights in the genomics context. Many of the participants in our research were concerned to clarify elements of the GDPR such as lawful bases, which—to a degree—arise prior to engaging with some of the challenges of data subject rights. In our view, this gives rise to an opportunity to proactively engage with regulators and the EDPB (as some data protection academics are)[179] and ensure that appropriate guidance and standards relating to data subject rights are put in place for the genomics context. In this chapter, we discuss the potential nature of the GDPR's data subject rights in the genomic context as well as further obligations to ensure the security, confidentiality and privacy of processing.

6.1 Data subject rights

In Articles 13—22 the GDPR provides a range of data subject rights and consequent data controller obligations. Some are new stand-alone rights, such as the right to be forgotten (Art 17) whereas others carry over from the Directive but may be enhanced, such as the right to information which is bolstered to support transparency.

These rights build on the overarching principles contained in Art 5 and provide more specific means of securing the fundamental right to data protection that the GDPR seeks to protect. As we discussed in chapter 2, this background is important because principles for processing personal data and the rights to data protection and privacy can guide the interpretation of uncertainties or ambiguities about how data subject rights should be accommodated in practice.

As we discuss below, not all the data subject rights apply in all forms of processing (in particular there are significant restrictions to rights where data is processed for scientific research in accordance with Art 89 in the UK). However, all those collaborating as data controllers in genomic data processing will need to agree when and how data subject rights apply as data passes through complicated processing operations.

Joint controllers will need to agree their respective responsibilities because they are all jointly and severally liable for compliance with data subject rights (Art 26(2)). Even if one controller is not given the responsibility for actively supporting the exercise of particular rights, it will be the responsibility of all joint controllers to pass on any data subject requests to the relevant collaborator.

For example, if a patient makes a request to their healthcare institution for data held by another controller in a research database, all controllers could agree that the research controller should be obliged to provide the data to the data subject, but that it is the responsibility of the healthcare institution to pass on that request to the relevant research controller. The GDPR also requires data processors to put in place measures (if it is possible) to enable controllers to respond to any requests related to data subject rights (Art 28(3)(e)).

It will require careful consideration to put in place the arrangements and systems that may be required to support data subject rights in the genomics context, particularly as within a family, there may be multiple interests in the same genetic information which will need to be managed and reconciled. Before we turn to the content of data subject rights, we consider when they are, or are not, likely to apply in the genomics context.

Determining which rights apply and when?

Understanding which rights apply and when depends on the nature of the processing, the choice of legal basis and whether derogations have been made in Member State law (as they have been in the DPA 2018). There are multiple ways in which data subject rights can be restricted in the genomics context. One potentially important restriction for controllers of de-identified data is found in Article 11 which restricts the rights under Arts 15-20. This applies if the data controller can show they are 'not in a position to identify the data subject' (unless the data subject provides further information to facilitate identification). We discuss this further below.

Some rights are inherently limited according to the chosen lawful basis for processing. For example, the right to erasure will only apply to processing based on consent under Art 6 or explicit consent under Art 9. Other rights have inherent limits to when they apply. For example, the Art 14 right to information does not apply insofar as it would be impossible or a 'disproportionate effort' to fulfil.

In the case of genomics research, there are a range of limitations that apply where data are processed in accordance with Art 89(1), both in the text of the rights themselves (discussed for each right below) and as limited by Member State law in accordance with Article 89(2). In the UK the Data Protection Act 2018 sets limits to the operation of Article 15(1) to (3), (confirmation of processing, access to data and safeguards for third country transfers) Article 16 (right to rectification), Article 18(1) (restriction of processing) and Article 21(1) (objections to processing).[180]

These limits only apply when research results do not identify an individual and only to the extent that the application of these provisions would 'prevent or seriously impair the achievement of the [research or statistical] purposes in question'[181] so the onus is on researchers to explain why these rights do not apply. Because data subject rights are themselves a means of safeguarding the fundamental rights to data protection and privacy, there must be careful consideration of whether the fulfilment of data subject rights would prevent or seriously impair the purposes of the research.

Finally, Member States are allowed to introduce further restrictions to the operation of data subject rights under Article 23 for specific purposes, for example to safeguard the rights and freedoms of others. We discuss these context specific restrictions in relation to each right below, after initial consideration of the potential role of Art 11 in the genomics context.*

Article 11 'Processing which does not require identification'

Article 11 is a new aspect of data protection law and had no equivalent under the Directive. It provides a way of balancing the principles of data minimisation, storage limitation and integrity and confidentiality with the burden of fulfilling some data subject rights. It also corresponds with the Regulation's new emphasis on 'pseudonymisation' of data where possible, as a means of safeguarding personal data. In general, Art 11 reduces a controller's burden of compliance with the Regulation where data have been legitimately and sufficiently de-identified (e.g. through pseudonymisation). For it to apply, data must have been de-identified so that they are no longer identifiable without other 'additional information'. This echoes the description of pseudonymisation in the Regulation (see chapter 4) and Art 11 clearly may apply to data which have undergone pseudonymisation, if the 'additional information—or 'key'—is sufficiently separated and the data controller is able to demonstrate that they are 'not in a position to identify the data subject'. For example, if an uploading healthcare controller has coded data and passed it on to a genomic research collaborator without the key. However, there are some ambiguities about how Art 11 should be interpreted and applied which require further consideration.

* For a full list of the exemptions provided by the DPA 2018 to rights and obligations under the GDPR see:
The Information Commissioner's Office. Exemptions. Available from: https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/exemptions/ [Accessed 11th May 2020]

The Article is split into two parts:

> Article 11: Processing which does not require identification
>
> (1) If the purposes for which a controller processes personal data do not or do no longer require the identification of a data subject by the controller, the controller shall not be obliged to maintain, acquire or process additional information in order to identify the data subject for the sole purpose of complying with this Regulation.
>
> (2) Where, in cases referred to in paragraph 1 of this Article, the controller is able to demonstrate that it is not in a position to identify the data subject, the controller shall inform the data subject accordingly, if possible. In such cases, Articles 15 to 20 shall not apply except where the data subject, for the purpose of exercising his or her rights under those articles, provides additional information enabling his or her identification.

It is not entirely clear how the two parts of the Article interact and in some ways, leaving aside some further obligations to inform a data subject, arguably all the second part (11(2)) does is provide a more specific version of 11(1): specifying that the GDPR obligations which to be exempted are the rights in Arts 15-20. Because this is not straightforward, we consider how the two parts of the Article work to determine what the likely conditions are for Art 11 to apply to a genomics data controller.

Art 11(1)

The first paragraph states that a data controller is not obliged to *maintain, acquire or process* additional information simply for compliance with the Regulation and recital 57 provides further background to the Article, namely that a controller should not be 'obliged to *acquire* additional information in order to identify the data subject' (emphasis added). Exempting controllers from 'acquiring' additional information to identify the data subject is straightforward and it means that a controller without a key is not obliged to ask collaborators for it.

However, it is less clear whether a data controller may rely on Art 11 where they hold both the pseudonymised data and the additional information, albeit separated securely and subject to organisational safeguards. This could frequently be the case in many genomics contexts as the data are safeguarded within projects or institutions as a matter of course, so this would be a significant restriction on data subject rights and associated obligations.

Art 11(1) also states that a controller is not obliged to 'maintain' or 'process' additional information solely to comply with the Regulation. Could a data controller who holds both pseudonymised data and its 'key' argue that they are not obliged to process that 'key' to comply with the Regulation? On the face of it, it could be argued that this is the case. However, it would be a surprisingly broad restriction of data controller obligations if it applied in circumstances where a controller could easily combine the data with the additional information. In effect this would disapply many obligations under the Regulation whenever data have been pseudonymised.

There are those who argue that the spirit of the GDPR demands that Art 11 should not apply, even where re-identification by the data controller is impossible (if profiling decisions about individuals are made).[182] So, it would be surprising if the restriction applied where identification was not only possible but in fact, the resources required to achieve identification were relatively trivial. If the first paragraph of Art 11 stands alone as a restriction of controllers' obligations this could be arguable, however, if it forms part of a single cumulative requirement, to be read in association with the second paragraph, a narrower interpretation seems more likely.

Art 11(2)

Art 11(2) makes clear that, where the data controller is able to demonstrate that they are not in a position to identify the data subject, they must inform the data subject (if possible) and that in such a case the rights in Arts 15-20 shall not apply. This more specific restriction also appears to apply in narrower circumstances: only where a controller is able to demonstrate that 'they are not in a position to identify the data subject'. This would rule out Art 11(2) applying to the situation where a controller is in possession of both pseudonymised data and the 'key' for example.

Although Art 11(1) could be read as independent of the further requirements and conditions in Art 11(2), another way of reading Art 11(2) is that it could help interpret the first part. If this was the preferred interpretation, it could be argued that the Article has been constructed cumulatively, with Art 11(1) setting out 'cases' when controllers should not be obliged to maintain, acquire or process additional information to identify the data subject and Art 11(2) setting two conditions that apply to 'such cases'.

The problem is that the wording of the first paragraph is not contingent on the second, so, although it is clear there may be further requirements if conditions in 11(2) are met (notifying the data subject and not refusing information provided by the data subject), they do not necessarily impinge on the requirements in Art 11(1). However, as we considered in chapter 2, the aim of the GDPR is to ensure the protection of the fundamental rights of the data subject and the data subject rights (and associated obligations) contained within it are integral to that aim. As Ausloos and colleagues emphasise[183] and the CJEU has affirmed, this is a high level of protection which should not be unduly restricted or limited. In their consideration of GDPR data subject rights, Ausloos and colleagues argue that restrictions or limitations should be interpreted narrowly and in light of the context.[184]

Furthermore, when the Article 29 Working Party considered Article 11 as part of other opinions and guidance, they stated that 'this article should be interpreted as a way to enforce 'genuine' data minimization, without however hindering the exercise of data subjects' rights',[185] and, 'The Working Party rejects any interpretation of Article 11 aiming at reducing the responsibility of the controller(s) for compliance with data protection obligations.'[186] Although this opinion has not been officially endorsed by the EDPB, it has been approved in WP29 guidelines on transparency[187] which have been endorsed by the EDPB so it can be assumed that these recommendations on Art 11 would remain the same if they were formally endorsed. Accordingly, this analysis seems to support a narrow interpretation of the restrictions on data subject rights in Art 11 as follows:

'the controller is able to demonstrate that it is not in a position to identify the data subject'

On the narrower interpretation of Art 11, it would only apply where the controller is able to prove that they are not in a position to identify the data subject. As Runshan Hu and colleagues note,[188] Art 11(2) refers to a category of data that are neither anonymised or pseudonymised in the way that is described in Art 4(5) and recital 26 (and as discussed in chapter 4) because it only focuses on the ability of the data controller to identify the data subject, not on third parties to be able to do so.

However, if the data controller is in possession of additional information (a key) they would be in a position to identify the data subject. So, interpreting Art 11 narrowly must mean that data controllers must not be obliged to acquire additional information to identify the data subject, but that if they are able to 'maintain' or 'process' a key they already possess to identify them, then they will be obliged to do so.

As Hu and colleagues highlight, being in a position to identify the data subject may involve other methods of identification, such as singling out or making inferences based on further information and background knowledge to identify an individual. The example they provide is of a hospital data controller who is in possession of pseudonymised dataset of patients but is unable to rely on Art 11 because they have background knowledge from other published datasets and media reports which places them in a position to identify a patient data subject. In the genomic context, an analogous example might be of a clinician or clinical scientist, who is able to identify a patient through analysis of a rare variant in a genomic test.

'the controller shall inform the data subject accordingly, if possible.'

Art 11(2) imposes a requirement to inform data subjects that are no longer sufficiently identifiable to fulfil certain obligations and facilitate the exercise of data subject rights, if possible. Bearing in mind the principles of transparency and accountability, if this is not possible (e.g. because they are no longer identifiable) it is likely that data controllers should seek to make this information available publicly or to the community of relevant data subjects.

'in such cases, Articles 15 to 20 shall not apply'

Although Art 11(1) refers to compliance with the whole regulation being contingent on identifying the data subject, the narrower interpretation limits the restriction of data controller obligations to compliance with Arts 15-20. We discuss these rights further below.

'except where the data subject, for the purpose of exercising his or her rights under those articles, provides additional information enabling his or her identification'

Finally, this part of Art 11 makes clear that a controller cannot refuse further information provided by the data subject for the purpose of exercising those rights (e.g. the right to access personal data in Art 15).

Again, on the narrow interpretation of Art 11 this cannot only mean 'additional information' resulting from the process of pseudonymisation, in the sense of a 'key' or cipher, because data subject would be very unlikely to possess such additional information: it must mean any further information provided by the data subject to assist identification. In fact, the Article 29 Working Party have determined that 'invoking art. 11 of the GDPR without specifying what additional data are necessary to enable identification of the data subjects, the exercise of data subjects rights (access, rectification, portability, etc.) is *de facto* prevented.'[189]

WP29 advised that controllers should specify what additional information is necessary to enable identification of data subjects, according to the context. Although this opinion has not been officially endorsed by the EDPB, we assume that if these recommendations on Art 11 were to be endorsed, that they would remain the same.

This means that genomic data controllers may need to determine what information could be provided by a data subject that would enable them to be identified from pseudonymised and de-identified data, in reliance on Art 11. The task of determining this hypothetically, and then ensuring that this is communicated to data subjects may be more challenging than for controllers to retain a key, code or other additional information in a more accessible form, allowing that identification to take place, and not seek to rely on Art 11. Perversely, Art 11 could act as a disincentive to data minimisation through encouraging less data minimisation than may otherwise be possible. As we noted in chapter 2, failing to comply with the principle of data minimisation could itself result in enforcement or liability so data controllers could find themselves on the horns a dilemma if this is applied to them.

Finally, Art 11 only applies where 'the purposes for which a controller processes personal data do not or do no longer require the identification of a data subject by the controller.' This means it will not apply if the controller plans to recontact the data subject or provide them with individual results for example. However, if the controller is only going to provide those results to another controller in pseudonymised form, for example if a research institution provides pseudonymised results to a genetics clinic, Art 11 could still apply.

Overall, despite the ambiguity in how Article 11 should be interpreted, this may still reduce the burden placed on genomics data controllers, particularly genomics researchers, who hold pseudonymised data and keep it separated it from additional information. However, further challenges remain in reconciling Article 11 and data subject rights. For example if a data controller is relying on Article 11 in regard to data it has directly collected from the data subject, there is a potential clash with an obligation to inform the data subject of certain changes, such as an intention to further process the data for research purposes (Art 13(3)—see further below). On the narrow interpretation of Art 11, this obligation remains but would be irreconcilable with the proposition that a data subject is not in a position to identify the data subject.

It is important that these difficulties are brought to the attention of regulators and the EDPB to provide clarity for genomics and other data controllers.

> Art 13 gives you a mandatory obligation if you repurpose the data to let the data subject directly know. What if … you are disclosing it to a research organisation for a new purpose - but you don't have the contact details. What happens then? Are you stuck then [between Arts 11 and 13]?
>
> **Workshop participant**

We now discuss the specifics of the data subject rights in the genomics context, considering other further restrictions that may arise in the genomics context.

Data subject rights in the genomics context

Requirement for transparent information (Art 12)

Building, in particular, on the general principles of accountability and transparency, the GDPR provides several data subject rights to information in Arts 13 and 14, and the right of access in Article 15 (considered below). Article 12 provides the general rules in relation to these key transparency rights. It sets standards for information provided in communication with data subjects, in accordance with Arts 13 and 14 and other Articles, to be concise, transparent, intelligible and easily accessible, using clear and plain language, particularly for information addressed to a child. These standards do not need emphasis in the healthcare and genomics context where professionals are already aware of the importance of clear and accessible communication. However, the added challenge of communication in the data protection context is communicating complex regulation that patients, research participants and health professionals may be even less familiar with than with complicated genetics.

As already noted in chapter 5, it may be a significant challenge explaining, for example, that consent is being sought for ethical and common law purposes but that it is not the lawful basis for processing personal data under the GDPR. What is clear in guidance and best practice is that written information can be provided in a range of means, tailored to the situation and potentially using a layered approach with short notices that have additional layers of more detailed information. If a data subject requests it, information should be communicated orally but this may be in automated form using pre-recordings.[190]

Rights to information (Arts 13 & 14)

The rights to information in Articles 13 and 14 have the fewest exceptions of all data subject rights and will need to be complied with in almost all circumstances when processing personal genomic data. Article 13 sets out the information to be provided where personal data are collected from the data subject. Article 14 applies where personal data have not been obtained directly from the data subject but have been obtained from another, even publicly accessible, source. The only exception to Article 13 within the GDPR applies where the data subject already has the relevant information (although other restrictions can be introduced by Member States under Art 23). However, Art 14 has a more significant exception which applies if compliance would be impossible or would involve disproportionate effort or would seriously impact the processing objectives. There are some challenges that are likely to arise in the genomics context as noted below.

The general principle of both Articles is that a data controller is obliged to provide data subjects with a range of information about how and why data are being processed. This is to support transparency and fairness, and so that data subjects can scrutinise and challenge the use of their data, as well as enabling them to secure their rights. These rights apply even where other obligations or principles have been limited. For example, although scientific research purposes are deemed not to be an incompatible purpose requiring a new legal basis under the GDPR (see chapter 5), the data controller is still obliged to provide information about that new purpose prior to processing (Art 13(3) & 14(4)). The rights apply ex ante, so information should be provided prior to processing and when certain relevant decisions have been made.

When is information required?

- Information is required either at the time of data collection of data from the data subject (Art 13), or if the data was obtained from another (even public) source (Art 14) then:
- Within 'a reasonable period' after obtaining the personal data (14(3)(a)) (not exceeding one month), or
- When the personal data is used to communicate with the data subject (14(3)(b)) (e.g. on recontact with new results), or
- Where disclosure to another recipient is envisaged (14(3)(c)), or
- Where the controller intends to further process the data for another purpose e.g. for scientific research purposes (Arts 13(3) & 14(4))

Controllers must provide a range of details to data subjects under both Articles 13 and 14. Communicating the information in the box below might pose challenges in the genomics context.

What information is required?

- The purposes of the processing and the legal basis for each set of purposes
- The categories of personal data involved
- Any (categories of) recipients of personal data
- If applicable, the basis on which data are being transferred outside the EU and safeguards involved
- The period personal data will be stored for and the criteria used to determine that period
- The legitimate interests being pursued if that is the legal basis for processing
- The existence of the other data subject rights
- The right to complain to a supervisory authority
- The existence of automated decision making as referred to in Art 22(1) & (4)
- The right to withdraw consent if consent is the legal basis for processing
- Information on any further purpose for which the controller intends to process the personal data

One challenge is that recipients, or categories of recipients of personal data should be disclosed. It is not clear how specifically the 'categories' of data recipients must be defined. This could be important for information about research projects. For example, when drawing on the principles of transparency, fairness and purpose limitation, it seems likely that simply describing 'researchers' as a category would be insufficient. At the very least, it seems likely that the type of recipients will be required, for example whether they are public sector or commercial organisations.

Another challenge already noted, is where a controller intends to make a change such as adding another recipient or category of recipients of data, or, where the controller intends process the data for further purposes, for example a new form of research. As discussed earlier, this may be difficult where data have already been significantly de-identified so that Art 11 applies and the data controller is no longer in a position to identify the data subject. In this scenario the differences between Art 13 and Art 14 become important. This is because Art 13 is only limited to the extent that the data subject already has the relevant information, whereas Art 14 contains a more significant exception which can be invoked if informing the data subject would be impossible or would involve disproportionate effort, or would be seriously impair the processing objectives. This means that a data controller who has directly obtained some personal data from a data subject (e.g. a patient/research participant) and has done their best to safeguard and minimise data in accordance with the Regulation will face a dilemma in fulfilling their obligations under Art 13. Whereas they may avoid the same dilemma using the following Art 14 exceptions simply because they have obtained the personal data indirectly, from another legitimate source. This distinction in burden placed on the controller is not entirely coherent.

Art 14 Exception: Proves impossible, disproportionate effort and serious impairment of objectives

Article 14 provides a significant exception to the right to information where personal data has not been obtained from the data subject if:

> 'the provision of such information proves impossible or would involve disproportionate effort … or in so far as [this would] render impossible or seriously impair the achievement of the objectives of that processing' (Art 14(5)(b))

Art 14 emphasises that this could apply, in particular, for scientific research conducted in accordance with Art 89(1) so it is clear that this is an important exception for genomics researchers. The WP29 advise that this (and the other exceptions—see below) should be interpreted and applied narrowly and sets a high bar for the exception.[191]

The WP29 considered the fact that there is no equivalent exception in Article 13, and concluded that the impossibility or disproportionate effort exception 'must be directly connected to the fact that the personal data was obtained other than from the data subject.'[192] They cite an example of a hospital processing data about next of kin obtained from their patients and conclude that it would involve a disproportionate effort to provide all next of kin with information in accordance with Art 14.

Another example is of researchers obtaining a large dataset which was collected over 50 years ago. In this case, the size and age of the dataset mean that it would be a disproportionate effort to trace all the data subjects. WP29 mandate a balancing exercise to assess the effort involved with the potential impact and effects on the data subject if they are not provided with the information. This assessment could lead to further measures to protect the data subject's rights and legitimate interests, and should be documented by the controller. According to WP29, one of the alternative measures that must always be taken to protect the data subject's rights and interests in such circumstances is making the information publicly available (Art 14(5)(b)).

The WP29 view is that this exception should not be routinely relied on by those who are not processing data for archiving, scientific or historical research purposes or statistical purposes (in accordance with Art 89(1)). The implication is that it is possible for genomics researchers to routinely rely on it if they can justify the impossibility or disproportionate effort required.

In summary, the exception in Art 14 may provide an important reduction in the burden on genomics data controllers, as part of scientific research. However, the inconsistency between Art 13 and 14 and the lack of exception where data have been obtained directly from the data subject may lead to dilemmas for genomics controllers. Some participants in our policy workshop highlighted this inconsistency and even argued that Art 13 should be revised.

> Art 13 isn't properly thought through. The same reasons that apply in cases of Art 14 - that it is a disproportionate effort - would apply in the same way in that situation in Art 13…. Art 13 needs to be revised.
>
> **Workshop participant**

Right of Access (Art 15)

One of the data subject rights that raises particular challenges in the genomics context is the right of access. This is a right to obtain a copy of the personal data undergoing processing from the data controller (Art 15(3)). Despite reports of one data protection authority interpreting this as more of a 'summary' than an exhaustive copy,[193] 'copy' is more generally understood to mean all the data that relate to a specific individual. As noted in chapter 4, personal data may include opinions and other subjective material if they 'relate' to an individual so this right will also apply to professional opinions and notes associated with genomic test results.

This is a major challenge in the genomic context, because a copy of personal data could extend to the full sequence of data resulting from whole-exome or whole-genome sequencing, in addition to the associated clinical or phenotypic information. Such a volume of data means that facilitating access is a significant challenge, and it may come as a surprise to patients/research participants to discover the volume of data involved.

Key impact
Research participants highlighted that a genome may be between 80-200Gb and it takes significant expense and time to create an encrypted file of this size, that sending a physical USB or hard drive may be the only practical option and that data subjects may be surprised at the scale and nature of 'their genome'.

The GDPR also stipulates that if the request is made electronically, a copy should be provided in a 'commonly used electronic form'. What this entails for genetic data is not precisely clear but at least a physical printout of data is not necessary following such a request. However, this is not the same as the requirements of the right to data portability which mandates structured machine-readable formats (see below).

Disclosing genomic information under Art 15

One limitation to the right of access that is particularly relevant to genetic information is that it should not adversely affect the rights and freedoms of others (Art 15(4)) and Member States are entitled to restrict the right of access for this purpose (Art 23(1)(i)).

To this end, the UK has implemented a clarification that data controllers are not obliged to disclose personal data under Art 15, 'to the extent that doing so would involve disclosing information relating to another individual who can be identified from the information',[194] unless the other individual has consented[195] or 'it is reasonable to disclose the information to the data subject without the consent of the individual'.[196] A range of considerations are mentioned as relevant to whether the disclosure is reasonable (e.g. any express refusal of consent by the other individual) so this decision is not entirely at the discretion of the controller.

Interestingly, the data controller is not asked to determine whether the information is 'personal data' relating to the other individual using the standard GDPR test, but instead must conduct a slightly different analysis to determine if it is 'information relating to another individual':* data which are either directly identifying or which could indirectly identify an individual from 'information that the controller reasonably believes the data subject is likely to possess or obtain.'[197]

This could clearly apply to genetic information of familial relevance. In this case, a controller faced with an access request for genetic or clinical data is required to focus on what other information the data subject is likely to possess or obtain which could identify another family member. This could be a significant challenge because data could relate to multiple family members and it would then become a controller's responsibility to seek consent from multiple people or to determine that disclosure is reasonable. As under the previous law[198] the DPA provides some factors that must be taken into account when determining whether it is reasonable to disclose information without consent:

> (a) the type of information that would be disclosed,
>
> (b) any duty of confidentiality owed to the other individual,
>
> (c) any steps taken by the controller with a view to seeking the consent of the other individual,
>
> (d) whether the other individual is capable of giving consent, and
>
> (e) any express refusal of consent by the other individual.

These are useful considerations but determining when disclosure is reasonable in this scenario may not be straightforward. The more common dilemma for controllers of genomic information is where information could cause harm to the individual (discussed below) or where the information is confidential patient information relating to a relative who has refused consent to disclosure.

In the Art 15 scenario, it is more complicated because an initial assessment must have been made that the data in question are personal data relating to the person making the request, so they are likely to perceive it as their data. Withholding a patient or participant's own genetic data because it also relates to another person appears different to beginning with proposition that this is another person's confidential information, yet the factual situation could be the same: there could be genetic results that are relevant to several first degree relatives, cared for by the same healthcare team.

The breadth of 'personal data' (as discussed in chapter 4) means that disclosing some information could indirectly identify another person, even with the addition of the slightly different test for 'information relating to another individual' discussed above. For example, disclosing a genetic result in the patient's records could indirectly identify the family member who was the source of that result if there is sufficient background knowledge, such as knowledge of recent interactions with the health system or even phenotypic information which is related to the result.

* This does not include another health professional. DPA 2018, sch 2 para 17(2).

In a clinical context, the challenge of balancing the interests of biological relatives within a family is addressed in best practice guidance. Because the common law duty of confidentiality will also be likely to apply to some of this information, it is likely that the conclusion that disclosure is reasonable would follow best practice for disclosing confidential information. However, it could be important to have clarity that the reasonableness test in such circumstances and guidance on disclosure of confidential information are indeed aligned.

The UK has also implemented another exemption from Art 15(1)-(3) specifically for health data known as the 'serious harm test'. This restricts a data controller from disclosing health data 'if the data would be likely to cause serious harm to the physical or mental health of the data subject or another individual.'[199] If the controller is a healthcare professional (HCP) they may withhold information on this 'serious harm' basis.[200] If the controller is not a HCP they must obtain, or have obtained within the previous 6 months, the opinion of an appropriate HCP[201] (e.g. the most recent responsible HCP or the most suitably qualified for the purposes) that the serious harm test is not met.[202] The exemption does not apply if the data subject is already aware of the information.

The implication for genomic data controllers is significant because this provides for a therapeutic exception to disclosure both in favour of the data subject but also in relation to another individual. This means that in the scenario discussed above, where information may relate to another family member for example, a HCP is entitled to consider if disclosure would cause either the data subject or the family member serious harm, and refuse to disclose on this basis.

For scientific research conducted in accordance with Article 89(1) and s19 of the DPA, the right of access will not apply to the extent that it would prevent or seriously impair the achievement of the research purposes, so long as no research results or statistics are published in an identifiable form.[203]

Right to Rectification (Art 16)

Another right that raises some specific questions for genetic or genomic data is the right to rectification of inaccurate data or, 'taking into account the purposes of processing', the completion of incomplete data. This is an important corollary to the principle of data accuracy, namely that data should be kept accurate and up to date, and raises the question of when genetic or genomic data can be considered inaccurate. The GDPR does not define accurate or inaccurate, but the DPA 2018 does describe inaccurate in relation to personal data as meaning 'incorrect or misleading as to any matter of fact'. (DPA 2018, s205(1)).

The reason this right, and the principle of data accuracy may be challenging in the genomics context is that data such as variant classifications frequently become out of date, and the rate at which new evidence is developed continues to grow. This has led to significant debate in the genomics field about whether there are obligations to re-analyse, update and 'correct' genetic results as evidence develops, and to recontact individuals with significant updates. The current guidance from professional organisations is generally that recontact is desirable if new significant information is available, but that the time and resources involved mean it is not feasible in all circumstances.[204, 205, 206]

Does the principle of data accuracy and right to rectification imply that all genetic records need updating as a matter of course, to correct inaccurate or misleading classifications? Part of the answer is that data are only required to be accurate on their own terms. For example, if they are the results of a test with a significant margin of error, so long as that margin of error is explained, the data will be accurate even if they may be erroneous. If there is no reference to the chance of error then the data may be misleading. Equally, if results are reported as accurate according to the current state of the evidence, this will not be misleading. If a conclusion about a genetic result is a matter of opinion, this should also be explained. Even if data are updated, it could be that earlier 'inaccuracies' should be retained as an accurate record of the analytical or decision-making process.[207]

In the longer term, it is an open question whether the right to rectification could be applied to require the updating of records if their results are clearly no longer accurate. On the one hand, it could be argued that the records are an accurate account of the results at the time of testing and analysis. On the other hand, if the results potentially influence the data subject's eligibility for screening or have an impact in other ways (perhaps even on insurance options in the future), then arguably, a failure to rectify results could result in records that are factually inaccurate or misleading about the data subject's true health status.

As with other rights, in scientific research the right to rectification is limited in the UK, to the extent that it would 'prevent or seriously impair the achievement of the [research or statistical] purposes in question.'[208] This could mean that the issue of data accuracy and rectification is likely to be a greater challenge for processing in clinical activities than genomics research. However, the research restriction will only apply if research results are not made available in a form which identifies a data subject.

If there is feedback of results or data into the clinical setting the right to rectification still applies. In many cases, there is a significant level of intended feedback of individual results from genomics research (and because these two activities are increasingly joined in large-scale initiatives) the right to rectification and the issue of data accuracy are likely to be relevant to a significant number of research and healthcare genomics data controllers.

Right to erasure or to be forgotten (Art 17)

The stand-alone right to erasure is a new addition to the GDPR with potentially far reaching impact, particularly if it can be applied to research databases. However, this right is largely restricted to circumstances where data are not processed for health, public health[209] or research purposes (if it would seriously impair the research)[210] and when processing is based on consent. If consent is relied on as a legal basis for processing, or if explicit consent is used to justify the processing of special category data, the right to erasure should be complied with unless there are other legal grounds for processing.[211] If a request for erasure is valid but the data has already been made public, the data controller is obliged to inform other controllers who are processing the data about the request.[212] This is particularly important if the data have been transferred internationally.[213]

An ambiguity that could have a significant impact on further genetic research is whether erasure can be validly achieved through anonymisation or if it requires complete deletion of data. Whilst some authorities agree that anonymisation could suffice,[214] others disagree and argue that this disempowers data subjects who may wish to prevent the future use of that data.[215] In the genomics context, this could mean the difference between requiring substantial efforts to remove data from a database such as ClinVar, or, accepting that 'anonymised' data can still be used for research.

Right to restriction of processing (Art 18)

According to Art 4(3) 'restriction of processing' means the marking of stored personal data with the aim of limiting their processing in the future. Art 18(1) provides data subjects with a right to restrict processing when one of the following applies: the data subject has requested that the accuracy of the data be verified by the controller; an objection under Art 21 is pending (see below); processing has been determined to be unlawful and the data subject opposes the erasure of the personal data and requests the restriction of their use instead, and; where the data subject wishes to ensure the data still exist for the establishment, exercise or defence of legal claims.

Our research suggests that no challenges have been raised about the impact of this right on genomics data processing but Ausloos and colleagues express concern in their recent submission to the EDPB on data subject rights that, '[d]espite many such requests, we have encountered not a single data controller that acknowledged, let alone accommodated, the right to restriction of processing.'[216] Given this it seems sensible to bring Art 18 to the attention of the genomics community. It is also important for data subjects to appreciate that they have a right to restrict processing while the veracity of information is being established, or while their objection is pending.

Notification obligation for Arts 16, 17 & 18 (Art 19)

Article 19 is not a data subject right as such but it is a notification obligation for controllers to communicate any rectification, erasure, or restriction of processing required under the preceding articles, to all previous (or ongoing) recipients of the relevant personal data. This applies unless this proves impossible or involves disproportionate effort. Again, in view of the fundamental importance of data subject rights, the assessment of disproportionate effort is likely to require a balancing exercise, with a significant weighting towards the importance of securing data subject rights. If it is impossible to directly notify all recipients, it is likely that measures such as publication of information on a website will be required.

Right to data portability (Art 20)

Another new addition in the GDPR is the right to data portability. A more limited form of the right to access personal data, portability is aimed at ensuring data are provided in 'commonly used and machine-readable' formats and that a data controller does not hinder the transmission of personal data to another controller.

This right is quite limited: it only applies to data that the data subject 'has provided to a controller', not data which have been obtained from a third party or which the data controller has generated themselves. This means that it applies to the 'input' data provided by the data subject and not the results of further analyses, although this distinction may not be so clear when the results of analyses contain or describe the input data.

Guidance from WP29 contrasts data provided by the data subject with 'inferred data' and 'derived data' created by the data controller through analysis of data 'provided by the data subject'.[217] Given that sequencing data and results are derived from an analysis of the material provided by the data subject, genetic data and results are not likely to fall within the right to data portability.[218] This right only applies when processing is on the basis of consent or contract, or where special category processing is justified via explicit consent, or, where processing is 'carried out by automated means' (Art 20(1)). Even if consent were the basis or justification for processing genetic data, neither the results of sequencing or further analysis need to be provided in machine readable format, because they are not the data which the data subject provided.

Right to object (Art 21)

The right to object may apply to any processing of genetic or health data (including profiling) based on legitimate interests or performance of a public task.[219] If a data subject objects, the controller may only continue processing if they demonstrate compelling legitimate grounds which override the rights and freedoms of the data subject. However, there is a specific exemption for scientific research conducted on the basis of the performance of a task in the public interest (Art 21(6)), so public research institutions are not obliged to comply with an objection. This means it is likely to only be private or charitable research bodies who have to demonstrate compelling grounds to continue processing. How 'compelling' is to be assessed is unclear but a high standard is likely to be applied to ensure sufficient respect for data subjects' fundamental rights.

Right not to be subject to solely automated processing (Art 22)

A final right that applies only in narrow circumstances is the right 'not to be subject to a decision based solely on automated processing, including profiling, which produces legal effects concerning him or her or similarly significantly affects him or her.' This will apply unless the processing is based on explicit consent (or contract, if the data are not special category data), in which case there should at least be the right to obtain human intervention, express a point of view or contest the decision.*

At present this is highly unlikely to apply in the health or genetic/genomic context, even with advanced AI-based processing, because there will almost certainly be a human health care professional who can intervene in the decision-making process if a decision has a serious and significant impact.

* Or if it is in the substantial public interest on the basis of Member State law (Art 22(4)) and safeguards are in place.

The principles of fairness, accountability and transparency all suggest that information may need to be provided, depending on the context, about how data are being processed using AI and other automated processing, even if the decision is not solely automated. The GDPR's provisions relating to automated processing are likely to become increasingly relevant in the medium to longer term as increasingly sophisticated AI-driven systems are applied in the genomics context. Some of these issues are considered in separate PHG Foundation research on Black box medicine and transparency.

6.2 Privacy and security of genetic data

Beyond data subject rights, the GDPR also requires data processors and controllers to comply with a range of further obligations, depending on the scope, context and purposes of processing, and the potential for harm. As we have already discussed at various points in this report, the GDPR generally requires proportionate responses to the sensitivity and risk of processing and these are likely to be elevated in the processing of genetic or genomic data, particularly on a large scale.

Article 24 requires data controllers to implement 'appropriate technical and organisational measures' according to the nature, scope and purposes of processing. Similarly, Article 25 requires 'data protection by design and default' which means a context-sensitive and proportionate implementation of safeguards, such as pseudonymisation, to implement data protection principles and protect the rights of data subjects.*

Article 32 adds further detail on the measures that are required to ensure security of processing and, where processing 'is likely to result in a high risk to the rights and freedoms of natural persons', a controller must carry out a data protection impact assessment (DPIA) in accordance with Art 35.

Although genetic data are particularly sensitive, the EDPB have made clear that the processing of genetic information does not automatically count as high risk in itself.† The WP29 guidance, endorsed by the EDPB, is that the processing of sensitive data is one criterion that could lead to a DPIA being required if another criterion is also met. These criteria potentially include large scale processing of data concerning vulnerable subjects, as in the case of patients or children, or, if the processing involves evaluation and scoring (including profiling and predicting) of data subjects' health.[220]

* For more detailed guidance on data protection by design and default see:

The Information Commissioner's Office. Data protection by design and default. Available from: https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/accountability-and-governance/data-protection-by-design-and-default/

† The EDPB made clear in its opinion on the ICO's DPIA proposals that the processing of genetic data alone is insufficient to automatically require a DPIA, and that another criterion must be present, see: European Data Protection Board. Opinion 22/2018 on the draft list of the competent supervisory authority of the United Kingdom regarding the processing operations subject to the requirement of a data protection impact assessment (Article 35.4 GDPR). 2018.

It is likely that the processing of genetic data will require a DPIA and the implementation of significant technical and organisational safeguards in many circumstances, including in research projects which process large amounts of personal data.* A key implication of these privacy and security obligations is that data controllers are required to consider future technology and the emergence of other data that could be used to identify individuals.

As discussed further in chapter 8, the rapidly developing approaches for identifying or protecting genomic data mean that fulfilling these obligations will require constant attention and review to ensure that state-of-the-art safeguards are in place to protect genomic data. The technical, organisational and legal safeguards that may be applied to mitigate risks in the genomic context are analysed in chapter 8.

6.3 Conclusions

Our research has highlighted that there are a range of challenges and ambiguities in relation to the fulfilment of data subject rights in the genomics context. First, there is the challenge of determining which rights apply in the specific context of processing within and between Member States. Then there are a number of ambiguities in interpreting and applying the exception under Art 11 for processing which does not require identification, and even the potential that this could reduce efforts to de-identify or minimise data rather than promote these efforts.

In terms of specific data subject rights, Art 15—determining how to deal with access to data which relate to several genetic relatives and Art 16—the right to rectification and assessing when genomic data are inaccurate—give rise to the most specific challenges for genomic data processing. In particular, more work is needed to clarify how best to manage and reconcile the interests of multiple family members in the same genomic information. As we have noted, other challenges, such as determining when research exemptions apply and whether erasure may be fulfilled through anonymisation are also likely to arise in the genomics context. The GDPR also requires data controllers and processors to approach data protection by design, to put in place safeguards to protect the privacy and security of data that are proportionate to the risks involved and to keep them under review as technologies and techniques progress. These are significant obligations which require technical, organisational and legal safeguards, as we discuss further in chapter 8. First, we turn to discuss another key impact of the GDPR on genomic data processing: its impact on genomic data sharing, both within the EU/EEA and internationally.

* For more detailed information on how to conduct a DPIA, see:

The Information Commissioner's Office. How do we do a DPIA? Available from: https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/data-protection-impact-assessments-dpias/how-do-we-do-a-dpia/#how9

7. Challenges for genomic data sharing

Some of the major concerns about the effect of the GDPR on genomic data relate to its impact on data sharing.

One aspect concerns the provisions for transfer of personal data outside the EU/EEA, which are subject to a specific set of rules under the GDPR. This has become particularly relevant for the UK genomics community because the UK is now technically a third country, outside the EU/EEA. However, as we discuss further below, EU legislation continues to apply by agreement to the UK during the current transition period and the UK has legislated to switch from the EU GDPR to a largely equivalent 'UK GDPR' thereafter.

While there is uncertainty about the legal mechanisms that may be required for data transfer between the EU and the UK, at present, it appears that the UK plans to remain closely aligned with the GDPR. However, as our research, in particular our interviews and workshop have emphasised, there are some very significant challenges for genomic data sharing within the EU/EEA and even within Member States raised by a lack of agreement about the requirements of the GDPR and Member State legislation.

In this chapter, we discuss the challenges for genomic data sharing within the EU/EEA under the GDPR and Member State legislation before assessing the additional challenge of international genomic data transfers outside the EU/EEA. We also consider the legal position in the UK following Brexit. Some of the sector-specific approaches that have been proposed in the literature and by participants in our research to address these challenges are discussed in the following chapter. First of all, we provide a brief overview of the current scale and multifaceted nature of genomic data sharing.

Genomic data sharing

Data sharing is a fundamental part of genomic medicine and research. Knowledge of genetics and genomics underpins our knowledge and understanding of the nature and development of most diseases including inherited conditions and cancers. The diversity of genetic conditions and the relatively small number of patients diagnosed with some genetic disorders, means that there has long been a practice of matching cases that share clinical symptoms and exchanging clinical and genetic information between clinical genetics services and researchers to enable very rare genetic changes to be correctly interpreted. This has frequently gone hand-in-hand with the generation of new insights from research based on such data so that clinical data are used for research which, in turn, will inform individual patients' care.[221]

The diversity and complexity of genetic conditions also means that local sample cohorts are frequently too small to generate useful or meaningful results for genomic research, which has therefore been advancing at a population level in the UK and around the world.[222] Research initiatives like the 100,000 Genomes Project and UK Biobank[223] are taking forward whole genome sequencing of hundreds of thousands of participants and around the world there is an increased fusion of genomics research and healthcare delivery as the costs of sequencing fall dramatically.

As Birney describes, there are many global examples of population scale whole genome sequencing. These include Genomics England (100,000 Genomes Project), Finnish population and health cohorts (FINNGEN), Australian Genomic Health Alliance (AGHA), the Estonia biobank, the AMED Umbrella project in Japan, Iceland as a single study population with Decode, Plan Ian France Génomique in France, the Andalucian Genome project in Spain, Geisnger and Kaiser Permanente, and the Centers for Mendelian Genetics in the US.[224] An EU project to facilitate access to at least 1 million genomes is underway in Europe[225] and leading scientists are predicting the sequencing of 47.5 million genomes and 83 million genome sequences for cancer worldwide by 2025.[226] Alongside the generation of such a volume of data, significant infrastructure will be required for storage and processing. This looks set to increasingly take place on the cloud,[227] further complicating flows of genomic and associated clinical data. Through science and research discovery these data have the potential to significantly advance knowledge of health and disease offering universal benefits. However, data sharing is being challenged by diverse and potentially diverging data protection laws both within and beyond Europe.[228]

7.1 Challenges for data sharing within the EU/EEA

Although the GDPR aims to increase the harmonisation of data protection law across the EU and EEA, there remains considerable scope for divergence in approaches between Member States, Supervisory Authorities and data controllers across Europe. The scope for such divergence is further increased in some of the key aspects of data protection relating to genomic healthcare or research: the processing of genetic data, the processing of health data and the processing of data for scientific research purposes. Many of these issues have been touched on in previous chapters. In this section we set out the aspects of data protection law that may require agreement between data controllers, or controllers and processors in order to enable data sharing within the EU/EEA and some of the challenges in reaching consensus.

Reaching agreement among collaborators

Some of the challenges for genomic data sharing stem from a lack of agreement about the requirements of the GDPR and what constitutes best practice in relation to genomic and health data.

The scope of personal data

One potentially fundamental challenge to data sharing within the EU/EEA is a divergence in views about the scope of personal data in the genomics context. As we discussed in chapter 4, different approaches may be taken to determining when genomic or health data are sufficiently identifiable to constitute 'personal data' and therefore fall within the scope of the GDPR. This may apply to assessment of whether data that have been de-identified are sufficiently anonymous, or, whether genome sequences are inherently identifiable. This may also apply to assessments of whether combinations of genomic and metadata are potentially identifying, or, whether additional information is available which could lead to the identification of an individual. We return to this topic, the technical debate about de-identification and some potential mitigations in the following chapter.

For now, we note that it is challenging for data controllers and regulators to agree when 'personal data' are being processed in the genomics context.

Agreement about the territorial scope of the GDPR

As discussed in chapter 3 it is also challenging in certain circumstances to determine when processing is taking place in the context of the activities of a controller or processor who is a healthcare or research institution established in the Union. If this is the case then the GDPR will apply. This has potentially profound implications for international genomics initiatives which will need to comply with the GDPR or, alternatively exclude collaborators from within the EU/EEA. Whilst this latter option may reduce the regulatory burden, it may be suboptimal from a scientific perspective.

Fulfilling rights and meeting legal obligations

Once there is consensus that personal data is being processed, collaborators must also agree what is required in order to satisfy the fundamental principles and requirements of the GDPR, and to facilitate and fulfil the exercise of data subjects' rights. This applies as much within a single Member State as it does between them. In chapter 5 we discussed how there are already different interpretations of consent, and the feasibility of broad consent as a lawful basis for processing personal data, and for processing special category genetic or health data under the GDPR. This is particularly important because Member State laws are required for most of the alternative options for processing genetic and health data. There may also be differences in opinion about what technical and organisational measures are appropriate to safeguard such sensitive data in accordance with the GDPR (e.g. Art 25 data protection by design and by default). These are some of the issues that must be transparently agreed by data controllers in accordance with Art 26.

In chapter 6 we highlighted the range of uncertainty that remains in relation to fulfilling data subject rights, such as the right to rectification and the need to ensure data accuracy, in the context of genomics. There is the potential for disagreement about what fulfilment of data subject rights require in this context and even disagreement about when they do or do not apply, because this varies according to the legal basis or Art 9 condition. These interpretative differences concerning data subject rights, along with a range of relevant aspects of the GDPR are problematic for genomic data sharing because they are subject Member State divergence.

Regulatory divergence within the EU/EEA

Throughout this report we have highlighted areas which are subject to actual and potential regulatory divergence between EU/EEA Member States. Although the GDPR harmonises data protection law across Europe to a greater degree than its predecessor, as we have noted, there is particular scope for Member State derogations and tailoring of rules in the genomics context: the processing of 'special category data' is subject to significant variation because most of the Art 9 conditions require implementation or authorisation in EU or Member State law. For example, the scientific research provision may not be available in all jurisdictions.

Where Member States have enabled a form of processing, such as scientific research, there may be significant differences in national requirements. For example, the choice of 'suitable and specific measures to safeguard the fundamental rights and interests of the data subject' in scientific research (Art 9(2)(j)) is left to the Member States so significant differences arise. The example of the Irish research measures which require explicit consent in health research unless a committee can be satisfied that the public interest in the research 'significantly outweighs' the public interest in requiring the explicit consent of the data subject is relevant in this context.[229]

National legal differences may also influence practice in genomics projects. For example, in Germany, 'research purposes' exclude processing that generates conclusions about specific individuals and in the UK, research results must not be made available in an identifiable form.[230] This challenges the increasingly blurred nature of genomics that is situated between healthcare and research practice and may lead to differences between individuals being offered testing and results between countries. Further challenges are raised by Article 9(4) which explicitly allows Member States to introduce (or maintain) further restrictions, conditions or limitations to the processing of genetic, health and biometric data, as countries such as France,[231] Finland and Italy[232] have done.

Member States are allowed to introduce further restrictions to the operation of data subjects' rights under Article 23 for specific purposes, for example to safeguard the rights and freedoms of others, and under Art 89(2) in the case of scientific research. This may significantly alter the application of data subjects' rights across the EU/EEA when processing genomic information.

> [Is it a challenge trying to ensure you've got the right safeguards for different MS in place?] Yes, that's a huge challenge … we are thinking of incorporating only in one MS as a solution
> **Interview participant**

Finally, it is not only hard law and regulation that is subject to divergence within the EU/EEA. The advice and governance of supervisory authorities and other national policymakers may also vary, and this can have significant impacts on data sharing practice.

Taken together, this means that all multi-party genomics projects—and especially cross-border— genomics projects are likely to face a complicated regulatory environment and will have to ensure that their approvals, policies, processes and patient/participant information meet the requirements of each relevant jurisdiction. All actors will need to agree their approach to processing in compliance with the GDPR and relevant MS law and they must agree what legal, organisational and technical measures are required to safeguard genomic and health data.

This requires the urgent attention of the genomics community and data protection authorities to ensure that there is not a negative impact on genomic and health data-sharing for healthcare and research. We discuss how the sector can help to ensure this is not the case in the following chapter.

Challenges for data-sharing post Brexit

A recurring question is whether, and to what extent the GDPR will still be relevant in the UK after Brexit. Pre-Brexit, the GDPR - as a regulation - was directly applicable and a part of UK law. After the end of the transition period, the GDPR will (all going to plan) become the 'applied GDPR' or 'UK GDPR,' being transferred under the authority of the European Union (Withdrawal) Act 2018 with the modifications listed in Schedule 1 of The Data Protection, Privacy and Electronic Communications (Amendments etc) (EU Exit) Regulations 2019.

At the time of writing, these modifications mostly concern procedure and not substantive content. However, on conclusion of the transition period there will be two forms of the GDPR governing processing in the UK and transfers between the UK and the EEA. It is currently unclear to what extent this may lead to conflict, in particular if the guidance of UK and European supervisory authorities diverge.

The UK Government has said that transfers from the UK to the EEA will not be restricted but any transfers from the EEA to the UK (as a third country) will need to ensure adequate protection of personal data under the GDPR, as we discuss below. Much will depend on the outcome of the negotiations but it is possible that there will be differences in approach to some of the aspects of the GDPR and how they apply to genomic data.

As a third country, the UK and the ICO will have a much reduced ability to press its case for a particular approach (and the UK will fall outside the consistency mechanism discussed in chapter 2) so it could be that UK based genomics professionals will have to follow different rules, for example in determining when data are identifiable 'personal data', when collaborating with European partners, from those that apply to purely domestic processing. This would have a profound impact on 'international transfers' of genomic data between the EU/EEA and the United Kingdom. These forms of transfer outside the EU/EEA are governed by specific rules and, as we discuss in the following chapter, they are already impacting genomic data sharing between the EU and the rest of the world.

7.2 To third countries and international organisations

One of the major challenges to data sharing under the GDPR is the transfer of personal data to third countries and international organisations. Notably, lawful transfer of personal data to a recipient in a third country or international organisation requires a legal mechanism, this mechanism often being in addition to a legal basis (Article 6) and a derogation for special category data (Article 9). This section considers the general principle for transfers and the relevant legal mechanisms to give effect to such transfers: adequacy decisions, appropriate safeguards, derogations, and compelling legitimate interests.

7.2.1 The general principle for transfers

> Article 44 Any transfer of personal data which are undergoing processing or are intended for processing after transfer to a third country or to an international organisation shall take place only if, subject to the other provisions of this Regulation, the conditions laid down in this chapter are complied with by the controller and processor, including for onward transfers of personal data from the third country or an international organisation to another third country or to another international organisation. All provisions in this chapter shall be applied in order to ensure that the level of protection of natural persons guaranteed by this Regulation is not undermined.

Article 44 states the general principle for transfers of personal data to third countries or international organisations. Article 44 provides the general context with which to interpret many of the subsequent legal mechanisms to lawfully transfer personal data out of the EU/EEA. There are three broad points to consider in regards to Article 44: what constitutes a 'transfer', the 'two-step' approach to data transfers, and the level of protection provided to personal data transferred outside the EU/EEA. Each interpretative element is analysed below.

1. What counts as a 'transfer'?

It is self-evident that if the general principle for transfers is to apply, there must be a transfer of some description envisioned. This begs the question: what is a 'transfer'?

The DPD and GDPR do not define 'transfer', nor do they define similar terms invoked such as 'disclosure' or 'transmission.'[233] The GDPR (and the DPD) does define 'cross-border processing.'[234] However, the 'border' in question here is between Member States rather than that between Member States and third countries.[235] Hence, the term is of little interpretative use in this context.

The starting point for adding interpretative depth to 'transfer' is to note that a transfer — keeping definitional uncertainties aside — is a 'processing' operation under Article 4(2). As indicated earlier in chapter 2, 'processing' as defined by Article 4(2) is extremely broad, encompassing most operations one could perform on data. In this way, transfers might be viewed as merely another form of personal data processing and so regulated by the GDPR. Indeed, transfers were confirmed as a form of processing under the Directive in the *Schrems I* (C-362/14) judgment, the CJEU opining: '...the operation consisting in having personal data transferred from a Member State to a third country constitutes, in itself, processing of personal data.'[236]

In this way, transfers are regulated as a processing operation and so are within the material scope of the GDPR. However, as discussed above, transfer of personal data within the EU/EEA can dramatically impact upon the controller's legal position. In this regard, transfer of personal data to another entity can impact upon the transferor's ability to comply with data subject rights such as the right to erasure or rectification if such rights are applicable. Moreover, as analysed below, transfers outside the EU/EEA are subject to special restrictions and rules under the GDPR.

'Transfer' is a form of processing but what distinguishes the concept from other processing operations that a controller/processor might perform? The borders of 'transfer', 'disclosure', and 'transmission' are notably blurred in two respects.

First, it is unclear whether the terms refer to the same concept or distinct concepts. For instance, the rights to information (Articles 13(1)(f) and 14(1)(f)) both speak in terms of 'transfer of personal data to a recipient' or – in Articles 13(1)(e) and 14(1)(e) – 'recipients or categories of recipients of the personal data.'[237] However, the right to access, specifically Article 15(1)(c) differs, speaking in terms of 'the recipients or categories of recipients to whom the personal data have been or will be disclosed.' At face value, 'disclosure' might read as a wider concept than 'transfer' - I can disclose facts about personal data without transferring data but all transfers (at least of personal data) constitute a disclosure of some kind. However, the distinction between these concepts is notably absent from authoritative interpretation of the GDPR. Moreover, this may be a distinction without a difference in light of the recent CJEU case *Nowak* (C-434/16), this case noting that opinions or inferences may themselves count as personal data.[238] This widening of personal data may leave less room for disclosure as a distinct concept from transfer.

Second, even where 'transfer' is clearly the term in question, the borders of even this singular term are unclear. For instance, consider the various examples of the ambiguity of 'transfer' below.

The seminal CJEU case *Lindqvist* (C-101/01) held that publication of data on a website did not constitute a third country transfer under the Directive.[239] In this instance, the Court reasoned that if publication to a website constituted a 'transfer to a third country', then every time data were loaded onto a website, there would necessarily be a transfer to all third countries that have the internet.[240] In the Court's view, this would give the rules on data transfer unacceptably broad extraterritorial effect. Accordingly, the CJEU made a general distinction between merely making data 'accessible' to those in a third country and transfer of that data, only the latter being subject to special rules relating to transfers outside the EU/EEA.[241]

Lindqvist drew a distinction between merely making accessible and transferring data. This distinction also appears in the observations offered by the UK Government to the CJEU in this case.[242] Indeed, the UK position builds on this distinction, adding another, this time between 'transfer' and 'transit'.[243] According to ICO's guidance, if data merely 'transits' through a server located in a non-EU/EEA country, the operation does not count as a third country transfer. Accordingly, the concept of transfer is narrowed with the introduction of mere 'transit'.

The *Lindqvist* 'mere accessibility' and the UK 'mere transit' are likely untenable distinctions to be made in the present climate. For instance, as discussed below, the appetite of the CJEU to extend data protection appears to have increased as evidenced by its recent rulings, notably: *Schrems I, Google Spain* (C-131/12), and *Wirtshaftsakademie* (C-210/16).[244] Further, ever present in the Court's mind are the fundamental rights found in the Charter of Fundamental Rights of the European Union (the Charter), these rights often taking precedence over careful, restrained drafting.[245]

As a consequence, the idea that the CJEU might again err on the side of emphasising proportionality, treating personal data published to the world at large on a website and data transferred differently, seems out of step with the direction of the Court's jurisprudence. Moreover, more granular distinctions, such as the ICO distinction between transit and transfer also seem at odds with the direction of CJEU jurisprudence. This is especially true as routing data through servers located in third countries is not a riskless endeavour, there being documented cases of data being skimmed in transit or being subject to court orders in the country in which the server is located.[246]

In blunt terms, the idea that there is a difference between mere 'accessibility' and 'transit' on one hand and 'transfer' on the other no longer seems sustainable - these distinctions being vulnerable to CJEU challenge.[247]

Any controller of genomic data should understand whether they 'transfer' data outside the EU/EEA. As outlined below, the presence of an international transfer of data complicates the legal position of the controller, especially as genomic data may counts as 'genetic data' under Article 9, and so is subject to additional restrictions. Given the variety of storage methods and access protocols in place in genomics, the question of whether a 'transfer' is being performed is made more vexed.

Arguably, the variety of storage and access methods commonly used in genomics exacerbates the interpretative ambiguities explored earlier in this section. For instance, genomic data where it is viewed as identifiable by the controller is often stored in secure online 'containers', meaning that data may be viewed but measures are taken to ensure data is not extracted and downloaded. This access protocol raises the question of whether the viewing of this data in a third country counts as a 'transfer' or 'disclosure' under the right to access. In short, the very first question a controller should ask of their genomic data with respect to data transfers is one of the most vexed they may have to answer.

2. The two-step approach to data transfers

Article 44 makes plain that other relevant provisions of the GDPR must be complied with before transfer of personal data outside the EU/EEA, hence the 'two-step approach.'[248] Notably, the following are particularly pertinent when transferring personal data to a third country or international organisation:

- The rights to information and access contain specific notification duties where transfer or disclosure to recipients in third countries (without adequacy decisions) or international organisations is envisioned.[249]

- Records of processing activities must be kept, specifically the categories of 'recipients to whom the personal data have been or will be disclosed including recipients in third countries or international organisations.'[250]
- Processing undertaken by a data processor should be governed by a contract or other legal act.[251] Processors should only process data on the documentation instructions of the controller, notably, transfers to third countries and international organisations are specifically mentioned as an instance of such processing.[252]

The two-step approach to transfers is important in practice. Transfers beyond the EU/EEA require more than just the selection of a legal mechanism to give effect to the transfer. As demonstrated, lawful international transfer often requires substantive action pre-transfer and potentially burdensome duties post-transfer.

For instance, in regards to research, scientific research can be granted more latitude in terms of derogating from data subject rights under Article 89(2). However, Article 89(2) does not allow for derogation from the duties found in the rights to information, the obligations to keep records, or the general obligations that attach to processors. Likewise, Article 11 often allows derogation from data subject rights where processing does not require identification of the data subject (although the data remains identifiable and so not anonymised).

Again, Article 11 does not provide for derogations from the duties found in the rights to information, the obligations to keep records, or the general obligations that attach to processors. Consequently, the two-step approach to transfers likely increases the regulatory burden on genomic research or healthcare applications where the data transferred is not anonymised.

3. The level of protection

The nub of Article 44 can be found in the Article's last sentence: 'All provisions in this chapter shall be applied in order to ensure that the level of protection of natural persons guaranteed by this Regulation is not undermined.' Recital 101 adds further clarification, noting:

> '...when personal data are transferred from the Union to controllers, processors or other recipients in third countries or to international organisations, the level of protection of natural persons ensured in the Union by this Regulation should not be undermined.'

What do these provisions mean? What consequences do they have for considering the standard of protection required to give effect to lawful data transfers outside the EU/EEA? The DPD equivalent of this provision received a modest amount of judicial attention, the CJEU adding interpretative clarifications across three cases (of which one is pending judgment).*

* The CJEU has only addressed the specific issue of international data transfers directly in three cases: *Lindqvist, Schrems I,* and the ongoing *Schrems II* case (the judgment due in the coming months).

Fundamental freedoms

Schrems I (C-362/14) clarified that the protection of personal data, insofar as it infringes fundamental freedoms, must be interpreted in the light of the Charter of Fundamental Rights of the European Union.[253] This interpretation has been made stronger as the Charter (since the Treaty of Lisbon came into force in 2009) includes a right to the protection of personal data (Article 8) alongside the right to respect for private and family life (Article 7).[254] Moreover, as mentioned in chapter 2, the treaty basis of the GDPR differs from that of the DPD, the GDPR relying upon Article 16 of the Treaty on the Functioning of the European Union (TFEU). Notably, Article 16 regards the power to lay down rules 'relating to the protection of individuals with regard to the processing of personal data… and the rules relating to the free movement of such data.'[255] As the CJEU identified with respect to the DPD, the Charter remains the standard for adequate protection for international data transfers under the GDPR.

'Essentially equivalent' protection

The CJEU in *Schrems I* also clarified that an 'adequate level of protection' required for third countries is to be interpreted as being 'essentially equivalent' as that guaranteed within the EU.[256] Indeed, the CJEU in its judgment points out that this could hardly be otherwise since a controller could send data outside the EU, thereby circumventing the high standard of protection typically afforded to personal data within the Union.[257] Kuner (2017) notes the judgment gave eight points of orientation to establish what 'essential equivalence' might look like, for example:[258]

I. As noted above, there must be a high level of fundamental rights protection (as interpreted under the Charter).
II. Moreover, as the CJEU highlights, the operation of fundamental rights restricts the interpretation of the DPD (and also the GDPR), as the EU legislature is bound by Charter.[259]
III. The CJEU specifically mentions that even though a third country's means to protect fundamental rights in question might differ, the means must be essentially equivalent to that guaranteed within the EU.[260]
IV. When considering means and the overall judgment of 'adequate protection', account must be taken of the country's domestic law and international commitments.[261]
V. Recourse to a mechanism like the Privacy Shield principles is not enough - self-certification must be supported by an effective detection and supervision system if there is to be essential equivalence.[262]

Additional considerations

In addition to the previous points, Kuner also notes:[263]

- Any adequacy decision must include a 'detailed explanation' of how the third country ensures an adequate level of protection.
- The third country must not limit protections, giving considerations like national security or public interest primacy over EU law: limitations must be placed on public authorities to not interfere with fundamental freedoms.

- Third countries must not violate the principle of data protection by design and by default, meaning data must not be made accessible without the individual's intervention to an indefinite number of natural persons.

Moreover, as noted above, the transfer outside the EU/EEA itself counts as 'processing.'[264] Given this, the standard of protection applies to the transfer itself.

All considered, the general principle for transfers must be kept in mind when considering the legal mechanisms for transfer outlined below. Indeed, many of the legal mechanisms rely on the concept of 'essential equivalence' for their interpretation. For instance, adequacy decisions straightforwardly require consideration of the third countries standard of protection for personal data.[265] Moreover, obliquely, standard contractual clauses as a legal mechanism for transfer also require consideration of how the clauses will function in the third country in question and the level of protection they will provide in that context.[266]

7.2.2 International organisations

In the standard case of international data transfers, data will be transferred to a public or private entity in a third country. However, data can also be transferred to an international organisation. International organisations are defined in Article 4(26) GDPR:

> "international organisation' means an organisation and its subordinate bodies governed by public international law, or any other body which is set up by, or on the basis of, an agreement between two or more countries.'

Notably, the large scale and constitution of many genomics projects means that some controllers will count as an 'international organisation' under Article 4(26). Moreover, transfer to international organisations also poses unique issues reconciling the GDPR with the international nature of the organisation. The constitution of genomics organisations and the problem they might pose for the GDPR are outlined below.

Genomics has become a global project, data sharing and shared bioinformatics pipelines underpinning ambitious genomics-based projects. Given this global nature, some institutions have sought to constitute themselves as international organisations. For instance, the European Molecular Biology Laboratory (EMBL) owes its existence to the Agreement establishing the European Molecular Biology Laboratory, Article I(1) designating the organisation as an 'intergovernmental institution.'*

* The Agreement establishing the European Molecular Biology Laboratory is also registered as an international agreement with the Secretariat of the United Nations, see: Agreement establishing the European Molecular Biology Laboratory (with annex) (No. 13668, 1973).

Accordingly, EMBL has legal personality as an intergovernmental organisation, being a subject of, and governed by international law.[267] As a consequence, EMBL, despite being headquartered in Heidelberg Germany, enjoys certain privileges and immunities under international law, the headquarters being established by international agreement between EMBL and its host state.[268] Needless to say, status as an international organisation brings with it some benefits, EMBL being immune from German jurisdiction, save where EMBL waives such immunity.[269] However, status as an international organisation also brings some disbenefits, especially with regard to data transfers.

Transfers of personal data to international organisations may prove more difficult than to third countries. There are two main challenges outlined below: the conflict of laws problem and, where the GDPR does apply, how the GDPR's provisions might fit with international organisations.

1. Conflict of laws

Generally, public international law has primacy over secondary law such as the GDPR.[270] On this interpretation, even where territorial scope would otherwise be established, the GDPR may not apply to international organisations. However, following an alternative interpretation, insofar as data protection counts as a fundamental right under the Charter, data protection may be an issue of primary law, prevailing over some international law.[271]

Currently, the European Commission's pragmatic (yet informal interpretation) is that the GDPR does not apply to international organisations apart from the transfer rules when they receive personal data from the EU/EEA.[272] However, the interpretation in related sectors not governed by the GDPR still seems unclear. For instance, as Kuner (2020) notes, the German administrative court had made a preliminary reference regarding data transfer adequacy to the International Criminal Policy Organisation (INTERPOL).[273]

2. Interpretation for international organisations

Insofar as the GDPR applies to international organisations with respect to data transfers, its interpretation and application is made more difficult. For instance, EMBL have Internal Policy No. 68 on General Data Protection as a form of self-regulation. One of the aims of Internal Policy 68 being to smooth the process resulting in an Article 45 declaration of adequacy, thereby allowing data transfers from the EU/EEA.[274] Internal Policy 68 closely mirrors elements of the GDPR. Indeed, the Policy perhaps represents one of the best attempts at reconciling the rigours of the GDPR with public international law.

However, it is still unclear how exactly the adequacy of an international organisation such as EMBL might be appraised. For example, as demonstrated above, the CJEU have emphasised the importance of background protections offered by third countries such as available judicial remedies when establishing 'essential equivalence.'[275] International organisations may struggle to evidence that there are 'essentially equivalent' means to secure the fundamental right to data protection.

To illustrate, Internal Policy 68, while broadly similar to the GDPR, lacks a judicial means of enforcement or a separate supervisory authority. Notably, disputes arising from data subjects are managed by a Data Protection Committee, an arbitration process being available if the dispute escalates.[276] However, full judicial remedies appear to be lacking, the arbitration process being governed by a restricted set of law.[277] It is unclear whether a tribunal process such as this is enough to underpin a finding of 'essential equivalence.' Moreover, the CJEU have also highlighted that considerations such as national security or law enforcement be premised above fundamental freedoms.[278]

In the context of international organisations, the fundamental right to data protection may conflict with the aims and constitution of the organisation in question. For example, the EMBL Internal Policy 68 deals with this difficult question by noting the need to also consider the fundamental right to scientific freedom under the Charter and Universal Declaration of Human Rights.[279] While the EMBL Policy may represent a sound strategy to 'square the circle' of respecting the fundamental right to data protection with the aims of scientific research, not all international organisations will be able to reconcile the rigours of the GDPR whilst also being consistent with their establishing instruments.

To summarise, data transfers to international organisations may be more difficult than third country transfers in two respects: they raise conflict of law issues and the interpretation of 'essential equivalence' may be more difficult to establish. The following analysis primarily considers the legal mechanisms available to entities in third countries rather than international organisations.

7.2.3 Adequacy decision (Article 45)

Transfer of personal data from the EU/EEA to an entity in a third country or an international organisation may be lawful where that third country or international organisation has a declaration of adequacy. This section considers the elements of adequacy decisions as they apply to transfers of genetic data or health-related data. There are six key points to note:

First, adequacy decisions are one of the favoured legal mechanisms for transfer, being perceived as often jurisdiction-wide solutions to lawful transfer of data. Indeed, adequacy decisions can function to ease the regulatory burden associated with data transfer, especially for small and medium-sized enterprises. However, as points two and three outline, the degree to which any given adequacy decision alleviates any burden of data transfer depends on what is being judged as adequate and does not result in a complete discharge of data transfer responsibilities.

Second, adequacy decisions are not just for third countries. On the contrary, declarations of adequacy can be issued for international organisations but also a territory, a specific sector, or a particular legal framework within a third country.[280] As discussed later in chapter 8 with respect to codes of conduct, there are advantages to wider adequacy declarations, these decisions cover more entities and so are useful to a broader range of organisations.

However, full adequacy decisions for jurisdictions are the result of a long process and intense negotiation. On the other hand, sector-specific findings for a third country or particular legal framework may be narrower, serving a smaller range of entities, and require more arrangements to provide a sufficient legal mechanism for transfer. However, narrower declarations of adequacy may be (at least marginally) easier to obtain than findings that apply to entire jurisdictions. Indeed, installing a voluntary regime, such as the EU-US Privacy Shield, to be declared adequate is likely much easier than trying to ensure the entire jurisdiction itself is adequate.

Third, adequacy decisions do not necessarily discharge many of the GDPR's data transfer requirements. For instance, while organisations which are self-certified and included in the 'Privacy Shield List' are 'deemed to provide adequate privacy protection,' WP29 has in the past (albeit under the predecessor Safe Harbour Framework) noted that self-certification was not enough in some contexts.[281] For instance, with respect to cloud processing, WP29 highlighted that mere self-certification under the Safe Harbour Framework was insufficient 'in the absence of robust enforcement of data principles in the cloud environment.'[282] In relation to genetic data, given that this data likely qualifies as 'sensitive information' under the EU-US Privacy Shield Adequacy Decision, extra considerations apply.[283] For instance, Annex II outlines the principles issued by the U.S. Department of Commerce, stipulating in Section 2(c) that where 'sensitive information' is disclosed to a third party, express opt-in consent must be obtained from the individual. In short, mere self-certification is not enough to discharge data transfer responsibilities.

Fourth, Article 45(2) specifies elements the Commission must take account of when assessing the level of protection offered.

> Article 45(2) When assessing the adequacy of the level of protection, the Commission shall, in particular, take account of the following elements:
>
> (a) the rule of law, respect for human rights and fundamental freedoms, relevant legislation, both general and sectoral, including concerning public security, defence, national security and criminal law and the access of public authorities to personal data, as well as the implementation of such legislation, data protection rules, professional rules and security measures, including rules for the onward transfer of personal data to another third country or international organisation which are complied with in that country or international organisation, case-law, as well as effective and enforceable data subject rights and effective administrative and judicial redress for the data subjects whose personal data are being transferred;
>
> (b) the existence and effective functioning of one or more independent supervisory authorities in the third country or to which an international organisation is subject, with responsibility for ensuring and enforcing compliance with the data protection rules, including adequate enforcement powers, for assisting and advising the data subjects in exercising their rights and for cooperation with the supervisory authorities of the Member States; and
>
> (c) the international commitments the third country or international organisation concerned has entered into, or other obligations arising from legally binding conventions or instruments as well as from its participation in multilateral or regional systems, in particular in relation to the protection of personal data.

Notably, the above list is not exhaustive and heavily influenced by the *Schrems I* (C-362/14) judgment.[284] The WP29 Working Document on Adequacy Referential (adopted by the EDPB), elaborates on these elements, adding the following basic content and procedural/enforcement data protection principles and mechanisms:[285]

I. Basic data protection concepts and/or principles that reflect and are consistent with European data protection law.
II. Legitimate bases under which personal data may be lawfully, fairly and legitimately processed should be set out in a sufficiently clear manner.
III. Data should be processed for a specific purpose and subsequently used only insofar as this is not incompatible with the purpose of processing.
IV. Data should be accurate and, where necessary, kept up to date.
V. Data should, as a general rule, be kept for no longer than is necessary for the purposes for which the personal data is processed.
VI. Any entity processing personal data should ensure that the data are processed in a manner that ensures security of the personal data.
VII. Each individual should be informed of all the main elements of the processing of his/her personal data in a clear, easily accessible, concise, transparent and intelligible form.
VIII. The data subject should generally be able to exercise the following rights, the exercise of these rights should not be excessively cumbersome: right of access, rectification, erasure, and objection.
IX. Further transfers of the personal data by the initial recipient of the original data transfer should be permitted only where the further recipient is also subject to rules affording an adequate level of protection.

All of these elements and extra content are to be interpreted in the light of the Article 44 general principle for data transfers discussed earlier.

Fifth, a declaration of adequacy also has procedural elements, these elements being outlined in the GDPR and elaborated on in the WP29 Adequacy Referential. In brief, the assessment begins with EDPB providing an opinion for the assessment of adequacy in the third country or international organisation in question.[286] Having considered this opinion, the Commission then passes an implementing act specifying the territorial and sectoral application of the adequacy decision.[287]

It is important to note that the Commission is required to periodically revisit the adequacy decision and monitor developments in the third country or international organisation in question. Significantly, the facts and findings of *Schrems I* resulted in the EU-US Safe Harbour Framework being declared inadequate.[288] In short, given the procedural requirements and complex assessment required, adequacy decisions can take years and may fail when the context changes.

Sixth, as discussed, the UK after the end of the transition period (currently scheduled to end on the 1st of January 2021), will become a third country. At the time of writing, the UK has committed to transpose the GDPR with minor amendments into UK law.[289] Given this, any adequacy decision is made easier since the UK will have a very similar version of the GDPR to ground the assessment of essential equivalence.

However, such a decision is far from a foregone conclusion. To highlight two issues that may hamper such a decision: the UK has declined to carry across the Charter, meaning that the finding of essential equivalence has to be held up by the UK GDPR, European Convention on Human Rights, and Convention 108+.[290] Moreover, the UK's respect for such fundamental rights has, in recent memory, been brought into question with the *Tele2/Watson* case, the case questioning the collection of communication data according to the Data Retention and Investigatory Powers Act 2014.[291] Accordingly, the path to a declaration of adequacy may not be straightforward.

What do these contextual facts about adequacy mean for the genomics community? A finding of adequacy for the UK may not be forthcoming. In this regard, the genomics community will either have to work for sector specific adequacy or find other means to lawfully conduct international data transfers. It is to these alternative legal mechanisms for transfer that we now turn.

7.2.4 Appropriate safeguards (Article 46)

Where there is no adequacy decision, safeguards may make lawful the transfer of data to a third country or international organisation. There are two broad methods by which safeguards might be relied upon to facilitate a transfer.

Article 46(1) and (3)

Following Article 46(1), (2) and (3), the controller or processor may rely upon 'appropriate safeguards' to conduct a lawful transfer. Article 46, nor the interpretative aid recital 108, provide an exhaustive list of what safeguards might count as 'appropriate.'

Nevertheless, Article 46(1) does stipulate that these safeguards must ensure that enforceable data subject rights and effective legal remedies are available. Moreover, Article 46(3) also provides that specific safeguards may be authorised by a competent supervisory authority, namely: contractual clauses (non-standard clauses that remain unadopted) and provisions to be inserted into administrative arrangements between public authorities. Recital 108 elaborates on how safeguards might be 'appropriate', outlining procedural elements: that the enforceability of data subject rights and access to legal remedies includes the ability to 'obtain effective administrative or judicial redress and to claim compensation.'

Regarding the substantive protections provided, Recital 108 also highlights that safeguards should relate to compliance with the general principles and data protection by design and by default. In short, 'appropriate safeguards' should be interpreted as requiring that there are effective remedies available to ensure that personal data are, for example, not 'made accessible without the individual's intervention to an infinite number of natural persons' and are processed only for specified, explicit and legitimate purposes.[292]

Article 46(2)

There are a set of appropriate safeguards that may provide a lawful mechanism for transfer without specific authorisation from a supervisory authority. These safeguards are listed in Article 46(2) and include:

> '(a) a legally binding and enforceable instrument between public authorities or bodies;
>
> (b) binding corporate rules in accordance with Article 47;
>
> (c) standard data protection clauses adopted by the Commission in accordance with the examination procedure referred to in Article 93(2);
>
> (d) standard data protection clauses adopted by a supervisory authority and approved by the Commission pursuant to the examination procedure referred to in Article 93(2);
>
> (e) an approved code of conduct pursuant to Article 40 together with binding and enforceable commitments of the controller or processor in the third country to apply the appropriate safeguards, including as regards data subjects' rights; or
>
> (f) an approved certification mechanism pursuant to Article 42 together with binding and enforceable commitments of the controller or processor in the third country to apply the appropriate safeguards, including as regards data subjects' rights.'

The safeguards that follow must be interpreted in light of the Article 44 general principle for transfers. In this respect the safeguard must operate to provide 'essential equivalence' in the protection provided within the EU/EEA.

Each safeguard that does not require authorisation from supervisory authorities as it relates to transfers of genomic data outside the EU/EEA is analysed below.

1. Legally binding and enforceable instrument between public authorities

> Article 46(2)(a) a legally binding and enforceable instrument between public authorities or bodies;

Many controllers of genomic data are large institutions that often serve a public function. Given this, Article 46(2)(a) as a safeguard may be a live option for transfer of genomic data between public authorities within the EU/EEA and those outside. This section considers the interpretation of this safeguard and its feasibility as a legal mechanism to share genomic data internationally. Broadly, three restrictions on the use of Article 46(2)(a) and two elements of flexibility (which may also address some challenges with standard contract clauses, discussed later in this chapter) are outlined below.

'Public authorities'

Both the entity transferring data and the entity receiving data must count as 'public authorities' or 'public bodies.' As the EDPB notes, the GDPR does not define 'public authority' or 'public body.' Despite this, the EDPB, interpreting Recital 108 does tell us that the terms are broad enough to capture both public bodies within third countries as well as international organisations.[293] As a consequence, Article 46(2)(a) may prove to be a valuable alternative to a full adequacy decision for international organisations.

The definition of 'public authority' or 'public body' is largely left to domestic law.[294] The EDPB does clarify that public bodies can include government authorities at multiple levels: national, regional, local as well as bodies governed by public law, for instance, executive agencies.[295] The UK's DPA 2018 defines the meaning of both terms in Section 7. Specifically for England and Wales, an entity is a 'public authority' or 'public body' for the purposes of Section 7(1) if that entity is a public authority as defined by the Freedom of Information Act 2000 (FOIA 2000) or is specified by the Secretary of State via regulations to be a public authority. In addition to this, the Secretary of State may also specify via regulation that an entity that would be a public authority for the purposes of FOIA 2000 is not a public authority for the purposes of the GDPR.[296]

In practice, an entity in England or Wales will be a 'public authority' where it is listed as such in Schedule 1 FOIA 2000 or listed as such by a regulation and is not specifically excluded from being a public authority by the Secretary of State. Schedule 1 FOIA 2000 contains the defined list of public authorities, Section 1 capturing any 'government department' (apart from the two explicitly excluded) and Part III capturing much of the NHS at most levels. It is also clear that UK Universities fall in the category of public authorities. Between these two provisions most of the genomics medicine service will likely count as a 'public authority.' The position of Genomics England may be slightly different, this entity being the trading name for a 'wholly owned' limited company, owned by the Department of Health and Social Care but not specifically listed in Schedule 1 FOIA 2000.[297]

'In the exercise of official authority'

Section 7(2) DPA 2018 clarifies that a public authority will only be public authority for the purposes of the GDPR 'when it performs a task carried out in the public interest or in the exercise of official authority.' That is, an entity may indeed be a public authority in some sense, but insofar as that authority is not performing a public function or exercising official authority, the entity will not count as a public authority. Consequently, an entity can be a public authority for Section 7 with respect to a set of purposes for processing data but not for other purposes.

Given this, Article 46(2)(a) cannot act as a legal mechanism for transfer if the transfer is not a task in the public interest or in the exercise of official authority. In practice, this means that some activities of public authorities will fall outside the scope of Article 46(2)(a). Accordingly, if a controller seeks to rely on Article 46(2)(a) as a legal mechanism for transfer in England or Wales, the controller should also check that the transfer will be a task in the public interest or an activity within their official authority. It is not enough that the controller identifies themselves as a public authority - they must be a public authority for that purpose for processing, in this case, a transfer.

A 'legally binding and enforceable instrument'

There must be a legally binding and enforceable instrument between the public authorities. The EDPB suggest that international treaties, public law treaties, or self-executing administrative agreements may all count as such instruments.[298] More onerous is what the instrument must include. Broadly, the EDPB stresses two points, namely points of substance and points of procedure. With respect to substance, the EDPB outlines that the instrument should encompass the 'core set of data protection principles and data subject rights' found in the GDPR.[299]

With respect to procedure, the instrument should be legally enforceable and commit the public authorities to apply the core principles and rights.[300] Moreover, where there is no possibility of ensuring effective judicial remedies, alternative 'redress mechanisms' may instead be adopted.[301] However, the EDPB specifically cautions that consultation with the competent supervisory authority is required where alternative redress mechanisms are sought.[302]

There are two main points of flexibility that Article 46(2)(a) provides.

An alternative to adequacy

Article 46(2)(a) provides a useful alternative to the use of adequacy decisions. As demonstrated, Article 46(2)(a) will in part depend upon there being judicial (or other forms) of enforcement. However, the focus of Article 46(2)(a) is more squarely on the public authorities and the instrument that facilitates the transfer rather than the ambient protection the jurisdiction provides. In this way, the EDPB seems to be primarily concerned with upholding the spirit of substantive protections the GDPR provides and ensuring there is, at the very least, some independent supervisory oversight to ensure these rights and principles are upheld.[303] Tentatively, this seems to be less stringent than the rubric of case law that applies to adequacy decisions. For example, perhaps international organisations would be better served by adopting the Article 46(2)(a) safeguard, as this sidesteps some of the problems with adequacy decisions and international organisations discussed earlier.

Article 46(3)(b)

Second, although our focus has been on the Article 46(2)(a) safeguard, Article 46(3)(b) provides another related and useful legal mechanism. It allows for 'provisions to be inserted into administrative arrangements between public authorities,' so long as these provisions 'include enforceable and effective data subject rights' and are authorised by the competent supervisory authority.[304] In short, where the instrument or arrangements may be in doubt under Article 46(2)(a), Article 46(3)(b) provides a route for supervisory authority oversight and checking.

To summarise, Articles 46(2)(a) and (3)(b) appear to offer a feasible route for data transfer where the entities in both jurisdictions are public authorities and a binding instrument between the two can be agreed. Nevertheless, these routes remain unavailable to private bodies performing private functions. Given this, Article 46(2)(a) and (3)(b) at best, offer a solution to a subset of controllers seeking to transfer personal data outside the EU/EEA.

2. Binding corporate rules

Binding corporate rules potentially provide a solution to data transfers outside the EU/EEA where data is to be transferred within a multinational entity. 'Binding corporate rules' are given the following definition in Article 4(20):

> "binding corporate rules' means personal data protection policies which are adhered to by a controller or processor established on the territory of a Member State for transfers or a set of transfers of personal data to a controller or processor in one or more third countries within a group of undertakings, or group of enterprises engaged in a joint economic activity;'

There are four key restrictions upon the use of binding corporate rules (BCRs) as a safeguard under Article 46(2)(b).

'Group of undertakings'

BCRs only facilitate transfers between members of a 'group of undertakings or group of enterprises engaged in a joint economic activity.'[305] 'Enterprise' according to Article 4(18) is defined as 'a natural or legal person engaged in an economic activity, irrespective of its legal form, including partnerships or associations regularly engaged in an economic activity.' 'Group of undertakings' is defined in Article 4(19) as 'a controlling undertaking and its controlled undertakings.'

Both of these terms are ultimately limited by the requirement that these groups must be 'engaged in a joint economic activity.' 'Economic activity' has meaning within EU competition law, the Grand Chamber in *FENIN v Commission* (C-205/03 P) noting that: 'the activity consisting in offering goods and services on a given market that is the characteristic feature of an economic activity.'[306] Accordingly, 'economic activity' stems from the activity that the entity is engaged in, not the legal status of the organisation or the way in which the entity is financed.[307] If this line of case law applies, it appears that multinational institutions that engage in genomics that do not commercialise their outputs will likely be beyond the definition of 'joint economic activity' and so unable to avail themselves of BCRs.

'Legally binding'

There are a number of procedural requirements that BCRs must fulfil to be approved by competent supervisory authorities. The most notable requirement being that BCRs must be 'legally binding and apply to and are enforced by every member concerned of the group of undertakings.'[308] In addition to this, Article 47(2) lists a number of elements that BCRs must specify, many of these elements being procedural. For example, BCRs must specify: 'their legally binding nature; both internally and externally,' the complaint procedures in place, the mechanisms to ensure verification amongst the group, the mechanism used to cooperate with the competent supervisory authority, and mechanisms to report third country rules that might have a 'substantial adverse effect on the guarantees provided' by the BCRs.[309] In short, to be a legal mechanism for data transfer, BCRs must have multiple procedures in place to ensure data protection principles are upheld and data subjects rights are vindicated.

The general principle

BCRs, like the other safeguards discussed, are to be interpreted in light of the Article 44 general principle for transfers. That is, the effect of BCRs should be to offer data subjects whose data have been transferred, essentially equivalent protection as that guaranteed under EU law. As a consequence, BCRs must implement data protection principles and confer enforceable rights on data subjects. In this regard, there are three main provisions that outline the substantive protections that BCRs must offer.

Article 47(1)(b) requires that BCRs must 'expressly confer enforceable rights on data subjects with regard to the processing of their personal data.' Article 47(2)(d) stipulates that BCRs must specify the application of the 'general data protection principles' outlined in Article 5. Article 47(2)(e) lays down the requirement for BCRs to stipulate the availability of, and ensure the means to exercise, data subject rights, including the presence of redress and liability mechanisms.

Supervisory authority approval

For BCRs to act as a legal mechanism for international transfer, the competent supervisory authority (there are rules establishing which supervisory authority is competent) must approve the BCRs according to the Article 63 consistency mechanism.[310] In practice, this means that the approval of BCRs is likely a lengthy process. For instance, the UK's ICO notes that even a 'straightforward application' could take 12 months to conclude.[311] In this regard, the institution should be sufficiently large and data transfers sufficiently valuable to justify the pursuit of BCRs.

The restrictions upon the use of BCRs to conduct lawful international transfers aside, there are a number of benefits that reliance on BCRs can bring.

Protections beyond adequacy

BCRs can facilitate data transfers partially independent from the level of ambient protection in a third country. For instance, many of the mechanisms required of BCRs will remain valid, even if protections within the jurisdiction change. In this way, BCRs do not fully rely on a jurisdiction having 'essential equivalence.' Nevertheless, as demonstrated below with respect to standard contractual clauses, if a jurisdiction lacks receptive courts or has, for example, surveillance programmes that impact upon fundamental freedoms, the viability of BCRs remaining valid will be impacted.[312]

Gold plating

BCRs, because of their required rigour and authorisation by a supervisory authority, have benefits beyond being a legal mechanism for transfer. For instance, BCRs can 'gold plate' an organisation's data protection policies and procedures - the thinking that BCR approval is indicative of a well-designed data protection system, a system that likely has proper safeguards in place.[313] In this way, BCRs, despite the time and effort required to receive approval, also demonstrate general compliance with the GDPR.

BCR flexibility

BCRs can be somewhat flexible. For instance, as evidenced by ICO's list of authorisations for transfers of personal data according to binding corporate rules, BCRs can be specific to only certain types of data. For instance, there are four listed corporations that have BCRs for specific purposes, namely 'employee data' or 'guest data.'[314]

Do BCRs offer a viable solution for transfer of genomic data to third countries or international organisations? In short, BCRs, by their very nature, are of limited value for the genomics community. The most common problem of data sharing in genomics is not data sharing within organisations but between organisations. That is, genomics in healthcare or research is collaborative - there being interdependent organisations rather than corporations that span the globe.[315]

Nevertheless, BCRs do represent valuable tools for those organisations in a position to use them. Indeed, some participants in our research reported exploring the potential of establishing subsidiaries of non-EEA based organisations within the EU/EEA in order to take advantage of the flexibility that BCRs provide.

3. Standard data protection clauses

What are standard data protection clauses? Standard data protection clauses (SSCs) are model contractual clauses that have been adopted or approved by the Commission. There are three broad routes to use contractual clauses as legal mechanisms for international transfers.

First, standard data protection clauses that have been adopted by the Commission may be inserted without variation and used as a safeguard without specific authorisation from a supervisory authority.[316] Currently there are three SSCs adopted by the Commission under the Directive that can be used under the GDPR.[317] Reliance on SSCs as a safeguard to facilitate international data transfer requires that the clauses be adopted in their entirety, without amendment, and not contradicted (directly or indirectly) by any additional provisions.[318]

Second, standard data protection clauses must have been adopted by a supervisory authority and approved by the Commission.[319] At the time of writing, the Danish Supervisory Authority have published draft standard contractual clauses between controllers and processors.[320] The draft received a generally favourable response from the EDPB in their Opinion 14/2019 on the draft clauses.[321] However, the clauses are not designed to facilitate international data transfer but clauses to assist in compliance with Article 28(3) on contracts between controllers and processors. In short, the Danish SSCs are not mechanisms for transfer according to Article 46(2)(c).

Third, ad hoc (non-standard) contractual clauses drafted for specific cases can be approved by supervisory authorities according to Article 46(3)(a). In this respect, approval may take some time and be bespoke to the arrangement outlined.

Regardless of the route taken, there are two broad challenges that all SSCs face: the interpretation of 'essential equivalence' with respect to *Schrems II* (C-311/18) and the limitations of their use in certain third countries.

Schrems II and SSCs

Schrems II considers the validity of SSCs in the context of the EU-US Privacy Shield and surveillance programmes in the US. The judgment (pending at the time of writing) will likely clarify a number of elements about the operation of SSCs in third countries. The Opinion of Advocate General Saugmandsgaard Øe (AG) provides some idea of what the judgment might contain.[322] The Opinion contains a number of notable points of clarification. The AG affirmed that SSCs are general mechanisms that can apply to transfers 'irrespective of the third country of destination and the level of protection guaranteed there.'[323] However, while this is technically true, the level of protection the third country provides impacts upon the protection that any SSC within its jurisdiction provides.[324]

The AG Opinion also considered the standard and method by which SCCs are assessed. The AG noted that SSCs are held to the same standard of 'essential equivalence' as adequacy decisions. Given the breadth of considerations examined under Article 45, for instance: rule of law, human rights protections, access of public authorities to personal data, the existence of an effective supervisory authority, international commitment made, and so on, this may be a difficult task. Further, it is also unclear how these considerations fit with the narrower assessment of how SSCs operate within that jurisdiction.[325] If the CJEU follow the AG Opinion, SCCs may be a more demanding safeguard to give effect to international data transfers.*

Limitations on the use of SSCs by international public bodies

One of the challenges our research has identified for international transfers of genomic (and associated health) data is due to the requirement that the current Commission-approved SCCs must be used in their entirety, with no tailoring or removal of certain provisions. For some third country public institutions, this has led to an intractable conflict with their domestic legal arrangements because they are legally barred from agreeing to some of the clauses on liability, jurisdiction and governing law provisions contained in the SCCs.

It appears that the EDPB are also conscious of this challenge because as we noted above, in their recent guidance on Art 46(2)(a) (legally binding and enforceable instruments between public authorities or bodies) they recognise that domestic laws, or the 'specific status of the receiving public body' may restrict the possibility of judicial redress mechanisms between EU/EEA and International public authorities.[326]

As noted above, the EDPB accordingly approved the more flexible alternative 'redress mechanisms' where judicial mechanisms of the type found in the SCCs are impossible.[327] The EDPB suggests that these could include quasi-judicial, binding mechanisms such as arbitration or alternative dispute resolution mechanisms such as mediation, which would guarantee an independent review, and a commitment to be liable for compensation of damages following such a review.

* Notably, the CJEU may issue a judgment that differs from the AG Opinion. Moreover, there is also another case T-738/16 *La Quadrature du Net and Others v Commission* also pending judgment that also examines the validity of the EU-US Privacy Shield more directly.

In light of this recent guidance, it may be that Art 46(2)(a) now provides a sufficiently flexible alternative to SCCs to address the challenges identified in our research. However, it is not yet clear if these developments fully address the concerns of all international public bodies.

4. Approved codes of conduct

Codes of conduct come in two broad forms. First, codes according to Article 40 that demonstrate compliance with the GDPR's provisions but do not act as a legal mechanism for data transfers. Second, codes of conduct as a safeguard according to Article 46(2)(e), that act as a legal mechanism to facilitate data transfers to third countries or international organisations. We discuss the nature of codes and their potential to set standards and facilitate compliance with the Regulation more generally in chapter 8. In this section we focus on the specific issues relating to codes of conduct as a safeguard for international data transfers.

In regards to codes of conduct to facilitate data transfers, no code of conduct to facilitate transfer has been approved nor is there specific guidance from the EDPB or WP29 on what such a code would look like. Under the Directive, there was the C-SIG Code of Conduct on Cloud Computing. Under the GDPR, the EDPB have released Guidelines 1/2019 on Codes of Conduct and Monitoring Bodies. However, the C-SIG Code and Guidelines 1/2019 both relate to codes to demonstrate general compliance with the GDPR not codes to facilitate international data transfer.[328]

It is likely that codes of conduct as a mechanism for international data transfer (according to Article 40(3) and Article 46(2)(e)) will be different from their general counterparts in a number of respects. Notably, Article 40(3) stipulates that where codes are designed to act as an appropriate safeguard under Article 46(2)(e), the 'controllers or processors shall make binding and enforceable commitments, via contractual or other legally binding instruments, to apply those appropriate safeguards including with regard to the rights of data subjects.' Accordingly, the problem of enforceability and the issues that accompany it remain.

For instance, as with standard contractual clauses, the ambient protections or intrusions that exist in a jurisdiction may hamper a code of conduct facilitating lawful transfer if they impact upon the enforceability of those clauses or impact upon the data subject rights and data protection principles contained within.[329] As we discuss in chapter 8, designers of codes of conduct should be mindful of what kind of code they are developing - codes sufficient to facilitate data transfers may be a more difficult task or require supporting mechanisms to ensure the standard of 'essential equivalence' is met.

5. Certification mechanisms

Certification mechanisms offer an opportunity to demonstrate compliance with the GDPR generally but can also act as a legal mechanism for data transfer outside the EU/EEA. As with codes of conduct, we consider certification mechanisms more fully in chapter 8 but we briefly note some challenges and opportunities specifically relating to certification as a legal mechanism for transfer here.

Certification schemes overlap significantly with codes of conduct under the GDPR. However, as we discuss in the following chapter, certification schemes that have a limited scope, applying to a limited set of processing operations within a stable technical context are perhaps most likely to comply with the GDPR. Like codes of conduct, certification mechanisms to facilitate data transfer do not sidestep many of the challenges posed by other legal mechanisms to transfer data to third countries or organisations. Moreover, certifications for this purpose are also likely more difficult to be approved, having the additional challenge of meeting the standard of 'essential equivalence.' For instance, certification has similar mechanisms to codes of conduct under Article 42(2) and Article 46(2)(f), Article 42(2) requiring that certification mechanisms for international transfer be binding and enforceable.

7.2.5 Derogations (Article 49)

Article 49 derogations offer a set of legal mechanisms to lawfully transfer personal data to third countries or international organisations. However, the circumstances under which Article 49 derogations can be invoked are extremely limited. There are 7 derogations that may be invoked:[330]

a. Where the data subject has explicitly consented to the transfer
b. Where the transfer is necessary for the performance of a contract between the data subject and controller
c. Where the transfer is necessary for the conclusion or performance of a contract in the interest of the data subject
d. Where the transfer is necessary for 'important reasons of public interest'
e. Where the transfer is 'necessary for the establishment, exercise or defence of a legal claim'
f. Where the transfer is necessary in order to protect the vital interests of the data subject
g. Where the transfer is made from a public register

In addition to these explicitly stated derogations, there is also a final 8th derogation listed in the second paragraph of Article 49(1) invoking 'compelling legitimate interests'. This 8th Article 49 derogation and derogations (a)-(d) and (f)-(g) are covered in turn below. However, before this, we first consider the conditions under which any derogation may be invoked.

There are four general restrictions to invoking an Article 49 derogation to transfer data outside the EU/EEA.

First, an Article 49 derogation can only be relied upon in the absence of an adequacy decision (Article 45(3)) or appropriate safeguards (Article 46).[331] Indeed, EDPB Guidelines 2/2018 on derogations of Article 49 under Regulation 2016/679 go further noting:

> 'data exporters should first endeavor possibilities to frame the transfer with one of the mechanisms included in Articles 45 and 46 GDPR, and only in their absence use the derogations provided in Article 49(1).'

Second, as noted by the EDPB Guidelines on Article 49, Article 49 derogations are 'exemptions from the general principle'. In this way, the derogations are to be interpreted strictly 'so that the exception does not become the rule.'[332]

Third, as the EDPB Guidelines highlight, recital 111 interprets some derogations as being limited to 'occasional' and 'not repetitive' transfers.[333] As the Guidelines go on to clarify, such transfers may happen more than once but 'not regularly', and not within the 'regular course of action.'[334] For example, the Guidelines note that a data transfer that occurs within a 'stable relationship' between the exporter or importer would be deemed as 'systematic and repeated' and so not count as 'occasional' or 'not repetitive.'[335]

The derogations subject to this restriction are: transfers necessary for the performance of a contract (Article 49(1)(b)), transfers necessary for the conclusion or performance of a contract in the interest of a data subject (Article 49(1)(c)), transfers necessary for the defence of a legal claim (Article 49(1)(e), and compelling legitimate interests (Article 49(1)(subpar 1)).[336] However, the EDPB are also quick to add that even where there is no express restriction with respect to the frequency of transfers, any derogation should not be interpreted in a way that would 'contradict the very nature of the derogations as being exceptions.'[337]

Fourth, some derogations are subject to a necessity test.[338] The derogations subject to this test are the same as those listed in the previous paragraph, with the addition of: transfers necessary for important reasons of public interest (Article 49(1)(d)) and transfers necessary to protect the vital interests of the data subject (Article 49(1)(f)).[339] * As the EDPB clarifies, this test requires 'an evaluation by the data exporter in the EU of whether a transfer of personal data can be considered necessary for the specific purpose of the derogation to be used.'[340] In this respect, the necessity test is best assessed with respect to each relevant derogation.

Fifth, the above derogations must also be interpreted in the light of the general principle of data transfers discussed earlier.[341] This means that the use of derogations must not undermine the level protection offered by the GDPR nor breach fundamental rights.[342]

The sections below consider each relevant derogation in turn, considering the feasibility of use for transfers of genomic data.

1. Explicit consent

> **Article 49(1)(a)**
>
> the data subject has explicitly consented to the proposed transfer, after having been informed of the possible risks of such transfers for the data subject due to the absence of an adequacy decision and appropriate safeguards;

* The EDPB Guidelines appear to miss that the necessity test also applies to compelling legitimate interests (Article 49(1)(subpar 1)).

Consent can act as a derogation to transfer of personal data to a third country or international organisation.[343] However, consent under the GDPR generally, and consent as a mechanism for data transfer are subject to a number of restrictions and requirements. In addition to these restrictions/requirements and as we discussed in chapter 5, consent in the genomics context can also be challenging. These barriers (as well as opportunities) to obtaining consent sufficient to conduct an international data transfer are outlined below.

Consent is defined in Article 4(11) of the GDPR:

> ' "consent" of the data subject means any freely given, specific, informed and unambiguous indication of the data subject's wishes by which he or she, by a statement or by a clear affirmative action, signifies agreement to the processing of personal data relating to him or her;'

Further, Article 7 provides four conditions for valid consent, namely:

1. The controller should be able to demonstrate consent for processing
2. Consent should be 'clearly distinguishable' from other matters and be in an 'intelligible and easily accessible form, using clear and plain language.'
3. The data subject should have the right to withdraw their consent at any time.
4. When assessing whether consent is freely given, special attention shall be paid where a contract or provision of a service is conditional on that consent.

The EDPB confirmed in their Guidelines 2/2018 on derogations of Article 49 under Regulation 2016/679 that this definition of consent and the conditions that attach to it also apply to consent as a derogation to transfer data.[344] The EDPB Guidelines also note a number of 'specific, additional elements required for consent to be considered a valid legal ground for international data transfers.'[345] There are two such extra elements.

'Explicit'

The EDPB notes that consent as a derogation is stricter than the Article 4(11) general definition of consent, Article 49(1)(a) requiring 'explicit' consent.[346] The EDPB highlights that this usage of 'explicit consent' is reserved for situations where there are particular data protection risks and where a high level of data subject control is required, international data transfer being such a risky situation.[347] As a consequence, consent as a mechanism for international transfer must be expressed in a particular way, 'explicit' requiring that the data subject 'give an express statement of consent.'[348]

Notably, this requirement goes beyond the Article 4(11) definition that consent be an 'unambiguous indication of the data subject's wishes.' There is no requirement that 'explicit' translate into written consent only but the WP29 do caution that explicit yet unwritten consent may be difficult to evidence: a two stage verification of consent may also be desirable to ensure the explicit consent is valid.[349]

'Informed of the possible risks of such transfers'

There is also the additional requirement that consent must be 'informed of the possible risks of such transfers for the data subject.'[350] Ordinarily, valid consent under the GDPR generally must be 'informed' and 'specific.'[351] In the context of the Article 49(1)(a), WP29 interprets 'specific' and 'informed' to require consent to be 'specifically given for the particular data transfer or set of transfers.'[352]

As WP29 points out, this poses problems for data exporters as often the circumstances for transfers are unknown or not specific enough to uphold this requirement prior to processing when consent would typically be collected.[353] Consequently, the specific risks associated with an international data transfer in the absence of adequacy or Article 46 safeguards cannot be properly disclosed to form a valid consent.[354] In response, WP29 notes that consent for specific transfers to be valid should be collected at the time when the transfer is 'envisaged.'[355] It is possible to collect this consent at the point of data collection but only when the specific transfer is made known to the data subject and if these circumstances do not change.[356]

This suggests that consent to facilitate data transfer is an extremely inflexible derogation to rely upon. This aside, it is unclear how the provisions allowing those processing data for 'scientific research' to construe their purposes more broadly, interact with the rigours of Article 49(1)(a).[357] For instance, these provisions are often interpreted as allowing a 'broader' form of consent in scientific research, allowing valid consent to state with less specificity the purposes for processing.[358] However, there is no specific exception or derogation for scientific research in relation to international data transfers listed in Article 89(2).

Further, where read liberally, the circumstances under which broad consent can be relied upon are narrow, where read conservatively (as by some supervisory authorities), the circumstances are emaciated indeed.[359]

In regards to the processing of genomic data, the difficulties of relying on consent generally in the context of research are well documented. Indeed, both the Human Research Authority and the Information Governance Alliance have cautioned against using consent as a legal basis or derogation.[360] Indeed, to summarise these concerns, consent as a legal basis or special category derogation is set a high bar under the GDPR and its interpretation by WP29, the validity of consent being especially vulnerable to challenge where a service is provided in return.[361] Moreover, even where the standard is met, the legal basis/derogation is thought to be restrictive and inflexible.[362]

Further, where personal data are not directly collected from a data subject, which is the case with many datasets in genomics, the ability to reconsent for a new purpose is often impaired - data subjects being difficult to track or elicit responses from. In this way, consent can be a poor fit for genomic data in research and healthcare.

Finally, Article 49(3) also restricts 'public authorities' from relying on consent in the 'exercise of their public powers' as an Article 49 derogation.* As noted earlier in Section 7.3.4, Section 7 of the DPA 2018 defines what entities count as 'public authorities', limiting the category to a defined set of lists. Notably, many NHS bodies are therefore highly likely to count as 'public authorities' when processing for healthcare purposes and so are barred from relying upon the Article 49(1)(a) consent derogation.[363] As noted earlier, the status of some entities, for instance those wholly owned by public authorities, is less clear.

In summary, consent as an Article 49 derogation poses a number of difficulties if it is to facilitate the lawful transfer of data. Namely, entities that count as 'public authorities' cannot rely on this derogation. Further, consent, even as a legal basis under Article 6(1)(a), is already a demanding standard to reach. Moreover, consent as a derogation also represents a more difficult standard to meet, the derogation requiring 'explicit' consent and specific risks regarding transfer to be disclosed.

2. Performance of a contract

Article 49 offers two derogations relating to contract Article 49(1)(b) and 49(1)(c):

> Article 49(1)
>
> (b) the transfer is necessary for the performance of a contract between the data subject and the controller or the implementation of pre-contractual measures taken at the data subject's request;
>
> (c) the transfer is necessary for the conclusion or performance of a contract concluded in the interest of the data subject between the controller and another natural or legal person;

The EDPB in their Guidelines 2/2018 on derogations of Article 49 under Regulation 2016/679 do not articulate the difference between the two derogations.[364] Nevertheless, the UK ICO's interpretation indicates that Article 49(1)(b) applies only to contracts between a controller and a data subject, whereas Article 49(1)(c) may also extend to third parties to a contract where they stand to benefit the data subject.[365]

Much of the EDPB's guidance in relation to both Article 49 derogations that relate to contract centres on the interpretation of 'necessary' covered above.[366] Notably, the EDPB has also issued interpretation of 'necessity' in relation to contract as a legal basis for processing in the context of provision of online services.[367] Four main conclusions flow from this adjunct guidance.

Necessity in contract

As expounded by *Heinz Huber v Bundesrepublik Deutschland* (C-524/06), 'necessity' has its own independent meaning in EU law, being interpreted in a way that reflects the objective on the law in question.[368] In the context of data protection, the fundamental right to privacy and protection of personal data as well as other data protection principles found in Article 5 are the main objectives in question.

* N.B. Public authorities are also restricted from relying on the two contractual derogations Article 49(1)(b) and (c) covered below.

Considering the 'necessity' of any given term includes a 'combined, fact-based assessment' of the purposes of processing and the term in question.[369] In this way, necessity of any given contractual term often turns critically upon the specific purpose of a contract. The EDPB explicitly acknowledged this, endorsing previous WP29 guidance that stated:

> 'There is a clear connection here between the assessment of necessity and compliance with the purpose limitation principle. It is important to determine the exact rationale of the contract, i.e. its substance and fundamental objective, as it is against this that it will be tested whether the data processing is necessary for its performance.'[370]

The EDPB, the CJEU, and the European Data Protection Supervisor (EDPS) all stress that this assessment must also consider whether there are other less intrusive means to achieve the same end.[371] In short, if there are 'realistic, less intrusive alternatives, the processing is not 'necessary.'[372]

In light of the previous finding, it is also clear that mere reference to a term being included in a contract is insufficient to establish that the term is 'objectively necessary' to perform or conclude that contract.[373] The EDPB again endorse WP29 comments on how to interpret 'necessary for the performance of a contract,' stating:

> '... must be interpreted strictly and does not cover situations where the processing is not genuinely necessary for the performance of a contract, but rather unilaterally imposed on the data subject by the controller. Also the fact that some processing is covered by a contract does not automatically mean that the processing is necessary for its performance. […] Even if these processing activities are specifically mentioned in the small print of the contract, this fact alone does not make them 'necessary' for the performance of the contract.'[374]

As a part of this assessment, the EDPB also highlights that where several services are offered, consideration of whether each service can be performed separately may indicate whether it is objectively necessary to deliver the individual services requested by the data subject.[375] To repeat the EDPB's words, you cannot bundle together services to create a 'take it or leave it situation for data subjects who may only be interested in one of the services.'[376]

Finally, similar to consent as a derogation, Articles 49(1)(b)-(c) are unavailable to public authorities in pursuit of their public powers.[377]

To summarise, the interpretation of 'necessity' makes it increasingly difficult to rely on contract as a derogation to facilitate transfer of data outside the EU/EEA. Specifically, since the bar for necessity is so strict, only a small number of contractual terms will ever be 'necessary.' Indeed, to facilitate data transfer outside the EU/EEA, the transfer will have to be core to the purposes for processing and not have other less intrusive options available. This seems an especially difficult task for a sector like genomics that often relies on reprocessing of personal data, data from third parties, and non-hypothesis driven research.

Further in relation to genomics for healthcare, contract can be a poor fit for the public provision of healthcare - contracts being uncommon between the NHS and its patients. Moreover, as EDPB Guidelines caution, contracts and consent concluded in return for provision of services, especially for services such as healthcare, can be exploitative and are held to a higher standard.[378]

3. Important reasons of public interest

Many of the purposes for processing genomic data have a public interest component. The Article 49(1)(d) derogation therefore represents a live mechanism for transferring personal data outside of the EU/EEA. Article 49(1)(d) states:

> **Article 49(1)**
>
> (d) the transfer is necessary for important reasons of public interest;

Broadly, there are four general restrictions and two points of flexibility when considering transfers that are 'necessary for important reasons of public interest' in the context of genomics.

Restrictions on the use of public interest

In regards to notable restrictions: Article 49(4) stipulates that the 'public interest' at stake must be laid down in EU or Member State law and apply to the controller in question if the Article 49(1)(d) derogation is to be relied upon. If there is no recognition of the interest in law, there is no legal mechanism under Article 49(1)(d) to transfer data.

It must be remembered that Article 49 derogations may only be selected in the absence of adequacy decisions or available appropriate safeguards, meaning that the controller must consider alternative safeguards available under Article 46 before relying on important reasons of public interest.

Much of the discussion of how to interpret 'necessity' under Articles 49(1)(b) and (c) also applies to 'transfer necessary for important reasons of public interest.' Namely, the transfer will have to be central to the public interest purpose for which the data is to be transferred. Moreover, if there is a less intrusive measure available, the public interest derogation may remain unavailable.

The EDPB (and its predecessor WP29) interpret Article 49(1)(d) as requiring something more than just a law requiring transfer of data that has public interest.[379] As described by the WP29, some kind of 'spirit of reciprocity for international cooperation' must be inferred or explicit in the law.[380] It is insufficient to have a law that serves a public function in both the EU Member State and the third country - there must be some kind of public international cooperation to rely on Article 49(1)(d). In this respect, the EDPB notes that an international agreement or convention recognising international cooperation being necessary for a given purpose can be indicative of public interest sufficient to rely upon Article 49(1)(d).

Public interest flexibility

Despite these potential restrictions, there is some interpretative flexibility with respect to this derogation. Unlike the Article 49 derogations consent and contract that are subject to the caveat that the transfers only be 'occasional' and 'not repetitive,' the public interest derogation is free of such restrictions, and may facilitate transfers that are more than occasional and repetitive.[381] However, the EDPB specifically notes that the general principle of transfer under Article 44 and the nature of derogations being exceptions not the rule must be respected. To quote the EDPB, in practice the use of public interest to facilitate data transfer needs to be 'restricted to specific situations and each data exporter needs to ensure that the transfer meets the strict necessity test.'[382]

There is the possibility that public interest under Article 49(1)(d) may be relied upon by private entities. The EDPB notes that recital 112 supports this interpretation mentioning both public and private entities with respect to public interest.[383] This is perhaps useful for wholly owned entities such as Genomics England that may not be public authorities for the purposes of selecting a legal basis but might perform some public functions and so benefit from some of the advantages of Article 49(1)(d) as a derogation.

To summarise, the public interest Article 49 derogation has multiple restrictions upon its use. However, it comes with flexibility that other derogations such as consent and contract lack. Moreover, where applicable, genomics, both for healthcare and research, may be in a unique position to take advantage of this derogation since data sharing in this sector is critical to facilitate good science and replicable studies. That is, genomic data sharing in particular and scientific research in general is supported by multiple international agreements and conventions outlining the benefits of sharing of data, underpinning the right to science. For example, Article 27 of the Universal Declaration of Human Rights outlines the right to science, a similar right being articulated by Article 15 of the International Covenant on Economic, Social and Cultural Rights, and Article 15 of the Oviedo Convention outlining that scientific research in the fields of biology and medicine shall be carried out freely. In this way, perhaps the most difficult hurdle of establishing a public interest basis that has some kind of 'spirit of reciprocity' already exists for scientific research and genomics in particular.

4. Vital interests

> **Article 49(1)(f)**
>
> the transfer is necessary in order to protect the vital interests of the data subject or of other persons, where the data subject is physically or legally incapable of giving consent;

Vital interests under Article 49(1)(f) may also facilitate data transfers to third countries or international organisations. When interpreting this derogation, the EDPB primarily considers a medical emergency necessitating transfer of personal data as being the kind of situation that would fall within 'vital interests.'[384] The upshot of the EDPB's interpretation is that this derogation has very narrow use for data controllers and should be invoked with special care. There are four key restrictions on the use of this derogation.

Incapable of giving consent

The data subject whose data is subject to transfer must be either physically or legally incapable of giving consent to rely on the Article 49(1)(f) derogation. Where the data subject's wishes can be 'solicited,' the derogation does not apply.[385] Notably, there are unexpected events such as natural disasters where the data subject is 'considered to be unable to provide their consent.'[386]

'Essential diagnosis'

Invocation of the Article 49(1)(f) derogation assumes that the 'imminent risk of serious harm to the data subject outweighs data protection concerns.'[387] The derogation allows for transfer of data from a data subject that is incapable of giving consent, to protect the vital interest of another but the harm must still be clear and apparent. As a consequence, where health data are concerned, the transfer must be necessary for an 'essential diagnosis.'[388] As the EDPB notes, this does not mean that only physical integrity of persons is concerned - mental integrity may indeed be sufficient under vital interests.[389] The EDPB clarifies that this derogation cannot be used to justify a transfer 'if the purpose of the transfer is not to treat the particular case of the data subject or that of another person's but, for example, to carry out general medical research that will not yield results until sometime in the future.'[390]

Elective or planned procedures

As with reliance on vital interests as a legal basis under Article 6(1)(d), vital interests as an Article 49 derogation are unlikely to apply where the procedure is planned in advance, where the processing is on a large scale, or where the procedure is elective and does not constitute an emergency.[391] In these cases, an alternative mechanism for transfer is likely to be appropriate.

Other appropriate mechanisms

Reliance on vital interests is generally only acceptable where other safeguards under Article 46 or other derogations under Article 49 remain unavailable. Indeed, recital 46 highlights with respect to vital interests as an Article 6 legal basis: 'Processing of personal data based on the vital interest of another natural person should in principle take place only where the processing cannot be manifestly based on another legal basis.' In some cases, such as epidemic surveillance, it is recognised that processing may simultaneously serve public interests and vital interests of the data subject. However, in these instances, the EDBP reaffirms that the data subject needs to be incapable of giving their consent, in this respect the controller should consider whether another basis such as Article 49(1)(f) might be more appropriate and, if not, whether the situation is akin to a natural disaster where the data subject can be deemed to be unable to provide their consent.[392]

The consequence of the above interpretation of the Article 49(1)(f) vital interests derogation is that the derogation will be difficult to lawfully rely upon. While the transfer of genomic data can often inform and underpin essential diagnosis, more often than not the transfer of genomic data is better supported by an alternate legal mechanism or does not constitute an emergency. Further, in the context of research, unless this research directly supports 'essential diagnosis,' vital interests likely do not apply. In short, controllers relying on Article 49(1)(f) should be certain that the vital interest protected by the transfer is indeed vital and is not better facilitated by another basis for transfer. Nevertheless, Article 49(1)(f) is an important (albeit infrequently available) tool, acting as a pressure release valve for the lawful transfer of data in emergency situations.

5. Public register

> **Article 49(1)(g)**
>
> the transfer is made from a register which according to Union or Member State law is intended to provide information to the public and which is open to consultation either by the public in general or by any person who can demonstrate a legitimate interest, but only to the extent that the conditions laid down by Union or Member State law for consultation are fulfilled in the particular case.

One of the core activities of genomics is the assembly of large variant databases used for reference. These databases, for example ClinVar, draw together pathogenic and suspected pathogenic variants. Although many of these databases purport to make available only anonymised data, to the extent that they store personal data Article 49(1)(g) may be a tempting derogation to facilitate data transfers. However, there are a number of restrictions on the use of this derogation that may make its use less appealing, less appropriate to certain institutions within the genomics community.

Definition of register

For a controller to invoke Article 49(1)(g), there needs to be some kind of 'register.' 'Register' is not defined within the GDPR but the EDPB defines the term according to its regular dictionary meaning: 'a (written) record containing regular entries of items or details' or as 'an official list or record of names or items.'[393] The EDPB adds that in this context a register may also be electronic.[394]

Public register open for consultation

The register must also be 'intended to provide information to the public' and must be 'open to consultation either by the public in general or by any person who can demonstrate a legitimate interest.'[395]

In regards to the 'public' nature of the register, EDPB Guidelines 2/2018 are slim, the Guidelines noting that 'private registers (those in the responsibility of private bodies)', for example, private registers that appraise credit-worthiness, are beyond the scope of the Article 49(1)(g) derogation.[396] This interpretation seems to rule out private registers for public ends.

The requirement for being 'open to consultation' also has unclear application in the realm of genomics. The EDPB provides a number of examples of registers that might meet the requirement of being 'open to consultation': 'registers of companies, registers of associations, registers of criminal convictions, (land) title registers or public vehicle registers.'[397] In this respect, if some genomics 'registers' did meet this requirement, they would hardly be the standard case and will likely have to be defended as being truly 'open for consultation.'

According to Union or Member State law

Reliance on Article 49(1)(g) is only possible where the register is 'according to Union or Member State law.' Accordingly, the controller purporting to rely on this derogation must have some form of EU or Member State law underpinning the register. This law, depending on its status within EU law, will be subject to challenge if it conflicts with the GDPR or the Charter of Fundamental Rights of European Union. Further, this law will also lay down the conditions under which personal data transfers from the register will take place.[398]

Restrictions upon transfers

The EDPB in their interpretation of Article 49(1)(g) highlight Article 49(2), cautioning that transfers according to this derogation cannot 'involve the entirety of the personal data or entire categories of the personal data.'[399] Further, Article 49(2) further stipulates that 'where the register is intended for consultation by persons having a legitimate interest, the transfer shall be made only at the request of those persons or if they are to be the recipients.' Consequently, where the register is open for consultation, transfers can only be made to the data subject themselves or upon their request, taking into 'full account the interests and fundamental rights of the data subject.'[400] The consequence of this being that Article 49(1)(g) has limited use for genomics in the context of healthcare or research, where the recipients (more often than not) are healthcare professionals or researchers.

Finally, as a reminder, Article 49(1)(g) is a derogation, being subject to the proviso that derogations are always exceptions to the rule, never the rule themselves.[401] All things considered, the restrictions upon the use of Article 49(1)(g) mean that only a subset of institutions (if any) transferring genomic data to third countries or international organisations will be able to safely rely upon this derogation.

7.2.6 Compelling legitimate interests (Article 49(§2))

Compelling legitimate interests according to Article 49(§2) is a derogation in a similar vein to the other Article 49 derogations covered above. Consequently, Article 49(§2) may facilitate the transfer of personal data to third countries or international organisations. However, the circumstances in which it may be invoked are extremely limited. The following restrictions on the use of Article 49(§2) are listed to emphasise its exceptional nature.

I. As a derogation, there must be no adequacy decision nor any appropriate Article 46 safeguard or Article 49 derogations listed in Article 49(1) available for use. In this respect, the EDPB stipulates the requirement for 'serious attempts' to transfer according to these alternative mechanisms.[402]

II. The transfer must not be 'repetitive' and can only concern a 'limited number of data subjects.

III. The transfer must be 'necessary for the purposes of compelling legitimate interests.' Under Article 6(1)(f), 'legitimate interests' may be relied upon as a legal basis. The EDPB notes that the standard for demonstrating 'compelling legitimate interests' is therefore higher than the Article 6(1)(f) legal basis. In this respect, the EDPB highlights that the legitimate purposes must be 'essential' for the controller, for example, 'to protect its organization or systems from serious immediate harm or from a severe penalty which would seriously affect its business.'[403]

IV. These 'compelling legitimate interests' must not be overridden by the 'interests or rights and freedoms of the data subject.'

V. The controller must assess 'in all the circumstances surrounding the data transfer and has on the basis of that assessment provided suitable safeguards with regard to the protection of personal data.'

VI. The controller must inform the competent supervisory authority of the transfer.

VII. The controller must also provide the information listed in the rights to information (Articles 13 and 14).

Following the above, lawful reliance on Article 49(§2) to conduct international data transfers requires exceptional circumstances to arise.

7.3 Conclusions

Genomic medicine and research rely on effective and proportionate data sharing. To interpret very rare genetic changes and to understand their significance for health requires large (typically international) datasets. As this chapter demonstrates, there are a number of core challenges with respect to the sharing of genomic data under the GDPR both within and outside the EU/EEA.

Sharing within the EU/EEA

The sharing of personal data between controllers and between controllers and processors often requires agreement between the parties which in turn, requires consensus on foundational issues, such as what data constitutes personal data. Without this consensus, divergent views between controllers even within the EU/EEA will likely hamper data sharing, regardless of whether the parties are within or outside the EU/EEA. The probability of finding divergent views amongst controllers is increased with MS SAs advocating different interpretations and implementations of the GDPR. Divergent views on the application of data subject rights and a lack of shared understanding of how to respond to such rights will diminish opportunities to share personal data regardless of where the data exporter and importer are located.

Sharing outside the EU/EEA

However, sharing data outside the EU/EEA poses additional challenges. If personal data is to be transferred to a third country or international organisation there must be a legal mechanism to facilitate its transfer. Uncertainty over what constitutes a 'transfer' could be problematic for the genomics sector given the presence of methods such as query-based research. Regardless of the legal mechanism selected, an adequate level of protection 'essentially equivalent' to that guaranteed within the EU must be provided. Legal mechanisms for transfer can be categorised as a hierarchy of three main categories, adequacy, Article 46 safeguards, and Article 49 derogations:

Adequacy decisions > Article 46 safeguards > Article 49 derogations

Broadly, a structured process mandates that adequacy must be relied upon first, Article 46 safeguards only in the absence of adequacy, and Article 49 derogations only in the absence of either adequacy or feasible use of safeguards. Nevertheless, each mechanism has a use, even if some are narrow, or unavailable to the genomics community.

Adequacy of a third country, a sector within a third country, or an international organisation represents a tempting solution to facilitate data transfer. However, adequacy decisions take time, can be a demanding standard to attain, and are not free from conditions nor impervious to challenge.

Article 46 safeguards provide a number of feasible mechanisms for the genomics community:

- Legally binding and enforceable instruments between public authorities are promising for the subset of organisations that process personal data in exercise of their official authority and have been given a favourable interpretation by recent EDPB guidelines.
- Standard data protection clauses or approved data protection clauses also offer genomics organisations a method of sharing data that can be standard or tailored to their circumstances.
- Codes of conduct and certification mechanisms offer genomics organisations an opportunity to standardise practice and demonstrate general compliance with the GDPR. Although the bar is high and there are no examples yet, codes or certification are both promising legal mechanisms for an organised sector that often has shared objectives such as genomics.

Given the warning from the EDPB that Article 49 derogations are necessarily exceptions to the rule, never the rule themselves, the following mechanisms seem most promising for the transfer of genomic data:

- Explicit consent can facilitate data transfer under most conditions, although it imposes a high standard (especially where obtained in return for provision of a service) and may be difficult to achieve in research where data is often not obtained directly from data subjects.

- Transfer necessary for important reasons of public interest, while also given a high bar, is promising for the genomics community, as many organisations are in a strong position to demonstrate tangible and critical benefits from international cooperation.

Our analysis suggests that the genomics community should explore a number of legal mechanisms for transfer. They should lobby for third country adequacy in their jurisdiction, craft codes of conduct and certification mechanisms to demonstrate general compliance with the GDPR, work toward developing these mechanisms to satisfy Article 46, and, if necessary, rely on a safeguard or derogation that fits their particular legal position. In short, controllers should pick the right mechanism for their situation but also work as a sector to develop sector-wide solutions to the challenge of international data transfer.

8. The reduction and mitigation of potential deleterious impacts

Our research has identified a wide array of impacts and challenges raised by the GDPR for the use of genomic data in healthcare and medical research. In the preceding chapters we have discussed some of the ways that legal requirements—or more frequently ambiguity about legal requirements—could have potentially deleterious effects on the use or sharing of genomic data for healthcare and research.

In this chapter we consider some of the potential mitigations or solutions that have been raised in the literature and through the course of our research. These fall into three main categories:

First, there are technical measures or approaches that may be taken to safeguard privacy and protect against identification whilst enabling processing of genomic data for healthcare or research.

Second, there are legal, procedural or organisational measures that can be put in place by processors and controllers to ensure compliance with the GDPR.

Third, we discuss broader approaches that the genomics community or specific sub-sectors could take to seek greater certainty about the requirements of the GDPR and Member State laws, and to set standards which protect individual rights without compromising valuable healthcare, public health or research activities.

In the final, subsequent chapter, we highlight general conclusions for the entire sector as well as more specific conclusions and recommendations for relevant audiences.

8.1 Identification risks and technical mitigations

There is now a rich literature and field of research focused on technical methods for breaching and protecting the privacy of genomic information, drawing heavily on computer science and cryptography.* In their influential review of technical privacy breaching techniques and privacy-preserving technologies for genetic data, Yaniv Erlich and Arvind Narayanan characterised several strategies for re-identification of an individual. 'Identity tracing attacks' describe strategies for re-identification based on the combination of genetic data with other 'auxiliary' information, such as records of Y chromosomes, surnames and family trees.[404] These form some of the most high-profile forms of re-identification attacks. For example, they include the methods used to identify the Governor of Massachusetts in a supposedly anonymised medical dataset using voter registration data,[405] and, most recently, the genealogical triangulation and inferential techniques used to identify the Golden State Killer in California.[406]

It has also been proposed that genetic data may be used to identify individuals through phenotypic prediction of visible traits, including face morphology.[407] The predictive power of such data is currently limited for most traits (excepting eye colour or age)[408] but in the future they may form a more powerful re-identification method. As Erlich and Narayanan discuss, other strategies can be used to identify newly generated personal data relating to an individual. For example, imputation techniques allow the 'completion' of regions of a genome from only partial DNA data. This approach was famously used to infer James Watson's predisposition to Alzheimer's disease despite the masking of the relevant gene.[409]

The personal information of family members may also be identified using DNA or genetic data from one family member. For example, in 2013 Humbert and colleagues demonstrated that the publication of Henrietta Lacks' genome allowed statistical inferences to be drawn about the DNA sequences of other members of the Lacks family.[410]

Mitigations: technical safeguards

In the face of these challenges, there is also an active field developing strategies to mitigate or reduce the likelihood of re-identification of individuals. There are a range of technical methods that may be deployed currently and in the near future to protect genomic information but they also have their own strengths and weaknesses. These have been dealt with at length in official guidance from the WP29[411] and are considered more fully in our previous work.[412]

* Although written under the previous regime, the ICO's anonymisation code of practice provides a useful overview of key techniques and approaches. See:

The Information Commissioner's Office. Anonymisation: managing data protection risk code of practice. 2012, 1-108.

For example, techniques such as k-anonymity can be used to ensure that no record in a dataset has a unique combination of quasi-identifiers by reducing the resolution of some data or suppressing unique data entirely. However, Erlich and Narayanan question its privacy properties and how useful it may be in the genomic context. Another method that is used to mitigate against identification risks is differential privacy. This method works by adding noise to results so that no one accessing a dataset can be sure that an individual's data are part of the results. The amount of noise that is necessary depends on the level that an individual's information contributes to the query that has been made; the more such information contributes, the greater the noise that has to be added to the released data.[413] Although this is a highly effective technique, as Erlich and Narayanan noted, the amount of noise that is required for the release of genetic results, such as GWAS summary statistics, is considerable and may be impractical for some uses.

Mitigations: bringing analysis to the data

Due to the inherent risks in the sharing of genetic data (and associated metadata), several approaches have been developed to reduce the sharing of locally held data in place of secure and targeted query-based research on local datasets resulting in non-identifying but high utility results.

An example of this is DataSHIELD (Data Aggregation Through Anonymous Summary-statistics from Harmonised Individual Level Databases),[414] an initiative to facilitate query based research of biomedical data collections. Under DataSHIELD, commands are sent from researchers to multiple local databases via a centralised analysis computer and summary statistics are returned to the researchers without revealing any locally held confidential data. This avoids the pooling of local databases but produces an analysis that is mathematically equivalent to pooling all the data centrally.[415]

Under the previous data protection regime, Wallace and her colleagues concluded that DataSHIELD analysis does not involve the sharing of any personal data because this process avoids any transfer (including cross-border transfer) of individual level data and because there are strict controls on the summary statistics which are produced. For example, these include preventing tables returning data including fewer than 5 observations and an approval process for the commands that can be sent through DataSHIELD.[416]

The developers of DataSHIELD acknowledge that there remains the residual risk of re-identification that applies to all aggregated summary statistics but the mitigations in place mean that DataSHIELD provides a robust mechanism to limit the exposure of identifiable personal data as far as possible in biomedical research.

At a global level, the Global Alliance for Genomics and Health (GA4GH) is also pursuing the query-response model across multiple large-scale datasets through their Beacon Project.[417] This was created in order to test the willingness of data holders ('Beacons') to share genetic data in the form of a simple binary response to a query for the presence of a specific nucleotide at a given position within a chromosome. For example, this allows clinicians to quickly search for existing case matches in participating clinical datasets without having to go through the secured-access procedures of systems like Matchmaker Exchange.[418]

If a Beacon is 'lit' it is then optional whether the Beacon discloses metadata such as pathogenicity scores and associated phenotypes along with a Yes response to a query, most likely on the basis of the requesting party identifies themselves as a registered user and under the terms of any prevailing consent agreements applying to the dataset.[419]

In addition, there are advanced cryptographic methods which can mitigate or reduce risk even further. These include homomorphic encryption, a technique that allows a query or algorithm to be executed on encrypted data so that the ensuing results are only visible by someone with a key.[420] This is a particularly promising method for third party genetic analysis, for example, by an interpretation service on genomes held in the cloud, without disclosure of private information to that third party.

Alternatively, cryptographic techniques may be used to allow multiple users to carry out computation on other users' private datasets without revealing their contents. Raisaro, Choi, and colleagues, have shown that techniques such as homomorphic encryption and differential privacy can be used efficiently to protect genomic privacy when conducting specific tasks. In their case, allowing researchers to query data for an aggregate number of patients who meet a given set of inclusion or exclusion criteria and conduct an exploratory analysis of a genetic cohort.[421]

Despite these mitigations, it is also the case that some identification risks remain. For example, Shringarpure and Bustamante demonstrated that it is possible to identify whether an already known individual (or their relatives) is part of a Beacon through repeated Yes/No queries alone.[422] Knowing membership of a Beacon is potentially highly sensitive information because Beacons are often condition specific and it can be inferred that an individual belongs to a family with a specific condition (for example autism spectrum disorder).[423]

More generally, it is also likely that methods to defeat de-identification or encryption will develop in time which means that the challenge for data controllers is to ensure that their methods do not become out of date and that they remain abreast of technological developments that can be used for re-identification or that pose a threat to privacy. As discussed in chapters 4 and 6, data controllers and processors are required by multiple provisions of the GDPR to keep up to date with the risks and safeguards required to protect the privacy and security of genomic data.

Reaching a consensus on identification risks

An awareness of the risks of identification has frequently led to increased protection and suppression of the data that are otherwise made openly available in many contexts, including medical and genetic contexts. However, there has been vigorous debate over the last decade about the nature and quantum of the risk of re-identification in different contexts, how risks can be mitigated and, more fundamentally, whether de-identification is capable of protecting privacy. Scholars like Paul Ohm[424] Arvind Narayanan[425] and Yves-Alexandre de Montjoye[426] have, with their colleagues, highlighted some of the 'broken promises' or failures of de-identification in the face of technological developments.

Others, like the former Information and Privacy Commissioner for the Canadian province of Ontario, Ann Cavoukian and her co-author Daniel Castro,[427] Khaled El Emam,[428] David Sánchez[429] and their colleagues, have emphasised that de-identification may be effective for many purposes and that risks of identification are, in their view, often exaggerated. In many cases these disagreements are related to specific technologies or statistical approaches and the precise nature of the context in question. However, it is possible to crudely characterise two approaches that may be taken to risks of re-identification.

On one hand, the risks discussed in the literature are often based on an attacker or adversary possessing a significant level of additional information about an individual and also possessing advanced computational skills. For example, the threat that Shringarpure and Bustamante identify of identification of an individual in a GA4GH Beacon is based on the 'attacker' possessing the individual's variant call format (VCF) file, which is a detailed list of all the SNP positions where the individual's genotype differs from a reference genome. Currently this is a high threshold to meet, given that a VCF file is generated from a genome sequence, and this technology is not yet used widely in clinical care and research, although this is set to change as genome sequencing technologies become part of routine clinical care.[430] One approach to this is therefore to focus on de-identification risks that present in real-world settings and manage them accordingly. On the other hand, technological advances are constantly diminishing the computational burden for re-identification and the increasing availability of 'high dimensional' datasets (such as individual level genetic or medical data) also increases the ability to single-out or cross-reference data to identify an individual.[431] Equally, although there is an increasing focus on privacy preservation in the use of artificial intelligence techniques for processing data,[432] as the European Commission recently noted in its White Paper on Artificial Intelligence, there is a risk that '[b]y analysing large amounts of data and identifying links among them, AI may also be used to retrace and de-anonymise data about persons, creating new personal data protection risks even in respect to datasets that *per se* do not include personal data.'[433] On this basis, an alternative approach to risk is to evaluate the near-to-medium term future developments and manage data in a precautionary manner.

In the context of the GDPR, the choice of approach to risk of re-identification is crucial because, as we discuss in chapter 4, the GDPR takes a risk-based approach to determining whether data are 'personal data' or whether they fall outside the scope of the Regulation. Disagreement about the risk of identification means that it can be very challenging to assess when means may be reasonably likely to be used to identify an individual (recital 26).

As we discuss further below, it may be that by narrowing the focus to analyse a particular form of processing within a specific sector, such as the management of genomic and phenotypic data for case-matching, consensus can be reached on the safeguards required and the residual risk of identification. If a specific community can come together to reach a consensus, it may be possible for them to set data protection standards that will be approved by the data protection authorities, for example in a code of conduct. If this is too ambitious, it may be that a sector or sub-sector of genomics activity can agree a process for evaluating identification risks, even if there may be some variation in opinion on its end results in a given circumstance.

For example, Mark Elliot and colleagues provide some suggestions for quantifying and formally expressing the risk of identification taking into the wider data environment as well as the data itself.[434]

On a more positive note, the methods discussed above, namely de-identifying data as far as possible, using advanced encryption and shifting to query-based systems rather than releasing data, should go a long way to ensuring that data remain reasonably unidentifiable. Combined with legal and environmental controls on data access and use (discussed below) it is highly likely that many uses of genomic data are capable of being sufficiently protected from re-identification that they fall outside the scope of the GPDR (for at least some of that processing).

However, as we emphasise in chapters 4 & 6, new developments need to be taken into consideration. For example, the potential for identification using molecular phenotype data, such as gene expression data, linking independent genotype and phenotype datasets and pinpointing an individual.[435] These require continual consideration to ensure that the risk of identification does not become too high. If the risk of identification increases, it is likely that technical measures can be taken in response to reduce any risks,[436] but ultimately, as the GA4GH acknowledges, this risk must be managed and balanced with the importance and utility of facilitating access to genomic data, including through secure access processes where the risk of identification is considered too high for public distribution.[437]

8.2 Legal and organisational measures

As well as technical safeguards, risks of identification and breach of privacy may also be addressed through legal and organisational safeguards. As we discuss in chapter 4, mechanisms such as restricted access procedures, data access agreements and legal obligations of confidentiality are all taken into account in the assessment of the risk of identifiability under the GDPR.

Mark Elliot and colleagues use the phrase 'data environment' to describe factors beyond the data itself that influence whether data are 'functionally anonymous' (i.e. anonymous in context) or not.[438] They highlight four key elements of a data environment: other data, data users, governance processes and infrastructure. Some of these may be addressed by legal and organisational controls but the first element, 'other data', is not likely to be capable of mitigation.

Indeed, the increasing popularity of genealogy and voluntary open sharing of genetic information by individuals is something that may require increased education and public awareness efforts to ensure that those individuals fully understand the implications for their—and their relatives—privacy.

The data users, governance processes and infrastructure are all aspects of the data environment where legal and organisational safeguards can be used to address identification risks. The users and their relationship with the data can be managed through data access controls, contracts, and internal policies which proscribe certain behaviour and provide sanctions for breach. The nature of the infrastructure, the software processes and security systems involved, can also be prescribed or required by agreements, contracts or policies.[439]

In terms of hard law, the criminal offence introduced in the Data Protection Act for knowingly or recklessly re-identifying information without the controllers consent is also a relevant safeguard.[440] Elliot and colleagues also propose a softer form of governance, licensing, to moderate a 'wild west', [441] and there may be various forms of agreement or self-regulation such as a code of conduct (discussed further below) which could also be used to reduce risk.

In practice, genomic data controllers will be required to adopt some or all of the mitigations discussed above in order to ensure data protection by design and default approach, as required by the GDPR (Art 25 and recital 78). Again, the challenge is to reach agreement on, and achieve greater certainty about, the level of risk of identification that meets the requirements of the GDPR. Adopting a structured approach which involves the formal expression and quantification of these risks, taking into account the broader data environment as well as technical nature of the data itself, could be a useful way of achieving a consensus on risks, although, there is still subjectivity involved in assessing the variables involved.

One of the difficulties of addressing these risks is that there is a pressing medical and scientific need for data to be as useful and as accessible as possible to those who are working to improve individual and population health. For example, in relation to matchmaking of clinical cases, the GA4GH are concerned that the need to register with matchmaking databases to search for a match creates a disincentive to data sharing which they are seeking to reduce. It may well be that there is an increasing need for intermediary services who can absorb the friction caused by data protection mechanisms for users whilst providing the usable data that is needed for healthcare or research purposes. In this regard, trusted third party processors who use homomorphic encryption and similar technologies to analyse data without confidential disclosure could form an important part of the genomic data ecosystem.

8.3 Sector-led approaches for improved standards and certainty

As we have emphasised throughout this report, a central challenge caused by the GDPR for genomic healthcare and research is that large parts of the Regulation apply directly to these sectors in much the same way that it applies to commerce, banking or other forms of processing. In particular, key concepts and most principles are applied regardless of whether they are applied to internet search engines or advanced genomic analysis.

This is not to say that the GDPR leaves no room for more context-sensitive application of rules which are proportionate to the aims and importance of the processing and the potential impact on data subjects rights. There are different provisions for healthcare, public health or research purposes which may (depending on the Member State) be used to legitimate processing.

The authors of the GDPR were also aware of the importance of contextual assessments and provided mechanisms for sector-specific standard setting to help ensure best practice and demonstrate compliance with the GDPR in the most appropriate way. These mechanisms could be a significant way in which some of the uncertainty and other potential negative impacts of the GDPR can be mitigated.

The two key mechanisms available to the genomics sector are codes of conduct (Art 40) and certification mechanisms (Art 42). As we noted in chapter 7, both of these are potential mechanisms to facilitate international transfers of personal data but they also have the potential to be used to set rules and facilitate compliance with the Regulation more generally. Currently it is codes of conduct that have been recognised by those in the genomics sector as potentially most useful[442] but as we consider further below, certification may also prove a useful tool for controllers and processors of genomic data.

Article 40 code of conduct

What is a code of conduct?

A code of conduct is a self-regulatory device which can be produced by a specific sector and voluntarily joined by members of that sector. In order to be effective under the Regulation, a code of conduct must be approved by a supervisory authority (such as the ICO) and, if it is intended to apply to processing beyond one Member State, the EDPB. Art 40 requires Member States, supervisory authorities, the EDPB and the Commission to 'encourage' the drawing up of codes of conduct 'intended to contribute to the proper application' of the Regulation. It envisages that associations and other bodies representing 'categories of controllers or processors' will prepare codes of conduct 'for the purpose of specifying the application of this Regulation'.

In particular, Art 40(2) makes clear that a code of conduct could specify the application of the Regulation in relation to:

> (a) fair and transparent processing;
>
> (b) the legitimate interests pursued by controllers in specific contexts;
>
> (c) the collection of personal data;
>
> (d) the pseudonymisation of personal data;
>
> (e) the information provided to the public and to data subjects;
>
> (f) the exercise of the rights of data subjects;
>
> (g) the information provided to, and the protection of, children, and the manner in which the consent of the holders of parental responsibility over children is to be obtained;
>
> (h) the measures and procedures referred to in Articles 24 and 25 and the measures to ensure security of processing referred to in Article 32;
>
> (i) the notification of personal data breaches to supervisory authorities and the communication of such personal data breaches to data subjects;
>
> (j) the transfer of personal data to third countries or international organisations; or
>
> (k) out-of-court proceedings and other dispute resolution procedures for resolving disputes between controllers and data subjects with regard to processing, without prejudice to the rights of data subjects pursuant to Articles 77 and 79.

This is not an exhaustive list but even so, many of the challenges highlighted through this Report could be the subject of a sector-specific code of conduct for genomics. As is clear from (j) in this list and Art 40(3) (discussed in chapter 7) a code of conduct may also address—and be relied on by—controllers and processers in third countries if they provide appropriate safeguards for the transfer of personal data outside the EU/EEA. The EDPB hope that such codes may serve as a mechanism to further develop and foster data subject trust in the processing of data outside the EEA.[443] However, the EDPB is planning to produce further guidelines on codes of conduct as a tool for international data transfers, meaning that there may well be further or quite different requirements for codes that potentially cover this form of processing. Moreover, to act as a safeguard for international data transfer, controllers or processors must make 'binding and enforceable commitments, via contractual or other legally binding instruments, to apply those appropriate safeguards, including with regard to the rights of data subjects' (Art 40(3)).

Benefits of a code of conduct

The EDPB produced guidelines on codes of conduct and monitoring bodies under the Regulation in 2019. They emphasised that codes of conduct represent a 'practical, potentially cost effective and meaningful method to achieve greater levels of consistency of protection for data protection rights', as well as acting as a mechanism to demonstrate compliance with the Regulation. The EDPB even suggest that codes may help to address harmonisation gaps between Member States although it is also possible that the development of different national codes governing the same form of processing could also lead to a divergence in approach.

In terms of acting as an accountability tool and demonstrating compliance with specific parts of the GDPR or the Regulation as a whole, the legislation makes clear that there are multiple ways in which a code of conduct may:

- provide guidance on appropriate measures and on the demonstration of compliance by the controller or processor 'especially as regards the identification of the risk related to the processing … and the identification of best practices to mitigate the risk, could be provided in particular by means of approved codes of conduct' (recital 77)
- be used as an element to demonstrate compliance with the obligations of the controller (Art 24(3)) or processors (Art 28(5))
- be used to demonstrate compliance with the requirements for security of processing (Art 32(3))
- be taken into account when assessing the impact of processing using a DPIA (Art 35(8))
- be taken into account by a supervisory authority when imposing an administrative fine (Art 83 (2)(j))

However, as a well as a means of demonstrating compliance with the Regulation, the EDPB suggests that they may serve as a means of establishing or tailoring data protection rules governing a particular sector. The EDPB refers to codes of conduct as 'an opportunity for specific sectors to reflect upon common data processing activities and to agree to bespoke and practical data protection rules.' Even more profoundly the EDPB define codes as 'voluntary accountability tools which set out specific data protection rules for categories of controllers and processors.' This language is more forthright than the Regulation itself which refers to codes of conduct as a means of specifying the application of the Regulation or as a means of contributing to the 'proper application' of the Regulation (Art 40) and demonstrating compliance with obligations (e.g. recital 81 & Art 32).

The potential benefit of a code for sectors like the genomics sector is that it would allow the sector itself to agree some of the data protection rules governing their processing operations, in co-regulation with the supervisory authorities. The EDPB see codes of conduct as a way of providing a degree of autonomy to controllers and processors to formulate or consolidate best practice in specific fields and provide much needed confidence and legal certainty.[444] The EDPB gives an example of 'micro enterprises involved in similar health research activities' coming together and developing a code 'rather than attempting to carry out such comprehensive data protection analysis on their own'.[445] Supervisory authorities may benefit as well, as the EDPB emphasise, through a potentially reduced demand for granular guidance for specific processing activities.[446]

What can a code of conduct cover?

The examples above make clear that a code of conduct could cover almost any aspect of the application of the GDPR to data processing in a sector like the genomics sector. In its guidance the EDPB discusses the example of a code of conduct for the processing of health data for research purposes. In this context the Board suggests that a code could 'outline in a fair and transparent manner' the following:

- the relevant safeguards to be applied regarding the information to be provided to data subjects;
- relevant safeguards to be applied in respect of the data collected from third parties;
- communication or dissemination of the data;
- the criteria to be implemented to ensure respect for the principle of data minimisation;
- the specific security measures;
- appropriate retention schedules; and
- the mechanisms to manage the data as a result of the exercise of data subjects' rights (as per Articles 32 and 89 of the GDPR).

The EDPB confirms that a code of conduct may be 'drafted in as narrow or as wide-ranging a manner as is befitting that particular sector, provided that the code contributes to the proper and effective application of the GDPR.'[447] However, they note that if a code is more focused it must be clear to data subjects that adherence with the code does not necessarily ensure compliance with all of the legislation.[448] The scope must be clearly and precisely defined according to the processing operations that it covers, as well as defining potential categories of controllers and processors.[449]

It is open to those in the genomics sector to determine how wide or narrow a code of conduct could or should be. As discussed further below, one of the challenges of a broad code of conduct in terms of both the breadth of the sector, and the range of activities it seeks to govern, is that it may be more challenging obtaining consensus on a need for a code and its contents. There may be additional challenges meeting the requirements for a code of conduct under the GDPR.

Justifying the need for a code

First, to obtain supervisory authority approval, the EDPB require the authors of the code to demonstrate that their draft code meets a particular need of that sector or processing activity, facilitates the application of the GDPR, specifies its application, provides sufficient safeguards and provides effective mechanisms for monitoring compliance.[450] Recital 99 indicates that a consultation should take place with the relevant stakeholders, including data subjects where feasible. The EDPB advise that a code should be submitted with supporting material such as a consultation summary, membership information or 'research that demonstrates the need for a code'.[451] Where no consultation has been carried out 'due to lack of feasibility' it will be a matter for the code owner to explain this position.[452] Moreover, the EDPB requires code owners to demonstrate that they are an effective representative body, suggesting this may be evidenced by e.g. the number or percentage of members from the sector that have subscribed to comply with the code.

Justifying the need for a code of conduct and also the representative nature of the code owner could involve considerable effort if, for example, it requires a broad sector-wide consultation with all relevant stakeholders. The EDPB give the example of the 'health research sector' identifying a need to formulate a code which provides consistency in approach by setting out standards to adequately meet explicit consent and accompanying accountability requirements.[453] The Board does not describe the evidence of need that may be provided in such a scenario but presumably evidence of a variation in approaches to explicit consent and accountability would provide a useful justification.

Challenge of monitoring and oversight requirements

Another challenge in developing a code of conduct is meeting the requirements for effective and efficient monitoring and oversight. The EDPB require clear, suitable, attainable, efficient and enforceable (testable) oversight mechanisms.[454]

Although the Regulation refers to oversight mechanisms 'which enable the body referred to in Art 41(1) to carry out mandatory monitoring', and Art 41(6) makes clear that no such monitoring body is required for processing carried out by public authorities, the EDPB guidance is that all codes (including those that apply to the public sector) should contain oversight mechanisms. The Regulation does not specify these but the EDPB suggests that such mechanisms could include regular audit and reporting requirements, clear and transparent complaint handling and dispute resolution procedures, concrete sanctions and remedies in cases of violations of the code, as well as policies for reporting breaches of its provisions (para 40).

Second, in the case of private sector processing, a code must specify an independent and expert monitoring body who is accredited for that purpose by the supervisory authority (Art 41). Accreditation is contingent on a body demonstrating its independence and expertise in relation to the subject-matter of the code as well as demonstrating that it does not give rise to a conflict of interest. If these conditions are no longer met by a body or where its actions infringe the Regulation, the supervisory authority may revoke its accreditation Article 41(5).

The EDPB require a monitoring body to demonstrate its independence from code members and the profession, industry or sector to which the code applies and it must be demonstrated that no conflict of interest will arise through the exercise of its tasks. It must also be sufficiently expert and well-resourced to perform its functions, and there must be transparent, accessible and effective complaints procedures in place. It is also important that the monitoring body has the legal status to perform its role and be the subject of a fine by a Supervisory Authority.[455]This could be interpreted as a very high standard, in particular the standards requiring independence and avoidance of potential conflicts of interest.

The Board's guidance tempers this by stating that a monitoring body may in fact be internal to a code owner (e.g. an ad hoc internal committee or internal department within the organisation) so long as its impartiality and independence can be demonstrated.[456]

This means that an independent monitoring body could either be a formally separate organisation or an internal division of a code owner. What matters is that impartiality and independence can be demonstrated through separate management structures, protection from sanctions or interference as a consequence of the fulfilment of the task, or, full autonomy for the management of a budget or resources. However, the EDPB advise that a 'monitoring body will need to identify risks to its impartiality on an ongoing basis' and that it must demonstrate how it safeguards against such risks.[457]

On the basis of the Board's guidance, the standard of independence for monitoring bodies, although high, is not insurmountable. Moreover, the EDPB considers it possible to appoint a number of monitoring bodies to carry out effective oversight[458] so there are a range of options open to code authors in the genomic sector.

Challenges of a transnational code

Where a code of conduct applies solely to processing within a single Member State, bodies can prepare a code of conduct but must then submit it to a supervisory authority who will issue an opinion on the code and whether it provides 'sufficient appropriate safeguards' (Art 40(5)). However, if the code relates to processing activities in several Member States then a different procedure applies. The code authors must choose a competent supervisory authority from whom approval can be sought. This choice is open but the EDPB advises that this could be based on the location of the code owner, the location of the largest density of the sector, the location of the largest density of the affected data subjects, or, initiatives developed by a supervisory authority in a specific field.[459]

The last of these aspects could be particularly important if, for example, a particular supervisory authority has developed initiatives or specific expertise in relation to genetic and genomic data. If a code is admitted by a competent supervisory authority, this will then trigger an 'informal cooperation procedure' whereby they will seek co-reviewers (a maximum of two) to assist with the assessment from other supervisory authorities.[460]

The competent supervisory authority will use their opinion to reach a final determination on whether to submit the code to the EDPB for an opinion under Art 40(7). The EDPB will either confirm or dispute that the code complies with the Regulation. If it confirms the code, then it will submit its opinion to the Commission who may determine that the approved code should have 'general validity within the Union' via an implementing act. This is required if a code is used to make binding and enforceable commitments which constitute appropriate safeguards for transfers to a third country or international organisation (under Art 46(2)(e)) (see chapter 7).

Although the potential benefits of a transnational code for the genomics sector are considerable in terms of the harmonisation of rules and improved legal certainty across the EU/EEA, the need for cross-EEA agreement and approval could prove a serious challenge for code authors.

Activity to date

Codes of conduct were encouraged under Art 27 of the Data Protection Directive but they were described in more open terms, with no suggestion of what topics they might cover (other than the encouragement of anonymisation codes of conduct in recital 26) and with only a requirement that national DPAs submit an opinion on sector led codes, rather than requiring approval. Furthermore, under the Directive there was not the explicit potential for codes of conduct to be used to demonstrate compliance in a way that Art 40 of the GDPR provides.

In their assessment of codes of conduct for research, Michal Koščík & Matěj Myška conclude that, although there has long been demand for improved clarity and harmonisation of data protection rules for research across Europe, no codes of conduct for research were established under the Directive because significantly lower sanctions resulted in commensurately lower motivation for research institutions to invest time and resources in the process.[461] Robert Bond and Alexander Dittel highlight a new feature, namely that adherence to a code of conduct may be used as an adequate safeguard for international transfers. In itself they argue that this new feature could spark considerable interest in codes of conduct under the Regulation.[462]

While no codes of conduct were developed for research under the Directive, a variety of codes were developed in multiple Member States, in some cases issued by the data protection authorities themselves.[463] These did not have to be approved by the DPAs and only one code was approved for community wide application by the Article 29 Working Party (a code of conduct issued by FEDMA relating to direct marketing).[464] Several others were in preparation during the transition from the Directive to GDPR, including some prepared by the cloud computing industry, and it may be that they will be taken forward for renewed approval under the GDPR, although this has not happened yet. (An EU Cloud of Code of Conduct has been published and is reportedly 'being submitted to the appropriate data protection authority' for approval but this has yet to be publicised).[465]

A code of conduct for research?

There have been recent significant developments towards a code of conduct relevant to the genomics community.

First, BBMRI-ERIC (the Biobanking and BioMolecular resources Research Infrastructure European Research Infrastructure Consortium, which operates and is developing a pan-European distributed research infrastructure of biobanks and biomolecular resources), with other collaborators, have been developing a code of conduct for health research.[466]

Jan-Eric Litton, BBMRI-ERIC's former Director General explained the motivation for such a code is so that scientists can ensure that 'hard-won concessions for research [in the GDPR] are not lost in translation.'[467]

Litton argues that the GDPR leaves too much room for interpretation, especially across countries, citing the different approaches that have been taken in the UK and Germany to pseudonymised data and when they may constitute identifiable personal data. Moreover, as Litton goes on to emphasise, a code can provide reassurance to citizens taking part in research that their sensitive data will be used carefully and protected from misuse.

The Code of Conduct for Health Research has been taken forward by BBMRI-ERIC and other European research organisations with a core drafting group and multiple topic-specific sub-groups covering challenges such as anonymisation and consent. The latest information available on the Code's website suggests that the development of the Code has been delayed due to the complexity of preparing a Europe-wide Code and the need to ensure compatibility with other national and sectoral initiatives.[468]

It is not surprising that the complexity of trying to develop a code of conduct across the broad sector of health research, addressing many different aspects of data protection regulation, is taking some time. The intention that this code is Europe-wide (and indeed that it will cover international transfers in a way that could be used by third country institutions as a sufficient and adequate safeguard) adds a considerable level of substantive and procedural difficulty.

The challenge for the development of such a Code is the drafters need to agree amongst themselves on key concepts, principles and data protection standards, before developing more granular, sector-specific rules and best practice. As we have noted throughout this report, the existing divergence across Europe in approaches to issues such as the scope of personal data, the appropriateness of different legal bases and the necessity of safeguards, need to be addressed if a code can be agreed. In terms of procedure, as discussed above, a transnational code of conduct must satisfy both the Member State supervisory authority chosen by the code owners (this appears to be the Austrian Supervisory Authority, in the case of the BBMRI-ERIC initiative)[469] and other relevant supervisory authorities through the EDPB opinion procedure.

A code of conduct for genomics?

Despite these challenges, the potential rewards of an approved Art 40 Code of Conduct for sectors such as health research are considerable. This is also something that the genomics community have recognised. David Townend, who is involved in the BBMRI-ERIC led initiative has written favourably about the potential for the scientific community—and particularly the genomics community—to take its self-regulation to a more formal level via a code of conduct. Townend proposes that this could include engaging with different publics across the world about the work of data sharing, the risks and safeguards involved and, in turn, could incorporate their responses in the governance structure.[470]

Most recently, an international group of lawyers and ethicists who were part of the global PCAWG (Pan-Cancer Analysis of Whole Genomes) consortium have called for an international code of conduct for genomic data sharing.[471] They report the difficulties of developing a single cloud of data accessible to researchers worldwide due to 'European regulators having concerns about genomics data from Europeans being held in the United States.'[472] In response they urgently call for clear data-sharing rules that are harmonized across jurisdictions and they specifically note the potential for a code of conduct under Art 40 and the benefits that the BBMRI-ERIC led initiative could bring. However, in the meantime, the PCAWG group advocate an international code that addresses how ethical and legal obligations can be satisfied in relation to international clouds. Whilst they propose such a code should outline the steps researchers must take to comply with the GDPR, amongst other key international regulations, such as the US Health Insurance Portability and Accountability Act, they do not propose that the international genomics community should currently aim for an Art 40 compliant code of conduct.

Molnár-Gábor and Korbel are also sanguine about a Europe-wide code of conduct for genomic data sharing. They seek to dispel 'false' short-term expectations about a code of conduct, in particular that it can provide for an immediate increase in the harmonisation of regulations on data processing within the EU.[473] They point out that most Member States have finalised their GDPR implementations and that this has already baked in a significant variation in approaches under the GDPR relating to the processing of special category data, processing for research purposes and the implementation of data subject rights. They highlight that any EU-wide code must respect Member State derogations—where they are allowed under the GDPR—and that the breadth of these means that a code of conduct does not provide a short-term solution for the harmonisation of a wide range of data protection requirements.

They also argue that a code of conduct is unlikely to offer significant advantages over other mechanisms for international transfer of data at present (such as standard contractual clauses and adequacy decisions) because it is still open for legal challenge to scrutinise aspects of the legal system in a third country on a case by case basis.[474]

However, Molnár-Gábor and Korbel are not negative about the value of work on a code of conduct for genomic data sharing (Molnár-Gábor is also a contributor to the BBMRI-ERIC drafting group). Like the PCAWG authors, they recognise the value of a voluntary code of conduct, even if it is not approved by the authorities, because it can lead towards further codification and integration of research ethics standards. They suggest that compliance with such standards could be considered an appropriate safeguard under data protection law even if the code is not Art 40 compliant. Molnár-Gábor and Korbel are most positive about the potential long-term benefits of codes of conduct. They emphasise that, in the long-term, codes could close any regulatory gaps left by the GDPR and lead to more consistent application of laws by supervisory authorities – even, potentially, influencing Member States' regulatory approaches and leading to a more coordinated understanding across Europe. However, they emphasise that no matter how 'awkward' it may seem, a code of conduct will only improve EU-wide data sharing 'if member state implementations of the GDPR are closely analyzed in advance and are taken into account when drafting the code.'[475]

Key impact
Although workshop participants were supportive of the concept of a sector-led code of conduct to help establish best practice and appropriate standards in the genomics context, many were cautious about the time and effort it would take and the difficulties there could be in reaching agreement across the sector.

A way ahead

We agree with Molnár-Gábor and Korbel that a pre-requisite for a comprehensive Europe-wide code of conduct for genomics (whether limited to research or including healthcare and other processing as well) will be an understanding of the variations in Member State law relating to processing of genetic and health data, scientific research, data subject rights and approaches to legal bases for processing.

It will not be easy to specify rules for genomic data in areas which are subject to diverse national approaches, for example on the safeguards required for research, the processing of health and genetic data, or, the requirements for fulfilling data subject rights. Where these apply, either the code would need to acknowledge and accommodate diverse legal requirements or, code authors could choose to provide a baseline set of requirements which supplementary national rules may vary depending on national context.

The complexity of preparing and agreeing a code of conduct that spans multiple areas of divergent Member State law is one of the reasons that codes of conduct may benefit from greater attention, if they are to successfully meet the requirements of Art 40. In addition, having a narrower focus might facilitate reaching consensus since the broader the sector and processing operations that a code aims to cover, the harder it may be to reach consensus.

As an alternative, the genomics sector could adopt a dual approach to codes of conduct. As the authors discussed above suggest, the genomics community could pursue a broad and sector wide code of conduct to establish and harmonise rules for genomic data as a self-regulatory code of conduct. Within, or alongside this, sub-sectors of specific processors and controllers could aim to crystallise best practice in a more formal code of conduct to be approved under Art 40, providing greater legal certainty and confidence in relation to those specific activities. For example, a more specific code of conduct could aim to address standards for de-identification of genomic data and set out processes and standards against which genomics researchers and other data controllers can assess whether data are 'personal data' or whether they are no longer reasonably identifiable.

This is not something that varies according to national law and even if there have been differences in approach to pseudonymised data and de-identification across Europe in the past, there is now a stronger need than ever to harmonise approaches if data-sharing is to succeed. This could be developed to apply to many uses of genomic data or perhaps just to a sub-set of processing activities, such as de-identifying data in genomic research. In parallel, the community could continue to develop standards and best practice on a wide range of data processing topics with the aim that guidance on discrete topics could be incorporated into, or be added alongside, an existing Art 40 code.

Article 42 certification

Although the focus of the genomics and research communities has mostly been on the potential for a code of conduct to help establish appropriate standards for compliance with the GDPR, the Regulation also encourages another mechanism for the 'purpose of demonstrating compliance with this Regulation of processing operations by controllers and processors' (Art 42(1)). This is through data protection certification and data protection seals and marks. In recital 100 it is explained that that such mechanisms can enhance transparency and allow data subjects to quickly assess the level of data protection associated with relevant products and services. The EDPB guidelines on certification[476] add that certification can also improve transparency for business-to-business relations, so, for example, this could apply to genomic data controllers and processors. As with codes of conduct, the GDPR also provides that 'an approved certification mechanism' can facilitate international transfers of personal data under Art 46(2)(f), together with binding and enforceable commitments of the controller or processor in the third country. However, we currently have limited guidance on the requirements for such certification mechanisms.

There are a number of different models for certification schemes under the GDPR. Supervisory authorities such as ICO may issue their own certification schemes and act as the certification body for them, determining whether controllers or processors meet the criteria they set out. Alternatively, supervisory authorities can delegate all or part of the assessment process to third parties, or they can encourage the market to develop their own certification schemes. In the genomics context, it is this model that holds the most promise for sector led-initiatives governing specific processing operations.

What certification can cover: 'Processing operations'

Unlike codes of conduct, certification schemes are aimed specifically at how controllers and processors may demonstrate compliance with the Regulation when carrying out 'processing operations'. However, as we note in chapter 2, 'processing' encompasses an extremely wide range of activities involving data and it is the EDPB's view that there is a similarly broad scope for what can be certified under the GDPR.[477] Likewise, the ICO consider that a certification scheme can be quite general and apply to a variety of products, processes or services.[478]

According to the EDPB, a set of processing operations can include governance processes and organisational measures but what is important is that a use case can be provided which explains the categories and amount of data, the technical infrastructure used and the nature, scope, content and purposes of processing, as well as the risks to the rights and freedoms of the data subjects involved.[479] This means that a certification scheme can have a relatively broad scope which overlaps to a degree with the breadth of activities that could be included within a code of conduct.

For example, it is possible that a certification scheme could apply to the processing of personal genetic data in direct-to-consumer contexts, even aiming to cover all aspects of the GDPR. However, what the guidance from the ICO and EDPB emphasises is that the 'object' of the certification scheme (also called a Target of Evaluation) can be described precisely and is capable of being assessed for compliance with the GDPR and audited by a certification body. This requires a significant level of granular detail and precision about the data, processes and technology involved so it is likely that it will be easier to develop certification schemes that have a limited scope, applying to a limited set of processing operations within a stable technical context. If a certification scheme is narrow, it should not mislead the user, consumer or data subject by suggesting that it applies to a broader set of processing operations.

It is also important how certification schemes or marks are presented to data subjects. The EDPB give an example of a "HealthPrivacyMark" which 'raises the expectation that data protection requirements in connection with health data have been examined' and therefore, the criteria should be 'adequate for assessing data protection requirements in this sector.'[480] As with codes of conduct, certification may also be developed or relied on to demonstrate the existence of appropriate safeguards to legitimate international transfers of personal data (Art 46(1)(f)), together with binding and enforceable commitments on the part of the third country recipients. However, the EDPB is also yet to produce guidance on this aspect of certification under the GDPR.

What criteria should be used in certification schemes?

Determining whether processing operations comply with the Regulation will depend heavily on context. In general, the EDPB has explained that the development of certification criteria should be focused on the 'verifiability, significance, and suitability of certification criteria to demonstrate compliance with the Regulation.'[481] Where applicable, this should include consideration of how processing operations should be assessed for compliance with (amongst others) lawfulness of processing (Art 6), data processing principles, data subject rights and the technical and organisational measures required by the GDPR. The EDPB also explain that the extent to which these are reflected in the criteria will also depend on the area of certification (e.g. health sector).

Another relevant consideration according to the EDPB is that certification criteria take account of, and (where appropriate) are inter-operable with other standards such as ISO standards or national level standards. As we discuss below, this could be important in deciding what topics a genomics related certification scheme could cover.

Who can act as a certification body?

Certification, seals or marks can only be issued by the supervisory authority or an approved certification body (Art 42(5)). Certification bodies themselves must be accredited by the supervisory authority or by approved national accreditation (in the UK this is the United Kingdom Accreditation Service or UKAS) which has demonstrated sufficient independence and expertise in relation to the subject-matter of the certification. The ICO stipulate that a certification body must be a formal legal entity that can be held responsible for its activities and that, for impartiality, there must be no connection (such as the provision of consultancy services) between the certification body and the applicant.[482]

In the UK, certification bodies will need to go through the UKAS accreditation process and be evaluated against relevant ISO standards (although the GDPR takes precedence over these if there are conflicts). This may take 6-18 months depending on the nature of the certification scheme so it is not a quick process. The EDPB emphasise that a certification body should be expert in the processes of review that will be necessary under the scheme. For example, this could include on-site inspections and review of codes and documents. Thus the requirements for certification bodies under the GDPR are considerable.

Certification in the genomics context

The relative lack of attention paid by scientific researchers and the genomics community to certification may be due to the fact that a large proportion of current certification schemes (i.e. not within the GDPR) are aimed at single issue or highly limited technical aspects of processing.[483] For example, this is generally the approach the International Organization for Standardisation (ISO) takes by encouraging a dedicated sectoral approach in the development of technical standards.[484]

However, more specific technical certification schemes could be very useful in setting standards for certain aspects of genomic data processing. For example, it may be possible to develop a certification scheme that sets standards for the pseudonymisation and de-identification of genomic data in compliance with GDPR provisions on data minimisation, scientific research and security of processing (amongst others). Such a standard would involve considerable technical specification and could take into account other existing standards such as the latest ISO standards on information security and privacy.* Although it could take considerable effort to develop a certification scheme specific to the genomic context that supplements existing standards, the benefit for the genomics community is that such a code (unlike existing standards) can explicitly be taken into account in assessing compliance under the GDPR.

Alternatively, it may be feasible for broader certification schemes to be developed which set standards for compliance with a wide range of GDPR requirements in a specific context. For example, the processing of genomic data in a specific healthcare context. However, a code of conduct may be a more appropriate mechanism for this approach.

A further potential of certification is that it can apply to EU-wide processing operations and the GDPR particularly encourages the establishment of certification mechanisms at Union level (Art 41(1)) and there is also the potential for a certification scheme with cross-border validity to be approved by the EDPB as a European Data Protection Seal (Art 42(5)).

It could also be that other specific sectors develop certification schemes which can be used and relied on by genomic data controllers and processors and which can ensure transparency and accountability for data subjects. For example, a certification scheme for cloud computing could help to demonstrate compliance with the GDPR when transferring and processing genomic data in the cloud, perhaps (in time) even as an adequate safeguard for international transfers of data to the cloud.

In summary, although they have been the subject of less focus from scientific researchers and the genomics community, Art 42 certification mechanisms could help streamline and harmonise aspects of compliance with the GDPR for genomic data processing. The technical nature of certification suggests that this is most likely to be the case for more specific aspects of genomic data processing, for example cloud-based processing of genomic data or, perhaps, pseudonymisation and de-identification of genomic data.

* For example, ISO/IEC 27701, see: British Standards Institution. ISO/IEC 27701:2019 Security techniques – Extension to ISO/IEC 27001 and ISO/IEC 27002 for privacy information management – Requirements and guidelines. London: BSI; 2019.

Code of conduct or certification mechanism?

There is considerable overlap between a code of conduct and certification as mechanisms for demonstrating compliance with the GDPR and this means it may be difficult for the genomics sector to decide which to pursue for certain purposes. For example, both may be broad or narrow in their scope. A code of conduct may apply to a whole sector such as the genomics healthcare and research sector, setting rules for compliance with a wide range of GDPR obligations, or it may focus on a specific sub-sector, such as genomics analysis and interpretation, setting specific standards for compliance with certain aspects of the GDPR (e.g. the information provided to the public and data subjects). Similarly, a certification scheme may broadly address a wide range of processing operations in a certain context, such as the handling of personal data by private enterprises,* or apply across multiple sectors (for example data flows in the cloud).

Both codes of conduct and certification may also be developed at a Union level (and contribute to the harmonisation of rules across the EU/EEA) but both must take into account considerable diversity in national derogations from certain aspects of the GDPR, such as the processing of genetic data and rules on scientific research. Furthermore, both codes of conduct and certification mechanisms may be developed as a tool for international transfer under the GDPR.

However, there are some significant differences. For example, as Kamara and colleagues highlight, Art 40 codes of conduct are aimed at tailoring rules for the needs of specific sectors, whereas certifications do not need to be so targeted under the GDPR. This may lead to codes of conduct being better suited to controllers and processors with sectoral, as opposed to multi-sectoral activity.[485] On the other hand, this difference may not be so marked in practice because it depends how 'sectors' are defined. For example, in the genomics context it could easily be the case that data controllers and processors fall across multiple 'sectors': the genomics sector, scientific research sector and possibly the healthcare sector.

In practice, there may not need to be an 'either/or' choice between codes of conduct and certification for the genomics sector. There is no reason why a suite of codes and certification mechanisms cannot be adopted and relied on to facilitate the processing of genomic and personal data in compliance with the GDPR but it is important that the genomics community and other relevant stakeholders are aware of the options available under the GPDR and are in a position to assess which approach is best suited to address challenges through further deliberation.

Our conclusion is that it would be most appropriate for the genomics community to develop codes of conduct in the manner considered above, both at the broad (even global) level to establish appropriate norms, and at more granular levels to set GDPR compliant standards for specific aspects of genomic healthcare or research activities.

* For example, the JIPDEC PrivacyMark System, see: JIPDEC. About the PrivacyMark. Available from: https://privacymark.org/about/outline_and_purpose.html

Controllers and processors could look to adopt approved, available certifications in relation to key aspects of genomic data processing (for example cloud based processing), to the extent that those standards are compatible with the specific needs and purposes of genomic data processing. It may also be the case that specific and technical processing operations with genomic data are best suited to being addressed via certification, particularly where they may build on other standards, such as ISO information security and privacy standards. There is nothing preventing compliance with specific certification mechanisms also forming part of appropriate best practice within a broader code of conduct for genomic data processing so it may be that both mechanisms can be combined to achieve greater legal certainty and harmonisation.

Given the considerable effort and resources that will be required in developing and operating GDPR compliant codes or certification mechanisms, it is important that the genomics sector consider how these tools may be best used to clarify and harmonise appropriate standards for processing. If discrete and achievable topics can be identified then both codes of conduct and certification could significantly assist the genomics sector to set appropriate and clear rules tailored to the particular challenges of this sector.

8.4 Conclusions

This research has identified a range of potentially challenging impacts that the GDPR may have on uses of genomic healthcare and research. However, as discussed in this chapter, there are some measures and approaches that the genomics community and other groups are taking, or could take, to mitigate negative impacts and ensure that appropriate standards are set. These standards should ensure a high level of protection for genomic data while facilitating data processing and sharing for health and research purposes.

There is an active field developing strategies to mitigate or reduce the likelihood of re-identification of individuals while facilitating the processing of genomic data. These include bringing analysis to data rather than sharing the data itself and using advanced computational methods such as homomorphic encryption to carry out secure analysis of sensitive data. By combining technical measures with legal and organisational safeguards (including data access controls, contracts, and internal policies) genomic data controllers can help to minimise the risks of misuse.

However, the high level of protection required for genomic data and the technical nature of the risks and mitigations involved may be challenging for some controllers. We suggest that there may be an increasing need for trusted third party processors who can process data applying technical and organisational best practice to generate the usable data that is needed for healthcare or research purposes.

Alongside this, there is scope for the genomics community or sub-sectors of genomics processors to develop best practice and appropriate standards for compliance with the GDPR, which ultimately, could become approved codes of conduct or certification mechanisms under the GDPR.

These may not be easy or quick to develop, and the broader the coverage of codes or certification schemes, the more they will require an understanding of the variations in Member State law relating to processing of genetic and health data and the more difficult it may be to secure support for, and reach agreement on, their substance.

To try and reach consensus more quickly and to establish some level of certainty for genomic data processing under the GDPR, we suggest that the genomics sector could adopt a dual approach to codes of conduct: simultaneously pursuing a broad and sector wide code of conduct to establish and harmonise rules for genomic data as a self-regulatory code of conduct and more specific sub-sector-led codes that aim to crystallise best practice into smaller formal codes of conduct to be approved under Art 40.

We make further recommendations for the genomics community, as well as for regulators and policymakers, in the following, concluding chapter.

9. Conclusions and recommendations

This research has identified a range of ways in which the introduction of the GDPR and associated Member State legislation impacts, or has the potential to impact, uses of genomic data for healthcare and health research purposes.

In this report we have analysed five main areas of challenge for genomic healthcare and research (chapters 3-7) and some of the potential mitigations and ways ahead that the genomics sector could take to improve legal certainty and ensure that the standards which are applied to the genomics context are appropriate, and tailored to the risks and benefits involved (chapter 8).

Throughout the report we have identified key impacts and made recommendations for data controllers and processors. In this chapter, first we draw some general conclusions about the impact of the GDPR on genomic medicine and research, and then draw some specific conclusions arising from the key impacts already identified in this research. We make recommendations for important groups of stakeholders, including researchers, data processors/controllers involved in healthcare and regulators/policymakers as part of these specific conclusions.

9.1 General conclusions

Genomics and data protection are **two highly dynamic spheres of activity**. Reconciling their aims and objectives is challenging. The GDPR is part of an ongoing process of updating and reinterpreting the legal framework for the processing of data to address modern technological revolutions and an increasingly automated, digital world. Genomics, and the techniques and approaches described in chapter 1, are advancing rapidly across healthcare and research. As novel uses and methods proliferate, these are generating increasing volumes of data of ever-improving utility. The scale and the nature of data that result from these genomic and other 'omic techniques give rise to significant challenges for data protection.

Our understanding of how these techniques can be used both alone, and in combination, is constantly improving. This, in turn, means such techniques are generating a greater level of data about individual and population health and physiology, for ever-increasing numbers of people. These data are frequently high-dimensional, containing many data points that are capable of being connected to both individuals and their relatives.

At one extreme, a full genome is unique to an individual and it is therefore a potentially powerful identifier. However, as we discussed in chapters 4 and 8, much more limited genomic information can also be powerfully identifying and there is a constantly evolving potential that data which currently appear to be highly technical and unrelated to an individual, may become individually identifying as technology advances. This also applies to areas of science, such as pathogen genomics, where the direct object of study is not human DNA but which involve 'incidental' processing of human genetic material or the capturing of information relating to individuals.

Whole genome sequencing, the ability to characterise the entire human genome, or the coding part of that genome (whole exome sequencing) are now being adopted at population scale for both research and healthcare purposes. This means that the data protection implications of processing such data are important for **a widening range of scientists and professionals**. In the healthcare context, mainstreaming of genome sequencing is set to increase the number of patients who undergo sequencing and expand the range of healthcare professionals who will require access to, and be responsible for, management of the resulting genetic and health data. As an increasingly wide range of healthcare and research professionals are utilising genomic techniques and data, they need to consider how this use gives rise to data protection implications and if so, how this impacts their work.

Because **data-sharing is a cornerstone of genomics** for research and the diagnosis of rare diseases, it is particularly impacted by any uncertainty or disagreements about data protection law. Our research has identified a number of areas that relate specifically to requirements for international data sharing, and to other aspects of the Regulation which are already causing significant disagreement among genomic data controllers and processors (chapter 7).

Brexit is set to further complicate and challenge data-sharing between the EU and the UK. The genomics and research communities are already pursuing ways of streamlining and clarifying data sharing rules but policymakers, regulators and legislators need to be aware of the potential costs of the legal ambiguities we have identified in our research, for healthcare and research. However, the particular sensitivity of genomics to any friction in data sharing also means that it may act as an early warning and provide an opportunity to pre-empt challenges for other sectors and categories of data processing. Moreover, addressing some of the impacts on genomic data processing identified in this research would be likely to benefit other related and valuable sectors of activity, including other areas of scientific research and parts of the developing 'personalised' or 'precision' medicine ecosystem.

There is **nothing categorically exceptional about genomics or genomic data** and, although the GDPR explicitly incorporates 'genetic data' for the first time in EU data protection law, it is welcome that it does not create a regime of restrictions and further obligations specific to genetic data. Instead, 'genetic data' are included alongside health data or biometric data as categories of data that deserve higher protection, and Member States retain the discretion to develop their own regimes.

However, the scope that the GDPR provides for Member States, national Supervisory Authorities and the EDPB to develop rules and guidance specific to genetic or genomic data means that there is the potential for the complexity of national data protection rules governing genomic data to increase. Legislators and regulators should be careful to avoid unnecessary complication by developing different national rules and guidance on genetic data instead of seeking consensus at a European level. Although genomic data are not exceptional, there are at least three features of genomic information that create particular challenges for compliance with data protection law.

One is that although they are not inherently identifiable, genome sequences and sub-sets of **genomic data are potentially highly identifying**. As discussed in this report, there is an active field of science assessing the identifiability of genomic data, and how its privacy can be preserved (chapter 8). However, this can be highly technical and it is not easy for those outside the field, including health care professionals, scientists, policymakers and regulators, to make an assessment of the identifiability of genomic data in context.

There is also a high degree of public sensitivity about genomic privacy and coverage of news stories (such as those about the 'Golden State Killer') reporting the identification of individuals, which can result in an increased focus on the protection of genomic data. Unfortunately for genomic healthcare and research, this public attention is not necessarily matched by an awareness of the implications of posting the results of consumer genetic testing, in particular for genealogy purposes, openly in online databases. Self-publication of genomic data is creating an increasingly challenging environment for de-identification of genetic data in medical and research contexts, particularly because it also enables identification of relatives by inference. There is a need for greater public awareness and education on this topic.

The **shared nature of genetic information** (99.9% of which is shared in common with others) highlighted in chapter 1 is another feature that creates challenges reconciling genomics with data protection law. This is partly due to the law's individualistic approach to data protection, which only protects as 'personal data' information which can be related to a specific individual as opposed to a group (even if that group is only two people). This means that data protection law is not equipped to protect group genetic interests. There will be moments, even in healthcare contexts, where genetic data are not sufficiently linked to a single person and therefore the data are not 'personal data'. Determining if this is the case will not be easy and it is likely in healthcare that there will usually be some other information alongside genetic data which can connect the data to at least one individual (chapter 4).

Conversely, if the same genomic data relates to multiple family members, managing this data can be complicated (chapter 6). For example, it may be that family members are entitled to access such data. In the UK, the data controller will have to determine whether another family member could be identified from the information, whether there is consent for such disclosure from those individuals, and whether it would be unreasonable or even cause serious harm to disclose this information. It will be complicated to develop automated systems for managing genomic data that can accommodate competing relatives' interests so individual decisions concerning disclosure are likely to be determined on a case by case basis for some time.

A third feature of genomic information that creates friction with data protection law is that its **utility is increasing over time**. This is the case for research as it is improved by the addition of more data and by technical and scientific advances. It is also true for healthcare as new insights are translated into new diagnoses, predictions and management of health conditions. This makes the long-term storage of genomic data and its re-use for health or research purposes increasingly desirable.

The GDPR starts from the opposite perspective, that data should be kept for no longer than necessary, that they should be used for specific purposes and that data should be minimised as far as possible, including the removal of identifying information. The Regulation does not prevent genomic data being retained, stored and used for health and research purposes but it does require genomic data controllers to justify the lawful basis for the retention and use of genomic data for longer durations (chapter 5).

Overall, we are at a **relatively early stage in the application of the GDPR** and the ambiguities and uncertainties about the correct interpretation of its provisions in the specific genomics context are perhaps to be expected. Given this, it is also unsurprising that some of our research participants were uncertain about how the GDPR and UK DPA 2018 interact with the rest of the legal framework (such as the common law duty of confidentiality) governing genomic medicine and research.

Urgency is required if we are to avoid a reduction in the processing and sharing of genomic data for health and research purposes in the UK, Europe and internationally. Some of our participants were concerned that continued challenges for international data sharing could lead to the EU being left behind in the international genomics landscape.

Our research found that participants were most uncertain about some of the fundamental principles and provisions of the GDPR, such as its territorial scope (chapter 3), material scope (when genomic data are 'personal data' – chapter 4), provisions on consent and the correct legal bases for different forms of processing (chapter 5). Some significant aspects of the GDPR, such as the status of pseudonymised data, are currently unresolved and once these are clarified the outcomes could have a positive or negative impact on genomic medicine and research.

Fundamentally, it is as yet unclear how consistent the approach to genomic data will be across the EU/EEA. The GDPR contains mechanisms which aim to promote consistency in the application of the GDPR in each Member State and the EDPB has a crucial role to play in issuing guidance which can harmonise approaches to the processing of genomic data (chapter 2). However, there is also considerable scope for divergence between Member States, particularly in relation to health data, genetic data, research processing (and the safeguards it requires) and the application of data subject rights.

However, being at an early stage in the GDPR's application also presents **an opportunity for the genomics community to advocate for standards** and rules that are appropriate to this context. As discussed in chapter 8, there are mechanisms in the form of codes of conduct and certification schemes that actively encourage specific sectors to engage with Supervisory Authorities and develop best practice for the application of the GDPR in their field.

There are significant hurdles to successfully developing codes or certification schemes but these are not insurmountable and they may be more likely to succeed if they begin by addressing discrete topics such as the de-identification and anonymisation of genomic data, as opposed to addressing compliance with all the requirements of the Regulation. It is also important that the courts, regulators and policymakers are aware of the nature of genomic healthcare and research and how it is practiced, and of the expectations of patients and participants for their genetic data.

If appropriate standards can be set in judgments, guidance and codes, the GDPR has the potential to achieve both the aim of securing a high level of protection of personal data while also promoting the processing and sharing of data for the benefit of all.

9.2 Specific conclusions

When and where does the GDPR Apply?

As discussed in chapter 3, our research has found that there is concern about the territorial scope of the GDPR, for example, to govern EEA-based university researchers who are involved in the processing of data from participants elsewhere in the world. In part this is because it is unclear how the application of the GDPR processing 'in the context of the activities of a controller or processor in the Union' (Art 3(1)) is to be interpreted in the genomics context.

For example, it is not clear if the GDPR would apply to govern overseas processing where an EEA-based collaborator is merely advising other collaborators, albeit having some influence over the means and purposes of processing. This is connected to a second uncertainty about how broad the notion of 'joint controllership' is in the genomic context. In both these aspects, the CJEU and authorities have previously taken a broad view to ensure protection of data subjects' fundamental rights and freedoms, and it is not yet clear if such a broad approach would be taken in the genomics context, and if so, the limitations on genomic data sharing that might ensue.

> It would assist the genomics community to have further guidance about the territorial application of the Regulation in this context and to obtain clarity about what activities are likely to meet the threshold of 'data controller'.

When are genetic or genomic data 'personal data'?

Determining when genetic/genomic data and associated health data are 'personal data' is a significant challenge. More precisely, uncertainty and disagreement about whether certain data are 'personal data' is challenging local, national and international flows of genomic and health data.

Our legal analysis (chapter 4) highlights a danger that the courts in England and Wales might conclude that individual genetic or genomic information which allow the singling-out or 'individuation' of an individual are 'personal data' even though there may be a very low likelihood that such information would ever be connected to an actual individual or additional information about them. This could lead to some genetic or genomic information being treated as inherently identifiable.

> It is our strong recommendation that genetic, and particularly genomic information should not be viewed in itself as inherently or directly identifying without some further link to or impact on an individual.
>
> There should be clarification of the correct approach to individuation under the Regulation in the genomic context.

We conclude that a distinction between direct and indirect identification is unhelpful in the genetic/genomic context. Because 'personal data' include information that could identify individuals indirectly, in combination with other information, it is rarely helpful to focus on the identifiability of the information held by the controller in isolation from other sources that could be available to indirectly identify an individual.

Determining when identification is 'reasonably likely' is not straightforward. Case law and guidance help to set the outer limits and demonstrate when identification is not reasonably likely. For example, WP29 have referred to 'mere hypothetical possibility',[486] and the UK courts to a 'remote' chance[487] of identification, which would be insufficient. However, there have been signs from WP29 and the CJEU that a relatively low level of risk of identification could result in data being classified as 'personal data'. What is clear is that an assessment of the risk of identification can only be made in context. In this light, we identified several factors from guidance, legal decisions and the literature, which are particularly relevant to the assessment of whether genetic/genomic and phenotypic data are reasonably likely to identify an individual. These are the 'richness' of the genetic and any associated phenotypic data, restrictions on who may access the data and the nature and availability of other information that may be used in conjunction with the genetic data to identify an individual.

> Data controllers should consider the richness of the data and how potentially identifying different categories are; whether users or third parties may be able to identify an individual; whether additional information could help to identify an individual, and; whether sufficient safeguards are in place to reduce the risk.

Technological developments and identifiability

The GDPR requires an assessment of identifiability that takes technological developments into account. As discussed in chapter 8, different views can be taken about how technological developments and the availability of additional information such as genealogy data are impacting on identification risks. This is something that the genomics community could aim to address in a code of conduct or certification scheme. Specific risks that data controllers should consider include risks of identification to forms of data that may not appear to currently relate to individuals. For example, the potential for identification using molecular phenotype data, such as gene expression data, linking independent genotype and phenotype datasets to pinpoint an individual.[488] Another developing risk associated with genomic data is due to the increasing availability of self-published genetic data on genealogy and other websites.

> Genomic data controllers should consider how external open sources such as genealogy websites impact on the identifiability of the data they hold.
>
> All stakeholders, including policymakers and regulators, should seek to communicate the consequences that posting genomic information online has for the identifiability of that data, and the downstream consequences for consumers and their relatives.
>
> Technical experts will play a key role in providing cross disciplinary expertise to assist genomics data controllers to determine when data are personal data. This could build on existing research ethics or data access committees, or through separate entities convened for this purpose.

Mitigations to reduce the risk of identification

A range of measures can be, and frequently are, adopted to reduce the risks of identification of genomic and health data. There are legal and organisational safeguards such as data access controls, contracts, and internal policies which proscribe certain behaviour and provide sanctions for breach. In terms of hard law, the criminal offence introduced in the Data Protection Act for knowingly or recklessly re-identifying information without the controller's consent is also a relevant safeguard.[489] Self-regulation such as a code of conduct could also be used to reduce this risk. There are also a range of technical approaches which can be taken to de-identify and protect genomic data. Such approaches will need to be chosen according to the nature of the data and its intended use and there is not likely to be a one-size fits all approach.

One way of reducing risks in collaborative research or data analysis is to avoid the transmission of datasets by bringing the analysis to the data. This includes target query-based research of locally held data (as is being developed at a global level in the GA4GH Beacon Project) and the use of advanced cryptographic techniques such as homomorphic encryption to ensure that the results of analysis executed on encrypted data are only visible to someone with a key.

Due to the complexity of de-identification and the need for an ongoing assessment of technological and societal threats, there is likely to be a need for intermediary services who can reduce the friction caused by data protection mechanisms for users whilst providing the usable data that is needed for healthcare or research purposes. In this regard, trusted third party processors who use homomorphic encryption and similar technologies to analyse data without confidential disclosure could form an important part of the genomic data ecosystem.

> As well as legal and organisational safeguards, such as data access agreements and internal governance policies, data controllers should consider what technical approaches could be taken to reduce risks of identification.
>
> The genomics sector could aim to develop guidance or a code of conduct on de-identification of genomic data under the GDPR, incorporating technical expertise to identify appropriate tools for different forms of genomic data processing.

Are pseudonymised data always personal data?

Our research suggests that there should not be a different standard of risk applied to pseudonymised data: pseudonymisation is primarily a safeguard for data in the Regulation, it is referred to as a process not a category of personal data and no rationale is provided to justify an exceptional standard for this form of de-identified data versus another (e.g. aggregated data). However, in the absence of certainty around the status of pseudonymised data, professionals who wish to adopt a precautionary approach could treat the data as 'personal data' unless technical and organisational safeguards make it highly unlikely for the key and the data to be recombined.

Supervisory authorities should provide urgent clarification on the standard of risk of identification that should be applied to data which have undergone pseudonymisation, and whether such data can be sufficiently de-identified to cease to be 'personal data'.

How is the category of 'genetic data' impacted by the GDPR?

The GDPR introduces a category of 'genetic data'. Data controllers should be aware that this is potentially a broad special category of data because it extends to genetic characteristics that provide information about the health or physiology of a person which are inherited or acquired over the course of their life.

Chapter 1 describes how this category can apply to a variety of substances and testing strategies, ranging from simple blood tests or analysis of other biological samples to complex genome sequencing and other 'omic technologies.

As chapter 4 demonstrates, because 'genetic data' must be 'personal data', they do not include all the information that clinicians or scientists would describe as genetic data. For these professional groups, this is a significant and important potential source of confusion.

It is also important to note that genomic and phenotypic data which are 'personal data' will be subject to a higher level of protection under the GDPR as they will almost certainly be part of the categories of 'genetic data' or 'data concerning health'.

Clarification of the definition of 'genetic data' is required to remove ambiguity about which data, resulting from what forms of analysis, may fall within its boundaries. Given the dynamic nature of these technologies, this will require interdisciplinary expertise.

How should shared, familial genetic data be managed under the GDPR?

The shared, familial dimension of genetic data creates particular challenges complying with data protection law. As analysed in chapter 4, many genetic testing results could relate to (and have implications for) multiple members of the same family.

In this scenario there may be insufficient 'other' information available to identify specific individuals, although it may be possible to single-out related individuals in a dataset (more likely in the research context). In this case data will not be 'personal data'. Alternatively, there may be further information, such as phenotypic data, which, together with the genetic data will enable the identification of one individual (as will be the case in medical records) so the data qualify as 'personal data'. A third possibility is that there is sufficient other information available to connect the same genetic test result to multiple individuals. In these circumstances the data are capable of being the 'personal data' of multiple individuals with competing rights and claims over the same data.

It will be difficult for genomic data controllers to determine which of these possibilities apply to their data in context, and data may move between these states depending on the other information available. However, it is important to assess the status of such data because it will influence if obligations under the GDPR apply and to whom they are owed (chapter 6).

For the time being, it is likely that managing competing interests will require case-by-case decisions, following the DPA 2018 provisions governing the right of access and disclosure of health data. If an access request is made for 'shared' genetic data that reveals health information, data controllers have to determine whether the person making the request could identify another family member from the information, whether there is consent for such disclosure from those individuals, and whether it would be unreasonable or even cause serious harm to disclose this information.

It is also important to recognise that although the GDPR only regulates the 'personal data' of living individuals, if genetic information from a deceased person can be connected to their living relatives, such data may need to be treated as 'personal data'. For example, if a tissue block or sample from a deceased relative is analysed to inform the treatment of a living individual, the data should be treated as 'personal data' relating to the living individual.

> Genomic data controllers need to consider whether data are 'personal data' that may identify a single individual, or, whether they allow the identification of multiple family members, depending on the context and other information available.
>
> Where there are multiple family members (or biological relatives, where the genetic relatedness does not map social relationships) who can be connected to the data, controllers will need to make case-by-case decisions about rights of access to such data and whether they may be disclosed.
>
> Going forwards, genomic professionals and policymakers, in the healthcare context in particular, should consider how to develop systems that allow multiple interests in the same genetic information to be managed and reconciled.

How does data protection law align with the wider legal framework?

For most research or healthcare professionals, data protection is likely to be considered as part of the wider information governance framework. Although this broader landscape is not directly within the scope of our research, one of the challenges repeatedly raised by some participants, was uncertainty and concern about the interaction between data protection law and the wider legal framework, in particular the common law duty of confidentiality.

As we note in chapter 2, the legality of processing within the wider legal framework is part of the assessment of the principle of legality under the GDPR, however, this is not explicitly within the remit of the data protection authorities. Because there are so many ambiguities about how data protection rules apply in the genomics context, it will take time to assess how they interact with the wider framework. However, regulators and guidance issuing authorities such as the National Data Guardian are already addressing how the frameworks interact to try and produce complementary guidance for professionals.

Specific aspects identified in this research, such as the relationship between the factors that must be considered when disclosing 'shared' genetic information under data protection law and the common law of confidentiality, require the input of regulators, statutory authorities and professional groups to provide clarity about what is required.

> Data protection authorities, regulators, statutory authorities, professional groups and other policymakers responsible for the governance of health and genetic information should work together to assess how different rules governing the same data interact, and to develop holistic guidance for professionals.

The challenge of consent as a basis for processing genetic or health data

Our research has identified that the GDPR's requirements in relation to consent are challenging for some data controllers for whom alternative legal bases are unavailable or undesirable (chapter 5).

The GDPR sets a high standard for consent and choosing consent under Art 5 or Art 9 has significant consequences for the rights and obligations that data controllers must fulfil, meaning that it is a challenging legal basis when applied to genomic data in healthcare and research.

For most uses of genetic or genomic data in healthcare and research, alternative legal bases and Art 9 conditions will be available. Indeed, the provisions for scientific research processing may significantly reduce the data controllers' burden in the UK. Despite this there are good reasons why controllers may still wish to rely on consent. One is that it is coherent with the wider ethical and regulatory framework, for example for health research, and that it is not easy to explain transparently, that consent is not a legal basis. Another is that alternatives, notably the provisions for scientific research, may not be available in other Member States, or, there may be uncertain and inconsistent safeguards and additional requirements associated with these alternatives, particularly in cross-border processing collaborations.

Despite the high standards for consent under the GDPR, our analysis has highlighted scope for the genomics community to advocate for consent standards that are appropriate to genomics activities. In particular, our work has suggested that it is not necessarily the case that there is a clear imbalance between data controllers and data subjects in the genomics context and that broader consent may be appropriate for certain areas of genomic research. Indeed, rare genetic disease patients may have more expertise in managing their disease than their health care professionals in certain circumstances suggesting that a more nuanced view should be taken of the possible imbalances of power potentially arising when relying on consent as a legal basis.

Moreover, for some areas of genomic research (e.g. for non-hypothesis driven research) it might not be possible to sufficiently specify the purposes at the outset of the research and that this requires the flexibility envisioned in recital 33 for broader consent.

> Data controllers should be aware that there are high thresholds for using consent as a legal basis under the GDPR and that, in the UK at least, there are likely to be alternate legal bases and Art 9 conditions which are more appropriate for their purposes.
>
> For data controllers who have to rely on consent or for whom consent is still the best option (for example for some cross-border genomic data processing) the genomics community should advocate for appropriate consent standards that recognise that there may not be an imbalance between controllers and data subjects, and that broad consent to processing of genomic data may be justified in certain circumstances.

Establishing a lawful basis and Art 9 condition for processing

There are a range of legal bases and Art 9 conditions that can be relied on for lawful processing of genetic and health data. Our analysis (chapter 5) has shown that it may not always be clear which is most appropriate, for example, where there are blurred boundaries between clinical, non-clinical (such as audit or service evaluation) or research activities. It is possible that the clinical safety testing of decision-support tools and apps falls within the Art 9(2)(h) condition for provision and management of healthcare systems.

Although the GDPR does not define scientific research, there is likely to be a close interaction between the requirement for ethical approval of certain activities and the appropriate Art 9 condition. The DPA 2018 requires that scientific research processing is in the 'public interest' but this is undefined in legislation. Art 89 requires that safeguards such as pseudonymisation are implemented in research and national law may mandate further technical and organisational measures for such processing.

An important safeguard in the DPA is that research must have valid ethical approval if processing is carried out for the purposes of measures or decisions with respect to a particular data subject, for example if participants are to be recontacted with new results.

> Controllers and authorities should disseminate best practice on communicating to data subjects that their consent is not the lawful basis or Art 9 condition for processing, to help comply fully with the principles of transparency and fairness.

Fulfilling data subjects' rights in genomic data processing

Our research suggests that we are at relatively early stage in terms of substantiating the requirements of some data subject rights in the genomics context. However, as our analysis in chapter 6 demonstrates, it can be complex for data subjects and data controllers to determine when rights and consequent obligations apply because they vary heavily according to the context and legal basis/Art 9 condition for processing.

We identify the potential for Art 11 to reduce data controllers' burdens when they have sufficiently de-identified data (for example through pseudonymisation) and where the controller is not in a position to identify the data subject. However, our analysis also highlights potential challenges if the opinion of WP29 is followed and controllers have to specify what information data subjects are able to provide to enable re-identification. In the context of genetics and genomics, the difficulty of determining and communicating this could outweigh the benefits to controllers of relying on Art 11. Perversely, this could act as a disincentive for minimising data as fully as possible, which in itself could constitute non-compliance with the Regulation.

Where data subject rights do apply, specific challenges arise in the genomics context. In particular, there are challenges determining how to deal with the right to access data relating to several genetic relatives (Art 15) and the right to rectification and assessing when genomic data are inaccurate (Art 16). These and other difficulties identified in our analysis of data subject rights require urgent focus from the genomics community to engage with data protection authorities, regulators and the EDPB prior to their production of specific guidance on data subject rights.

> Regulators and the EDPB should clarify the requirements of Art 11 bearing in mind the potential that setting certain standards may disincentivise the GDPR's goal of data minimisation, or, give rise to an irreconcilable clash with other GDPR obligations.

> The genomics community should consider when, and how, data subject rights will apply in their context and develop suggestions for determining when compliance with rights would involve disproportionate effort, prevent or seriously impair the achievement of the objectives of processing, or render it impossible for scientific research in particular.

The potential impact of data subject rights on genomic data processing

It is not yet clear how data subjects might seek to enforce all of the data subject rights in the context of genetics and genomics. Some rights, such as Art 13 requirements for data subjects to be informed about how their data will be used when data have been directly obtained from them, have no current exceptions. Our research highlights the need for data controllers to be aware of this distinction between Art 13 and Art 14, namely that there is no derogation for the right to be limited if data is obtained directly from data subjects even if it would require disproportionate effort or seriously impair the objectives of the processing. Participants in this research further highlighted the potential for a number of rights, particularly when used in combination, to challenge genomic medicine and genomic research. These rights are the right of access by the data subject (Art 15), the right to erasure (Art 17) and the right to data portability (Art 20).

> The genomics community, policymakers and regulators should evaluate the cumulative impact of data subject rights on genetic and genomic healthcare and research.

Impact of the GDPR on genomic data-sharing within the EU/EEA

A major impact that our research has identified is on genomic data-sharing (chapter 7). Participants in our research have reported challenges reaching agreement between collaborators about the application of the GDPR in the genomics context, particularly between international partners. Disagreement can stem from any of the ambiguities and uncertainties identified in this report, from the fundamental application of the territorial and material scope, to the choice of legal basis and what is required to meet obligations and fulfil data subject rights. Our analysis has also highlighted how regulatory divergence between Member States in relation to multiple aspects of processing might have a significant deleterious impact on genetic and health data transfers, and processing for scientific research purposes.

> The genomics community, the EDPB and SAs should seek to develop consensus in relation to key issues such as the scope of 'personal data', the appropriate legal bases for specific forms of processing and how data subject rights should be fulfilled in the genomic context.

> Although MS law is likely to vary in relation to scientific research and processing of genetic and health data, efforts should be made to reach harmonised approaches across the EU/EEA in the medium to long term.

Impact of the GDPR on data-sharing outside the EU/EEA and to international organisations.

Under the GDPR a lawful transfer of genomic data to a third country or international organisations requires a legal mechanism. One challenge is that it is currently unclear precisely what constitutes a transfer. Our analysis in chapter 7 highlights that the selection of the appropriate legal mechanism is highly contingent upon the context and legal position of the controller. There is a structured process to selecting a legal mechanism for transfer. If an adequacy decision exists, this must be used before reaching for Article 46 safeguards or, failing those, Article 49 derogations. In all cases, an adequate level of protection 'essentially equivalent' to that guaranteed within the EU must be provided. Further, data controllers should be aware that there are specific duties that apply where transfer to a third country or international organisation is envisaged.

> We recommend that data controllers explore multiple mechanisms to facilitate specific international transfers, first relying on adequacy, then Article 46 safeguards only in the absence of adequacy and Article 49 derogations only in the absence of adequacy or safeguards.
>
> Data controllers should be cognizant of the 'essentially equivalent' standard by which all legal mechanisms for transfer are assessed.

Which Article 46 safeguards or Art 49 derogations could facilitate international transfer of genomic data?

The 'legally binding and enforceable instruments between public authorities' safeguard (Article 46(2)(a)) has been given a flexible interpretation by the EDPB, representing a promising mechanism to facilitate transfer between those organisations that qualify. The safeguards on contractual clauses (Articles 46(2)(c) and (d)) also represent feasible solutions to transfer genomic data internationally, especially if a standard set of clauses could be secured for the sector as a whole. However, as demonstrated by the *Schrems II* (C-311/18) AG Opinion, these clauses are vulnerable to challenge.

The safeguards on codes of conduct and certification mechanisms (Articles 46(2)(e) and (f)) also represent sector-wide methods to facilitate international data transfer. These mechanisms, although lacking precedent and difficult to secure, could have a key role in demonstrating general compliance with the GDPR for the sector as a whole, even if they are insufficient to act as a legal mechanism for international transfer. In terms of Article 49 derogations, the EDPB warns that these may only ever act as 'exceptions' and never the rule themselves.

The derogation on consent (Article 49(1)(a)) will facilitate international data transfer under most conditions. However, consent for the processing of data under the GDPR in healthcare and research settings is set a high bar and often difficult to secure, especially in the research context. The derogation on important reasons of public interest (Article 49(1)(d)), while set at a high bar, is a promising mechanism for the genomics community, the community often being in a strong position to demonstrate this through the tangible and critical benefits from international cooperation.

> We recommend that the genomic sector should work toward sector-specific solutions such as standard contractual clauses, codes of conduct, and certification mechanisms to facilitate international transfer of genomic data.

> Caution is required where Article 49 derogations are invoked, the EDPB has specifically warned that these derogations are exceptions to the rule and never the rule in and of themselves.

Mitigations and the reduction of potential deleterious impacts

Alongside impacts and challenges, our research aimed to identify mitigations and best practice which could prevent or reduce potential deleterious impacts on genomic data processing. Throughout this report, we have suggested approaches that data controllers or processors could take to specific challenges, and have highlighted a wide range of areas that would benefit from greater deliberation among genomics data controllers, regulators, policymakers and other stakeholders. We have summarised the technical, legal and organisational mitigations that can be taken to safeguard data. The final, significant category of mitigation identified in this report are sector-led codes of conduct or certification mechanisms. These hold the promise for genomic data controllers and processors to co-develop appropriate rules and standards for aspects of genomic data processing under the GDPR. These may not be easy or quick to develop: the broader the scope of codes or certification schemes, the more they will require an understanding of the variations in Member State law relating to processing of genetic and health data and the more difficult it may be to achieve consensus. To facilitate consensus and certainty for genomic data processing under the GDPR, we recommend that the genomics sector could adopt a dual approach to codes of conduct: simultaneously pursuing a broad and sector wide code of conduct to establish and harmonise rules for genomic data as a self-regulatory code of conduct and more specific sub-sector-led codes that aim to crystallise best practice into smaller formal codes of conduct to be approved under Art 40.

> The genomics community and sub-sectors of genomics data controllers and processors should consider developing codes of conduct or certification schemes for aspects of processing which would benefit from tailoring of rules and increased legal certainty for their context.

To try and reach consensus more quickly and to establish some level of certainty for genomic data processing under the GDPR, we suggest that the genomics sector could adopt a dual approach to codes of conduct: simultaneously pursuing a broad and sector wide code of conduct to establish and harmonise rules for genomic data as a self-regulatory code of conduct and more specific sub-sector-led codes that aim to crystallise best practice into smaller formal codes of conduct to be approved under Art 40.

Whilst a suite of potential codes of conduct and certification mechanisms might be available for the sector, the breadth of possible applications and cross cutting nature of the technologies means that there will be a need for clarity about which codes/certification are most meaningful for particular applications.

There are opportunity costs involved in having too much duplication and uncertainty within the sector as to which codes/certification schemes apply, so multidisciplinary engagement, including representation from professional organisations will be crucial in determining who should be responsible for developing which codes/certification schemes.

9.3 Concluding remarks

Our research has identified a significant range of current and potential impacts of the new EU data protection regime on the processing of genomic data in healthcare and medical research. However, there is considerable scope for mitigation, clarification or the development of harmonised approaches and consensus, and throughout our report we have suggested how this may be achieved. Although there are dangers that divergent views across the sector and across borders will hinder important genomic data sharing, this report has highlighted the importance of efforts to find agreement within the research and genomics communities. The GDPR provides considerable opportunities for sector-led approaches. Moreover, addressing challenges for genomic data processing could provide a valuable exemplar for other sectors and thereby benefit other sectors and related areas of activity.

Developing appropriate standards for genomic data processing under the GDPR also has the potential to have a positive impact on genomics: reassuring data subjects or patients/participants that their rights to privacy and data protection are safeguarded at the same time as facilitating ethical and lawful uses of genomic data. By requiring transparent and accountable processing, and upholding data subject rights and related obligations, the GDPR provides some of the tools to ensure that individuals retain trust in an increasingly complex, data-driven healthcare and research environment.

There are other factors which could influence public trust that have not been directly addressed within the scope of this report. These include the impact of the GDPR on direct-to-consumer genomics data processing, or, on patient or citizen-led genomic health research initiatives. Another has been the subject of separate PHG Foundation research: the regulation and governance of AI-driven tools in healthcare and research. The impact of the GDPR on these areas is likely to be a significant influence how genomic data might be processed for healthcare and medical research in the future, and what might be perceived as an appropriate and proportionate regulatory response. It is clear that ensuring that the GDPR has the most positive possible impact on genomic medicine and research will require sustained and resourced engagement by all stakeholders, including health care professionals, researchers, policymakers, regulators, research and genomics institutions, in the UK, the EU and internationally.

References

Chapter 1

1 Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation), OJ 2016 L 119/1.
Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data, OJ 1995 L 281/31.

2 (UK) Data Protection Act 2018 (DPA 2018).

3 The Joint Initiative for Metrology in Biology. Genome in a Bottle. https://jimb.stanford.edu/giab

4 NHS. National Genomic Test Directory. https://www.england.nhs.uk/publication/national-genomic-test-directories/

5 NHS. DNA testing on the NHS to fast track diagnosis for critically ill babies and children. https://www.england.nhs.uk/2020/01/dna-testing-on-the-nhs/

6 Department of Health and Social Care. UK launches whole genome sequence alliance to map spread of coronavirus. https://www.gov.uk/government/news/uk-launches-whole-genome-sequence-alliance-to-map-spread-of-coronavirus

7 University of Birmingham. University of Birmingham joins COVID-19 genome sequencing alliance to map spread of coronavirus. https://www.birmingham.ac.uk/news/latest/2020/03/birmingham-joins-covid-19-genome-sequencing-alliance-to-map-spread-of-coronavirus.aspx

8 Human Cell Atlas. About Human Cell Atlas. https://www.humancellatlas.org/

9 Gawad C, Koh W, Quake SR. Single-cell genome sequencing: current state of the science. Nature Reviews Genetics. 2016; 17: 175-188.

10 Bell CG, Lowe R, Adams PD, et al. DNA methylation aging clocks: challenges and recommendations. Genome Biology. 2019; 20: 1-24.

Chapter 2

11 Case C-39/72 *Commission v Italy* [1973] ECR 13, para 17.

12 Charter of the Fundamental Rights of the European Union [2012] OJ C326/391, art 8(1).

13 Consolidated Version of the Treaty on the Functioning of the European Union [2012] OJ C326/01, art 16.

14 Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data 1981, ETS 108. Chart of signatures and ratifications of Treaty 108. Available from: https://www.coe.int/en/web/conventions/fulllist/-/conventions/treaty/108/signatures?p_auth=xk6DvN0315

16 Charter of the Fundamental Rights of the European Union [2012] OJ C326/391, art 8(1).

17 European Commission, 'Safeguarding Privacy in a Connected World: A European Data Protection Framework for the 21st Century' (Communication) COM (2012) 9 final, 2-5.

18 GDPR, art 2(2)(c).

19 GDPR, arts 51(1), 57(1).

20 The Information Commissioner's Office. Legislation we cover.
Available from: https://ico.org.uk/about-the-ico/what-we-do/legislation-we-cover/

21 European Data Protection Board. GDPR: Guidelines, Recommendations, Best Practices.
Available from: https://edpb.europa.eu/our-work-tools/general-guidance/gdpr-guidelines-recommendations-best-practices_en

22 European Data Protection Supervisor. About. Available from: https://edps.europa.eu/about-edps_en

23 European Data Protection Board. European Data Protection Board Rules of Procedure. 2018, art 8.

24 GDPR, recital 159.

25 GDPR, art 89(1)-(2).

26 DPA 2018, s 19(2).

27 DPA 2018, s 19(3).

28 DPA 2018, s 19(4).

Chapter 3

29 de Hert P, Czerniawski M. Expanding the European data protection scope beyond territory: Article 3 of the General Data Protection Regulation in its wider context. International Data Privacy Law. 2016; 6(3): 230-243.

30 Case C-230/14 *Weltimmo v Nemzeti Adatvédelmi és Információszabadság Hatóság* [2015] ECR I-639, para 29.

31 European Data Protection Board. Guidelines 3/2018 on the territorial scope of the GDPR (Article 3). 2018.

32 Case C-230/14 *Weltimmo v Nemzeti Adatvédelmi és Információszabadság Hatóság* [2015] ECR I-639.

33 de Hert P, Czerniawski M. Expanding the European data protection scope beyond territory: Article 3 of the General Data Protection Regulation in its wider context. International Data Privacy Law. 2016; 6(3): 230-243.
Maduro MP. Interpreting European Law: Judicial Adjudication in a Context of Constitutional Pluralism. European Journal of Legal Studies. 2008; 1(2): 137-152.

34 Case C-230/14 *Weltimmo v Nemzeti Adatvédelmi és Információszabadság Hatóság* [2015] ECR I-639, para 30.
Case C-131/12 *Google Spain and Google v AEPD and Costeja González* [2014] ECR I-317, paras 53-54.

35 European Data Protection Board. Guidelines 3/2018 on the territorial scope of the GDPR (Article 3). 2018, 6.

36 Case C-131/12 *Google Spain and Google v AEPD and Costeja González* [2014] ECR I-317, paras 56.

37 European Data Protection Board. Guidelines 3/2018 on the territorial scope of the GDPR (Article 3). 2018, 10.

38 Kuner C, Cate FH, Lynskey O, et al. If the legislature had been serious about data privacy …. International Data Privacy Law. 2019; 9(2): 75-77.

39 Article 29 Data Protection Working Party. Opinion 1/2010 on the concepts of "controller" and "processor". 2010, 4.

40 Case C-40/17 F*ashion ID v Verbraucherzentrale NRW* [2019] ECR I-629, para 66.
Case C-131/12 *Google Spain and Google v AEPD and Costeja González* [2014] ECR I-317, para 34.
Case C-210/16 *Unabhängiges Landeszentrum für Datenschutz Schleswig-Holstein v Wirtschaftsakademie Schleswig-Holstein* [2018] ECR I-388, para 28.

41 Bygrave LA, Tosoni L. Article 4(1): Personal Data. In Bygrave LA, Kuner C, Docksey C. (eds.) The EU General Data Protection Regulation (GDPR). Oxford University Press, Oxford; 2020, 103-115.

42 GDPR, Art 4(7).

43 Article 29 Data Protection Working Party. Opinion 1/2010 on the concepts of "controller" and "processor". 2010, 18.

44 GDPR, Art 26(2).

Chapter 4

45 Article 29 Data Protection Working Party. Opinion 4/2007 on the concept of personal data. 2007, 9.

46 Ibid, 10.

47 Case C-434/16 *Peter Nowak v Data Protection Commissioner* [2017] ECR I-994, [35].

48 The Information Commissioner's Office. What is the meaning of 'relates to'?. Available from: https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/what-is-personal-data/what-is-the-meaning-of-relates-to/

49 *Durant v Financial Services Authority* [2003] EWCA Civ 1746, [2003] WLUK 204.

50 Case C-434/16 *Peter Nowak v Data Protection Commissioner* [2017] ECR I-994, [33]-[39].

51 Ibid, [34].

52 Article 29 Data Protection Working Party. Opinion 4/2007 on the concept of personal data. 2007, 12.

53 Ibid, 8-9.

54 Finnegan T, Hall A. Identification and genomic data. PHG Foundation. 2017.

55 Article 29 Data Protection Working Party. Opinion 4/2007 on the concept of personal data. 2007, 14.

56 The Information Commissioner's Office. Can we identify an individual directly from the information we have? Available from: https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/what-is-personal-data/can-we-identify-an-individual-directly-from-the-information-we-have/

57 *Vidal-Hall v Google Inc* [2015] EWCA Civ 311, [2016] Q.B. 1003 [115].

58 Case C-212/13 *Rynes v Urad pro ochranu osobnich udaju* [2015] 1 W.L.R. 2607.

59 *R (Bridges) v Chief Constable of South Wales* [2019] EWHC 2341 (Admin), [2020] 1 W.L.R. 672 [125].

60 Ibid.

61 Skopek JM. Reasonable expectations of anonymity. Virginia Law Review. 2015; 101(3): 722.

62 Ibid.

63 Article 29 Data Protection Working Party. Opinion 4/2007 on the concept of personal data. 2007, 14.

64 Council of Europe. Explanatory Report to the Protocol amending the Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data. Council of Europe Treaty Series No. 223. 2018, para 18.

65 Article 29 Data Protection Working Party. Opinion 4/2007 on the concept of personal data. 2007, 10.

66 Hallinan D, Friedewald M, de Hert P. Genetic Data and the Data Protection Regulation: Anonymity, multiple subjects, sensitivity and a prohibitionary logic regarding genetic data? Computer Law Security Review. 2013; 23(4): 322.

67 Shabani M, Marelli L. Re-identifiability of genomic data and the GDPR: Assessing the re-identifiability of genomic data in light of the EU General Data Protection Regulation. EMBO Reports. 2019; 20: 2.

68 The Information Commissioner's Office. What is personal data? Available from: https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/key-definitions/what-is-personal-data/

69 Article 29 Data Protection Working Party. Opinion 4/2007 on the concept of personal data. 2007, 1-26.

70 *Department of Health v Information Commissioner* [2011] EWHC 1430 (Admin), [2011] WL 1151213.

71 Case C-582/14 *Patrick Breyer v Bundesrepublik Deutschland* [2016] ECR I-769, para 46.

72 Finck M, Pallas F. They Who Must Not Be Identified - Distinguishing Personal from Non-Personal Data Under the GDPR. SSRN. 2019.

73 *Mircom International Content Management and Consulting v Virgin Media* [2019] EWHC 1827 (Ch), [2019] 7 WLUK 245.

74 Stalla-Bourdillon S. Anonymising personal data: where do we stand now?. Privacy and Data Protection. 2019; 19(4): 3-5.

75 Shabani M, Borry P. Rules for processing genetic data for research purposes in view of the new EU General Data Protection Regulation. European Journal of Human Genetics. 2018; 26(2): 149-156.
Lin Z, Owen A.B. Altman R.B. Genomic Research and Human Subject Privacy. Science. 2004; 305(5681): 183-183.

76 Data Protection Act 1998, s.1(1).

77 Article 29 Data Protection Working Party. Opinion 4/2007 on the concept of personal data. 2007, 19-20.

78 *Common Services Agency v Scottish Information Commissioner* [2008] UKHL 47, [2008] 1 W.L.R. 1550.

79 *Department of Health v Information Commissioner* [2011] EWHC 1430 (Admin), [2011] WL 1151213, per Cranston J;
All Party Parliamentary Group on Extraordinary Rendition v Information Commissioner [2011] UKUT 153 (AAC), [128].

80 The Information Commissioner's Office. Anonymisation: managing data protection risk code of practice. 2012, 1-108.

81 Article 29 Data Protection Working Party. Opinion 06/2013 on open data and public sector information ('PSI') reuse. 2013, 1-28.

82 *University of Bristol v John Peters* [2018] EA/2018/0142 [42].

83 Article 29 Data Protection Working Party. Opinion 06/2013 on open data and public sector information ('PSI') reuse. 2013, 1-28.

84 The Information Commissioner's Office. What is personal data? Available from: https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/key-definitions/what-is-personal-data/

85 *Queen Mary University of London v (1) The Information Commissioner and (2) Alem Matthees* EA/2015/0269.

86 Erlich Y, Shor T, Pe'er I, et al. Identity inference of genomic data using long-range familial searches. Science. 2018; 362(6415): 690–694.

87 The Information Commissioner's Office. Anonymisation: managing data protection risk code of practice. 2012, 1-108.

88 Case C-582/14 *Patrick Breyer v Bundesrepublik Deutschland* [2016] ECR I-769, para 46.

89 Queen Mary University of London v (1) The Information Commissioner and (2) Alem Matthees EA/2015/0269. University of Bristol v John Peters [2018] EA/2018/0142.

90 Mourby M, Mackey E, Elliot M, et al. Are 'pseudonymised' data always personal data? Implications of the GDPR for administrative data research in the UK. Computer Law & Security Review. 2018; 34(2): 222-233.

91 GDPR, recital 28.

92 GDPR, art 4(5).

93 GDPR, recital 29.

94 Rumbold JMM, Pierscionek B. The effect of the general data protection regulation on medical research. Journal of Medical Internet Research. 2017; 19(2): 1-6.
Shabani M, Borry P. Rules for processing genetic data for research purposes in view of the new EU General Data Protection Regulation. European Journal of Human Genetics. 2018; 26(2): 149-156.
Donnelly M, McDonagh M. Health research, consent and the GDPR exemption. European Journal of Health Law. 2019; 26(2): 97-119.
Quinn P. The Anonymisation of Research Data — A Pyric Victory for Privacy that Should Not Be Pushed Too Hard by the EU Data Protection Framework? European Journal of Health Law. 2017; 24(4): 347-367.

95 The Information Commissioner's Office. Anonymisation: managing data protection risk code of practice. 2012, 1-108.

96 Phillips M. Can Genomic Data Be Anonymised? Available from: https://www.ga4gh.org/news/can-genomic-data-be-anonymised/

97 Article 29 Data Protection Working Party. Opinion 05/2014 on Anonymisation Techniques. 2014.

98 The Information Commissioner's Office. What is personal data? Available from: https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/what-is-personal-data/what-is-personal-data/#pd4

99 Mourby M, Mackey E, Elliot M, et al. Are 'pseudonymised' data always personal data? Implications of the GDPR for administrative data research in the UK. Computer Law & Security Review. 2018; 34(2): 222-223.

100 Case C-582/14 *Patrick Breyer v Bundesrepublik Deutschland* [2016] ECR I-769.

101 van Veen EB. Observational health research in Europe: understanding the General Data Protection Regulation and underlying debate. European Journal of Cancer. 2018; 104: 70-80.

102 GDPR, art 4(13).

103 Hallinan D, Friedewald M, de Hert P. Genetic Data and the Data Protection Regulation: Anonymity, multiple subjects, sensitivity and a prohibitionary logic regarding genetic data?. Computer Law & Security Review. 2013; 29(4): 317-329.

104 Council of Europe. Additional Protocol to the Convention on Human Rights and Biomedicine concerning Genetic Testing for Health Purposes (CETS No.203, 2008).

105 Hallinan D, de Hert P. Genetic Classes and Genetic Categories: Protecting Genetic Groups through Data Protection Law. In: Taylor L, Floridi L, van der Sloot B. (eds.) Group Privacy: New Challenges of Data Technologies. Berlin, Springer; 2017, 175-196.

106 Dove, ES. Collection and protection of genomic data. In: Gibbon S, Prainsack B, Hilgartner S, et al. (eds.) Routledge Handbook of Genomics, Health and Society (2nd Edition). London, Routledge; 2019, 161-168.

107 Article 29 Data Protection Working Party. Opinion 4/2007 on the concept of personal data. 2007, 12.

108 Taylor, Mark, Genetic discrimination and the draft European Union General Data Protection Regulation, in Genetic Discrimination: Transatlantic perspectives on the case for a European-level legal response, Gerard Quinn, Aisling de Paor, and Peter Blanck (eds.), Abingdon, Routledge, 2015, pp. 211-226, at p. 222.

109 Hallinan D. Feeding Biobanks With Genetic Data. Brussels, Vrije Universiteit Brussel; 2018, 299.

110 Frederik E. Ultrasensitive protein method lets scientists ID someone from a single strand of hair. Available from: https://www.sciencemag.org/news/2019/11/scientists-can-now-identify-someone-single-strand-hair?utm_campaign=news_daily_2019-11-21&et_rid=603089970&et_cid=3084988

111 GDPR, art 4(1).

112 GDPR, recital 53.

113 GDPR, art 4(15).

114 Article 29 Data Protection Working Party. Guidelines on Data Protection Impact Assessment (DPIA) and determining whether processing is "likely to result in a high risk" for the purposes of Regulation 2016/679. 2017.

115 European Data Protection Board. Opinion 22/2018 on the draft list of the competent supervisory authority of the United Kingdom regarding the processing operations subject to the requirement of a data protection impact assessment (Article 35.4 GDPR). 2018, 6-7.

116 European Parliamentary Research Service Scientific Foresight Unit. How the General Data Protection Regulation changes the rules for scientific research. STOA Options Brief. 2019.

Chapter 5

117 European Data Protection Board. Opinion 3/2019 concerning the Questions and Answers on the interplay between the Clinical Trials Regulation (CTR) and the General Data Protection Regulation (GDPR) (art. 70.1.b)). 2019, para 20.

118 Hallinan D. Broad consent under the GDPR: An optimistic perspective on a bright future. Life Sciences, Society, and Policy. 2020; 16(1): 1-18.

119 European Data Protection Board. Opinion 3/2019 concerning the Questions and Answers on the interplay between the Clinical Trials Regulation (CTR) and the General Data Protection Regulation (GDPR) (art. 70.1.b)). 2019, para 21.

120 Article 29 Data Protection Working Party. Guidelines on consent under Regulation 2016/679. 2018, 12.

121 Article 29 Data Protection Working Party. Opinion 03/2013 on purpose limitation. 2013, 15-16.

122 Ibid.

123 Article 29 Data Protection Working Party. Guidelines on consent under Regulation 2016/679. 2018, 28.

124 Ibid.

125 Ibid, 29.

126 Kaye J, Whitley EA, Lund D, et al. Dynamic Consent: A Patient Interface for Twenty-First Century Research Networks. European Journal of Human Genetics. 2015; 23: 141-146.

127 van Quathem K, de Meneses AO. Association of German Supervisory Authorities Issues Paper on Broad Consent for Research. Available from: https://www.insideprivacy.com/data-privacy/association-of-german-supervisory-authorities-issues-paper-on-broad-consent-for-research/

128 BBMRI-ERIC. BBMRI-ERIC joint comments to the Article 29 Working Party Guidelines on Consent under Regulation 2016/ 679 (wp259) and Transparency under Regulation 2016/679 (wp260). Available from: https://www.bbmri-eric.eu/wp-content/uploads/WP29_consent-joint-comments_BBMRI-ERIC_as-submitted.pdf
Kubetin R, Barnes M, Massey R, et al. New draft guidelines on GDPR consent Requirement's application to scientific research. Available from: https://news.bloomberglaw.com/pharma-and-life-sciences/new-draft-guidelines-on-gdpr-consent-requirements-application-to-scientific-research

129 Thompson B. Data protection: how medical researchers persuaded the European Parliament to compromise. Available from: http://eprints.lse.ac.uk/73315/1/blogs.lse.ac.uk-Data%20protection%20how%20medical%20researchers%20persuaded%20the%20European%20Parliament%20to%20compromise.pdf

130 Hallinan D. Broad consent under the GDPR: An optimistic perspective on a bright future. Life Sciences, Society, and Policy. 2020; 16(1): 1-18.

131 Ibid.

132 NHS Health Research Authority. Consent in research. Available from: https://www.hra.nhs.uk/planning-and-improving-research/policies-standards-legislation/data-protection-and-information-governance/gdpr-guidance/what-law-says/consent-research/

133 Article 29 Data Protection Working Party. Opinion 2/2009 on the protection of children's personal data (General Guidelines and the special case of schools). 2009.

134 Family Law Reform Act 1969 (England and Wales). Age of Legal Capacity (Scotland) Act 1991. Age of Majority Act (Northern Ireland) 1969.

135 NHS Health Research Authority. Research involving children. Available from: https://www.hra.nhs.uk/planning-and-improving-research/policies-standards-legislation/research-involving-children/
Gillick v West Norfolk and Wisbech AHA [1986] A.C. 112, [1985] 3 W.L.R. 830.

136 The Information Commissioner's Office. What is valid consent? Available from: https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/consent/what-is-valid-consent/

137 Taylor MJ, Dove ES, Laurie G, et al. When can the child speak for herself? The limits of parental consent in data protection law for health research. Medical Law Review. 2018; 26(3): 369-391.

138 The Information Commissioner's Office. What is valid consent? Available from: https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/consent/what-is-valid-consent/

139 Taylor MJ, Dove ES, Laurie G, et al. When can the child speak for herself? The limits of parental consent in data protection law for health research. Medical Law Review. 2018; 26(3): 369-391.

140 The Information Commissioner's Office. How should we obtain record and manage consent?. https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/consent/how-should-we-obtain-record-and-manage-consent/

141 Article 29 Data Protection Working Party. Guidelines on consent under Regulation 2016/679. 2018, 31.

142 Bombard Y, Brothers KB, Fitzgerald-Butt S, et al. The Responsibility to Recontact Research Participants after Reinterpretation of Genetic and Genomic Research Results. The American Journal of Human Genetics. 2019; 104(4): 578-595.

143 European Data Protection Board. Guidelines 2/2019 on the processing of personal data under Article 6(1)(b) GDPR in the context of the provision of online services to data subjects. 2019, para 37.

144 Case C-524/06 *Heinz Huber v Bundesrepublik Deutschland* [2008] ECR I-724.

145 Article 29 Data Protection Working Party. Opinion 06/2014 on legitimate interests of the data controller under Article 7 of Directive 95/46/EC. 2014, 17.

146 Ibid, 16-17.

147 European Data Protection Board. Guidelines 2/2019 on the processing of personal data under Article 6(1)(b) GDPR in the context of the provision of online services to data subjects. 2019, para 10.

148 *ABC v. St George's Healthcare NHS Trust & Ors* [2020] EWHC 455 (QB), [2020] 2 WLUK 400.

149 The Information Commissioner's Office. Vital interests. Available from: https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/lawful-basis-for-processing/vital-interests/

150 The Information Commissioner's Office. Public task. Available from: https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/lawful-basis-for-processing/public-task/

151 DPA 2018, s. 7.

152 Explanatory Notes to the Data Protection Act 2018, para 85.

153 The Information Commissioner's Office. What is the 'legitimate interests' basis? Available from: https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/legitimate-interests/what-is-the-legitimate-interests-basis/

154 Ibid.

155 Article 29 Data Protection Working Party. Guidelines on consent under Regulation 2016/679. 2018, 19.

156 Ibid, 21.

157 DPA 2018, s 10(2), Part 1 sch 1.

158 DPA 2018, s 11(1).

159 Luheshi L, Raza S, Moorthie S, et al. Pathogen Genomics Into Practice. PHG Foundation. 2015.

160 DPA 2018, s.10(2), sch 1 para 3.

161 Elizabeth Denham. Letter sent to: Sir David Sloman. 3rd July 2017. Available from: https://ico.org.uk/media/action-weve-taken/undertakings/2014353/undertaking-cover-letter-revised-04072017-to-first-person.pdf

162 The Information Commissioner's Office. Royal Free NHS Foundation Trust update, July 2019. Available from: https://ico.org.uk/about-the-ico/news-and-events/news-and-blogs/2019/07/royal-free-nhs-foundation-trust-update-july-2019/

163 Royal Free London NHS Foundation Trust. Our Work with Google Health UK. Available from: https://www.royalfree.nhs.uk/patients-visitors/how-we-use-patient-information/our-work-with-deepmind/

164 DPA 2018, s 10(2), sch 1 para 4.

165 Taylor MJ, Whitton T. Public Interest, Health Research and Data Protection Law: Establishing a Legitimate Trade-Off between Individual Control and Research Access to Health Data. Laws. 2020; 9(1): 6.

166 DPA 2018 s 19(2).

167 European Data Protection Supervisor. Preliminary Opinion on data protection and scientific research. 2020, 20.

168 Article 29 Data Protection Working Party. Opinion on some key issues of the Law Enforcement Directive (EU 2016/680). 2017, 10.

169 European Data Protection Supervisor. Guidelines 2/2018 on derogations of Article 49 under Regulation 2016/679. 2018, 13.

170 DPA 2018, s 10(3), Part 2 sch 1.

171 Proust O. Post-GDPR French Data Protection Law adopted. Available from: https://www.fieldfisher.com/en/services/privacy-security-and-information/privacy-security-and-information-law-blog/post-gdpr-french-data-protection-law-adopted

172 European Parliamentary Research Service Scientific Foresight Unit. How the General Data Protection Regulation changes the rules for scientific research. STOA Options Brief. 2019.

173 Proust O. Post-GDPR French Data Protection Law adopted. Available from: https://www.fieldfisher.com/en/services/privacy-security-and-information/privacy-security-and-information-law-blog/post-gdpr-french-data-protection-law-adopted

174 van Quathem K, Cooper D. European Commission Issues Updated Q&A on Interplay between the GDPR and the Clinical Trials Regulation. Available from: https://www.insideprivacy.com/international/european-union/european-commission-issues-updated-qa-on-interplay-between-the-gdpr-and-the-clinical-trials-regulation/

175 (Ireland) Data Protection Act 2018 (Section 36(2)) (Health Research) Regulations 2018, SI 2018/314, regulations 3(1)(e), 5(5).

176 Hallinan D. Broad consent under the GDPR: An optimistic perspective on a bright future. Life Sciences, Society, and Policy. 2020; 16(1): 1-18.

177 World Medical Association. Declaration of Taipei on Ethical Considerations Regarding Health Databases and Biobanks. Available from: https://www.wma.net/policies-post/wma-declaration-of-taipei-on-ethical-considerations-regarding-health-databases-and-biobanks/

178 Human Tissue Authority. Code of Practice A: Guiding Principles and the Fundamental Principle of Consent. 2017.

Chapter 6

179 Ausloos J, Veale M, Mahieu R. Getting Data Subject Rights Right. Journal of Intellectual Property, Information Technology and Electronic Commerce Law. 2019; 10: 283.

180 DPA 2018, sch 2 para 27(2).

181 DPA 2018, sch 2 para 27(1).

182 Hu R, Stalla-Bourdillon M, Yang M, et al. Bridging Policy, Regulation, and Practice? A Techno-Legal Analysis of Three Types of Data in the GDPR. In: Leenes R, van Brakel R, Gutwirth S, et al. (eds.) The Age of Intelligent Machines. Oxford, Hart Publishing; 2017.

183 Ausloos J, Veale M, Mahieu R. Getting Data Subject Rights Right. Journal of Intellectual Property, Information Technology and Electronic Commerce Law. 2019; 10: para 1.

184 Ibid.

185 Article 29 Data Protection Working Party. Opinion 03/2017 on processing personal data in the context of Cooperative Intelligent Transport Systems (C-ITS). 2017, 6-8.

186 Ibid.

187 Article 29 Data Protection Working Party. Guidelines on transparency under Regulation 2016/679. 2018, para 68.

188 Hu R, Stalla-Bourdillon M, Yang M, et al. Bridging Policy, Regulation, and Practice? A Techno-Legal Analysis of Three Types of Data in the GDPR. In: Leenes R, van Brakel R, Gutwirth S, et al. (eds.) The Age of Intelligent Machines. Oxford, Hart Publishing; 2017.

189 Article 29 Data Protection Working Party. Opinion 03/2017 on processing personal data in the context of Cooperative Intelligent Transport Systems (C-ITS). 2017, 6-8.

190 Article 29 Data Protection Working Party. Guidelines on transparency under Regulation 2016/679. 2018, para 20.

191 Article 29 Data Protection Working Party. Guidelines on transparency under Regulation 2016/679. 2018, para 57.

192 Ibid, para 62.

193 Elteste U, van Quathem K. German court decides on the scope of GDPR right of access. Available from: https://www.insideprivacy.com/international/european-union/german-court-decides-on-the-scope-of-gdpr-right-of-access/?_ga=2.219280411.323359227.1573063203-1840120643.1573063203

194 DPA 2018, sch 2 para 16(1).

195 Ibid, sch 2 para 16(2)(a).

196 Ibid, sch 2 para 16(2)(b).

197 DPA 2018, sch 2 para 16(4)(b)(ii).

198 Data Protection Act 1998, s 7(4).

199 DPA 2018, sch 3 para 2(2).

200 Ibid, para 5.

201 Ibid, sch 3, para 2(2).

202 Ibid para 6.

203 DPA 2018, sch 2 para 27 (1), (3).

204 Deignan JL, Chung WK, Kearney HM, et al. Points to Consider in the Revaluation and Reanalysis of Genomic Test Results: A Statement of the American College of Medical Genetic and Genomics (ACMG). Genetics in Medicine. 2019; 21: 1267.

205 Bombard Y, Brothers KB, Fitzgerald-Butt S, et al. The Responsibility to Recontact Research Participants after Reinterpretation of Genetic and Genomic Research Results. The American Journal of Human Genetics. 2019; 104(4): 578-595.

206 Carrieri D, Howard HC, Benjamin C, et al. Recontacting patients in clinical genetics services: Recommendations of the European Society of Human Genetics. European Journal of Human Genetics. 2019; 27: 169-182.

207 The Information Commissioner's Office. Right to rectification. Available from: https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/individual-rights/right-to-rectification/

208 DPA 2018, sch 2, para 27(1).

209 GDPR, art 17(3)(c).

210 GDPR, art 17(3)(d).

211 GDPR, art 17(1)(b).

212 GDPR, art 17(2).

213 Taylor MJ, Wallace SE, Prictor M. United Kingdom: transfers of genomic data to third countries. Human Genetics. 2018; 137(8): 637-645.

214 Hackl E. GDPR: Decision of the DPA on the erasure of personal data. Available from: https://www.lexology.com/library/detail.aspx?g=e776689f-84f6-466b-aad1-f30294d55a16

215 Ausloos J, Mahieu R, Veale, M. Getting Data Subject Rights Right. LawArXiv. 2019: 19-21.

216 Ausloos J, Veale M, Mahieu R. Getting Data Subject Rights Right. Journal of Intellectual Property, Information Technology and Electronic Commerce Law. 2019; 10: para 73.

217 Article 29 Data Protection Working Party. Guidelines on the right to data portability. 2016, 8.

218 Taylor MJ, Wallace SE, Prictor M. United Kingdom: transfers of genomic data to third countries. Human Genetics. 2018; 137(8): 637-645.

219 GDPR, art 21(1).

220 Data Protection Working Party. Guidelines on Data Protection Impact Assessment (DPIA) and determining whether processing is "likely to result in a high risk" for the purposes of Regulation 2016/679. 2017.

Chapter 7

221 Bertier G, Cambon-Thomsen A, Joly Y. Is It Research or Is It Clinical? Revisiting an Old Frontier through the Lens of next-Generation Sequencing Technologies. European Journal of Medical Genetics. 2018; 61(10): 634-641.

222 Molnár-Gábor F, Korbel JO. Genomic data sharing in Europe is Stumbling — Could a code of conduct prevent its fall? EMBO Molecular Medicine. 2020; 12: 1-7.

223 UK Biobank. UK Biobank leads the way in genetic research. Available from: https://www.ukbiobank.ac.uk/2019/09/uk-biobank-leads-the-way-in-genetics-research-to-tackle-chronic-diseases/

224 Birney E. The Convergence of Research and Clinical Genomics. The American Journal of Human Genetics. 2019; 104(5): 781-783.

225 European Commission. European '1+ Million Genomes' Initiative. Available from: https://ec.europa.eu/digital-single-market/en/european-1-million-genomes-initiative

226 Birney E, Vamathevan J, Goodhand P. Genomics in healthcare: GA4GH looks to 2022. BioRxiv. 2017; 4.

227 Molnár-Gábor F, Lueck R, Yakneen S, et al. Computing patient data in the cloud: Practical and legal considerations for genetics and genomics research in Europe and internationally. Genome Medicine. 2017; 9(1): 1-12.

228 Molnár-Gábor F, Korbel JO. Genomic data sharing in Europe is Stumbling — Could a code of conduct prevent its fall? EMBO Molecular Medicine. 2020; 12: 1-7.

229 (Ireland) Data Protection Act 2018 (Section 36(2)) (Health Research) Regulations 2018, SI 2018/314, regulations 3(1)(e), 5(5).

230 Molnár-Gábor F, Korbel JO. Genomic data sharing in Europe is Stumbling — Could a code of conduct prevent its fall? EMBO Molecular Medicine. 2020; 12: 1-7.

231 Proust O. Post-GDPR French Data Protection Law adopted. Available from: https://www.fieldfisher.com/en/services/privacy-security-and-information/privacy-security-and-information-law-blog/post-gdpr-french-data-protection-law-adopted

232 European Parliamentary Research Service Scientific Foresight Unit. How the General Data Protection Regulation changes the rules for scientific research. STOA Options Brief. 2019.

233 Kuner C. Article 44 General principles for transfers. In: Kuner C, Bygrave LA, Docksey C. (eds.) The EU General Data Protection Regulation (GDPR): A Commentary. Oxford, Oxford University Press; 2020: 762.

234 GDPR, art 4(23).

235 Kuner C. Article 44 General principles for transfers. In: Kuner C, Bygrave LA, Docksey C. (eds.) The EU General Data Protection Regulation (GDPR): A Commentary. Oxford, Oxford University Press; 2020: 762.

236 Case C-362/14 *Maximillian Schrems v Data Protection Commissioner (Schrems I)* [2015] ECR I-650, para 45.

237 GDPR, arts 13(1)(f), 14(1)(f), 13(1)(e), 14(1)(e).

238 Case C-434/16 *Peter Nowak v Data Protection Commissioner* [2017] ECR I-994, para 35. Ausloos J, Mahieu R, Veale M. Getting Data Subject Rights Right: A Submission to the European Data Protection Board from Data Protection Academics. 2019, paras 31-35.

239 Case C-101/01 Bodil Lindqvist [2003] ECR I-12992, para 71.

240 Ibid, para 69.

241 Ibid, para 57.

242 Ibid, para 55.

243 The Information Commissioner's Office. International transfers. Available from: https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/international-transfers/

244 Case C-362/14 *Maximillian Schrems v Data Protection Commissioner (Schrems I)* [2015] ECR I-650.

Case C-131/12 *Google Spain and Google v AEPD and Costeja González* [2014] ECR I-317.

Case C-210/16 *Unabhängiges Landeszentrum für Datenschutz Schleswig-Holstein v Wirtschaftsakademie Schleswig-Holstein* [2018] ECR I-388.

245 Case C-362/14 *Maximillian Schrems v Data Protection Commissioner (Schrems I)* [2015] ECR I-650, para 38.

246 Du YYS. The impact of the GDPR and China's data protection regime towards Chinese cloud service providers with regards to cross-border data transfers. Tilburg University LLM. 2018, 30-44.

Case C-362/14 *Maximillian Schrems v Data Protection Commissioner (Schrems I)* [2015] ECR I-650.

247 Kuner C. Reality and Illusion in EU Data Transfer Regulation Post Schrems. German Law Journal. 2017; 18(4): 893.

248 Kuner C. Article 44 General principles for transfers. In: Kuner C, Bygrave LA, Docksey C. (eds.) The EU General Data Protection Regulation (GDPR): A Commentary. Oxford, Oxford University Press; 2020: 757.

249 GDPR, arts 13(1)(f), 14(1)(f), 15(1)(c), 15(1)(c), 15(2).

250 GDPR, arts 30(1)(d), 30(2)(c).

251 GDPR, art 28(3).

252 GDPR, art 28(3)(a).

253 Case C-362/14 *Maximillian Schrems v Data Protection Commissioner (Schrems I)* [2015] ECR I-650, para 38.

254 European Union Agency for Fundamental Rights, Council of Europe. Handbook on European data protection law (2018 edition). 2018, 28.

255 Consolidated Version of the Treaty on the Functioning of the European Union [2012] OJ C326/01, art 16.

256 Case C-362/14 *Maximillian Schrems v Data Protection Commissioner (Schrems I)* [2015] ECR I-650, para 73.

257 Ibid.

258 Kuner C. Reality and Illusion in EU Data Transfer Regulation Post Schrems. German Law Journal. 2017; 18(4): 899-900.

259 Joined Cases C-293/12 and C- 594/12 *Digital Rights Ireland v Minister for Communications, Marine and Natural Resources* [2014] ECR I-238, para 48.
Case C-362/14 *Maximillian Schrems v Data Protection Commissioner (Schrems I)* [2015] ECR I-650, para 73.

260 Ibid, para 74.

261 Ibid, para 71.

262 Ibid, para 81.

263 Kuner C. Reality and Illusion in EU Data Transfer Regulation Post Schrems. German Law Journal. 2017; 18(4): 899-900.

264 Case C-362/14 *Maximillian Schrems v Data Protection Commissioner (Schrems I)* [2015] ECR I-650, para 45.

265 GDPR, art 45(2).

266 Kuner C. International data transfers, standard contractual clauses, and the Privacy Shield: the AG Opinion in Schrems II. Available from: https://europeanlawblog.eu/2020/01/07/international-data-transfers-standard-contractual-clauses-and-the-privacy-shield-the-ag-opinion-in-schrems-ii/

267 European Molecular Biology Laboratory. Agreement establishing the European Molecular Biology Laboratory (2012) art XI. Shaw MN. International Law. Cambridge, Cambridge University Press; 2018, 1000.

268 Headquarters Agreement between the Government of the Federal Republic of Germany and the European Molecular Biology Laboratory (1974).
Shaw MN. International Law. Cambridge, Cambridge University Press; 2018, 1007-1016.

269 Headquarters Agreement between the Government of the Federal Republic of Germany and the European Molecular Biology Laboratory (1974), art 6.

270 Kuner C. Article 44 General principles for transfers. In: Kuner C, Bygrave LA, Docksey C. (eds.) The EU General Data Protection Regulation (GDPR): A Commentary. Oxford, Oxford University Press; 2020: 764.

271 Ibid.

272 Ibid.

273 Ibid. Case C-505/19 *WS v Federal Republic of Germany* [2019] OJ C 357.

274 European Molecular Biology Laboratory. Internal Policy No. 68 on General Data Protection (68, 2018).

275 Case C-362/14 *Maximillian Schrems v Data Protection Commissioner (Schrems I)* [2015] ECR I-650.

276 European Molecular Biology Laboratory. Internal Policy No. 68 on General Data Protection (68, 2018), arts 22-23.

277 Ibid, 23(2)(d).

278 Case C-362/14 *Maximillian Schrems v Data Protection Commissioner (Schrems I)* [2015] ECR I-650.

279 European Molecular Biology Laboratory. Internal Policy No. 68 on General Data Protection (68, 2018), recital 5.

280 GDPR, art 45. Kuner C. Article 45 Transfers on the basis of an adequacy decision. In: In: Kuner C, Bygrave LA, Docksey C. (eds.) The EU General Data Protection Regulation (GDPR): A Commentary. Oxford, Oxford University Press; 2020: 786.

281 Article 29 Data Protection Working Party. Opinion 05/2015 on C-SIG Code of Conduct on Cloud Computing. 2015, 17.

282 Ibid.

283 Commission Implementing Decision (EU) 2016/1250 on the adequacy of the protection provided by the EU-U.S. Privacy Shield [2016] L 207/1, annex II.

284 Kuner C. Article 45 Transfers on the basis of an adequacy decision. In: In: Kuner C, Bygrave LA, Docksey C. (eds.) The EU General Data Protection Regulation (GDPR): A Commentary. Oxford, Oxford University Press; 2020: 788.

285 Article 29 Data Protection Working Party. Adequacy Referential. 2018, 5-6.

286 GDPR, art 70(1)(s).

287 GDPR, art 45(3).

288 Case C-362/14 *Maximillian Schrems v Data Protection Commissioner (Schrems I)* [2015] ECR I-650.

289 European Union (Withdrawal) Act 2018. The Data Protection, Privacy and Electronic Communications (Amendments etc) (EU Exit) Regulations 2019, SI 2019/419, sch 1.

290 McCorkindale C. Brexit and Human Rights. Edinburgh Law Review. 2018; 22(1): 126-132.

291 Joined Cases C-203/15 and C-698/15 *Tele2 Sverige v Post-och telestyrelsen and Secretary for the Home Department* [2016] ECR I-970.

292 GDPR, arts 25, 5(1)(b).

293 European Data Protection Board. Guidelines 2/2020 on articles 46(2)(a) and 46(3)(b) of Regulation 2016/679 for transfers of personal data between EEA and non-EEA public authorities and bodies. 2020, 5.

294 Ibid.

295 Ibid.

296 DPA 2018, s 7(4).

297 Genomics England Limited. Genomics England. Available from: https://www.genomicsengland.co.uk/

298 European Data Protection Board. Guidelines 2/2020 on articles 46(2)(a) and 46(3)(b) of Regulation 2016/679 for transfers of personal data between EEA and non-EEA public authorities and bodies. 2020, 13.

299 Ibid.

300 Ibid.

301 Ibid, 13-14.

302 Ibid, 14.

303 Ibid, 13.

304 GDPR, art 46(3).

305 GDPR, art 47(1)(a).

306 Case C-205/03 *P Federación Española de Empresas de Tecnología Sanitaria (FENIN) v Commission* [2006] ECR II-453, para 25.

307 Case C-41/90 *Klaus Höfner and Fritz Elser v Macrotron* [1991] ECR I-01979, para 21.

308 GDPR, art 47(1)(a).

309 GDPR, arts 47(2)(c), (i), (j), (l), (m).

310 Article 29 Data Protection Working Party. Working Document Setting Forth a Co-Operation Procedure for the approval of "binding corporate rules" for controller and processors under the GDPR. 2018.

311 The Information Commissioner's Office. Binding Corporate Rules. Available from: https://ico.org.uk/for-organisations/binding-corporate-rules/

312 Case C-362/14 *Maximillian Schrems v Data Protection Commissioner (Schrems I)* [2015] ECR I-650.

313 van Eecke P, Umhoefer C, Haie AG. EU: Binding Corporate Rules are Generating Greater Interest. Available from: https://blogs.dlapiper.com/privacymatters/eu-binding-corporate-rules-are-generating-greater-interest/

314 The Information Commissioner's Office. Binding Corporate Rules. Available from: https://ico.org.uk/for-organisations/binding-corporate-rules/

315 Molnár-Gábor F, Korbel JO. Genomic data sharing in Europe is Stumbling — Could a code of conduct prevent its fall? EMBO Molecular Medicine. 2020; 12: 1-7.

316 GDPR, art 46(2)(c).

317 Commission Decision (EU) 2001/467/EC on standard contractual clauses for the transfer of personal data to third countries, under Directive 95/46/EC [2001] L 181/19.
Commission Decision (EU) 2004/915/EC as regards the introduction of an alternative set of standard contractual clauses for the transfer of personal data to third countries [2004] L 385/74.
Commission Decision (EU) 2010/87/EU on standard contractual clauses for the transfer of personal data to processors established in third countries under Directive 95/46/EC [2010] L 39/5.

318 The Information Commissioner's Office. International transfers. Available from: https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/international-transfers/ GDPR, recital 109.

319 The Information Commissioner's Office. International transfers. Available from: https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/international-transfers/

320 Datatilsynet. Standard Contractual Clauses for the purposes of Article 28(3) of Regulation 2016/679. 2020.

321 European Data Protection Board. Opinion 14/2019 on the draft Standard Contractual Clauses submitted by the DK SA (Article 28(8) GDPR). 2019.

322 Case C-311/18 *Data Protection Commissioner v Facebook Ireland (Schrems II)* [2019] ECR I-1145, Opinion of AG Saugmandsgaard Øe.

323 Ibid, para 120.

324 Kuner C. International data transfers, standard contractual clauses, and the Privacy Shield: the AG Opinion in Schrems II. Available from: https://europeanlawblog.eu/2020/01/07/international-data-transfers-standard-contractual-clauses-and-the-privacy-shield-the-ag-opinion-in-schrems-ii/

325 Kuner C. International data transfers, standard contractual clauses, and the Privacy Shield: the AG Opinion in Schrems II. Available from: https://europeanlawblog.eu/2020/01/07/international-data-transfers-standard-contractual-clauses-and-the-privacy-shield-the-ag-opinion-in-schrems-ii/

326 European Data Protection Board. Guidelines 2/2020 on articles 46(2)(a) and 46(3)(b) of Regulation 2016/679 for transfers of personal data between EEA and non-EEA public authorities and bodies. 2020, 47.

327 Ibid, 13-14.

328 Article 29 Data Protection Working Party. Opinion 05/2015 on C-SIG Code of Conduct on Cloud Computing. 2015. European Data Protection Board. Guidelines 1/2019 on Codes of Conduct and Monitoring Bodies under Regulation 2016/679. 2019, 5.

329 Case C-311/18 *Data Protection Commissioner v Facebook Ireland (Schrems II)* [2019] ECR I-1145, Opinion of AG Saugmandsgaard Øe.

Kuner C. International data transfers, standard contractual clauses, and the Privacy Shield: the AG Opinion in Schrems II. Available from: https://europeanlawblog.eu/2020/01/07/international-data-transfers-standard-contractual-clauses-and-the-privacy-shield-the-ag-opinion-in-schrems-ii/

330 GDPR, art 49(1).

331 GDPR, art 49(1).

332 European Data Protection Board. Guidelines 2/2018 on derogations of Article 49 under Regulation 2016/679. 2018, 4.
Case C-73/07 *Tietosuojavaltuutettu v Satakunnan Markkinapörssi Oy and Satamedia Oy* [2008] ECR I-09831, para 56,

333 GDPR, recital 111. European Data Protection Board. Guidelines 2/2018 on derogations of Article 49 under Regulation 2016/679. 2018, 4.

334 Ibid.

335 Ibid.

336 Ibid.

337 Ibid, 4-5.

338 Ibid, 5.

339 Ibid.

340 European Data Protection Board. Guidelines 2/2018 on derogations of Article 49 under Regulation 2016/679. 2018, 5.

341 Ibid, 3.

342 Ibid.

343 GDPR, art 49(1)(a).

344 European Data Protection Board. Guidelines 2/2018 on derogations of Article 49 under Regulation 2016/679. 2018, 6. Article 29 Data Protection Working Party. Guidelines on consent under Regulation 2016/679. 2018, 18.

345 European Data Protection Board. Guidelines 2/2018 on derogations of Article 49 under Regulation 2016/679. 2018, 6.

346 Ibid.

347 Ibid. Article 29 Data Protection Working Party. Guidelines on consent under Regulation 2016/679. 2018, 18.

348 Ibid.

349 Ibid, 18-19.

350 GDPR, art 49(1)(a).

351 GDPR, art 4(11).

352 European Data Protection Board. Guidelines 2/2018 on derogations of Article 49 under Regulation 2016/679. 2018, 7.

353 Ibid.

354 Ibid.

355 Ibid.

356 Ibid.

357 GDPR, recital 159.

358 Hallinan D. Broad consent under the GDPR: An optimistic perspective on a bright future. Life Sciences, Society, and Policy. 2020; 16(1): 1-18.

359 GDPR, art 89(2). van Quathem K, de Meneses AO. Association of German Supervisory Authorities Issues Paper on Broad Consent for Research. Available from: https://www.insideprivacy.com/data-privacy/association-of-german-supervisory-authorities-issues-paper-on-broad-consent-for-research/

360 NHS Health Research Authority. Consent in research. Available from: https://www.hra.nhs.uk/planning-and-improving-research/policies-standards-legislation/data-protection-and-information-governance/gdpr-guidance/what-law-says/consent-research/ Information Governance Alliance. The General Data Protection Regulation: Guidance on Consent. 2018.

361 GDPR, art 7(4). Article 29 Data Protection Working Party. Guidelines on consent under Regulation 2016/679. 2018.

362 Information Governance Alliance. The General Data Protection Regulation: Guidance on Consent. 2018.

363 Freedom of Information Act 2000, s 7.

364 European Data Protection Board. Guidelines 2/2018 on derogations of Article 49 under Regulation 2016/679. 2018, 8-10.

365 The Information Commissioner's Office. International transfers. Available from: https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/international-transfers/

366 European Data Protection Board. Guidelines 2/2018 on derogations of Article 49 under Regulation 2016/679. 2018, 8-10.

367 European Data Protection Board. Guidelines 2/2019 on the processing of personal data under Article 6(1)(b) GDPR in the context of the provision of online services to data subjects. 2019, 8-9.

368 Case C-524/06 *Heinz Huber v Bundesrepublik Deutschland* [2008] ECR I-724, para 52.

369 European Data Protection Board. Guidelines 2/2019 on the processing of personal data under Article 6(1)(b) GDPR in the context of the provision of online services to data subjects. 2019, 7.

370 Article 29 Data Protection Working Party. Opinion 06/2014 on legitimate interests of the data controller under Article 7 of Directive 95/46/EC. 2014, 17.

371 European Data Protection Board. Guidelines 2/2019 on the processing of personal data under Article 6(1)(b) GDPR in the context of the provision of online services to data subjects. 2019, 7-8.

Joined Cases C-92/09 and C-93/09 *Volker und Markus Schecke and Hartmut Eifert v Land Hessen* [2010] ECR I-662.
European Data Protection Supervisor. Assessing the necessity of measures that limit the fundamental right to the protection of personal data: A Toolkit. 2017, 5.

372 European Data Protection Board. Guidelines 2/2019 on the processing of personal data under Article 6(1)(b) GDPR in the context of the provision of online services to data subjects. 2019, 7.

373 Ibid, 8.

374 Article 29 Data Protection Working Party. Opinion 06/2014 on legitimate interests of the data controller under Article 7 of Directive 95/46/EC. 2014, 16-17.

375 European Data Protection Board. Guidelines 2/2019 on the processing of personal data under Article 6(1)(b) GDPR in the context of the provision of online services to data subjects. 2019, 10.

376 Ibid.

377 GDPR, art 49(3).

378 European Data Protection Board. Guidelines 2/2019 on the processing of personal data under Article 6(1)(b) GDPR in the context of the provision of online services to data subjects. 2019.

379 European Data Protection Supervisor. Guidelines 2/2018 on derogations of Article 49 under Regulation 2016/679. 2018, 10.

380 Article 29 Data Protection Working Party. Opinion 10/2006 on the processing of personal data by the Society for Worldwide Interbank Financial Telecommunication (SWIFT). 2006, 25.

381 European Data Protection Supervisor. Guidelines 2/2018 on derogations of Article 49 under Regulation 2016/679. 2018, 11.

382 Ibid.

383 Ibid.

384 Ibid.

385 Ibid, 13.

386 Ibid.

387 Ibid, 13.

388 Ibid.

389 Ibid.

390 Ibid.

391 The Information Commissioner's Office. Vital interests. Available from: https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/lawful-basis-for-processing/vital-interests/

392 European Data Protection Supervisor. Guidelines 2/2018 on derogations of Article 49 under Regulation 2016/679. 2018, 13.

393 European Data Protection Supervisor. Guidelines 2/2018 on derogations of Article 49 under Regulation 2016/679. 2018, 13.

394 Ibid.

395 GDPR, art 49(1)(g).

396 European Data Protection Supervisor. Guidelines 2/2018 on derogations of Article 49 under Regulation 2016/679. 2018, 13-14.

397 Ibid, 14.

398 Ibid.

399 GDPR, art 49(2). European Data Protection Supervisor. Guidelines 2/2018 on derogations of Article 49 under Regulation 2016/679. 2018, 14.

400 GDPR, recital 111.

401 European Data Protection Supervisor. Guidelines 2/2018 on derogations of Article 49 under Regulation 2016/679. 2018, 4.

402 Ibid, 14-15.

403 Ibid, 15.

Chapter 8

404 Erlich Y, Narayanan A. Routes for breaching and protecting genetic privacy. Nature Reviews Genetics. 2014; 15(6): 409-421.

405 Sweeney L. Weaving Technology and Policy Together to Maintain Confidentiality. Journal of Law, Medicine & Ethics. 1997; 25(2-3): 98-110.

406 Erlich Y, Shor T, Pe'er I, et al. Identity inference of genomic data using long-range familial searches. Science. 2018; 362(6415): 690-694.

407 Lippert C, Sabatini R, Maher MC, et al. Identification of individuals by trait prediction using whole-genome sequencing data. Proceedings of the National Academy of Sciences of the United States of America. 2017; 114(38): 10166-10171.

408 Erlich Y. Major flaws in "Identification of individuals by trait prediction using whole-genome sequencing data." BioRxiv. 2017.

409 Nyholt DR, Chang-En Y, Visscher PM. On Jim Watson's APOE status: genetic information is hard to hide. European Journal of Human Genetics. 2009; 17: 147-149.

410 Humbert M, Ayday E, Hubaux JP, et al. Addressing the concerns of the Lacks family: Quantification of kin genomic privacy. Proceedings of the 2013 ACM SIGSAC conference on computer & communications security. 2013; 1141-1152.

411 Article 29 Data Protection Working Party. Opinion 05/2014 on Anonymisation Techniques. 2014.

412 Finnegan T, Hall A. Identification and genomic data. PHG Foundation. 2017.

413 Dwork C, McSherry F, Nissim K, et al. Calibrating Noise to Sensitivity in Private Data Analysis. In: Halevi S, Rabin T. (eds.) Theory of Cryptography. Third Theory of Cryptography Conference. New York, Springer; 2006: 265-284.

414 Budin-Ljøsne I, Burton P, Isaeva J, et al. DataSHIELD: An Ethically Robust Solution to Multiple-Site Individual-Level Data Analysis. Public Health Genomics. 2015; 18(2): 87-96.

415 Wallace SE, Gaye A, Shoush O, et al. Protecting Personal Data in Epidemiological Research: DataSHIELD and UK Law. Public Health Genomics. 2014; 17(3): 149-157.

416 Ibid.

417 Beacon Project, Global Alliance for Genomics and Health. Beacon. Available from: https://github.com/ga4gh-beacon/

418 Matchmaker Exchange. Matchmaker Exchange. Available from: https://www.matchmakerexchange.org/

419 Fiume M, Cupak M, Keenan S, et al. Federated discovery and sharing of genomic data using Beacons. Nature Biotechnology. 2019; 37(3): 220-224.

420 Zhou TP, Li NB, Yang XY, et al. Secure Testing for Genetic Diseases on Encrypted Genomes with Homomorphic Encryption Scheme. Security and Communication Networks. 2018; 1-13.

421 Raisaro JL, Choi G, Pradervand S, et al. Protecting Privacy and Security of Genomic Data in i2b2. IEEE/ACM Transactions on Computational Biology and Bioinformatics. 2017; 1-17.

422 Shringarpure SS, Bustamante CD. Privacy risks from genomic data-sharing beacons. American Journal of Human Genetics. 2015; 97(5): 631-646.

423 Ibid.

424 Ohm P. Broken Promises of Privacy: Responding to the Surprising Failure of Anonymization. UCLA Law Review. 2010; 57: 1701-1777.

425 Narayanan A, Felten EW. No silver bullet: De-identification still doesn't work. Available from: https://www.cs.princeton.edu/~arvindn/publications/no-silver-bullet-de-identification.pdf

426 de Montjoye Y-A, Pentland AS. Response to Comment on "Unique in the shopping mall: On the reidentifiability of credit card metadata". Science. 2016; 351(6279): 1274.

427 Cavoukian A, Castro D. Big Data and Innovation, Setting the Record Straight: De-identification Does Work. Information and Privacy Commissioner Ontario, Canada. 2014; 18.

428 El Emam K, Jonker E, Arbuckle L, et al. A Systematic Review of Re-Identification Attacks on Health Data. PLoS One. 2011; 6(12): 1-12.

429 Sanchez D, Martinez S, Domingo-Ferrer J. Comment on "unique in the shopping mall: On the reidentifiability of credit card metadata." Science. 2016;351(6279):1274.

430 The Test Directory contains details of the tests that are available in the NHS from 1st April 2020, see: NHS. National Genomic Test Directory. Available from: https://www.england.nhs.uk/publication/national-genomic-test-directories/

431 de Montjoye YA, Hidalgo CA, Verleysen M, et al. Unique in the Crowd: The privacy bounds of human mobility. Scientific Reports. 2013; 3: 1-5.

Harmanci A, Gerstein M. Quantification of private information leakage from phenotype-genotype data: Linking attacks. Nature Methods. 2016; 13: 251–256.

432 Thaine P. Perfectly Privacy-Preserving AI: What is it and how do we achieve it?. Available from: https://towardsdatascience.com/perfectly-privacy-preserving-ai-c14698f322f5

433 European Commission. White Paper: On Artificial Intelligence – A European approach to excellence and trust (COM(2020) 65 final), 11.

434 Elliot M, O'Hara K, Raab C, O'Keefe CM, Mackey E, Dibben C, et al. Functional anonymisation: Personal data and the data environment. Computer Law & Security Review. 2018; 34(2): 204-221.

435 Harmanci A, Gerstein M. Quantification of private information leakage from phenotype-genotype data: Linking attacks. Nature Methods. 2016; 13: 251–256.

436 Fiume M, Cupak M, Keenan S, et al. Federated discovery and sharing of genomic data using Beacons. Nature Biotechnology. 2019; 37: 220-224.

437 Ibid.

438 Elliot M, O'Hara K, Raab C, O'Keefe CM, Mackey E, Dibben C, et al. Functional anonymisation: Personal data and the data environment. Computer Law & Security Review. 2018; 34(2): 204-221.

439 Ibid.

440 DPA 2018 s 171.

441 Elliot M, O'Hara K, Raab C, O'Keefe CM, Mackey E, Dibben C, et al. Functional anonymisation: Personal data and the data environment. Computer Law & Security Review. 2018; 34(2): 204-221.

442 Phillips M, Molnár-Gábor F, Korbel JO, et al. Genomics: data sharing needs an international code of conduct. Nature. 2020; 578; 31-33.

443 European Data Protection Board. Guidelines 1/2019 on Codes of Conduct and Monitoring Bodies under Regulation 2016/679. 2019, 9.

444 European Data Protection Board. Guidelines 1/2019 on Codes of Conduct and Monitoring Bodies under Regulation 2016/679. 2019, paras 14-15.

445 Ibid, para 11.

446 Ibid, para 13.

447 Ibid, para 12

448 Ibid, 7.

449 Ibid, para 23.

450 Ibid, para 32.

451 Ibid, 10.

452 Ibid, para 28.

453 Ibid, para 33.

454 Ibid, para 41.

455 Ibid, para 81.

456 Ibid, para 64.

457 Ibid, para 66.

458 Ibid, para 41.

459 Ibid, 27.

460 Ibid, para 50.

461 Michal Koščík & Matěj Myška (2018) Data protection and codes of conduct in collaborative research. International Review of Law, Computers & Technology. 2018; 32(1): 141-154.

462 Bond R, Dittel A. Data protection codes of conduct hitting the fast lane under GDPR. PDP Journals. 2017; 6(3).

463 Ibid.

464 Article 29 Data Protection Working Party. Opinion 4/2010 on the European code of conduct of FEDMA for the use of personal data in direct marketing. 2010.

465 EU CLOUD COC. EU Data Protection Code of Conduct for Cloud Service Providers (Version 2.2). 2019. Available from: https://eucoc.cloud/en/about/about-eu-cloud-coc.html

466 Litton JE. We must urgently clarify data-sharing rules. Nature. 2017; 541: 437.

467 Ibid.

468 Code of Conduct for Health Research. Learn More. Available from: http://code-of-conduct-for-health-research.eu/faq

469 Ibid.

470 Townend D. Conclusion: harmonisation in genomic and health data sharing for research: an impossible dream? Human Genetics. 2018; 137: 657-664.

471 Phillips M, Molnár-Gábor F, Korbel JO, et al. Genomics: data sharing needs an international code of conduct. Nature. 2020; 578; 31-33.

472 Ibid.

473 Molnár-Gábor F, Korbel JO. Genomic data sharing in Europe is Stumbling — Could a code of conduct prevent its fall? EMBO Molecular Medicine. 2020; 12: 1-7.

474 Ibid.

475 Ibid.

476 European Data Protection Board. Guidelines 1/2018 on certification and identifying certification criteria in accordance with Articles 42 and 43 of the Regulation. 2019.

477 Ibid, para 51.

478 The Information Commissioner's Office. Certification. Available from: https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/accountability-and-governance/certification/

479 European Data Protection Board. Guidelines 1/2018 on certification and identifying certification criteria in accordance with Articles 42 and 43 of the Regulation. 2019, 15-16.

480 Ibid, para 72.

481 Ibid, para 48.

482 The Information Commissioner's Office. How do we become a certification body? Available from: https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/certification-schemes-detailed-guidance/how-do-we-become-a-certification-body/

483 Kamara I, Leenes R, Lachaud E, et al. Data protection certification mechanisms: Study on Articles 42 and 43 of the Regulation (EU) 2016/679. European Commission Directorate Justice and Consumers. 2019.

484 Ibid, 49.

485 Kamara I, Leenes R, Lachaud E, et al. Data protection certification mechanisms: Study on Articles 42 and 43 of the Regulation (EU) 2016/679. European Commission Directorate Justice and Consumers. 2019.

Chapter 9

486 Article 29 Data Protection Working Party. Opinion 4/2007 on the concept of personal data. 2007, 9.

487 *Department of Health v Information Commissioner* [2011] EWHC 1430 (Admin), [2011] WL 1151213.

488 Harmanci A, Gerstein M. Quantification of private information leakage from phenotype-genotype data: Linking attacks. Nature Methods. 2016; 13: 251–256.

489 DPA 2018, s 171.

The PHG Foundation is a non-profit think tank with a special focus on how genomics and other emerging health technologies can provide more effective, personalised healthcare and deliver improvements in health for patients and citizens.

For more information contact:
intelligence@phgfoundation.org

UNIVERSITY OF CAMBRIDGE

phg foundation
making science work for health