An Ethics Framework for Big Data in Health and Research: An Overview

SHAPES Big Data Ethics Working Group

Co-chairs:

Associate Professor Tai E Shyong
Division of Endocrinology,
National University Hospital and
Saw Swee Hock School of Public Health,
National University of Singapore

Professor Graeme Laurie
School of Law and JK Mason
Ins1tute for Medicine,
Life Sciences and the Law,
University of Edinburgh

Members in alphabetical order:

Associate Professor Angela Ballantyne
Department of Primary Health Care
& General Practice, University of Otago

Dr Iain Brassington
Centre for Social Ethics and Policy/
School of Law, University of Manchester

Mr Markus Labude
Centre for Biomedical Ethics,
National University of Singapore

Associate Professor Hannah Yeefen Lim
Division of Business Law,
College of Business,
Nanyang Technological University

Associate Professor Wendy Lipworth
Sydney Health Ethics,
The University of Sydney

Assistant Professor Tamra Lysaght
Centre for Biomedical Ethics,
National University of Singapore

Dr Owen Schaefer
Centre for Biomedical Ethics,
National University of Singapore

Professor Cameron Stewart
Sydney Law School,
The University of Sydney

Associate Professor Shirley Sun
School of Social Sciences, College of
Humanities, Arts, & Social Sciences
Nanyang Technological University

Dr Vicki Xafis
Centre for Biomedical Ethics,
National University of Singapore

CONTENTS

Introduction

The *Science, Health and Policy-relevant Ethics in Singapore* (SHAPES) Initiative is a programme at the Centre for Biomedical Ethics, Yong Loo Lin School of Medicine, National University of Singapore. SHAPES received the Singapore National Medical Research Council *Research, Innovation and Enterprise 2020 Grant* to build bioethics capacity in Singapore and provide academic support and expertise to policy-makers, government agencies, clinician-scientists, and other stakeholders.

In 2018-2019, SHAPES launched a Big Data Project and established an international Working Group to produce a framework for deliberating on the ethical issues arising in health and research activities involving big data.

This booklet summarises the Framework and its key features. It also presents aspects of the six Domains where ethical issues that arise with the use of big data in health and research are examined.

To access the Framework in its entirety, please visit the *Asian Bioethics Review* 2019: 11(3) at www.springer.com/journal/41649 or www.abrjournal.wordpress.com.

01
Volume

The sheer quantity of data, taking into account the number of persons whose data is contained in given datasets and the level of detail about each individual

What is 'big data'?

The term 'big data' has been defined by a number of scholars, practitioners, and policymakers in various ways but there are three key characteristics of big data that exist across these different accounts.

03
Velocity

The great speed at which data can be transmitted and analysed

02
Variety

The substantial diversity of data forms about individuals (e.g. structured, unstructured, images, audio etc.) as well as the diversity of sources for that data (e.g. scientific data, user-generated data, web data etc.)

Why develop the Framework?

The topic of big data has been explored extensively in academic, technical, government, legal and private sector literature for a number of decades. However, there is a lack of **practical guidance in the form of a framework** that considers the ethical issues that arise from the use of big data in a variety of health and research contexts.

The *Ethics Framework for Big Data in Health and Research* addresses this gap.

Who is it for?

The Framework is intended for a wide-range of professional audiences including:

- biomedical researchers
- clinician-researchers
- data scientists
- policymakers
- those involved in the governance of big data activities in health and research (including ethics committees and data access committees)
- data controllers

The Framework may also be useful to lay people with an interest in big data, patients, and research participants.

What are the aims of this Framework?

1. To support decision-makers in the identification of underlying values at stake relating to a range of big data uses, such as sharing, linkage, granting access to third parties

2. To provide decision-makers with examples of using a balancing approach to weigh up the relevant values when making decisions about big data (illustrated through case studies in the various domains where big data is commonly used)

3. To demonstrate how decision-makers can be more robust and transparent in their decision-making, thereby better equipping them to justify their decisions about the use and sharing of big data.

What does the Framework comprise?

01 Central to a number of big data contexts that Framework users can consider when making decisions are **16 values**.

02 A **six-step** deliberative decision-making process to help guide users' thinking during the decision-making process.

03 Articulation of **three broader considerations** that underlie decisions made using the Framework: the need to respect persons, to take account of community expectations, and to consider issues of vulnerability which can arise in uses of big data.

What are the 16 Values?

The Framework identifies **9 substantive** and **7 procedural** values as relevant to big data. This is not an exhaustive list.

• *Substantive values* are those considerations that should be realised through the outcome of a decision.

• *Procedural values* are the values that guide the process of deliberation and decision making itself.

These sets of values are distinct but there is an important relationship between them in that procedural values may assist in realising certain substantive values.

Substantive values considered:
Harm minimisation
Integrity
Justice
Liberty/Autonomy
Privacy
Proportionality
Public benefit
Solidarity
Stewardship

Procedural values considered:
Accountability
Consistency
Engagement
Reasonableness
Reflexivity
Transparency
Trustworthiness

Three overarching issues

There are three ethical issues that are central to the consideration of all Domains in which big data is used. These are the concepts of **Respect for Persons** and **Social Licence** and the concept of **Vulnerability** in the context of big data.

Key points about *Respect for Persons* and *Social Licence*

▶ 1. *Respect for Persons* is a moral attitude that individuals, groups, or institutions hold and display towards others.

▶ 2. *Social Licence* relates to the positive public expectations associated with the perceived legitimacy of activities that have broad societal impacts.

▶ 3. The level of respect we hold towards others is often evident through interpersonal communication.

▶ 4. Showing respect towards publics in relation to the use of big data entails engaging in a variety of communicative exchanges to share information about big data activities and to receive input from the publics.

▶ 5. Engaging with publics in such a way is a process that should be ongoing if public trust is to be achieved.
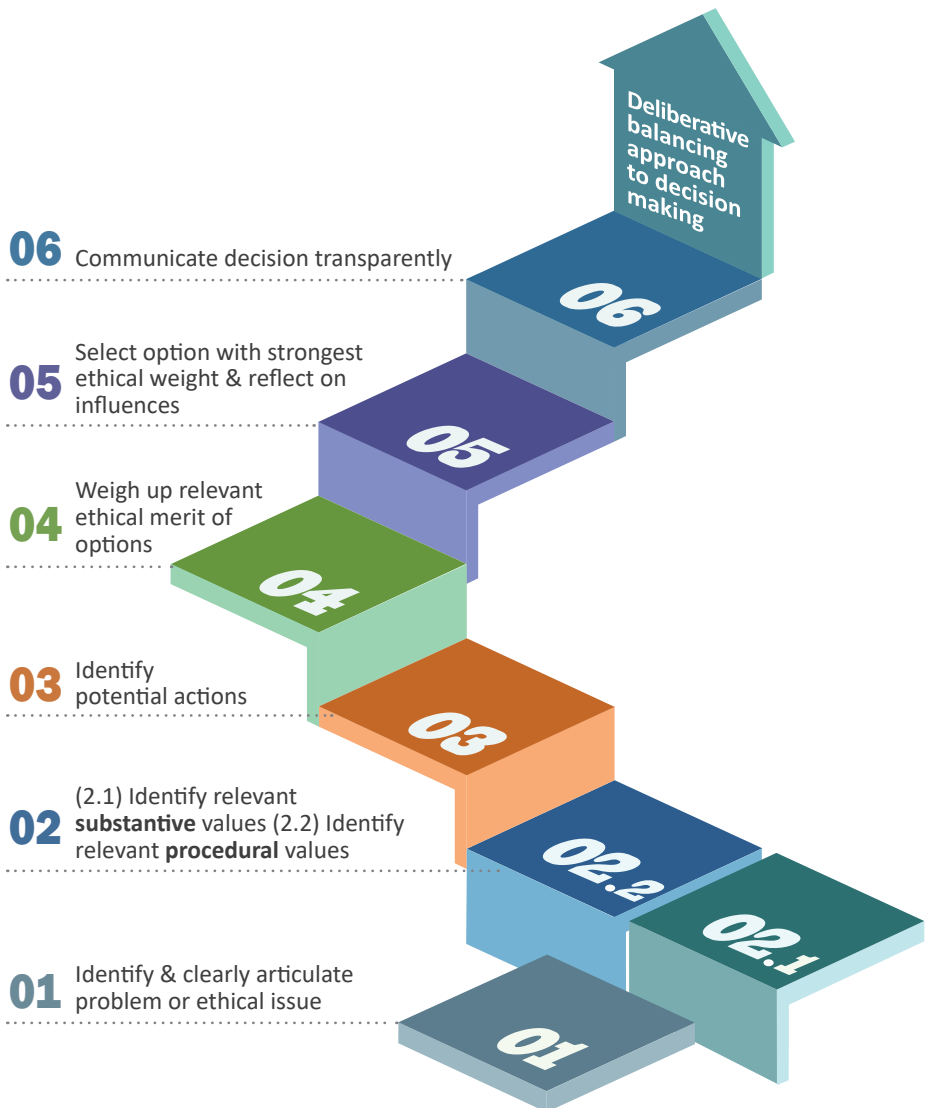
Key points about *Vulnerability and Power* in big data

▶ 1. Vulnerability takes several forms and is often contextual; a person may be vulnerable in one situation but not another.

▶ 2. Using big data in health research provides a way to relieve some vulnerabilities but it might generate or exacerbate others.

▶ 3. Those handling big data should be aware of this, and consider ways in which possible harms and wrongs may be mitigated or avoided entirely.

A deliberative balancing approach to decision-making

Broadly speaking, in the Domains we apply a process of reasoning that has been adapted from systematic, procedural approaches to ethical decision-making used by a variety of scholars and practitioners in healthcare ethics. We have adopted the following steps for the purposes of this Framework (represented in the adjacent diagram):

01 Identify and clearly articulate the presumptive ethical issue or problem at hand

02 Identify the relevant values pertinent to the issue or problem (from the list of 16 Key Ethical Values, noting that this may not be an exhaustive list). This is a two-step process (see adjacent diagram)

03 Identify potential actions (including consideration of policies, legal issues, etc.) that could be taken in response

04 In light of the values and context, weigh up the relative ethical merit of the different options

05 Select the option that has the strongest ethical weight attached to it and reflect on how your personal or the group's position and interests have influenced the decision, noting that the decision may require further consideration

06 Communicate the decision transparently to all stakeholders.

Deliberative balancing approach to decision making

06 Communicate decision transparently

06

05 Select option with strongest ethical weight & reflect on influences

05

04 Weigh up relevant ethical merit of options

04

03 Identify potential actions

03

02 (2.1) Identify relevant **substantive** values (2.2) Identify relevant **procedural** values

02.2

02.1

01 Identify & clearly articulate problem or ethical issue

01

How do we identify which values are relevant?

The three overarching consider-ations (respect for persons, social licence, vulnerability) assist in iden-tifying some of the values we need to consider in each case but there is likely to be **an iterative process** in steps 1, 2 and 3 in order to fully identify all the relevant issues and values. Also assisting in identifying the relevant values is the specific is-sue/problem and the context with-in which it arises.

When we begin to consider some possible solutions, it is likely that additional values will become obvious and will lead us to continue to consider the specific solution, amend it, or discard it, as a result of the deliberative process.

How do we resolve conflicts between values?

The resolution of conflicts between values in step 4 is arguably **the most difficult aspect** of any ethical decision-making framework and it is difficult to articulate a single process for balancing conflicting values, particularly outside a spe-cific context.

Deliberative balancing is the process we engage in when trying to deter-mine and justify which value carries greater weight than another rele-vant value in a particular case. The justifications and reasoning provid-ed for considering one value to hold greater importance than another help promote and further moral debate and provide a clear basis on which decisions have been reached.

Application of the Framework to six Domains

In the following sections, we present a very brief description of the **six Domains and excerpts** from each Domain paper and the associated case studies. A full consideration of the issues can be found in the respective publications in the Special Issue of the *Asian Bioethics Review* 2019: 11(3) at www.springer.com/journal/41649 or www.abrjournal.wordpress.com.

Openness in Big Data and Data Repositories

Authors: Vicki XAFIS, Markus LABUDE

Background: Sharing data through well-curated online data repositories presents opportunities as well as challenges. These repositories create a central 'pool' of data and make the data easily discoverable for bona fide researchers. The storage of data in research data repositories also ensures the long-term availability of data. A challenge concerns questions about the appropriate governance mechanisms for data repositories, including questions about who will be able to access the data and what (if any) levels of restriction should be applied. This Domain examines some of the ethical considerations that arise in relation to these opportunities and challenges.

Key issues
- Access restrictions and transparent governance
- Data ownership and the fair attribution of credit
- Concerns regarding privacy and confidentiality

Relevant values
Substantive: Autonomy/liberty, privacy, public benefit, justice

Procedural: Transparency, accountability, trustworthiness

Case Study: A clinician-researcher has completed a study on the prevalence and risk factors for colonization by antimicrobial drug resistant bacteria in 2000 adult hospital inpatients. Data collected included biological material and information on participants' history of healthcare contact, antibiotic use, travel, and information on housing and occupation. Informed consent was obtained. The consent form stated that participants' de-identified research data may be "shared for research and teaching purposes". The approving IRB understood this to mean conferences, journal papers, workshops, and teaching activities. Making the data accessible for future research has been strongly encouraged by the funder and is mandated by the journal selected for publication of findings. The central question for this case is the appropriateness of uploading the research data to an online research data repository.

Precision Medicine and Big Data

Authors: Owen SCHAEFER, TAI E Shyong, Shirley SUN

Background: As opposed to a "one size fits all" approach, precision medicine uses relevant biological (including genetic), medical, behavioural and environmental information about a person to further personalize their healthcare. This could mean better prediction of someone's disease risk and more effective diagnosis and treatment if they have a condition. Big data allows for far more precision and tailoring than was ever before possible by linking together diverse datasets to reveal hitherto-unknown correlations and causal pathways.

Key issues
- Balancing risks and benefits
- Addressing challenge of anonymization
- Preventing group harms
- Addressing potential genetic discrimination

Relevant values
Substantive: Harm minimization, justice, public benefit

Procedural: Transparency, engagement, reflexivity

Case Study: A large hospital is considering the deployment of routine whole genome sequencing in its clinics as part of continued service improvement. Whole genome sequencing is being conducted to facilitate additional discoveries and also allow for future insights to be implemented for these patients without the need for re-testing. The aim is both to reduce costs and to improve quality of care. The hospital would also like sequenced data to be collated into a database that can be used for secondary data research, particularly by collaborators at other institutions, domestically and abroad. This would raise the hospital's reputation as a cutting-edge centre of research, as well as support the development of innovations that the hospital itself relies upon. Before proceeding, the administration wishes to consider the ethical ramifications of making genomic data available for secondary research use.

Real-world Data to Generate Evidence about Healthcare Interventions

Authors: Wendy LIPWORTH

Background: This Domain illustrates the ethical and governance challenges that arise in the use of "real-world data". The phrase "real-world data" (henceforth RWD) refers here to data that is collected outside the laboratory or conventional randomised controlled trials (RCTs). In this Domain, the focus will be on the use of RWD to generate real-world evidence (henceforth RWE) about healthcare interventions, which is then used to guide their regulation, financing or clinical use, as well as broader health service design.

Key issues
• The quality of RWD, and of the scientific methods used to generate RWE from it
• The potential for those gathering and using RWD and RWE to be driven by commercial, political, professional or personal self-interest and, closely related to this, the potential limitations of self-governance.

Relevant values
Substantive: Public benefit, integrity

Procedural: Reflexivity, transparency, accountability

Case Study: A health service that has a well-established electronic medical record system is approached by a pharmaceutical company that wants to use these records to assess the "real-world" safety and effectiveness of a Hepatitis C treatment that it sells. The company will pay clinician-researchers from the health service to conduct the research. The researchers will not be paid personally; rather, the money will be paid into academic departmental accounts. The company and researchers submit an application to the health service's ethics committee in order to gain access to the records. One of the issues that the committee wishes to focus on in detail is how to address the influence of the pharmaceutical company over the research.

AI-assisted Decision Making in Healthcare

Authors: Tamra LYSAGHT, Hannah Yeefen LIM, Vicki XAFIS, NGIAM Kee Yuan

Background: Artificial Intelligence (AI) is set to transform healthcare. Key ethical issues to emerge with this transformation encompass the accountability and transparency of the decisions made by AI-based systems, the potential for group harms arising from algorithmic bias and the professional roles and integrity of clinicians. These concerns must be balanced against the imperatives of generating public benefit with more efficient healthcare systems from the vastly higher and accurate computational power of AI. This Domain focuses on how these issues arise with the development and implementation of AI-assisted clinical decision support systems (CDSS). Our analysis suggests it is important that developers and implementers of AI-assisted CDSS put in place control mechanisms to protect individuals and groups from harms arising through these tools. At the same time, these controls must not be so restrictive as to prevent the public benefits they can also deliver for health systems.

Key issues
• Accountability and transparency
• The potential for group harms
• Public interest in generating more efficient healthcare from AI-assisted systems
• The potential conflicting professional roles and duties of clinicians

Relevant values
Substantive: Integrity, justice, public benefit

Procedural: Transparency, accountability

Case Study: A software developer approaches the head of an intensive care unit (ICU) in a large university hospital to build an AI-assisted CDSS that can predict in real time which patients have or do not have a significant chance of survival to discharge and ability to recover functionally. The Application ("App") will not only help tertiary care physicians predict the outcomes for patients admitted to the ICU with higher accuracy but will help hospital administrators better manage scarce resources according to the volume of patients that the system predicts will be admitted to the unit at any one time. Savings in those resources can then be reallocated to another area of the hospital for improved patient care.

Big Data and Public-Private Partnerships in Healthcare and Research

Authors: Angela BALLANTYNE, Cameron STEWART

Background: Public-Private Partnerships (PPPs) are established to specifically harness the potential of big data in healthcare; and can include partners working across the data chain – producing health data, analysing data, using research results or creating value from data. This can happen in a range of ways, for example:

• Private sector organisations can apply to use public sector data for research
• Electronic health records can include data generated in the private sector and data produced by public health agencies
• Public and private sector agencies may form partnerships to pool resources and/or expertise to provide clinical care, or support research, innovation and product development.

Key issues
• Working within a social licence
• Public antipathy to the commercialization of public sector health data
• Questions of ownership, both of the data and any resulting intellectual property or products

Relevant values
Substantive: Public benefit, stewardship

Procedural: Transparency, engagement

Case Study: In 2015 the NHS provided Google DeepMind with 1.6 million identifiable and complete medical records (including sensitive information regarding, for example, abortion, drug overdoses, mental health and HIV status) to test a smartphone app called "Streams" that could help detect people with acute kidney disease. Google DeepMind is an artificial intelligence division within Google, which was created after Google bought University College London's DeepMind in 2014. One of the justifications for DeepMind joining Google in 2014 was the potential to use Google's scale and experience to achieve rapid progress in AI health interventions. Up to 25% of kidney deaths may be preventable if detected early, so DeepMind planned to use the NHS data to develop an algorithm to spot early signs of the disease and thereby save lives. The key question here is whether the NHS behaved as responsible data stewards.

Cross-Sectoral Big Data

Authors: Graeme LAURIE

Background: It is clear that any distinction between health data and other types of data is becoming increasingly blurred as notions of health and well-being expand into multiples areas of denizens' lives; and it is reasonable to assert that the promises of big data will only be fully realised if data can be safely and effectively linked and shared across a range of sources to address multiple objectives in the name of maximising benefit. While such benefit can include addressing unmet health and social care needs, it can also encompass improvements in efficiency of services and reduction in costs. This is what is meant by "cross-sectoral" sharing.

Key issues
• The expanding definition of 'health'

• Uneven standards of protection

• Loss of control of data

• Trustworthiness concerns between data controllers

Relevant values
Substantive: Privacy, public benefit, privacy, proportionality, liberty/autonomy, stewardship

Procedural: Transparency, trustworthiness, accountability, engagement, reasonableness, reflexivity

Case Study: The Veteris Foundation is a non-profit charitable organisation that seeks to offer a data-gathering device free to groups aged 65+ that will support better care at home. It wants to make the health and mobility data available to social services to improve efficiency in community care, such as concentrating on day care centres for the elderly to provide some respite cover for carers. In turn, Veteris will seek access to social care records in order to improve its own algorithms. Eventually, the Foundation wants to explore linkage of the device to electronic health records (EHRs) and other data it holds to make a new dataset openly available for further research into ageing.

Concluding remarks

This booklet provides a summary of the *Ethics Framework for Big Data in Health and Research* and the Framework's application to six Domains. The Domains provide an overview of some of the key considerations but readers should access the published version of the Framework and the individual Domain papers for a full understanding of the issues and the decision-making process.

Acknowledgments

SHAPES & Centre for Biomedical Ethics
Blk MD 11, 10 Medical Drive
#02-03 Singapore 117597

SHAPES
An NUS Centre for Biomedical Ethics initiative
supported by the Singapore Ministry of Health's
National Medical Research Council

NUS
National University
of Singapore

Centre for Biomedical Ethics
Yong Loo Lin School of Medicine

+65 6601 5173 | shapes@nus.edu.sg