

# Title: Digital Mega-Studies as a New Research Paradigm: Governing the Health Research of the Future

---

## INTRODUCTION

Large-scale research projects that rely on advances in digital technologies to address hard-to-solve problems in population health are emerging across the globe. In Australia, the Generation Victoria (GenV) initiative is under way. This whole-of-state cohort study aims to target the parents of all 160,000 babies born in the state of Victoria over a two-year period from mid 2021, to create a multi-faceted data- and bio-bank from children and their parents over many years. With consent, the collection will comprise newly-collected data, blood and other biosamples (enabling genetic and other analyses), linked with large existing health and education datasets about the research participants, from multiple sources.

Studies like GenV are ambitious in scale and scope, seeking to harness the power of ‘big data’ to achieve unprecedented outcomes. What is missing is conceptualisation of whether, and how, governance approaches might need to adapt to the demands of studies that are like GenV, which we describe as ‘Digital Mega-Studies’ (DMS). We posit that these are distinctly new types of studies presenting novel challenges in terms of the complex legal and regulatory framework within which they operate, the appropriate consent and privacy-protecting mechanisms they demand, the need to overcome the ‘digital divide’, and the level of transparency and accountability that must be inherent in the governance practices they adopt.

Both to contribute to the development of the GenV project specifically, and to establish a forward-looking research agenda for the planning, governance and implementation of DMS more broadly, we designed and undertook in mid-2019 a multidisciplinary, multi-stakeholder process to define DMS and to map their key characteristics, their inherent legal and governance challenges, and likely solutions. This paper reports on the methods and findings of this process, and the progress we have made towards establishing priority governance principles for high-quality large-scale digital research in health.

## Background

GenV is a major health research project in the state of Victoria, Australia, under the auspices of the Murdoch Children’s Research Institute at the Melbourne Children’s Campus in Melbourne. It has government and philanthropic funding. The project aims to recruit very large parallel consented birth and parent cohorts for the whole state, targeting all 160,000 babies born over two full years (2021-22) across a geographic region of 227,000 square kilometres. The project will leverage big data in order “to improve the health, development and wellbeing of Victoria’s children” and parents (Gold et al., 2019), exploring “the critical links between environmental exposures, genome (genetics), physical characteristics and later outcomes across the life course” (*Generation Victoria (GenV) | Murdoch Children’s Research Institute*, n.d.). Important features of GenV include: recruitment offered to all families (unrestricted by location, language or other criteria) on the basis

of parental consent; the gathering of existing health, educational and other datasets from state and federal jurisdictions and the collection of new data; the use of existing biomedical samples and the collection of new samples; and linkage with geospatial, clinical and administrative data. The resulting research dataset will form the foundation of observational and interventional research, in particular targeting hard-to-solve issues such as obesity, developmental disorders, mental ill-health and allergies.

Large birth cohort studies (numbering approximately 10 to 20,000) are not new, at least in higher income countries. In the United Kingdom (UK), examples date back to at least the 1940s (Lawlor et al., 2009). They include the National Survey of Health & Development (recruited essentially all babies born in Great Britain in a single week in 1946), the National Child Development Study (a similar design in 1958), the ALSPAC study (1992) and the Millennium Cohort Study (2001). Australian examples include the Longitudinal Study of Australian Children (Sanson & Johnstone, 2004; Edwards, 2012) and the ORIGINS Project (Hagemann et al., 2019) in Western Australia, with Generation R a European example (Kooijman et al., 2016). There are similar, albeit smaller, studies in low- and middle-income countries such as South Africa's Birth to Twenty study of 3,273 babies born in 1990 (Richter et al., 2007) and the Sri Lankan Twin Registry that commenced in 1997 (Sumathipala et al., 2000). Core to these cohorts are face-to-face research assessment visits and surveys. These cohorts have become international assets supporting wide-ranging research into, for example: health and environmental factors; genomics; microbiomics; intergenerational issues; and the impact of public health interventions.

This millennium, attention has turned to the possibility of much larger cohorts of 100,000 participants or more. These offer new potential for precision research capturing the diverse needs of subgroups rather than populations as a whole. A handful of larger studies has involved whole-population cohorts, on a scale not dissimilar to GenV. The Danish National Birth Cohort recruited 100,000 participants in the late 1990s (Danish National Birth Cohort, 2019; Olsen et al., 2001); the Norwegian Mother, Father and Child Cohort Study (MoBa) commenced in 1999, recruiting 100,000 pregnant women as well as 70,000 fathers (Magnus et al., 2006; *Norwegian Mother, Father and Child Cohort Study (MoBa)*, n.d.); and the UK Biobank recruited 500,000 40 to 60 year old participants commencing in 2007. The strategy for all these studies was a major face-to-face consented recruitment effort followed by light-touch mail surveys and extensive linkage to national registries. Very large cohorts that have subsequently attempted greater depth or diversity of data at recruitment or follow-up have shown the risk inherent in bigger ambitions. For example, the National Children's Study in the US (Kaiser, 2014) and the Life Study in the UK (University College London, 2015) also aimed to recruit 100,000 mothers and their babies, but both were controversially de-funded (Dezateux et al., 2016; Pearson, 2015). While the reasons for their failure to recruit are complex, participant burden appeared to be a major factor.

The proposed very large GenV cohort seeks to achieve all of (1) very large numbers, (2) depth of data, spanning biosamples, participant assessment and linked administrative and service data, and (3) low participant and study burden, within (4) a consented framework. Low burden appears attainable if the cohort is essentially wholly on-line – a DMS. It is hoped that this could solve both new and recalcitrant problems at scale and speed.

Helpful to this endeavour is the burgeoning literature addressing 'Big Data' issues of ethics and governance. It is widely acknowledged that Big Data, i.e. large-scale collection and use of data characterised in terms of its volume, variety, velocity, and veracity (IBM 2016), raises ethical, legal, and social challenges. Big Data practices are disrupting individual decision-making paradigms, moving from a position where individual decisions are made with specific, foreseeable consented outcomes, towards actions with unintended, unforeseen outcomes. This has been posited to have fundamentally challenged ethical premises of human agency, with Zwitter (Zwitter 2014) arguing that the "internet of things" has increased the power imbalance between 'one actor's knowledge and will and the other actor's source of information and power.' This power imbalance, and the likelihood of data use beyond those purposes for which the data has been provided, is one of the key reasons behind calls for 'big data ethics' (Richards & King, 2014), and data governance policies, which have proliferated over the last decade and which have been adopted by policy makers at the national (Central Digital and Data Office, UK Government, 2020), regional (European Data Governance Act 2020) and international levels (OECD, 2019). Through these policies, there is emerging consensus around important data governance principles such as transparency, fairness and accountability. However, one key principle occasionally overlooked (Abraham et al., 2019) is effective involvement of individuals and the public in data-driven initiatives.

The challenge of unforeseen future uses of data is not entirely new. In fact, while large-scale data-driven health research typically takes place within the parameters of a broad purpose, the scope of 'health research' is potentially very wide and, in the case of longitudinal studies in particular, it will not always be possible to anticipate the kinds of research projects that will be deemed to fit within this remit. To tackle this problem, population biobanks like UK Biobank adopted 'broad consent' models, premised on the idea that consent to the purposes of the research endeavour as a whole (i.e. health research for public benefit) is sufficient consent, so long as it is supplemented by adequate governance safeguards (Koenig, 2014).

Those advocating for more granular consent choices in research participation have challenged the concept of broad consent (Kaye & Prictor, 2021). There has also been longstanding debate around best practice approaches to patient and public involvement in health research, as a means of providing transparency and honouring participant consent, and to inspire trust and confidence in the studies over time. New participatory models of governance have emerged to tackle the potentially divergent interests of the various stakeholders involved, including those 'donating' data and those concerned with the scientific outputs of the project, such as public deliberation exercises to develop ethical norms and inform policy for biobanks (O'Doherty & Burgess, 2013).

Together, this literature only partly addresses central tenets of consented cohort studies which are not only confined to their scale and digitised nature but also span very long time horizons, broad and deep scope and the involvement of children. Nor do they take into account the views of diverse professionals involved in governance in the same community as the study will be conducted. In the health research space, rapid technological development is constantly providing new categories of data of interest to new categories of users. Individual models of participation may not be sufficient and there is a growing recognition for the need to move away from individual approaches towards a focus on the collective (Erikainen et al., 2021). The challenge of moving towards increased

participation, and moving from the individual to the collective, is particularly acute in health research projects involving multiple generations of family members.

We wanted to understand whether DMS must also grapple with novel difficulties. Drawing upon our own experience across law, governance and large-scale health research, together with relevant literature, we anticipated that difficulties would lie primarily in the following areas.

### *Complex legal framework*

It will likely be difficult to navigate the diverse and complex legislative and policy frameworks applicable to data that are currently collected in different jurisdictions and for different purposes, to be repurposed for a DMS (Gajanayake et al., 2012; Pictor, Hemsley, et al., 2018). How can datasets be interoperable when each has its own technical, legal and ethical characteristics?

### *Appropriate consent*

While adopting a consent-based approach in DMS is ethically appealing, it is challenging both conceptually and operationally. DMS, like biobanks, are likely to be premised upon a long or potentially indefinite lifespan. Their constituent research materials are therefore collected to be put to unknown future research use. Obtaining consent from members of a culturally, linguistically, cognitively, educationally and socially diverse mega-cohort in an inclusive manner is resource intensive and costly (Pictor, Teare, et al., 2018). For longitudinal studies involving children, it is conceivable and appropriate that re-consent is obtained at the point when minors become adults during the study (Hein et al., 2015; Taylor et al., 2018). Researchers establishing a DMS may require a consent model that is durable, flexible and participant-centric, whilst neither impinging on recruitment targets nor unduly burdening a project's participants, financial resources or research staff.

### *Inclusivity*

To ensure the wide applicability of research findings and to avoid entrenching further disadvantage, it would be important that DMS be as inclusive as possible—not only in terms of recruitment, but also, for instance, in identifying and adjusting for bias in datasets and in the choices that are made as to what data are collected (and thence what research enabled) (Bentley et al., 2017; Landry et al., 2018; Oh et al., 2015). DMS researchers will need to identify risks to inclusivity and plans to address these risks, in laying the foundations of such studies.

### *Privacy*

We envisaged that DMS would raise complex questions around participant privacy, including, for example, where the data are inherently identifiable (e.g., photographs of participants), where genomic data collected have potential relevance for genetic relatives, and where data linkage increases re-identification risk (Narayanan et al., 2016; Phillips & Knoppers, 2016).

## *Governance*

DMS will require governance processes that assure the transparent and accountable access to existing data and addition of new data, for use by both internal and external stakeholders (Murtagh et al., 2016; The British Academy and the Royal Society, 2017; Vayena & Blasimme, 2017). How can DMS ensure that such a governance framework is fit for purpose over time, reflects public opinion and interacts appropriately with the evolving external regulatory environment for data collection, storage and use?

### **Research aims**

For this research, led from GenV and the University of Melbourne Law School, we consulted with diverse professionals involved in research and data governance in the community in which GenV will be conducted (ie the state of Victoria, Australia).

The primary aim of this research was to develop a consensus definition of DMS. Our secondary aims were to:

- explore the ways in which DMS represent a new research paradigm with unique features compared to traditional biomedical research;
- identify the main ethical, legal and social challenges for such studies; and
- map appropriate regulatory and governance responses to these.

In order to achieve these aims, we also needed to design, pilot and report on a novel participatory methodology to explore health research governance issues in collaboration between health and legal researchers. We posited a connection between certain design features of DMS and the existence of specific challenges requiring tailored governance responses. We intended this to be a contribution both to the planning and long-term success of GenV and other DMS, and to the academic discourse around the governance of new health research paradigms.

### **METHODS**

We hypothesised that, to reach an agreed definition of DMS – necessary in order to frame future considerations of this new form of research – input from a broad community of stakeholders would be important. The potential challenges we outlined above would need to be explicitly investigated, and appropriate governance mechanisms identified iteratively, to provide the context for the effective and ethical conduct of future DMS. As such, we designed a robust multi-stage process including framing and elicitation (in a one-day symposium held in April 2019), synthesis and prioritisation (post-symposium) and the development of governance recommendations. We report on both the process and the outputs in order to facilitate future related research.

#### *Planning*

The symposium intentionally included certain features to maximise the breadth and richness of participants' contributions. First, participation by diverse stakeholders was considered important. We designed the symposium to accommodate up to 60 participants, and distributed invitations

based on the professional contacts of the investigators, and to representatives of state and national organisations with a likely interest in, or an established connection with, the GenV project. Snowballing elicited increased participation.

Second, we invited experts in youth engagement, information technology, health policy, law and governance to give presentations at the symposium exploring the likely risks and benefits inherent in DMS broadly. Presenters had broad scope to cover areas they considered likely to be relevant, with a focus on 'future gazing'. Pre-symposium briefings with presenters outlined the key features of DMS with GenV as the exemplar and invited their views on the likely issues and solutions to ensure the long-term success of DMS.

Third, we decided to elicit rapid-fire input of participants and to utilise two phases to synthesise their responses: one at the symposium and another following that event. In the symposium, and immediately after the presentations, we conducted back-to-back sessions involving brainstorming and affinity mapping (Lepley, 1998) using paper sticky notes to scope the unique elements of DMS and the diverse challenges that participants identified in the arena of DMS (see program at <<LINK TO SUPPLEMENTARY DIGITAL CONTENT>>). The process of affinity mapping was used to group the ideas into themes that were then prioritised amid the symposium. Participants were asked to further consider the prioritised themes describing key areas of challenge for DMS, in a session using a digital interactive brainstorming platform, GroupMap (www.groupmap.com), to capture responses or tools that could address the challenges.

Finally, an important part of the planning for this project included allowing for further directed participant engagement after the symposium to further develop and refine the questions and potential solutions that were emerging.

#### *Symposium: Framing and elicitation*

Approximately 60 people attended the day-long symposium in Melbourne to frame the research questions and elicit input, including academic researchers representing a breadth of disciplines (including law, medicine, public health, health informatics, information technology), graduate students, staff of government departments and agencies, research directors and managers and consumer representatives.

Presentations provided an overview of the GenV project and reflections on it from their diverse perspectives, namely information technology, youth engagement, policy, privacy and governance. Presenters highlighted technical advances (such as virtual laboratories, rapid data transfer and high connectivity) that will support DMS and reflected on communication and consent with children and young people as long-term study participants. Interspersed with presentations, interactive sessions provided opportunity for all participants to consider three key issues in depth: DMS' characteristics, challenges and potential responses to these.

#### Activity 1: Mapping unique *characteristics* of Digital Mega-Studies

Symposium attendees were asked to work in table groups (approximately 8 groups of 6 to 8 participants per group) in a 45-minute structured inductive session to brainstorm responses to the question 'What do you think is unique about digital mega-studies?'. Discussion at each table was lightly guided by a nominated attendee. We then created an affinity map (American Society for Quality (ASQ), 2020) of the unique characteristics of DMS. Figure 1 is a photograph of one whiteboard used in the event where we collected responses on the unique features of DMS.

<<INSERT FIGURE 1>>

Activity 2: Mapping main **challenges** associated with Digital Mega-Studies' characteristics

We repeated the above process using the same groups over a further 30-minute period to address the question 'What do you foresee to be the main challenges associated with digital mega-studies?'

Activity 3: Mapping the main potential **responses** to challenges using GroupMap

During the lunch break, the authors jointly synthesised participants' ideas through a further processing of affinity mapping based on key connections and relationships, considering the key challenges for DMS. We identified three overarching themes: Data, Participation and Trust. The following three questions were then developed to aid in the second brainstorming session focused on solutions to overcome key challenges:

- How do we ensure useful data?
- How do we ensure individual and public(s) wide participation?
- How can digital mega-studies build and maintain trust?

In the afternoon, these three questions were uploaded onto GroupMap and for 45 minutes, 39 participants individually and anonymously brainstormed a total of 213 solutions under each theme, using digital devices.

*Post-event: Synthesis and Prioritisation*

After the symposium we again collated and synthesised the ideas that had been entered onto GroupMap to remove duplication. Eighty discrete ideas were then uploaded back onto GroupMap and returned to participants approximately 8 weeks after the face-to-face event, and 28 participants rated each idea based on value (0=not valuable, 4=extremely valuable). The results of this process are described below.

## Results

We report below on the main results of the multi-stage deliberations according to the aims of the project as specified earlier. These were: identification of DMS' characteristics (which included developing a wording definition); mapping ethical, legal and social challenges; and developing potential governance responses.

*Unique characteristics of Digital Mega-Studies*

During the symposium, participants identified a wide range of characteristics that were specific to DMS. Several of these overlapped with the challenges described in the next section.

The identified characteristics are shown in Table 1. We loosely grouped the characteristics into three categories: data, participants, and governance and structure, while acknowledging that different groupings were possible.

<<INSERT TABLE 1>>

From the above characteristics we developed a working definition of DMS, as follows:

*Digital Mega-Studies are entirely or extensively digitised, longitudinal, population-scale data initiatives, collecting, storing and making available individual-level data of different types and from multiple (linked) sources, shaped by new technological developments and unforeseeable risks over time.*

We note that individual-level data may be identified or reasonably identifiable (ie. 'personal data' at law) or may be de-identified.

#### *Key challenges of Digital Mega-Studies*

Our elicitation and framing process enabled participants to identify the main challenges associated with DMS, which were initially mapped as follows.

1. **Data quality**, such as ensuring fitness for purpose, managing error and bias, data volume, and data linkage.
2. **Data security and privacy**, including the shifting regulatory landscape, data breach response planning, and connectivity with other health and data systems.
3. **Participation**, including ensuring participant welfare, taking into account the changing preferences of caregivers and children for the child's participation over time, and achieving truly informed and voluntary consent.
4. **Equity**, including cultural and linguistic diversity, supporting participation by those traditionally excluded; indigenous participation.
5. **Dissemination**, including return of results (with special considerations around genetic test results), communication of findings, engagement with policy-makers.
6. **Trust**, including who holds power, the role of commercial entities, and how governance is exercised.

We further grouped this list of key challenges into three principal areas with consensus with the symposium participants: Data, Participation and Trust. Participants were invited to provide their creative responses via GroupMap to the following questions.

- **Data:** How do we ensure useful data?
- **Participation:** How do we ensure individual and public(s) wide participation?

- **Trust:** How can Digital Mega-Studies build and maintain trust?

After the symposium, we completed the de-duplication, synthesis and ranking process described in *Methods* above. In Table 2 we present the most highly rated six ideas in each of the above three categories. We consider these to be high-priority governance principles that have an important function in addressing the key challenges facing DMS. In the *Discussion* section below we suggest some of the ways in which these principles can be translated into practice.

<< INSERT TABLE 2 >>

## Discussion

The process we designed and enacted yielded rich data from multidisciplinary stakeholders to better characterise DMS as a new research paradigm, to map the challenges and to develop governance solutions. This underpinned our development of 18 governance recommendations (below) to guide future DMS, spread across the three overarching areas of Data, Participation and Trust. These recommendations are key to enabling some of the more intrusive aspects of DMS projects to proceed, for example: collection of facial development data via photographs; learning and development monitoring via smart phone technology; and also dealing with the question of consent, as children who are part of longitudinal studies like GenV progress from being minors to adults.

Our recommendations contribute to, and further develop, a number of proposals in the literature, particularly in the field of biobanking, that seek to re-conceptualise governance models and move individuals from passive participants, to involved ‘stakeholders’ with a say in the running of a research project. For example, the ‘Stakeholder Model’ proposed by Hunter and Laurie requires mechanisms for on-going dialogue and communication between the stakeholders of a project, which, in the case of UK Biobank for example, includes donors and those responsible for the administration and management of the resource (Hunter & Laurie, 2009). More broadly, ‘stakeholders’ have been defined as ‘any group or individual who can affect or be affected by the achievement of the organisation’s ‘objective’ (Freeman, 2010, p. 46). For example, GenV stakeholders include child and parent participants, collaborators, researchers, communities, the public, services, policymakers and funders.

### *Governance recommendations*

**Data policies** should be drafted at the outset of a project to achieve the **six prioritised Data recommendations**, for collection of valid, relevant data that are necessary for the defined purposes of a DMS study (see Table 2). It is recommended that data policies promote principles of ‘purpose limitation’ and ‘minimisation’ as seen in the European data protection law framework, and ensure that only necessary data are collected, in the least identifying format possible to achieve project aims. DMS should not aim to collect as much data as possible, but should set clear goals at the outset. Defined purposes should justify the nature and scale of data to be collected from research participants, in line with their expectations both at the time of consent and that evolve with regular research updates throughout the project. This also extends to future data linkage projects, and applications for access to data held within DMS. Only data that have been validated should be

collected, used and made available within DMS, to adhere to examples of legislative approaches seen in both the General Data Protection Regulation (Article 5(1)(d)) and the Australian Privacy Principles (*Privacy Act 1988* (Cth) sch. 1 cl. 10(b)).

Data access criteria that specify who gets access, to what type of data, and on what basis, can be developed to steer review committees responsible for granting access to data (Teare et al., 2018). We suggest that DMS make access policies transparent at the outset. These criteria should be iterative, so that they can be changed to reflect participant and public attitudes over time. Criteria from large-scale data initiatives like UK Biobank include the requirement for only ‘bona fide’ researchers to be given access (Kaye et al., 2016; UK Biobank, 2007). GenV is exploring a tiered approach with access policies varying by the sensitivity of the data; it is hoped that this may also allow citizen scientists safe access to some data.

With regard to the **six prioritised Participation recommendations**, DMS could move beyond one-off consent towards co-developed governance regimes that are iterative and informed by **participant involvement** at every stage (see Table 2). While logistically challenging, we emphasise the impact and value of participant input for the trustworthiness of a DMS (, 2018). This may go beyond participants in the DMS itself – for example, incorporating rigorously-developed published design or measurement recommendations to which consumers are central. For example, patients and families are central to the elicitation and prioritisation of Core Outcome Sets that are designed to ensure that researchers measure and report those outcomes that are most likely to yield greatest benefit (Kirkham et al., 2017).

There are a number of examples of mechanisms for enhanced participant ‘control’ over the uses of data that can be adopted instead of static consent models, including data access committees and approvals panels comprising of ‘lay’, ‘participant’ and ‘public’ representatives, as well as cooperative structures (Vayena & Blasimme, 2017). Newer consent approaches such as ‘Tiered Consent’ (Bunnik et al., 2013) and ‘Dynamic Consent’ (Teare et al., 2020) seek to give research participants control over the uses of their data in a more granular and ongoing way. These are currently being trialed in large-scale research projects in Australia, for example the Australian Genomics Health Alliance (Haas et al., 2021). GenV is exploring these, weighing up their benefits vs complexities for whole-population cohorts aiming for equal inclusion across all languages and all levels of ability.

It is also crucial to consider how to make participation in decision making inclusive, and especially how to involve those communities facing challenges in navigating complex issues, which are often under-represented or excluded entirely (Prictor et al., 2020; Prictor, Teare, et al., 2018). Specific legal forms, such as models derived from the charitable sector (eg. Winickoff & Winickoff, 2003), can also be used to ensure decision makers have duties and responsibilities to run a DMS in pursuit of public benefit, subject to oversight via external regulatory bodies. Future DMS could consider adopting legal structures that build in public benefit aims, because inherent legal protections can operate to (legally) empower participants in the governance of the project (Bell, 2020).

With regard to the **six prioritised Trust recommendations**, we advocate that **trust, and trustworthy governance**, are enablers rather than barriers to DMS research activity. While compliance with the legislative framework for privacy will go some way to creating a trustworthy regulatory environment, transparency, communication and feedback can help ensure there are ‘no surprises’ (National Data

Guardian & Caldicott, 2020). DMS should endorse this principle by aligning participant expectations with DMS research objectives and outcomes, and committing to ongoing dialogue regarding data use and reuse (Taylor & Wilson, 2019), drawing upon current research into measures known to demonstrate trustworthiness (Milne et al., 2021). This will further honour participants' involvement in DMS, as it will be impossible to anticipate all future uses of participants' information at the time of the original consent, and will further substantiate their right of withdrawal. Feedback on how DMS research is benefitting health, wellbeing or other valued outcomes can be a way of giving back to participants, and regular updates via online platforms, newsletters and engagement 'roadshows' are effective means of encouraging active, rather than passive, participation in research.

### *Strengths and limitations*

This project included wide participation from diverse disciplines across academic research, government policy and research management. It was particularly enhanced by the participation of young people who were members of a youth advisory group for a relevant government agency. The range of experience and expertise amongst participants was appropriate to this project which is focused on developing an overarching approach to future DMS. As is usual with qualitative research, the project was not intended to involve a representative sample. However we acknowledge that consumer participation was lacking and that this will be particularly important to further reflections on governance approaches at the level of individual projects – particularly given the focus on engaging with harder to reach populations in studies of this nature (Nilsen et al., 2006; Ocloo & Matthews, 2016). We came to a consensus based on the rich and diverse expertise and views of the attendees. While the prioritisation process was rigorous, it reflected the views only of those who were present and thus should not be considered to be empirically representative of the Victorian or other communities. Replicability and other forms of formal validity remain to be determined.

Another strength of our work is its identification of DMS as a new research paradigm and the mapping of high-level principles for governance which are suitable for consideration and application in multiple settings for these study types. DMS involving very large participant numbers and deep and complex data whilst relying on consent and minimising participant burden will require well-designed governance mechanisms, such as those outlined herein, to underpin these studies' effectiveness and longevity.

Finally, the methodological approach involving expert presentations together with synchronous and asynchronous participation via both analogue and digital means is innovative, flexible and proved to be an effective method to capture data that were both rich and wide-ranging. In that regard, this paper makes a useful addition to the empirical ethics literature.

### *Considerations for the future*

Overall, the findings of our conceptualisation and prioritisation work support continued review of governance approaches, informed by engagement with DMS participants, to promote trust and confidence in the ongoing use of their information in different contexts over time. Perhaps most crucially, further work is needed to improve the literacy of participants and the public about data use in health research. We therefore propose that a key priority for future DMS is the development of plain language descriptions of information governance and the purposes and research goals of DMS. Examples of such activity in the UK includes the Wellcome Trust 'Understanding Patient Data' project (*Understanding Patient Data*, n.d.), which has sought to engage members of the public on their preferred language choices, including whether 'patient data' is accessible language in the first instance. Other mechanisms to enhance awareness about data use in health research could include

social media campaigns and could draw upon public-facing materials developed by relevant government-led entities such as Information Commissioners and Data Guardians. Reference to an existing framework (eg. *Building Digital Capability*, 2018) for enhancing data literacy could be useful. Arguably only when there is a shared language can participants be truly informed and engaged in the ambitious and important research in DMS of the future.

### *Best practices*

This research makes clear the need for, and benefit of, early consideration of the governance risks pertaining to DMS, and the design and implementation of governance models than adequately involve participants to future-proof such data-intensive research projects. The identification of pertinent risks and responses is assisted by wide consultation across disciplines, including computer science, medicine and public health, bioinformatics, law, governance and regulation, consumer engagement and communication. Our project provides a useful example of how empirical research relevant to legal issues can be embedded in projects alongside their practical legal and governance activities, to expand the identification and mitigation of challenges for DMS.

### *Research agenda*

Springboarding from the research presented in this article, we suggest there is a need for evaluation of the impact of a chosen governance model for a given DMS. Outcomes such as participation (recruitment, retention), trust and satisfaction with the research will be important, as well as financial implications of particular approaches to governance. We are also supportive of efforts to report on similar deliberative processes with stakeholders, to map and test the acceptability of different approaches to health data governance.

### *Educational implications*

This article has shown that affinity mapping and group-based prioritisation processes can be applied using analogue and digital methods in real time and asynchronously, in a single multidisciplinary project. This is a useful method that yields both rich and broad data about different facets of health research governance and could be adapted easily for a range of purposes. This article defines DMS and describes what we know about them so far, examining the challenges and potential risk mitigation strategies for data governance.

## **Conclusion**

We are in the midst of a new era of health research where entirely digital research initiatives are collecting, storing and making available individual-level data of different types and from multiple (linked) sources at population scale. This new wave of ambitious research projects, which we have called Digital Mega-Studies, risks swamping existing governance approaches. DMS hold unprecedented research potential, but require high levels of participant trust, especially if DMS use entirely digitised methods. Importantly, without the face-to-face participation that has characterised population scale research to date, DMS need to devise ways to develop trustworthy, ongoing relationships with participants who are actively involved in the research, in order to ensure the

levels of involvement necessary for rich, inclusive datasets. Our research findings, based on an expert stakeholder workshop involving an affinity mapping and prioritisation process with Australian participants, suggest that this will not simply be a case of transposing existing governance mechanisms at scale, and that mere legal compliance is not enough. Whilst future similar work could benefit from increased public participation, our project identified that DMS raise unique governance challenges and require new governance models that move passive participants to involved stakeholders. We hope that the principles and recommendations in this article, which have been informed by a diverse range of perspectives, will assist GenV as it enters its recruitment phase, and can be used as a benchmark for future DMS.

## References

- Abraham, R., Schneider, J., & vom Brocke, J. (2019). Data governance: A conceptual framework, structured review, and research agenda. *International Journal of Information Management*, 49, 424–438. <https://doi.org/10.1016/j.ijinfomgt.2019.07.008>
- American Society for Quality (ASQ). (2020). *What is an Affinity Diagram? K-J Method*. <https://asq.org/quality-resources/affinity>
- Bell, J. L. (2020). Governing commercial access to health data for public benefit: Charity law solutions. *Medical Law Review*, 28(2), 247–269. <https://doi.org/10.1093/medlaw/fwz022>
- Bentley, A. R., Callier, S., & Rotimi, C. N. (2017). Diversity and inclusion in genomic research: Why the uneven progress? *Journal of Community Genetics*, 8(4), 255–266. <https://doi.org/10.1007/s12687-017-0316-6>
- Building digital capability*. (2018, October 31). Jisc. <https://www.jisc.ac.uk/rd/projects/building-digital-capability>
- Bunnik, E. M., Janssens, A. C. J. W., & Schermer, M. H. N. (2013). A tiered-layered-staged model for informed consent in personal genome testing. *European Journal of Human Genetics*, 21(6), 596–601.
- Cameron, C. M., Scuffham, P. A., Shibl, R., Ng, S., Scott, R., Spinks, A., Mihala, G., Wilson, A., Kendall, E., Sipe, N., & McClure, R. J. (2012). Environments For Healthy Living (EFHL) Griffith birth cohort study: Characteristics of sample and profile of antenatal exposures. *BMC Public Health*, 12(1), 1080. <https://doi.org/10.1186/1471-2458-12-1080>
- Central Digital and Data Office, UK Government. (2020). *Data Ethics Framework*. <https://www.gov.uk/government/publications/data-ethics-framework>
- Danish National Birth Cohort. (2019, July 31). *About the DNBC*. <https://www.dnbc.dk/about-the-dnbc>
- Dezateux, C., Colson, D., Brocklehurst, P., & Elias, P. (2016). *'Life After Life Study': Report of a Scientific Meeting held at The Royal College of Physicians*. University College London. <https://discovery.ucl.ac.uk/id/eprint/1485681/7/Life%20Study%20Report%20of%20Life%20After%20Life%20Study%20meeting%20January%202016%201485681.pdf>
- Edwards, B. (2012). Growing Up in Australia: The longitudinal study of Australian children. *Family Matters*, 91, 6–17.
- Erikainen, S., Friesen, P., Rand, L., Jongsma, K., Dunn, M., Sorbie, A., McCoy, M., Bell, J., Burgess, M., Chen, H., Chico, V., Cunningham-Burley, S., Darbyshire, J., Dawson, R., Evans, A., Fahy, N., Finlay, T., Frith, L., Goldenberg, A., ... Sheehan, M. (2021). Public involvement in the governance of population-level biomedical research: Unresolved questions and future directions. *Journal of Medical Ethics*, 47(7), 522–525. <https://doi.org/10.1136/medethics-2020-106530>
- Freeman, R. E. (2010). *Strategic Management: A Stakeholder Approach*. Cambridge University Press.
- Gajanayake, R., Lane, B., Iannella, T., & Sahama, T. (2012). *Legal issues related to accountable-ehealth systems in Australia*. 1st Australian eHealth Informatics and Security Conference,

- held on the 3rd-5th December, 2012 at Novotel Langley Hotel, Perth, Western Australia, Perth, Western Australia. <https://doi.org/10.4225/75/5796f5b740a96>
- Generation Victoria (GenV) | Murdoch Children's Research Institute.* (n.d.). Retrieved 5 June 2020, from <https://genv.org.au/>
- Gold, J., Davies, S., Davenport, L., Perera, P., Wake, M., & Goldfeld, S. (2019). *GenV Rapid Evidence Assessment Report: Large research-led partnerships*. Murdoch Childrens Research Institute. [https://mcri.figshare.com/articles/GenV\\_Rapid\\_Evidence\\_Assessment\\_Report\\_Large\\_research-led\\_partnerships/10452185](https://mcri.figshare.com/articles/GenV_Rapid_Evidence_Assessment_Report_Large_research-led_partnerships/10452185)
- Haas, M. A., Teare, H., Prictor, M., Ceregra, G., Vidgen, M. E., Bunker, D., Kaye, J., & Boughtwood, T. (2021). 'CTRL': An online, Dynamic Consent and participant engagement platform working towards solving the complexities of consent in genomic research. *European Journal of Human Genetics: EJHG*. <https://doi.org/10.1038/s41431-020-00782-w>
- Hagemann, E., Colvin, L., Gibson, L. Y., Miller, S. J., Palmer, D. J., Srinivas Jois, R., Silva, D. T., & Prescott, S. L. (2019). The ORIGINS Project. In F. Sata, H. Fukuoka, & M. Hanson (Eds.), *Pre-emptive Medicine: Public Health Aspects of Developmental Origins of Health and Disease* (pp. 99–116). Springer. [https://doi.org/10.1007/978-981-13-2194-8\\_6](https://doi.org/10.1007/978-981-13-2194-8_6)
- Hein, I. M., De Vries, M. C., Troost, P. W., Meynen, G., Van Goudoever, J. B., & Lindauer, R. J. L. (2015). Informed consent instead of assent is appropriate in children from the age of twelve: Policy implications of new findings on children's competence to consent to clinical research. *BMC Medical Ethics*, *16*. <https://doi.org/10.1186/s12910-015-0067-z>
- Hunter, K. G., & Laurie, G. T. (2009). Involving publics in biobank governance: Moving beyond existing approaches. In C. Mullen & H. Widdows (Eds.), *The Governance of Genetic Information: Who Decides?* (pp. 151–177). Cambridge University Press. <https://doi.org/10.1017/CBO9780511635274.009>
- Kaiser, J. (2014, December 12). NIH cancels massive U.S. children's study. *Science*. <https://www.sciencemag.org/news/2014/12/nih-cancels-massive-us-children-s-study>
- Kaye, J., Briceño Morais, L., Mitchell, C., Bell, J., Bovenberg, J. A., Tassé, A.-M., & Knoppers, B. M. (2016). Access governance for biobanks: The case of the BioSHaRE-EU cohorts. *Biopreservation And Biobanking*, *14*(3), 201–206. <https://doi.org/10.1089/bio.2015.0124>
- Kaye, J., & Prictor, M. (2021). Consent. In A. Ganguli-Mitra, A. Sorbie, C. McMillan, E. Dove, E. Postan, G. Laurie, & N. Sethi (Eds.), *The Cambridge Handbook of Health Research Regulation* (pp. 103–111). Cambridge University Press. <https://doi.org/10.1017/9781108620024.014>
- Kirkham, J. J., Davis, K., Altman, D. G., Blazeby, J. M., Clarke, M., Tunis, S., & Williamson, P. R. (2017). Core Outcome Set-STANDards for Development: The COS-STAD recommendations. *PLOS Medicine*, *14*(11), e1002447. <https://doi.org/10.1371/journal.pmed.1002447>
- Koenig, B. A. (2014). Have we asked too much of consent? *The Hastings Center Report*, *44*(4), 33–34. <https://doi.org/10.1002/hast.329>
- Kooijman, M. N., Kruithof, C. J., van Duijn, C. M., Duijts, L., Franco, O. H., van IJendoorn, M. H., de Jongste, J. C., Klaver, C. C. W., van der Lugt, A., Mackenbach, J. P., Moll, H. A., Peeters, R. P., Raat, H., Rings, E. H. H. M., Rivadeneira, F., van der Schroeff, M. P., Steegers, E. A. P., Tiemeier, H., Uitterlinden, A. G., ... Jaddoe, V. W. V. (2016). The Generation R Study: Design and cohort update 2017. *European Journal of Epidemiology*, *31*(12), 1243–1264. <https://doi.org/10.1007/s10654-016-0224-9>
- Landry, L. G., Ali, N., Williams, D. R., Rehm, H. L., & Bonham, V. L. (2018). Lack of diversity in genomic databases is a barrier to translating precision medicine research into practice. *Health Affairs*, *37*(5), 780–785. <https://doi.org/10.1377/hlthaff.2017.1595>
- Lawlor, D. A., Andersen, A.-M. N., & Batty, G. D. (2009). Birth cohort studies: Past, present and future. *International Journal of Epidemiology*, *38*(4), 897–902. <https://doi.org/10.1093/ije/dyp240>
- Lepley, C. J. (1998). Problem-solving tools for analyzing system problems: The Affinity Map and the Relationship Diagram. *JONA: The Journal of Nursing Administration*, *28*(12), 44–50.

- Magnus, P., Haug, K., Nystad, W., & Skjaerven, R. (2006). [The Mother and Child Cohort Study will give new answers]. *Tidsskrift for Den Norske Laegeforening: Tidsskrift for Praktisk Medicin, Ny Raekke*, 126(13), 1747–1749.
- McKnight, C. M., Newnham, J. P., Stanley, F. J., Mountain, J. A., Landau, L. I., Beilin, L. J., Puddey, I. B., Pennell, C. E., & Mackey, D. A. (2012). Birth of a cohort—The first 20 years of the Raine study. *Medical Journal of Australia*, 197(11–12), 608–610.  
<https://doi.org/10.5694/mja12.10698>
- Milne, R., Morley, K. I., Almarri, M. A., Anwer, S., Atutornu, J., Baranova, E. E., Bevan, P., Cerezo, M., Cong, Y., Costa, A., Critchley, C., Fernow, J., Goodhand, P., Hasan, Q., Hibino, A., Houeland, G., Howard, H. C., Hussain, S. Z., Malmgren, C. I., ... Middleton, A. (2021). Demonstrating trustworthiness when collecting and sharing genomic data: Public views across 22 countries. *Genome Medicine*, 13(1), 92. <https://doi.org/10.1186/s13073-021-00903-0>
- Murtagh, M. J., Turner, A., Minion, J. T., Fay, M., & Burton, P. R. (2016). International data sharing in practice: New technologies meet old governance. *Biopreservation and Biobanking*, 14(3), 231–240. <https://doi.org/10.1089/bio.2016.0002>
- Narayanan, A., Huey, J., & Felten, E. W. (2016). A precautionary approach to big data privacy. In S. Gutwirth, R. Leenes, & P. De Hert (Eds.), *Data Protection on the Move* (Vol. 24, pp. 357–385). Springer Netherlands. [https://doi.org/10.1007/978-94-017-7376-8\\_13](https://doi.org/10.1007/978-94-017-7376-8_13)
- National Data Guardian, & Caldicott, F. (2020, May 27). No surprises. *GOV.UK*.  
<https://www.gov.uk/government/speeches/no-surprises>
- Nilsen, E. S., Myrhaug, H. T., Johansen, M., Oliver, S., & Oxman, A. D. (2006). Methods of consumer involvement in developing healthcare policy and research, clinical practice guidelines and patient information material. *Cochrane Database of Systematic Reviews*, 3.  
<https://doi.org/10.1002/14651858.CD004563.pub2>
- Norwegian Mother, Father and Child Cohort Study (MoBa)*. (n.d.). Norwegian Institute of Public Health. Retrieved 5 June 2020, from <https://www.fhi.no/en/studies/moba/>
- Ocloo, J., & Matthews, R. (2016). From tokenism to empowerment: Progressing patient and public involvement in healthcare improvement. *BMJ Quality & Safety*, 25(8), 626–632.  
<https://doi.org/10.1136/bmjqs-2015-004839>
- O’Doherty, K. C., & Burgess, M. M. (2013). Public deliberation to develop ethical norms and inform policy for biobanks: Lessons learnt and challenges remaining. *Research Ethics*, 9(2), 55–77.  
<https://doi.org/10.1177/1747016113488858>
- OECD. (2019). *Enhancing access to and sharing of data: Reconciling risks and benefits for data re-use across societies*. OECD Publishing. <https://www-oecd-ilibrary-org.eu1.proxy.openathens.net/sites/276aaca8-en/./sti-2019-1215-en/index.html>
- Oh, S. S., Galanter, J., Thakur, N., Pino-Yanes, M., Barcelo, N. E., White, M. J., de Bruin, D. M., Greenblatt, R. M., Bibbins-Domingo, K., Wu, A. H. B., Borrell, L. N., Gunter, C., Powe, N. R., & Burchard, E. G. (2015). Diversity in clinical and biomedical research: A promise yet to be fulfilled. *PLoS Medicine*, 12(12). <https://doi.org/10.1371/journal.pmed.1001918>
- Olsen, J., Melbye, M., Olsen, S. F., Sorensen, T. I., Aaby, P., Andersen, A.-M. N., Taxbøl, D., Hansen, K. D., Juhl, M., Schow, T. B., Sørensen, H. T., Andresen, J., Mortensen, E. L., Olesen, A. W., & Søndergaard, C. (2001). The Danish National Birth Cohort – its background, structure and aim. *Scandinavian Journal of Public Health*, 29(4), 300–307.
- Pearson, H. (2015). Massive UK baby study cancelled. *Nature*, 526(7575), 620–621.  
<https://doi.org/10.1038/526620a>
- Phillips, M., & Knoppers, B. M. (2016). The discombobulation of de-identification. *Nature Biotechnology*, 34, 1102.
- Pictor, M., Hemsley, B., Taylor, M., & McCarthy, S. (2018, September 21). If privacy is increasing for My Health Record data, it should apply to all medical records. *The Conversation*.  
<http://theconversation.com/if-privacy-is-increasing-for-my-health-record-data-it-should-apply-to-all-medical-records-102654>

- Prictor, M., Huebner, S., Teare, H. J. A., Burchill, L., & Kaye, J. (2020). Australian Aboriginal and Torres Strait Islander collections of genetic heritage: The legal, ethical and practical considerations of a dynamic consent approach to decision making. *The Journal of Law, Medicine & Ethics*, 48(1), 205–217.
- Prictor, M., Teare, H. J. A., & Kaye, J. (2018). Equitable participation in biobanks: The risks and benefits of a ‘dynamic consent’ approach. *Frontiers in Public Health*, 6. <https://doi.org/10.3389/fpubh.2018.00253>
- Richards, N. M., & King, J. (2014). Big data ethics. *Wake Forest Law Review*, 49, 393–432.
- Richter, L., Norris, S., Pettifor, J., Yach, D., & Cameron, N. (2007). Cohort profile: Mandela’s children: The 1990 birth to twenty study in South Africa. *International Journal of Epidemiology*, 36(3), 504–511. <https://doi.org/10.1093/ije/dym016>
- Sanson, A., & Johnstone, R. (2004). Growing Up in Australia takes its first steps. *Family Matters*, 67, 46–53.
- Sumathipala, A., Fernando, D. J. S., Siribaddana, S. H., Abeysingha, M. R. N., Jayasekare, R. W., Dissanayake, V. H. W., & Silva, N. D. (2000). Establishing a twin register in Sri Lanka. *Twin Research and Human Genetics*, 3(4), 202–204. <https://doi.org/10.1375/twin.3.4.202>
- Taylor, M. J., Dove, E. S., Laurie, G., & Townend, D. (2018). When can the child speak for herself? The limits of parental consent in data protection law for health research. *Medical Law Review*, 26(3), 369–391. <https://doi.org/10.1093/medlaw/fwx052>
- Taylor, M. J., & Wilson, J. (2019). Reasonable expectations of privacy and disclosure of health data. *Medical Law Review*, 27(3), 432–460. <https://doi.org/10.1093/medlaw/fwz009>
- Teare, H. J. A., de Masi, F., Banasik, K., Barnett, A., Herrgard, S., Jablonka, B., Postma, J. W. M., McDonald, T. J., Forgie, I., Chmura, P. J., Rydzka, E. K., Gupta, R., Brunak, S., Pearson, E., & Kaye, J. (2018). The governance structure for data access in the DIRECT consortium: An innovative medicines initiative (IMI) project. *Life Sciences, Society and Policy*, 14(1), 20. <https://doi.org/10.1186/s40504-018-0083-0>
- Teare, H. J. A., Prictor, M., & Kaye, J. (2020). Reflections on dynamic consent in biomedical research: The story so far. *European Journal of Human Genetics*, 1–8. <https://doi.org/10.1038/s41431-020-00771-z>
- The British Academy and the Royal Society. (2017). *Data Management and Use: Governance in the 21st Century*. <https://royalsociety.org/~media/policy/projects/data-governance/data-management-governance.pdf>
- UK Biobank. (2007). *UK Biobank Ethics and Governance Framework (Version 3.0)*. <https://www.ukbiobank.ac.uk/wp-content/uploads/2011/05/EGF20082.pdf>
- Understanding Patient Data*. (n.d.). Retrieved 4 June 2020, from <http://understandingpatientdata.org.uk/>
- University College London. (2015). *Why did Life Study close?* Life Study. <http://www.lifestudy.ac.uk/about/faqs.html>
- Vayena, E., & Blasimme, A. (2017). Biomedical big data: New models of control over access, use and governance. *Journal of Bioethical Inquiry*, 14(4), 501–513.
- Wilkins, C. H. (2018). Effective engagement requires trust and being trustworthy. *Medical Care*, 56(10 Suppl 1), S6–S8. <https://doi.org/10.1097/MLR.0000000000000953>
- Winickoff, D. E., & Winickoff, R. N. (2003). The charitable trust as a model for genomic biobanks. *New England Journal of Medicine*, 12, 1180.



# Digital Mega-Studies of the Future

Room 920, Melbourne Law School,  
185 Pelham Street, Carlton

Tuesday 2 April 2019

Twitter #digimegastudies

We are in a new era of health research. Technological innovation is enabling digital-mega studies of unprecedented scale and scope. Entirely digitised research endeavours are now commencing across Australia, reaching across whole populations and involving children as well as adults. The potential to link existing data sets and capture and integrate new types of sensitive functional and physiological data raises novel and urgent questions. How do we inspire and maintain public confidence and build trust in the research use of data in an untrusting age? How do we ensure these studies are fit for purpose into the future? Are existing models sufficient, or do we need new approaches?

This one-day Symposium will bring together leaders from across the research, digital, consumer, legal and policy spheres, to identify the unique characteristics of these studies from different perspectives and forecast future opportunities and challenges for digital mega-studies. The Symposium's forward-looking approach will provide the foundation for informed, robust and sustainable governance of digital mega-studies. True to the topic, we hope all participants will engage in digital brainstorming and prioritising, to capture the cutting-edge thinking from the day.

9:00	Registration
9:15 – 9:30	Welcome and overview: <i>Jessica Bell and Megan Prictor</i> Melbourne Law School

## SESSION 1:

9:30 – 9:45	GenV perspective: <i>Melissa Wake</i> Generation Victoria
9:45 – 10:00	Tech perspective: <i>Andrew Turpin</i> University of Melbourne
10:00 – 10:15	Youth engagement perspective: Youth Advisory Group, Office of the Victorian Information Commissioner
10:15 – 10:30	Policy perspective: <i>Sharon Goldfield</i> Royal Children's Hospital
10:30 – 10:40	Q & A

10:40 – 11:00	<i>Morning tea</i>
---------------	--------------------

11 – 12:15	Idea mapping
------------	--------------

12:15 – 12:45	<i>Lunch</i>
---------------	--------------

## SESSION 2:

12:45 – 13:30	Legal lightning round <ul style="list-style-type: none"> <li>• <i>Veronica Scott Minter Ellison</i></li> <li>• <i>Mark Taylor Melbourne Law School</i></li> <li>• <i>Jessica Bell Melbourne Law School</i></li> </ul>
13:30 – 14:15	Structured discussion and prioritisation
14:15 – 14:30	Wrapping up

14:30	<i>Finish</i>
-------	---------------