Publishing and sharing health data:

Informed consent matters

A/Prof Craig Fry
Research Program Leader, Culture and Values in Health
Centre for Cultural Diversity and Wellbeing
My Perspectives

• Human Research Ethics Committees since 2003 (Hospital, University, Community, Chair, Dep Chair, Expert member, low risk, high risk)
• Researching and publishing on research ethics since 2000
• NHMRC Public Health Fellowship (2008-11): Public health research ethics
• NHMRC Career Development Fellowship (2011-14): E-health and e-research ethics
Why publish and share health data?

- Publicly funded research should be open access
- Open science is a good thing
- Govt agencies increasingly expecting it (ARC, NHMRC)
- < costs, > research efficiency, > impact
- Aids scientific discovery and disease prevention and treatment
- > transparency, scrutiny, accountability
- > collaboration
- Economics

Research data curation and sharing may well be worth at least $1.8 billion and possibly up to $5.5 billion a year, of which $1.4 billion to $4.9 billion annually is potentially yet to be realised (Victoria Institute of Strategic Economic Studies, 2014)
Sharing and publishing data: The horse is already bolting...
Australian National Data Service

Our Vision: More valuable data for Australian research

ANDS is enabling the transformation of:

Data that are: | to |
--- | --- |
Unmanaged | Managed |
Disconnected | Connected |
Invisible | Findable |
Single-use | Reusable |

Australian Research Data Commons

ANDS is building the Australian Research Data Commons: a cohesive collection of research resources from all research institutions, to make better use of Australia's research data outputs.

Research Data Australia

Research Data Australia, ANDS' flagship service, provides a comprehensive window into the Australian Research Data Commons. It is an Internet-based discovery service designed to provide rich connections between data, projects, researchers and institutions, and promote visibility of Australian research data collections in search engines.

vu.edu.au

CRICOS Provider No. 00124K
Maps and Data

Maps

The latest Social Health Atlas of Australia is presented in interactive maps by multiple geographical structures for the whole of Australia, including totals and comparison charts for Capital City / Rest of state and State Territory areas.

Data

The latest Social Health Atlas of Australia is presented in data spreadsheets by multiple geographical structures for the whole of Australia, including totals for Capital City / Rest of State and State Territory areas.

Graphs

The latest Social Health Atlas of Australia is presented by Quintile of Socio-economic Disadvantage of Area, referred to as Inequality Graphs, and by Remoteness Area - for Australia and the State Territory areas.

Data Archive

The data archive presents earlier Social Health Atlas of Australia data in spreadsheets by multiple geographical areas. The data archive also presents earlier Aboriginal & Torres Strait Islander Social Health Atlas data and other topic-specific or South Australian Social Health Atlas data.

Welcome to the Australian Data Archive

The Australian Data Archive (ADA) provides a national service for the collection and preservation of digital research data and to make these data available for secondary analysis by academic researchers and other users.

The ADA is comprised of seven sub-archives:

- ADA Social Science
- ADA Historical
- ADA Indigenous
- ADA Longitudinal
- ADA Qualitative
- ADA Crime & Justice
- ADA International

The archive is provided by a consortium of leading national universities, managed by the Australian National University (ANU) and including nodes around Australia, at the University of Melbourne, University of Queensland, University of Technology Sydney and University of Western Australia.

Accessing Data

All access to the ADA Data Catalogue is browse and search the catalogue, view study and variable documentation (including frequencies) and download related material (questionnaires, codebooks, etc). Registered users can also analyse and visualise most data and users who have completed the relevant training form) can download online studies or subsets of variables in a range of formats. For more information, explore the Data Access menu above.

Note for former ARSDA users:

Access to the ARSDA website has now been closed, and all queries and requests for data are now made through this site. You can continue to use your user registration from the ARSDA website here, by logging in with your details at the login screen. You should continue to have access to any restricted access data that you requested previously through the ARSDA site, but will need to request new access to any general access datasets through the “Request Access to General Datasets” button in your account, due to changes in our access policy. Please contact us at admin@australian-data-archive.org for more information.

Recent News

A new home for ADA at the Australian National University
23 Aug 2011

Release of new Data Management framework from the UK Data Archive
03 Apr 2014

New release of LSAY data
13 Dec 2013

Recent Datasets

Deposits in January 2013
- Indonesian Surgeon, 2011
- Indonesian Surgeon, 2012
- AMDP 2014: Indigenous Affairs
- Staff in Australia’s Schools, 2010
- The Australian Longitudinal Study of Ageing
- The Australian Longitudinal Study of Parenting and Child Health
SO....the rationale, technical capacities, infrastructure, and policy support for data sharing exists

BUT...it is a complex environment
- Data types and sources
- Data uses
‘Health’ data

• Datasets / databases / databanks
• Primary data
  • Collected by researcher for research purposes with ethics approval
  • ‘Owned’ / controlled by the researcher / institution
• Secondary data
  • Routinely collected in health services not necessarily for research purposes
  • ‘Owned’ / controlled by the data custodian / govt department
• Biobanks (human tissue, digital data)
• PCEHR system data (and similar)
• Social media content (text, images, sounds…)
‘Digital’ data

Digital data are created in a variety of ways that concern researchers, and include:

- survey data from purposive collection of research data in online environments
- data routinely generated through contact with health professionals, hospitalisations, vaccinations and social service providers and increasingly through the use of direct-to-consumer services such as pathology and genetic testing services
- self-generated “lifelogging” data (including metadata) emitted from mobile phones and other “smart” appliances (e.g. Google Glass), generated through educational and lifestyle applications such as fitness monitoring devices and web-based games, gambling, dating, and posts on social media
- transactional and geospatial data including data generated from online records of retail purchases and the use of educational and financial services and roads and transport systems, as well as location sensing devices in public places
- administrative and legal data about births, deaths, marriages; credit ratings, criminal convictions, immigration and customs records.
‘E-Research’ (incl e-Health, Tele-health, m-Health etc)

• Range of e-research methods now used with human participants (hard to reach vulnerable populations)
  • Internet / online participant recruitment
  • Smart-phone data collection and interventions
  • Internet-based survey and interview
  • Social media data collection (e.g. Twitter, Facebook, Instagram, online chat-rooms)
  • Digital photo / video methods
  • ICT-enabled data linkage
  • Data sharing

• In intervention research (i.e. prevention, treatment, education)
• In non-intervention research (e.g. epidemiological, social and behavioural, humanities, etc)
Ethical considerations

• Because of the diversity and complexity
• ‘E-research’ differs from traditional means of human participant access, recruitment, data collection, and analysis
• E-research alters the nature, dynamics and potential consequences of participation
  • ‘Share-ability’ of digital (control of data use and impact)
  • Duty of care in virtual space / participants removed from research moments
  • Unpredictability of future uses

• E-research methods evolving faster than ethics guidelines
GOOD guidelines do exist
Chapter 3.2: Databanks

Introduction

This chapter covers a wide range of data types and methodologies. Given that the nature of data, data collection, research methodologies and data usage may change over time, the chapter presents principles rather than prescriptions.

Types of research that commonly make use of databanks include epidemiology, pathology, genetics and social sciences.

The term 'databanks', as used in this National Statement, includes databases.

What are data?

Data are pieces of information, for example:

- what people say in interviews, focus groups, questionnaires, personal histories and biographies;
- analysis of existing information (clinical, social, observational or other);
- information derived from human tissue such as blood, bone, muscle and urine.

Data identifiability

Data may be collected, stored or disclosed in three mutually exclusive forms:

- Individually identifiable data, where the identity of a specific individual can reasonably be ascertained. Examples of identifiers include the individual’s name, image, date of birth or address;
Publishing and Sharing Sensitive Data

Key messages
- The advantages of publishing your sensitive data will probably far outweigh any potential disadvantages when simple and appropriate steps are taken.
- Publishing your data, or just a description of your data (i.e., the metadata), means that others can discover it and cite it.
- You can publish a description of your data without making the data itself openly accessible.
- You can place conditions around access to published data.
- Sensitive data that has been confidentialised can be shared.

Author: Dr. Sarah Olesen, Australian National Data Service
GUIDELINES
FOR THE ETHICAL
USE OF DIGITAL
DATA IN HUMAN
RESEARCH

(Authors in alphabetical order)
Karin Clark,
Matt Duckham,
Marilys Guillemin,
Assunta Hunter,
Jodie McVernon,
Christine O’Keefe,
Cathy Pitkin,
Steven Prawer,
Richard Bennett,
Deborah Warr &
Jenny Waycott

Guidelines

• CONSENT
• PRIVACY AND
  CONFIDENTIALITY
• OWNERSHIP AND
  AUTHORSHIP
• GOVERNANCE AND
  CUSTODIANSHIP
• DATA SHARING: ASSESSING
  THE SOCIAL BENEFITS OF
  RESEARCH
## CREATE & MANAGE DATA

### OVERVIEW

Much research data about people—even sensitive data—can be shared ethically and legally if researchers employ strategies of informed consent, anonymisation and controlling access to data.

- Researchers obtaining data from people are expected to maintain high ethical standards and comply with relevant legislation. Ethical guidance is provided by professional bodies, host institutions and funding organisations.
- Informed consent is an ethical requirement for most research. It must be considered and implemented throughout the research lifecycle, from planning to publication and should include making provisions for sharing and future use of data.
- Before data obtained from research with people can be shared with other researchers, archived or published, data may need to be anonymised so that individuals, organisations or businesses cannot be identified.
- Sensitive and confidential data may be safeguarded effectively by regulating or controlling access to data or use of them. As well as securely storing such data, such safeguards enable data to be shared with others for research and educational purposes.
- Ethical guidelines for research are issued by professional bodies, host institutions and funding organisations. We have collated a number of useful resources on research ethics, consent and anonymisation.
Information Privacy Principles
1. Manner and purpose of collection
2. Collecting information directly from individuals
3. Collecting information generally
4. Storage and security
5-7. Access and amendment
8-10. Information use
11. Disclosure

www.oaic.gov.au
BUT...there are challenges in interpreting the guidelines in a still evolving complex environment

Despite our technical capacities important ethical questions exist

Especially around INFORMED CONSENT
Informed consent in publishing and sharing data

• Existing guidelines all have good advice around:
  • Form of data storage (identifiable, re-identifiable, non-identifiable)
  • Type of consent (specific, extended, unspecified, waived)
  • Timing of consent (if data sharing / re-use is likely)

• These are largely administrative / practical guidelines for good research conduct / governance
Informed consent in publishing and sharing data

Some unresolved ethical questions

• What if participants were told their data would NOT be available to others outside the research team, re-used, retained after 5-7 years?

• What consent requirements for special groups who have greater interest in uses of their data (Children? Cultures/collectives? Rare illness/disease? Engaged in illegal activity?)

• All future uses are impossible to predict for digital data, so what specificity of consent is reasonable?

• What does informed consent mean here?

What does the public / participant want?
Privacy and the use of health data for research

Christine M O’Keefe and Chris J Connolly

Analysts of the vast archives of clinical and health system data can yield information that is vital to effective health policy development and evaluation. It can also lead to enhanced clinical care through evidence-based practice and safety and quality monitoring. However, analysts must be conducted in such a way that standards of privacy and confidentiality are not compromised for individual health care consumers. In recognition of Australia’s international leadership in scope and extent of health-related data collected at the population level, the Population Health Research Network (PHRN) (http://www.phrn.org.au) has been established to provide Australian researchers with access to linkable, de-identified data from a wide range of health data sets, across jurisdictions and sectors.

The use of sophisticated data analysis and data-mining tools can increase the risks of privacy breaches occurring. This topic is becoming more important in the context of the PHRN investments designed to improve accessibility to Australian health-related data for the research sector.

In this article, we review privacy regulation and privacy perception in Australia, in the context of health data used in research. We also provide a summary of resources for researchers interested in privacy issues related to secondary use of health data.

Privacy regulation in Australia

ABSTRACT

Objective: We reviewed resources for researchers interested in privacy issues surrounding secondary use of health data for research. These include applicable privacy regulations and available information on privacy perception in Australia. The review is timely because the current Australian Population Health Research Network infrastructure investments are likely to attract new researchers to the field.

Data sources: We used Australian federal, state and territory regulations and programs, polls and surveys, public speeches and academic literature, and some international resources.

Data synthesis: We identify four themes (de-identification, consent, bias and participation) emerging as areas of concern from the review, and discuss issues relevant to these themes. We provide arguments that excessive privacy regulation has a negative effect on public health research.

Conclusions: There is little evidence of privacy complaints or breaches in health research, but significant concerns about consent and de-identification appear to persist in the community. New researchers need to take account of privacy regulation and may wish to take account of privacy perception when designing study and consent processes.

MJA 2010, 193: 537–541

“There is little evidence of privacy complaints or breaches in health research, but significant concerns about consent and de-identification appear to persist in the community”
The Impact of Commercialisation and Genetic Data Sharing Arrangements on Public Trust and the Intention to Participate in Biobank Research

Critchley C.\(^{a,b}\), Nicol D.\(^{c}\), Otolowski M.\(^{c}\)

Departments of \(^{a}\)Psychological Sciences and \(^{b}\)Statistics, Data Science and Epidemiology, Faculty of Health, Arts and Design, Swinburne University of Technology, Hawthorn, Vic., and \(^{c}\)Centre for Law and Genetics, Faculty of Law, University of Tasmania, Hobart, Tas., Australia

Abstract

Objectives: The necessity for biobanks to share their resources with third parties poses potential risks to public trust and the intention to participate in genetic research. We explore the effects of data sharing and the type of third-party access (public vs. private) on public trust and, in turn, the intention to participate in biobank research. Methods: An experimental design was used to assess a national sample of 1,701 Australians via a computer-assisted telephone interview. Results: The results revealed that trust and the intention to participate significantly decreased in relation to private compared to public biobanks, and when access to third-party researchers was allowed compared to when it was not. Somewhat surprisingly, no differences were found in relation to the third party being international compared to Australian, but trust and the intention to participate were significantly eroded when private third parties were allowed access. Those with a university education were particularly distrustful of private biobanks and biobanks that allowed access, while those who were more aware of genetic databases appeared more confident with biobanks sharing with private-sector third parties. Conclusion: The pattern of results suggests that public awareness of the need for biobanks to share their resources widely needs to be increased to maintain public trust and support.
“Consumers are in favor of electronic data sharing but elements of transparency are important: individual control, who has access, and the purpose for use of data.”
“...the public report currently seeing little benefit from sharing their data and little confidence that they will see benefits in future.

The public is particularly concerned about losing control of their personal data, with fear that they will become a victim of fraud or identity theft, and that their data will be shared with others without their knowledge or agreement.”
“...the public are supportive of data sharing when personal or public benefit can be demonstrated and these are communicated effectively;

...any objections to the use of personal data are largely related to security and privacy concerns.”

**ABSTRACT**

**Background:** The creation of biobanks for storage of human specimens and use in health and medical research is expanding globally. Biobanks hold significant potential to facilitate such research, however important ethical questions arise particularly in the paediatric setting in relation to consent, future use, and the balance of benefits against potential harms. To-date these ethical issues have been under-theorised and under-researched in the biobanks domain. The aim of this study was to examine stakeholder attitudes regarding the ethical responsibilities of researchers, biobank custodians, Human Research Ethics Committees (HRECs), research institutions and parents.

**Methods:** Qualitative study using semi-structured interviews with a purposive sample of N=14 key informants with relevant expertise. Participants were interviewed about their paediatric biobanks experiences, the main ethical issues observed as HREC member, researcher, or custodian, and future needs.

**Results:** Participants expressed concerns about consent processes in the biobanking context, including issues such as the provision of information, level of understanding, voluntariness, and the point at which children have a role or can consent for themselves. Other major issues raised were biobank quality management, the return of results, and the idea of human tissue as a scarce precious resource. Key informants also highlighted uncertainties about the custodianship of biobank samples and reasonable limits on the custodian’s role.

“Participants expressed concerns about consent processes in the biobanking context, including issues such as the provision of information, level of understanding, voluntariness, and the point at which children have a role or can consent for themselves”
People support data sharing for public good...

BUT they want clear information about future use, benefits, and risks...
Informed consent matters

- Complex and fast evolving digital data / e-research environment
- Innovation impacts data types, sources, uses = unknowns
- High ‘public good’ potential of digital data / e-research
  - Due to share-ability, link-ability, scale, permanence, problem solving applications
- Identity issues can (mostly) be addressed technically
  - Easy for numbers and non-public text
  - Harder for public text / images / recordings, biological/genetic samples
- Our informed consent responsibility for digital data sharing and use should be addressed
Best informed consent practices for all?

• Specify known and likely future research contact and data use
• Specify likely research types and bodies that could access
• Website / community notices possible if direct contact is not
• Re-negotiated consent with vulnerable groups / community reps
• Clarify access conditions (confidential? open? commercial?)
In conclusion

- Meeting our ethical responsibilities in data sharing is about more than identity protection, confidentiality, privacy
- Must also meet informed consent responsibilities re
  - What participants want to know about access, use, and by whom
  - Vulnerable groups and sensitive data
- Digital / ICT capacities that facilitate ‘Big Data’ innovations / data sharing ought to also make it easier to get required consent
- Is it ever really “impracticable” to obtain informed consent today?
QUESTIONS?
Consent to future use of data and tissue in research

2.2.14 Consent may be:

a. ‘specific’: limited to the specific project under consideration;
   b. ‘extended’: given for the use of data or tissue in future research projects that are:
      i. an extension of, or closely related to, the original project; or
      ii. in the same general area of research (for example, genealogical, ethnographical, epidemiological, or chronic illness research);
   c. ‘unspecified’: given for the use of data or tissue in any future research. The necessarily limited information and understanding about research for which extended or unspecified consent is given can still be sufficient and adequate for the purpose of consent (see paragraph 2.2.2).

2.2.15 Extended or unspecified consent may sometimes need to include permission to enter the original data or tissue into a databank or tissuebank (see paragraph 3.2.9).

2.2.16 When unspecified consent is sought, its terms and wide-ranging implications should be clearly explained to potential participants. When such consent is given, its terms should be clearly recorded.

2.2.17 Subsequent reliance, in a research proposal, on existing unspecified consent should describe the terms of that unspecified consent.

2.2.18 Data or tissue additional to those covered by the original extended or unspecified consent will sometimes be needed for research. Consent for access to such additional data or tissue must be sought from potential participants unless the need for this consent is waived by an ethical review body.
Waiver

2.3.9 Only an HREC may grant waiver of consent for research using personal information in medical research, or personal health information. Other review bodies may grant waiver of consent for other research.

2.3.10 Before deciding to waive the requirement for consent (other than in the case of research aiming to expose illegal activity), an HREC or other review body must be satisfied that:

a. involvement in the research carries no more than low risk (see paragraphs 2.1.6 and 2.1.7) to participants
b. the benefits from the research justify any risks of harm associated with not seeking consent
c. it is impracticable to obtain consent (for example, due to the quantity, age or accessibility of records)
d. there is no known or likely reason for thinking that participants would not have consented if they had been asked
e. there is sufficient protection of their privacy
f. there is an adequate plan to protect the confidentiality of data
g. in case the results have significance for the participants' welfare there is, where practicable, a plan for making information arising from the research available to them (for example, via a disease-specific website or regional news media)
h. the possibility of commercial exploitation of derivatives of the data or tissue will not deprive the participants of any financial benefits to which they would be entitled
i. the waiver is not prohibited by State, federal, or international law.
Constraints on data openness (privacy and confidentiality)

All parties must realise that, while the default is open access, there may be privacy, ethical, security, commercial or other constraints on the open release of research data. Addressing these constraints at an early stage in the research process is crucial, and it is essential to ensure that there are clear data access and management guidelines (e.g. clear processes for meeting any privacy, confidentiality and security concerns, whilst imposing the minimum obstacles on worthwhile research being conducted).

In this regard, the focus should be on protecting data providers and those whose data is used from foreseeable injury, rather than obtaining consents from them for each and every research use of their data. Building research freedoms around consents will necessarily foreclose many opportunities for re-use and the discovery of serendipitous uses for data which, though they may generate huge benefits, were not contemplated at the time the data were collected.