

# **SPEAR**

**Spear Data Saving and Sharing**

## **Discussion Document**

### **Challenges, Risks and Solutions for Administrative Data and Researcher Sourced Data**

Martin Tolich PhD\*

19 April 2007

# Challenges, Risks and Solutions for Administrative Data and Researcher Sourced Data

Martin Tolich PhD\*

*Data saving and data sharing* is a cost effective reuse of either administrative data or data collected by researchers. The practice is not currently embedded in either the New Zealand or international research cultures yet sufficient national and international pathways are now recognizable identifying an embryonic data saving and data sharing research infrastructure. The advent of data sharing in New Zealand involves as much a procedural shift as a cultural shift, and is unlikely to occur if left to individuals. The reuse of administrative data and data sourced by researchers have different starting points and different solutions and each is considered independently in this report#.

1) Strict enforcement of New Zealand privacy laws by senior government department managers currently limit the reuse of administrative data automatically saved by their departments. Innovative developments in the UK and guidelines on Observational Studies recently announced by the Ministry of Health provide pathways for researchers to access data sets, inclusive of identifiable data, without individual informed consent. The next step forward is to trial the applicability of these Ministry of Health guidelines for social science purposes.

2) Access to data sourced by researchers themselves meet different impediments to sharing data. Historically, researcher-sourced data saving is manifest as data hoarding: the storage of data (often mandated by an ethics committee) in locked file cabinets or claimed as intellectual property rather than with any longer term or collective purpose. A second set of Ministry of Health guidelines on tissue banking focus on gaining Broad Consent from participants for both the saving and the reuse of data for future unspecified research. Curation of these researcher-sourced data-sets remains a stumbling block for data sharing. The next step forward is to secure a site to curate this saved data for its eventual reuse.

**Key words:** Data sharing, Data saving, Funding, Curating, Confidentiality, Disclosure Review Boards, Ethics review, Broad Consent, Public Good

\* A discussion document commissioned by “the SPEaR Secretariat as part of the SPEaR Data Saving and Sharing Working Party”, December 1, 2006. Martin Tolich PhD (University of California, Davis) is Programme Coordinator of the Sociology Programme at the University of Otago. He also Chairs the Ministry of Health’s Multi-region ethics committee

## Table of Contents

<b>Section 1: Executive Summary</b> .....	<b>5</b>
1.1 The Challenge.....	5
1.2 Solutions 2007.....	5
1.3 International Best Practice.....	8
<b>Section 2: Ten Challenges and Solutions</b> .....	<b>10</b>
2.1 Data sharing infrastructure: who are the stakeholders? .....	10
2.2 What are the risks in data sharing? Motive and opportunity.....	10
2.3 Should data be curated in one central institution, or housed in the researcher’s institution?.....	11
2.4 How can the data’s context be maintained during curation? .....	11
2.5 Can confidentiality be maintained in data sharing?.....	12
2.6 Is the issue of ownership of data a fundamental impediment to widespread data sharing? .....	13
2.7 Can data sharing happen without informed consent? .....	13
2.8 Funding data sharing for specific research projects or all research.....	13
2.9 Is data hoarding data saving?.....	14
<b>Section 3: United Kingdom: MRC, ESRC and PIU Data Sharing Strategies</b> .....	<b>15</b>
3.1 Medical Research Council policy on data sharing and preservation?.....	15
3.2 Economic and Social Research Council commercial exploitation policy .....	15
<b>Table 1</b> “Data Sharing and Curation Policies across the UK Research Councils.”.....	16
3.3 UK government privacy and data sharing?.....	18
<b>Section 4: USA Data Sharing, NIH, ISCPDR and Disclosure Review Boards</b> .....	<b>21</b>
4.1 NIH data sharing policy.....	21
4.2 The Inter-university Consortium for Political and Social Research .....	22
4.3 Disclosure review boards.....	22
4.4. DRB: Managing Risk.....	24
<b>Section 5 Australia: The Australian Social Science Data Archive</b> .....	<b>26</b>
<b>Section 6: The NZ Situation: Exemplars and Transmission Lines</b> .....	<b>28</b>

6.1 Exemplars.....	28
6.2 The Health Research Council.....	28
6.3 Oral historians.....	28
6.4 Linked employer-employee data.....	29
6.5 The New Zealand Family Violence Clearinghouse.....	29
6.6 Enhanced data transmission lines.....	29
<b>Section 7: Qualitative Research: Challenges and Solutions.....</b>	<b>31</b>
7.1 The international attempts to share qualitative data .....	31
7.2 Six perceived barriers for data sharing qualitative data .....	32
7.3 Two major issues for successful qualitative data sharing .....	32
7.4 Emergent Best Practice.....	34
<b>Section 8: Three wind shifts on informed consent.....</b>	<b>35</b>
<i>a) Broad Consent</i>	
<i>b) Research without Informed Consent</i>	
<i>c) Observational Guidelines</i>	
<b>Section 9: Conclusions and Possible Next Steps.....</b>	<b>40</b>
<b>Endnotes.....</b>	<b>43</b>

## **Section 1: Executive Summary**

### **1.1. The Challenge**

The Social Policy Evaluation and Research Bulletin of April 2004<sup>1</sup> reported that New Zealand is being asked to treat with urgency the issue of saving and sharing publicly funded social research data. At a one-day seminar on sharing data, hosted by the Health Research Council (HRC) and supported by the SPEaR Linkages Programme, HRC Chief Executive Dr Bruce Scoggins stated that it is vital to ensure that data collections arising from public sector investments “are accessible, affordable and generate the maximum possible benefit to New Zealand”. However, while those present at the conference recognised the advantages to New Zealand of implementing a policy on data sharing, Dr. Scoggins noted that, “they were less clear on how to proceed from here”.

### **1.2: Solutions 2007**

The advent of data sharing in New Zealand involves as much a procedural shift as a cultural shift, and is unlikely to occur if left to individuals. The reuse of administrative data and data sourced by researchers have different starting points and different solutions and each is considered independently in this report.

1) Strict enforcement of New Zealand privacy laws by senior government department managers currently limit the reuse of administrative data automatically saved by their departments. Innovative developments in the UK and guidelines on Observational Studies recently announced by the Ministry of Health provide pathways for researchers to access data sets, inclusive of identifiable data, without individual informed consent. The next step forward is to trial the applicability of these Ministry of Health guidelines for social science purposes.

2) Access to data sourced by researchers themselves meet different impediments to sharing data. Historically, researcher-sourced data saving is manifest as data hoarding: the storage of data (often mandated by an ethics committee) in locked file cabinets or claimed as intellectual property rather than with any longer term or collective purpose. A second set of Ministry of Health guidelines on tissue banking focus on gaining Broad Consent from participants for both the saving and the reuse of data for future unspecified research. Curation of these researcher-sourced data-sets remains a stumbling block for data sharing. The next step forward is to secure a site to curate this saved data for its eventual reuse. There is little point in

agreeing to share data if there is no budget for the cost of doing so properly.<sup>2</sup> Researchers and research administrators, including ethics committees, need to anticipate the costs of data sharing, i.e. data documentation, the preparation of restricted data and monitoring of risk.

It would be a mistake to assume that the path to data sharing is only in the future. Existing exemplars are present and they need to be recognised and funding sought to monitor and report their practice. For example, exemplars of researcher-sourced data sharing include the long-term integrated practice by historians to create Greenfield conditions, in which permission to save and share data for *unspecified* future reuse is standard. The Alexander Turnbull Library is best practice for those wanting to replicate data sharing:

- The Alexander Turnbull Library, the Hocken Library and the National Library represent exemplars for the type of infrastructure needed for any data sharing initiative. Both are *funded to provide curation for data sets, allowing researchers access to context-rich documents under restricted and ethical circumstances*. Successful data sharing requires more than a willingness to share data: it is incumbent on each of the phases highlighted above.

Other, more limited exemplars provide a glimpse of the NZ data sharing infrastructure:

- In 2006 the Health Research Council<sup>3</sup> leads the way in New Zealand by supporting large-scale research and setting aside funding for major longitudinal studies, in which the researcher, in conjunction with an ethics committee, determines the specifics of long term data sharing. The HRC guidelines on data sharing are not prescriptive, but reliant on the researcher in conjunction with the ethics committee in determining best practice for the collection, storage, data preparation and redistribution of the data.
- The New Zealand Family Violence Clearing House<sup>4</sup> has plans to curate datasets, yet there is no guarantee that the original participants consented for the data sets to be shared.
- There are winds of change. Ministry of Health guidelines<sup>5</sup> (forthcoming) promoting “broad” consent for future unspecified research can easily be

modified to facilitate participant consent involvement in long term data sharing projects.

- E-Government,<sup>6</sup> Karen,<sup>7</sup> BRCCS<sup>8</sup> and the Digital Strategy<sup>9</sup> facilitate the transmission of electronic data sharing.

Thus, while parts of the infrastructure for researcher-sourced data sharing are present in New Zealand, there is no unified strategy or legislation promoting the various stages of data sharing and data saving inclusive of funding, ethical or legal oversight and curation.

#### *Administrative data*

Use of administrative data for data sharing requires weighing up two competing goals: maintaining the data's context while providing for confidentiality.<sup>10</sup> Confidentiality considerations are essential and can take place at two distinct phases: either ethics review oversight prior to data collection or more likely post-collection by disclosure review boards prior to the distribution of data to external researchers. Internationally de Wolf, Sieber et al.<sup>11</sup> claim two major drivers for administrative data sharing. These are:

- a) government funding agencies requiring, rather than urging, data sharing when it can reasonably occur. In such cases, grant proposals must contain evidence of a well-planned sharing arrangement that provides data in a useful form to other researchers without breaching promises of confidentiality.
- b) a clear role for external ethics committees or internal Disclosure Review Boards reviewing protocols when research funding is contingent on an agreement to share data.

The CYF Research Access Committee is an example of a quasi disclosure review board (see section 4.3). CYF RAC are custodians of access to CYF administrative data and staff and access requests for research purposes meet criteria and standards placed on access. How the information is used is also monitored.

Section 8 details the newly released Ministry of Health's "Observational Studies Guidelines" which have the potential to significantly impact the implementation of data sharing of administrative data. The guidelines permit New Zealand Health and

Disability Ethics Committees to give ethics approval to researchers accessing administrative health databases e.g. New Zealand Health Information Service (NZHIS) and the Institute of Environmental Science and research (ESR). As stated above, the next step forward is to trial the applicability of these Ministry of Health guidelines for social science purposes.

### 1.3 International Best Practice

- a) The Inter-university Consortium for Political and Social Research (ICPSR) established in 1962, housed at the University of Michigan, represents the gold standard for data sharing of quantitative data sets. Its **core mission is to archive social science data, provide open and equitable access to data, and promote the effective use of data.** ICPSR provides extensive instructions for persons who plan to deposit data in its repository. These excellent instructions include details on how to plan for sharing at the research stage, how to ensure confidentiality, and how to document the data.<sup>12</sup>
- b) In the UK the Social Science Research Council (SSRC) established a UK data archive (UKDA) in 1967. In 1996 the SSRC's successor, the Economic and Social Research Council (ESRC)<sup>13</sup>, introduced a formalised Datasets Policy that contracted all award holders to submit data for possible accession to the UK Data Archive (see Section 3: Table 1).
- c) Since 2001 the Australian Social Science Data Archive (ASSDA<sup>14</sup>) has been located at the Australian National University.
- d) Inter-government data sharing best practice is documented in the UK government report published by the Performance and Innovation Unit (PIU):<sup>15</sup> *'Privacy and data-sharing: The way forward for public services'*. This document provides innovative solutions to data share, while managing the twin objectives of enhancing privacy and making better use of personal data to deliver smarter, more trusted public services. However, the report's applicability for government "research" needs close scrutiny by government researchers, as its focus is ensuring the delivery of high quality services using case-by-case joined-up matching of administrative data, rather than a focus on research.
- e) Qualitative research data sharing is by no means as advanced. Currently funded by the Economic and Social Research Council, the UK-based

Qualitative Archiving and Data Sharing (QUADS)<sup>16</sup> group is attempting to find the means to share qualitative data. If successful, this innovation could be replicated in New Zealand in an existing or future New Zealand qualitative research project. Qualitative data sharing is inherently problematic as the data cannot be readily anonymised to the initial researcher, and it is doubtful that it can be for future data sharing, as context within the text may self report a person's identity.

## Section 2: Ten Challenges and Solutions

### 2.1 Data sharing infrastructure: Who are the stakeholders?

The infrastructure needed to successfully create a culture of data sharing goes beyond the original researcher and the person reusing the data. Stakeholders include some, if not all, of the following:

1. The *funding body* may urge mandated data sharing as a condition of funding, requiring the production of reusable data.
2. The *research participant* provides a broad consent to reuse the data for unspecified purposes.
3. An *ethics committee* reviews both the data collection and its storage, and the conditions under which the raw data can be reused.
4. A *disclosure review board* reviews an application from those wishing to re-use existing data, thus minimising threats of disclosure.
5. *Host institutions* may need to be co-opted to house the stored data.
6. The *re-used data* itself may serve as a new data set, and this too can be reused.
7. *Journal editors* could publish journal articles that feature secondary data.
8. *Ethics codes* must be written, providing details on how original researchers and secondary data researchers should re/use the data.
9. *Methodology textbooks* need to be written detailing the purpose and practice of data sharing.
10. *Teachers* of research methods courses need to incorporate data saving and data sharing into their curriculum.
11. In a PBRF environment, *Masters students* complete their thesis within twelve months, hobbled by data collection and ethics approval. Normally risks in this low level research are usually not balanced by any benefit. Ethically, data sharing would be sound.
12. Data sharing would thwart *the overuse of under represented samples*, i.e. Maori and Pacifica.

### 2.2 What are the risks in data sharing? Motive and opportunity<sup>17</sup>

- *Motive*. Data are a commodity. Thus, for certain items there may already be an active, often completely legitimate market for such a commodity. Employers and insurance companies can save millions of dollars by knowing the health or genetic status of prospective employees or insurees. Reidentified data could

show HIV+ status, cancer diagnoses, bankruptcy, criminal behaviour, or mental illness.

- *Opportunity.* Record linkage technology continues to advance. The reidentification of data that a decade ago took an unreasonable amount of time and resources can now be accomplished quickly and cheaply.

### **2.3 Should data be curated in one central institution, or housed in the researcher's institution?**

- Who will fund curating data documentation and the preparation of restricted data for release? Will monies for data sharing become part of the routine funding of all research, or will data sharing funding be targeted to specific types of research? If so, what is the criterion?
- The IPSCR is the gold standard for curating and preparation. Should this type of curation be replicated in New Zealand.
- What aspects of data sharing will be funded? Curation involves both the storage of data and data preparation. Whose responsibility is it to clean or prepare the data for distribution?

### **2.4 How can the data's context be maintained when curating data?**

- Data can be distributed in various forms, ranging from sanitised and anonymous data available freely on the Internet, to sensitive, raw data held in data enclaves and restricted by specified criterion. Should all data be treated in a similar manner and rendered anonymous? This is a fundamental question. Or should a hierarchy of data release be established i.e.
  - i. public data files, freely available to all who register;
  - ii. health data files, with somewhat more stringent registration requirements than the public data files;
  - iii. restricted data files, available only to those who apply and meet strict requirements; or
  - iv. a data enclave that provides access to restricted as well as public data.<sup>18</sup>
- If restricted data enclaves were created, would these enclaves be housed in the host institution i.e. a government department, or in a central location? Note:

Restricted data enclaves already exist – e.g. LEED (DOL, SNZ); Living Standards (MSD), and Crime & Safety Survey (MOJ).

### **2.5 Can confidentiality be maintained in data sharing?**

- Confidentiality considerations are essential and take place at two distinct phases: ethics review oversight prior to data collection, and disclosure review board prior to the distribution of data to external researchers.
- Ethics oversight is unevenly distributed across the New Zealand social research sector. Health and university sectors routinely submit their research for ethical review. Government researchers and NGOs remain outside mandatory ethical regulation yet most have informal processes with peer review by senior researchers/managers. Perhaps there needs to be more formal approaches developed. The draft SPEaR Guidelines are a start and should be extended to address data sharing aspects for the future. Recent Royal Society of New Zealand attempts to develop such a guide was limited to a brief half page statement about ‘humans’ and was not sufficiently robust to supercede most of the social science associations codes.
- Should Disclosure Review Boards<sup>19</sup> sanitise data before it is reused? Should these be standalone committees, or should they be integrated into the current ethics review committee?
- Forthcoming MOH guidelines of tissue banking<sup>20</sup> advocate the widespread use of broad consent for the storage of human tissue in bio-banks for later unspecified research. Noteworthy in these guidelines is the assumption that autonomous persons *can and will* provide their human tissue for research processes not yet conceived of. There should be no doubt that similar arrangements can be given to social research data.
- In Canada<sup>21</sup> and Australia<sup>22</sup> all researchers across all sectors follow a national set of ethical guidelines. Is this initiative part of the future data sharing infrastructure?

### **2.6 Is the issue of ownership of data a fundamental impediment to widespread implementation of data sharing?**

- Who owns the data is a fundamental hurdle for data sharing. In clinical drug trials, CRI, market or government research data may be collected and access to it restricted at the whim of the sponsor. In any data sharing the ownership of this data and exclusive use restrictions need to be addressed at the outset.
- How will ownership be both recognised to protect investment and to promote data sharing?<sup>23</sup> Under what conditions can data be suppressed and not shared?
- The absence of a culture of ethical oversight in government research creates an inherent danger for those wanting to data share (see historical example about grave robbing<sup>24</sup>). A guide to using data without ethical oversight may be to consider the researcher as an outsider. For example, how would an outsider to the organisation (I.e. government department) gain access to any data set? Is this route advisable? Note: In health research, medical practitioners have the same insider access to data as government researchers have to administrative data.

## **2.7 Can data sharing happen without consent? Brownfield vs. Greenfield Solutions**

- Data sharing initiatives can be split into two distinct categories: new data and existing data. All data sharing exemplars could use a Greenfield approach in setting up the right conditions at the outset, i.e. funding research to facilitate data sharing inclusive of cleaning the data, curating it and redistributing the data. Should future funding of data sharing initiatives focus on a Greenfield strategy?
- Brownfield research data sets (existing research data) are inherently problematic in terms of confidentiality, as invariably no initial consent to share the data was gained at time of data collection. Should SPEaR's data sharing initiative be focused on new data or existing data?
- In the USA, oversight of the release of administrative data sets is governed by Disclosure Review Boards.<sup>25</sup>

## **2.8 Funding specific or all research?**

- Should all social research in New Zealand be funded to data share, or should criterion be set to fund specific types of research? What would the criterion be?<sup>26</sup>

- Should government research agencies and universities set aside funds for data sharing? If so, should funding bodies urge researchers to data share, or should this be mandated?
- Beyond the HRC, what other NZ institutions specifically fund data sharing?
- Should the HRC follow the MRC guidelines and fund data sharing for all research?

## **2.9 Is data hoarding data saving?**

New Zealand is blessed with infinite amounts of data, but little of which was collected to be shared with other researchers. Encouraged by ethics review committees, New Zealand researchers routinely store data for between five to 17 years. However, the sharing of this data is prohibited, as usually no informed consent was gained at the time of data collection for its reuse. To gain a re-consent of this data would be a cumbersome process that few would attempt. Internationally and nationally there has been a sea change in the informed consent process. For example, two separate sets of Ministry of Health guidelines are outlined in Section 8 below. The new Observational Studies Guidelines balance the traditionally sacrosanct mandating of individual consent against the interests of the public good. The Ministry of Health's draft guidelines on Tissue Banking go beyond the traditional one-off direct consent process providing for a broad consent allowing research participants to consent both to a particular research project at the same time as consenting to their data being saved and shared with other researchers conducting unspecified future research in the future. Note: broad consent by definition cannot be enacted retrospectively.

## **Section 3: United Kingdom: MRC, ESRC and PIU Data Sharing Strategies**

### **3.1 Medical Research Council Policy on Data Sharing and Preservation<sup>27</sup>**

From 1 January 2006 **all MRC funding proposals must include a strategy for data preservation and sharing**. Any applicants who consider that the data arising from their proposals will not be suitable for sharing must provide clear reasons for not making it available. The MRC data sharing and preservation policy does not prescribe when or how researchers should preserve and share data, but requires them to make clear provision for doing so when planning and executing their research.

The MRC expects valuable data arising from MRC-funded research to be made available to the scientific community with as few restrictions as possible. Such data must be shared in a timely and responsible manner.

Data arising from MRC-funded research must be properly curated throughout its life cycle, and released with the appropriate high-quality metadata. This is the responsibility of the data custodians, who are usually those individuals or institutes that receive MRC funding to create or collect the data. A limited, defined period of exclusive use of data for primary research is reasonable, according to the nature and value of the data and the way they are generated and used.

### **3.2 Economic and Social Research Council Commercial Exploitation Policy<sup>28</sup>**

It is the ESRC's policy that potentially valuable results or products arising from ESRC-funded research should, where practicable, be exploited for the benefit of the UK economy, the quality of life of the nation, the institution, the award-holder and the UK social science community. This policy applies to all forms of results or products, regardless of the nature of the intellectual property rights involved. The ownership of the intellectual property arising from a research project should be clear from the outset. The ESRC will assume that this will rest with the university or institution receiving the ESRC award, unless otherwise stated.

**Table 1** documents Mark Thornley’s “Data Sharing and Curation Policies across the UK Research Councils.”<sup>29</sup> The ESRC and MRC policies are in bold.

### UK Cross council summary

AHRC – Arts & Humanities	Detailed in funding guidance – a policy all but in name.
BBSRC – Biotechnology & Biological Sciences	Data sharing policy statement in draft. Formal policy in place by mid-2006.
EPSRC – Engineering & Physical Sciences	No formal policy, as does not overly intervene in the research dissemination process.
<b>ESRC – Economic &amp; Social</b>	<b>Formal data policy. Current version dates from 2000, but to be updated soon.</b>
<b>MRC – Medical</b>	<b>Data sharing and preservation policy – applies to new grants awarded from January 2006.</b>
NERC – Natural Environment	Data policy handbook. Current version from 2002, currently being updated.
PPARC – Particle Physics & Astronomy	Data curation policy has been agreed in principle from April 2005.
<i>CCLRC – Council for the Central Laboratory of the Research Councils</i>	<i>A number of policies dealing with the curation of CCLRC data on a per-project basis.</i>

### Curation

AHRC	Joint JISC and AHRC supported Arts and Humanities Data Service.
BBSRC	No curation facilities supported as yet – however, recognise need for resources to support long-term, sustainable management.
EPSRC	Data managed in a durable form under control of the institution of origin.
<b>ESRC</b>	<b>Joint JISC and ESRC supported UK Date Archive, including the Economic and Social Data Service.</b>
<b>MRC</b>	<b>Encourages curation and long-term management. Onus is on the PI and their institution. Considering data curation facilities.</b>

NERC	NERC supported Designated Data Centres.
PPARC	Supports long-term curation of selected data sets. Driven by the science, data centres supported as projects. No appropriate to maintain all data.
CCLRC	<i>A number of curation activities – including the Atlas Data Store.</i>

### Obligations on researchers

AHRC	For grants awarded where a significant product is the creation of an electronic resource, data and documentation must be offered to AHDS within three months.
BBSRC	Under consideration as part of policy development.
EPSRC	Encourages PIs to manage primary data as the basis for publications securely and for an appropriate time in a durable form under the control of the institution of their origin.
ESRC	Applicants must carry out a data review to ensure funds not requested for data that are already available. Data must be offered to the data archive within 3 months of end of award.
MRC	From January 2006 applicants must produce a plan for data sharing and preservation.
NERC	Programmes must have data management plans. PIs offered reasonable time for first use, all data must be offered to a NERC data centre.
PPARC	Under consideration as part of policy development. PIs will have to consider data curation and sharing activities.

### Monitoring and enforcement

AHRC	Each AHRC grant panel has a data specialist in an advisory capacity. AHDS sees proposals and advises PIs as necessary. End loop not closed.
ES	Data archive is made aware of all ESRC grants awarded. ESRC will not sign off award until data archive indicates it is happy with data situation.
NERC	Directed Mode programmes have data management plans and strong links to data centres. Need stronger linkage between non-directed PIs

	and data centres – too many grants to monitor everyone. End loop not closed.
--	--

### 3.3 UK Government Privacy and Data-sharing

On 11 April 2002 the Performance and Innovation Unit<sup>30</sup> (PIU) at the Cabinet Office published a report entitled “Privacy and Data-sharing: The way forward for public services”. The twin objectives of the report were to improve public services through better use of personal data: and to secure public trust in the handling of personal data by safeguarding personal privacy. In line with its existing responsibilities for privacy, data protection, human rights and freedom of information, the Department for Constitutional Affairs assumed overall responsibility for championing and overseeing implementation of the conclusions of the report, and the outcome of the consultation process.

The PIU report sets out a five-point strategy, detailed in 25 recommendations for achieving five objectives:

- building public trust in the way that public services handle personal data;
- ensuring that the data used by public services meets high standards of accuracy, integrity and is up-to-date;
- using technology more effectively to deliver more secure, more joined-up public services;
- improving the management of information and personal data in the public sector, and ensuring greater consistency in the way that data are handled; and
- achieving greater clarity over how the law regulates data-sharing and consulting on options for legislative change.

The two key recommendations created 1) a Public Trust Charter and 18) the position of Chief Knowledge Officer:

*Recommendation 1:*

A draft **Public Services Trust Charter** ....sets out the guiding principles and key commitments made to the citizen in protecting their privacy and personal data in their interactions with public services. All public sector organisations should look to embody these principles in service-level privacy statements *describing precisely in each case* how personal information will be shared in support of service delivery or research and evaluation, and how individuals can get access to their personal data. In

turn, these privacy statements will be key instruments to help inform the public and secure consent where information is shared to support delivery of public services. They must therefore be easily and readily available to the public where appropriate at physical outlets and websites. To ensure implementation of these privacy principles and undertakings, each service-level privacy statement will need to be embodied in working-level codes of practice and information sharing protocols, themselves underpinned by management guidance. These should also be made publicly available.

*Recommendation 18:*

Public service bodies should consider integrating the functions set out in Box 9.2 below, including through an evaluation of the appointment of a board level Chief Knowledge Officer as a means to ensure integration of information issues into decision-making processes. Ideally, **Chief Knowledge Officers** would be responsible for integrating, over time, the disparate functions of legislative compliance and business planning.

*Box 9.2: Chief Knowledge Officer – range of responsibilities. Increasing Integration*

<ul style="list-style-type: none"> <li>• DPA/HRA compliance</li> <li>• FoI</li> <li>• Data standards</li> <li>• Data-sharing protocols</li> <li>• Records and information management</li> </ul>	<ul style="list-style-type: none"> <li>• Business design and e-business action plans</li> <li>• Customer relationship management</li> <li>• IS/IT strategy</li> </ul>	<ul style="list-style-type: none"> <li>• Analytical, research and statistics services</li> <li>• Research and evaluation to feed into decision-making</li> </ul>
---	---	--

The Chief Knowledge Officer should bring together a number of relevant functions (as determined by the needs of the organisation, including, for example, legal compliance and business redesign) to ensure effective integration of data management and privacy protection into mainstream decision making.

**The PIU report produced a public sector toolkit on data sharing** that includes guidance and examples of good practice on information handling, such as data sharing protocols. Their long term goal is to add to this toolkit until it contains a complete set of guidance on whether to data share and how to do it with the maximum respect for people’s privacy.

The PIU report's annex also listed a range of insightful examples demonstrating wide variance in the national experience of privacy. For example, certain countries appear to have attitudes radically different from those of the UK in this area:

- In **Sweden** it is accepted that everyone's tax return can be inspected by anyone who cares to do so.
- In **the US**, it is accepted that drivers should carry their licence with them at all times when driving.
- In **the Netherlands**, Dutch government files listing religious affiliations were used by the Nazis to identify Jews.
- In the mid 1980s *The Australia Card*, a national ID card designed to reduce tax evasion and social security fraud, was withdrawn after public protest, having originally received a reasonable level of public support.
- In **France**, National ID cards are compulsory, and in theory must be carried at all times.

## Section 4: USA Data Sharing: NIH, ISCPDR and Disclosure Review Boards

The US data sharing experience has three exemplars: the NIH funding model, the ICPSR, the gold standard for curation, and third, the use of disclosure review boards (DRB).

### 4.1 National Institute of Health Data Sharing Policy

In the US, data sharing was given a tremendous boost as of 1 October 2003, when the National Institutes of Health (NIH) required all investigator-initiated proposals for grants with direct costs greater than \$500,000 in any single year to specify plans for sharing research data, or explain why data sharing is not possible.

*Supporting documentation:*

- [Data Sharing Regulations/Policy/Guidance Chart for NIH Awards \(08/30/2006\)](#) - (MS Word) - This chart is designed as a quick guide only for the purpose of identifying various data sharing regulation/policy/guidance documents applicable to NIH funding.
- [NIH Guide Notice \(02/26/2003\)](#) - Final NIH Statement on Sharing Research Data.
- [NIH Guide Notice \(03/01/2002\)](#) - NIH announces a Draft Statement on Sharing Research Data.
- [NIH Data Sharing Policy and Implementation Guidance \(03/05/2003\)](#) - Guidance providing the NIH policy statement on data sharing and additional information on the implementation of this policy.
- [Frequently Asked Questions - Data Sharing \(02/16/2004\)](#) - Listing of Frequently Asked Questions that will be updated as new questions are received.
- [Data Sharing Workbook \(PDF or MS Word\)](#) - (02/16/2004) - Workbook to show how investigators working in a variety of scientific areas have shared their data.
- [NIH Data Sharing Brochure \(PDF\)](#) - (05/20/2003) - Printable brochure that summarizes main elements of the NIH Data Sharing Policy.
- [Testimonials \(MS Word\)](#)- (03/05/2003) - First-hand accounts from researchers who have shared data.
- [Other Data Sharing Documents and Resources \(02/19/2004\)](#) - Additional resources relating to data sharing.

## **4.2 The Inter-university Consortium for Political and Social Research<sup>31</sup>**

Created in 1962 at the University of Michigan, The Inter-university Consortium for Political and Social Research's (ICPSR) **core mission is to archive social science data, provide open and equitable access to data, and promote the effective use of data.** ICPSR provides extensive instructions for persons who plan to deposit data in its repository. These excellent instructions include details on how to plan for sharing at the research stage, how to ensure confidentiality, and how to document the data. As its website indicates, ICPSR prefers that investigators take initial responsibility for the removal of all identifiers and other kinds of information that could possibly lead to deductive identification (reidentification) of subjects. Because there is no formula for making this determination, each data set offers its own challenges.

Some ICPSR data collections have restricted-use versions, whereby an interested user must apply or submit a proposal to obtain the data collection. These are usually data that are thought to be so sensitive that public-use versions cannot be made, or that public use versions would require so much alteration that the collection would be rendered useless. Distributing restricted-use data is labour intensive because it requires proposal review and evaluation, as well as monitoring of the use and return of the data. Of ICPSR's 6,300 collections, less than one percent has restricted versions available.

Before accessing any data collection, all ICPSR users are required to electronically sign a pledge that states they will not attempt to identify anyone in the data collection and that if they do so inadvertently, they will provide immediate notification.

## **4.3 Disclosure Review Boards**

DRB members function as disclosure-risk sleuths, and many kinds of knowledges and experience are required if they are to be effective in that role.<sup>32</sup> Within the DRB, there must be diverse membership with intimate knowledge of the particular data products that the agency releases, and of related data products available from elsewhere that might be used for matching purposes, as described above, to re-identify data. Members must fully understand their organization's obligations to preserve both confidentiality, and the organization's reputation for assuring confidentiality. Members must be as knowledgeable as possible of cutting-edge techniques for

preventing re-identification, and aware of emerging computer technology that can be used for harmful purposes to re-identify data.

Much of the DRB deliberation centres around six broad areas:<sup>33</sup>

1. *the justification for providing data users with a high level of detail, and the use to which it would be put;*
2. *the possibility that data needed for certain studies (e.g., studies requiring very detailed codes for medical conditions) can be made available under restricted conditions in data enclaves;*
3. *potential “tradeoffs” between detail for certain variables with correspondingly less detail for other variables with lower analytical priority;*
4. *the degree of difficulty likely to be encountered in the use of given data by an intruder;*
5. *the potential for harm or degree of sensitivity of certain types of data; and*
6. *the types of masking or disclosure protection procedures needed.*

These techniques are not foolproof. DRBs attempt to *reduce risk*, but acknowledge that risk is never completely eliminated. To this end, all users of restricted data are **periodically audited** in order to ensure that all conditions of the restricted data agreement are being met. To date, these audits have not revealed any serious violations of a restricted data agreement.

Four modes of data release are used, with different restrictions placed on the types of variables included in a particular mode, and different restrictions placed on those persons given access to that mode:<sup>34</sup>

1. public data files, freely available to all who register;
2. health data files, with somewhat more stringent registration requirements than the public data files;
3. restricted data files, available only to those who apply and meet strict requirements; and
4. a data enclave that provides access to restricted as well as public data but is strictly monitored to prevent the removal of any electronic or paper output with sensitive information.

#### 4.4 DRB: Managing Risk

All risk cannot be averted from data sharing, but below are some of the general steps in disclosure risk the ICSPR or DRB perform to limit harm to persons while attempting to maintain a robust data set:

*Removing Direct Identifiers is the most obvious disclosure risk procedure, yet its utility is limited<sup>35</sup>. Direct identifiers include names, addresses, employers' names or addresses, relatives' names or addresses, dates, telephone and fax numbers, email addresses, Social Security numbers, medical record numbers, account numbers, photos, and so on. Removing these identifiers renders data anonymous, yet record linkage technology continues to advance. Note: Using data matching software snoopers can access private and internet-based many data sets available today<sup>36</sup> to determine with a high degree of accuracy the identity of some research participants whose data are technically "anonymous" because they were stripped of names, addresses, and other direct identifiers.*

Other disclosure risk procedures include deleting some variables;<sup>37</sup> recoding categorical variables into larger categories; rounding or truncating continuous variables; "masking" outliers; and enlarging geographical areas. Under certain circumstances more statistically sophisticated methods may be employed, such as adding "noise" to the data, data swapping, rank swapping, *coarsening*, *blurring*, *blunting the data*, micro aggregation, and most importantly, the suppression of detailed geographic data. **The more geographic specificity in a public-use file, the more attention must be paid to disclosure risk.** Geographic codes immediately narrow the search for a specific record to the most detailed geography on the file.

The most common disclosure risks are outliers and geographic areas with small populations. From such information it may be easy to pinpoint, for example, which participant is the 15-year old with a PhD, the city mayor, or the 105-year-old man, and from that inference to learn other kinds of sensitive information about the person from the data. However, the very detail that creates the greatest risk also provides insight into differences that are of greatest interest to analysts.

Two useful ballpark techniques are:<sup>38</sup>

- Estimate to assess disclosure risk is based on the combination of key variables, i.e., the size of the population must be more than three times the product of the cross-classified key variables.
- **A general rule is that only geographic areas with populations of 100,000 or more are identified** helps assure that the population from which a study sample or administrative data set is drawn is large enough to support a minimum level of diversity in the sample. (Few NZ cities are of this size.)

Several federal agencies, e.g., the National Center for Health Statistics (NCHS), the U.S. Census Bureau, and the Agency for Health Care Research and Quality, maintain facilities where researchers can analyse data not released to the general public.<sup>39</sup> In these data centres, researchers work under controlled conditions and their research output is subject to disclosure analysis to ensure that what is taken away cannot lead to the reidentification of research participants.

## **Section 5: Australia: The Australian Social Science Data Archive<sup>40</sup>**

In 2001 the Australian Social Science Data Archive (ASSDA), located in the Research School of Social Sciences (RSSH) at The Australian National University (ANU), was incorporated into the ACSPRI Centre for Social Research (ACSR), established through a joint initiative by the Research School of Social Sciences and the Australian Consortium for Social and Political Research Incorporated (ACSPRI). ASSDA collects data files from all parts of Australia, and from many different types of organisations, including universities, market research companies, and government organisations. Since its establishment, ASSDA has collected over 1050 datasets from Australian surveys and opinion polls. ASSDA also holds Australian population Census data and data from other countries within the Asia Pacific region. The uniqueness of ASSDA as a repository for machine-readable data makes it an attractive storage place for many important national surveys. Data stored in the archives can usually be made available for secondary analysis, depending on any access restrictions set by the depositor. The catalogue of ASSDA holdings is available through the NESSTAR facility.

The Australian Social Science Data Archive is a member of the International Federation of Data Organisations (IFDO) through which it maintains contacts with data organisations abroad that are actively engaged in providing the social science community with computerised data and documentation. ASSDA receives newsletters and catalogues from IFDO member archives, and data and codebooks can be obtained from these archives on behalf of Australian researchers upon request.

One of the consequences of depositing data in the Australian Social Science Data Archive is that the data can be made available to other researchers who may wish to analyse them. Data are expensive to collect, and the principal investigator(s) (PIs) often do not have the time or money to analyse all aspects of the data they have obtained. When data are made available through a data archive, maximum utilisation of the data is possible. Other researchers may test different questions on the data, or may provide feedback about the original results. Data sharing is one avenue through which a researcher may develop contacts with others who are studying similar variables.

The conditions under which the data may be made available by the Archive to other researchers are determined by the PIs. The conditions of access may be stated in one of the standard forms below:

- There are no restrictions on access to the data or publication of results, and the depositor does not wish to be informed of the use being made of the data.
- The depositor wishes to be informed by the Archive of use being made of the data, in order to comment on that use and make contact with colleagues of similar interests.
- The user is required to obtain the permission in writing of the original depositor of the data, or an authorised representative, before publishing any interpretation of such materials.
- The depositor wishes to be informed by the Archive of each request to use the data in order to give or withhold permission.
- There are special access conditions to the data set in question.

In addition, an embargo period may be imposed. No access to the data would be permitted until after the date specified by the PIs. At the end of the embargo period, the data may be released under the access conditions set by the PIs.

The PIs can create access conditions to suit their particular concerns about how the data should be used in the future. For example, if the principal investigators told the respondents that the data would only be used for academic purposes, they could stipulate that the data may only be made available to university employees. If they want to ensure that other researchers cannot pre-empt their findings, they can stipulate that the data be made available only after they have had time to publish their results.

## **Section 6: The NZ Situation: Exemplars and Transmission Lines**

### **6.1 Exemplars**

In New Zealand data sharing there are few exemplars of data sharing, yet the infrastructure to promote data sharing does exist. What follows identifies both exemplars and the infrastructure, often legislative, that could promote larger scale data sharing initiatives. To the untrained eye, this essential infrastructure may go undetected. The HRC and Oral Historians are two exemplars.

### **6.2 The Health Research Council**

In 2006 the HRC plans to follow a limited form of the MRC model of funding data sharing initiatives in specific longitudinal projects. Like the MRC, the data sharing directive is not prescriptive, but will be negotiated between the researcher and the ethics committee. Funding specific to data sharing is estimated to be three percent of total costs.

### **6.3 Oral historians**

Oral historians have confronted problems associated with data sharing for some time, and developed protocols for the re-use of “raw” data. Researchers routinely read personal letters and diaries that lay bare the souls of those gone before, with these original documents allowing restricted access to their intimate details.

Confidentiality issues are dealt with at the time of deposit. Institutional custodians can protect the person’s original consent for his or her life story to be shared by researchers. The Oral History Society’s ethical guidelines are a template for those planning to consent the donor and creating standards for access to the data.

Three features are prominent in the NOHANZ code and ethics and serve as the basis for any data sharing guidelines:

1. consent issues are sorted at the time of deposit or collection;
2. funding is provided for the curating and distribution of the data; and
3. the context of the original document is in most cases preserved.

Oral historians are role models!

#### **6.4 Linked Employer-Employee Data<sup>41</sup>**

Other New Zealand examples of data sharing include LEED (Linked Employer-Employee Data tables), created by linking a longitudinal series of Inland Revenue's Employer Monthly Schedule (EMS) to business data from Statistics New Zealand's Business Frame (BF). The dataset covers all industries and people who receive income from which tax is deducted at source. Statistics from LEED measure labour market dynamics, and provide an insight into the operation of New Zealand's labour market.

LEED comprises unit record data that has been processed in accordance with **strict privacy and confidentiality protocols**. For confidentiality purposes, some individuals are withheld from data provided to Statistics NZ by Inland Revenue. Statistics NZ undertakes the data integration, and any information released is in the form of summary statistics or statistical research. No information is released from the data that would allow for the identification of any individual or business. The categories for data release are established so that each cell in a table complies with Statistics NZ confidentiality rules.

#### **6.5 The New Zealand Family Violence Clearinghouse<sup>42</sup>**

The New Zealand Family Violence Clearinghouse is the national centre for collating and disseminating information about domestic and family violence in Aotearoa New Zealand.

#### **6.6 Enhanced data transmission lines**

Various government initiatives have been created to facilitate the transmission of digital data sharing, for example:

- E-government<sup>43</sup> delivers better results by adapting government to the environment of the information age and the Internet.
- BRCSS<sup>44</sup> (Building Research Capability in the Social Sciences, a five-year TEC funded programme) have a BRCSS-Network and focus on the creation of "virtual" research networks/communities.

- The National Library Digital Strategy<sup>45</sup> enhances data sharing. It produces standards of presentation and deals with ethical issues when the local guardian is no longer present.
- KAREN<sup>46</sup> provides high capacity, ultra high-speed connectivity between New Zealand's tertiary institutions, research organisations, libraries, schools and museums, and the rest of the world. Any member can connect through KAREN to any other member or to collaborators on other advanced networks internationally.

## Section 7: Qualitative Research: Challenges and Solutions

In New Zealand the Alexander Turnbull Library and the National Library Digital Strategy both curate data (diaries, oral histories) that are used by qualitative researchers. Restrictions on the use of their data collections are created at the time of depositing the data, promoting both ethical considerations and access to the data's context.

**7.1 Internationally** the data sharing of qualitative data is a present day issue, and a number of non-research projects focussing on the methodological and ethical considerations involved in data sharing have been funded by the ESRC.<sup>47</sup>

Data sharing qualitative research data is inherently problematic given that by its very nature the data is rarely anonymous, whereby the researcher collecting the data in an observation or an interview knows the identity of the participant. In addition, the data is contextual, rich and unique and likely to reveal key aspects of the informant's identity. The ESRC Quads website<sup>48</sup> outlines the unique problems in data sharing qualitative data.

Archived qualitative data is often an unexploited source of research material compared to the better-established tradition in social science of reanalysing quantitative data. Until recently there has not been a well-developed paradigm, or a pervasive research culture of sharing or secondary analysis of qualitative data. Moreover, there is still a noticeable imbalance in attitudes towards sharing and re-using data across disciplines and types of methodological approaches.

Current attempts are being made to render qualitative data suitable for data sharing. The Qualidata Centre,<sup>49</sup> set up in 1994 in the Sociology Department at Essex, complemented the UK Data Archive with a joint mission to actively acquire, curate, disseminate and promote the raw data from social science research. At the UK Data Archive, where some 150 qualitative datasets are catalogued, user figures have soared, particularly for use in research methods teaching. Nevertheless, there are still barriers.

**7.2 The six main perceived barriers<sup>50</sup>** for data sharing qualitative data that have been identified through contact with researchers in the UK over the past ten years can be summarised as follows:

1. The practice of secondary analysis of qualitative data is not yet a commonplace research activity. The literature is not forthcoming on methodological guidance on how to approach the revisiting of data. A cultural shift is required, and this has been progressively happening since 1994.
2. There are problems with the implicit nature of qualitative data collection and analysis, of context and reflexivity, which are sometimes proclaimed to be indefinable.
3. There is a lack of time to get fully acquainted with research materials created by someone else. New ways and tools that more efficiently expose the content and context of digital data sources need to be developed, in order to reduce such researcher burden.
4. Constraints of informed consent mean informed consent must be thought through at the time of research proposal planning and writing, and be tailored towards the specific research questions and the sample.
5. There can be insecurity about the exposure of one's research practice, or threat of misinterpretation.
6. Lack of a wide range of publicly available catalogued qualitative as opposed to quantitative research data.

**7.3 Two major issues appear to be at the heart of making data fully shareable.**

The first is producing rich and full documentation about the data and the research processes used to conceptualise, collect, manage, process and analyse data. Full documentation enables effective resource discovery (i.e. catalogues) of distributed data sources and enables more informed re-use. The second challenge for sharing qualitative data is that of exposing data in the most flexible way possible, so as to enable multiple methods of accessibility and innovative uses, for example, combine and link: activities that are the very core of some of the initial considerations of e-social scientists.

#### **7.4 Emergent Best Practice: Qualitative Archiving and Data Sharing (QUADS) Demonstrator Scheme<sup>51</sup>**

QUADS is the ESRC Qualitative Archiving and Data Sharing Scheme, running from April 2005 until October 2006. The aim of the scheme is to develop and promote innovative methodological approaches to the archiving, sharing, re-use and secondary analysis of qualitative research and data. A range of new models for increasing access to qualitative data resources, and for extending the reach and impact of qualitative studies will be explored.

Five small exploratory projects have been funded together with a co-ordination role:

- **Representing context in a research archive of educational evaluation studies** (18 months) P. Carmicheal, M. James, J. Elliot and D. Bridges (Cambridge & UEA).
- **Smart qualitative data: methods and community tools for data mark-up (SQUAD)** (18 months) L. Corti and C. Grover (Essex and Edinburgh).
- **Negotiating the long view: archiving, representing and sharing a qualitative longitudinal resource** (18 months) S. Henderson, J. Holland and R. Thomson (Southbank).
- **Methodological issues in qualitative data sharing and archiving** (18 months) A. Coffee, B. Dicks and M. Williams (Cardiff).
- **Collating and preserving primary material on the Northern Ireland conflict** (18 months) R. Miller and M. Melaugh (QUB & Ulster).

Four common challenges can be identified for the QUADS scheme.<sup>52</sup> These are: defining and capturing data context; audio-visual archiving; consent, confidentiality and IPR; and web and metadata standards.

- Capturing degrees of context enables informed re-use of data. QUADS aims to devise and recommend a minimum set of contextual constructs that would be necessary to document a collection of qualitative data to enable informed secondary use.
- The archiving and sharing of digital audio-visual data from qualitative research is fairly new. As many of the QUADS projects are handling these kinds of data, the scheme is providing an opportunity to share expertise on presenting and re-using such sources.

- Consent, confidentiality and copyright perhaps provide the greatest challenges for re-using qualitative data.
- In order to approach primary data now and in future years, we need that data to be accurately, richly and contextually described. And in turn, re-presentation of original data, methods and analytic interpretation and their interweaving requires agreed and exemplary standards and procedures. We must also anticipate emerging innovations in qualitative methods, including new data forms, sources, possibilities for research archiving and data mining, and the potential for increased participation and access.

## **Section 8: Risks and opportunities for data sharing in New Zealand: Identifying an ethical infrastructure essential for data sharing**

*(Note to reader) The first draft of this report was critiqued by the Data Saving and Data Sharing Working Group as being sanguine toward the possibility of data sharing in New Zealand. Section 8 provides evidence for this sanguine perspective. It demonstrates both international and national examples of the changing balance in informed consent where the traditional sacrosanct reliance on individual consent is weighed up within a wider social context focusing on the public good.*

*Recent Ministry of Health Guidelines in response to Tissue Banking and the 2001 Gisborne Cervical Cancer inquiry have produced a wider scope for researchers to conduct their research without the direct consent of individuals. These Ministry of Health innovations provide an opportunity for researchers to access identifiable and/or deidentifiable administrative data under specific circumstances. These new guidelines should be read with caution as they may also produce risk (a public outcry) as the data sought by researchers is administrative data gathered compulsorily.*

### **a) Broad Consent (not applicable to administrative data)**

The Ministry of Health is currently reviewing an example of good practice broad consent.<sup>53</sup> The first draft of the **proposed guidelines on use of human tissue for future unspecified research has been released for consultation and the revised guidelines are expected in 2007.**

The proposed guidelines say it is acceptable, with suitable consent, to send human tissue samples overseas to bio-banks for use by research collaborations and for potential use by third parties. New research, particularly in areas such as cancer and genetic diseases, is often reliant on tissue collected for other clinical, diagnostic or research purposes. The Nuffield Council on Bioethics (2003) outlines the benefits of broad consent in *Pharmacogenetics: Ethical issues*:

Allowing broad consent may be of significant benefit to researchers and to society's interest in the acquisition of knowledge about health and disease. Researchers may not be able to predict at the start of a study whether the information gathered may subsequently be useful in additional research. If this proves to be the case, the practical difficulties of contacting participants and obtaining new consent for the use of their data in a different project, perhaps a number of years later, may be prohibitive.

Tissue banking is a form of data sharing, and if the guidelines adopt a workable definition of broad consent, other researchers will be able to adopt similar strategies. Currently New Zealand ethics committees treat informed consent as sacrosanct. Yet here too researchers have some flexibility, as existing legislation permits data sharing without a person's informed consent.

Organ donation is essentially data sharing based on a weak form of broad consent. Even though donor consent information is recorded on the driver's license, at the time of death the family is always asked if they know the wishes of their relative, and for their agreement to the donation of organs and tissues. Note: The 1964 Human Tissue Act is currently being revised to facilitate the donor's broad consent.

Note: An individual can only give broad consent in advance it cannot be given retrospectively. Broad consent only applies to data collected by researchers; it does not refer to administrative data.

#### ***b) Research without Informed Consent***

Health and Disability Ethics Committees routinely permit data sharing **without informed consent**. The most common example of data sharing is personal information data stored by District Health Boards or held by the New Zealand Health Information Service. The exclusion is permitted under recent amendments to the Right 7(10) of the Code of Health and Disability Services Consumers' Rights, providing for limited exceptions to the requirement to obtain explicit re-consent to use body specimens removed as part of a health care procedure:<sup>54</sup>

No body part or bodily substance removed or obtained in the course of a health care procedure may be stored, preserved, or used otherwise than—

- (a) With the informed consent of the consumer;
- (b) Or for the purposes of research that has received the approval of an ethics committee.

The Health and Disability Commissioner has stated<sup>55</sup> that accredited ethics committees may only waive the need for a researcher to obtain individual consent in limited circumstances where it would be significantly problematic. The committee

must be satisfied that the potential public benefit of allowing the research to proceed outweighs the very strong need to protect an individual's autonomy.

*The USA has an equivalent to Right 7(10).* The Health Insurance Portability and Accountability Act of 1996 (HIPAA), the Department of Health and Human Services issued the Privacy Rule, which permits research use of individually identifiable health information without the individual's authorization when an IRB approves a waiver of authorization. Criteria for waiving authorization include the treatment of identifiable data and restrictions on their reuse and disclosure to others.

**c) The Ethical Guidelines for Observational Studies: Observational Research, Audits and Related Activities** (released by the Ministry of Health in December 2006. [www.neac.health.govt.nz/moh.nsf/pagescm/520/\\$File/ethicalguidelines.pdf](http://www.neac.health.govt.nz/moh.nsf/pagescm/520/$File/ethicalguidelines.pdf))

These guidelines recognise observational studies (i.e. non-interventional) are relatively low risk and as such they waiver, in some circumstances, the requirement for researchers to seek informed consent from persons to use their data. In Section 6.40 of these guidelines a traditional form of consent is listed as:

The consent of participants should generally be obtained for using identifiable or potentially identifiable data for research.

In Section 6.41 the scope for the "collection of health information without consent" is widened permitting access without informed consent under the following circumstances.

- a) the procedures required to obtain consent are likely to cause unnecessary anxiety for those whose consent would be sought;
- or the requirement for consent would prejudice the scientific value of the study;
- or it is impossible in practice to obtain consent due to the quantity or age of the records
- and
- b) there would be no disadvantage to the participants or their relatives or to any collectivities involved
- and
- c) the public interest in the study outweighs the public interest in privacy.

The Observational Studies guidelines permit New Zealand Health and Disability Ethics Committees to give ethics approval to researchers accessing administrative health databases e.g. New Zealand Health Information Service (NZHIS) and the Institute of Environmental Science and research (ESR).

*In sum*, the three guidelines outlined above provide opportunities for data saving and data sharing by researchers researching data either they have collected or collected by others with broad consent, or collected as part of routine administrative data collection. These guidelines, like their UK equivalents involve a case-by-case risk rationale.

#### *Essential Further Reading*

W. Lowrance's<sup>56</sup> '*Learning from experience: privacy and the secondary use of data in health research*' provides a useful background both to the opportunities and the risks accessing administrative data in the UK. Lowrance's key suggestion is to deidentify the data making the file an anonymous "case" rather than representing personal data. Lowrance (2002, 31-6) states (*my emphasis*):

A widely relied upon solution has been to allow data to be studied absent express consent if researchers make formal promises to take precautions, use the data judiciously, and not disclose information that can be linked with particular subjects. It is a risk rationale. But this rationale has had to be defended study-by-study, and, at least until recently in many contexts, it has been surrounded by uncertainty as to whether anonymised data are considered still "personal" under law. Against it, too, rights rationales may be raised. Another solution has been to recognise such activities as research on data in properly safeguarded cancer registries or public-sector healthcare reimbursement databases as being clearly in the public interest. But recently this rationale has come under pressure, and the public-interest mandate urgently deserves to be reinforced globally. Probably the default stance at present should remain that whenever consent can reasonably be sought, it should be sought. Urgency, practicality, cost, and other factors should be considered in appraising the "reasonably can be." *The general movement in many countries seems to be towards informing the public/patients*

*in a general though serious way that data about their experience may be studied for a variety of common-good purposes, assuring them that safeguards and governance are in place, then proceeding openly, being responsive to inquiries, and so on. **This would amount to a cultural change.***

## **Section 9: Next Steps for Data Sharing Research Data and Administrative Data**

*Administrative data* collected compulsorily under various acts of parliament (i.e. Statistics, Tax, Social Welfare) and automatically curated by government departments provides both an opportunity and a risk for their Chief Executive Officers, who act as the custodians of the administrative data. The reuse of this administrative data is subject to strict privacy laws, and up until now these laws have been restrictive. In Section 8 c) above the new Ministry of Health's Guidelines on Observational Studies provide a precedent for the reuse of health information collected compulsorily for one purpose and reused for another.

Various next steps are now available. First, a comparison could be made between data held by government departments and the types of health data bases (i.e. NZHIS) referred to in the Observational Studies guidelines. If the Observation Guidelines permit access to these databases featuring compulsory collected data can they be used to access government department administrative data? Second, government department CEO's, acting as the custodians of this administrative information could begin public consultation to debate the reuse of this information for research purposes. The public consultation would need to weigh up the rights of individual privacy against the use of administrative data for a public good. Various supporting oversight regimes were mentioned in Sections 3 and 4.

If, and when, risks to privacy are satisfactorily addressed by departmental heads and the Ministry of Health's Guidelines on Observational Studies are deemed applicable for the reuse of government administrative data this reuse would require the types of best practice oversight (of data sharing of administrative data) detailed in Section 3.3. above i.e. the Performance and Innovation Unit<sup>57</sup> toolkit micro-managed by an independent *Chief Knowledge Officer*. In addition, the use of Disclosure Review Boards (outlined in Section 4.3) permitting data oversight on a case by case basis.

As data sharing guidelines for inter-departmental government emerge, any attempts to practice data sharing in either field needs to be encouraged, *monitored and the experience written up* to contribute to a growing body of knowledge. This scanning

project should be ongoing, updated annually and presented to the Data Saving and Data Sharing working group.

#### *Researcher sourced data*

Data sharing of research data is both simple and complex. Its simplicity is based on a willingness to fund, provide oversight and curate data for future unspecified research; the Alexander Turnbull Library is an example of best practice. The Health Research Council funds some longitudinal research projects where oversight is provided gratis by the Health and Disability Ethics Committees and the individual researcher both curates and cleans the data. However, data saving and data sharing are not embedded in the New Zealand research culture. Current research funders (i.e. FORST, Marsden) do not require researchers to save and share their data or to justify why they are not planning to extract more from the initial investment. To require researchers to both use and reuse the data would be to promote a new culture in the research community going some ways to addressing problems with intellectual property rights. Publically funded data sharing initiatives make the claim that the data belongs to the funder, not the researcher. Following the analysis of the data and the writing up of results the data would be deposited to create further returns on the funders' investment. Part of that funding would provide monies to prepare data for storage. The stumbling block for data saving and data sharing in New Zealand is funding the curation of the research data.

The next steps are

- 1) To identify what agency or University will act to curate data sets along the lines of the University of Michigan's Inter-university Consortium for Political and Social Research (see Section 4.2) providing both a depository and oversight of the research data?
  
- 2) Replicate the Medical Research Council Policy on Data Sharing and Preservation<sup>58</sup> (Section 3.1 above) for all publicly funded research in New Zealand. It states, "all MRC funding proposals must include a strategy for data preservation and sharing. Any applicants who consider that the data arising from their proposals will not be suitable for sharing must provide clear reasons for not making it available. "

- 3) Context-rich qualitative research data sets based on transcriptions of participant interviews have distinct confidentiality problems and are not currently suitable for data sharing. ESRC funded research in the UK is ongoing and it is essential to remain networked to the roll out of the ground breaking Qualitative Archiving and Data Sharing projects to address the four major barriers to successful qualitative data sharing listed in Section 7.4: defining and capturing data context; audio-visual archiving; consent, confidentiality; and web and metadata standards.
- 4) As data sharing guidelines qualitative research emerge, any attempts to practice data sharing needs to be encouraged, *monitored and the experience written up* to contribute to a growing body of knowledge. For example, a New Zealand replication of the best practice Economic and Social Research Council financial investment in *methodological research of qualitative research archiving* would fund independent observation of the ethical arrangements made between the researchers and ethics review committee in the HRC funded longitudinal studies.
- 5) Promote the notion of broad consent as the ethical basis for data sharing of research data.

## Endnotes

---

- <sup>1</sup> <http://www.spear.govt.nz/news/archive-2005-earlier.html>
- <sup>2</sup> de Wolfe et al., “Phase III: Meeting the challenge when data sharing is required”, *IRB: Ethics & Human Research* 28, No. 1 (2006).
- <sup>3</sup> Personal communication with Dr. Bruce Scoggins, CEO, Health Research Council.
- <sup>4</sup> <http://www.nzfvc.org.nz/>
- <sup>5</sup> Ministry of Health. 2006. Guidelines on the Use of Human Tissue for Future Unspecified Research Purposes: Discussion document. Wellington: Ministry of Health. <http://www.moh.govt.nz>
- <sup>6</sup> <http://www.e.govt.nz/>
- <sup>7</sup> <http://www.reannz.co.nz/home>
- <sup>8</sup> <http://www.brccs.net/>
- <sup>9</sup> <http://www.digitalstrategy.govt.nz/>
- <sup>10</sup> W. Rodgers and M. Nolte, “Solving problems of disclosure risk in an academic setting: Using a combination of restricted data and restricted access methods”, *Journal of Empirical Research on Human Ethics* 1(3) (2006): 85 – 98.
- <sup>11</sup> de Wolf et al., “Part I: What is the requirement for data sharing?” *IRB: Ethics & Human Research* 27 No. 6 (2005): 12-16.
- <sup>12</sup> J. O’Rourke, S. Roehrig, S. Heeringa, E. Reed, W. Birdsall, M. Overcashier and K. Zidar, “Solving problems of disclosure risk while retaining key analytical uses of publicly released microdata”, *Journal of Empirical Research on Human Ethics* 1(3) (2006): 63 – 84.
- <sup>13</sup> <http://www.esrc.ac.uk/ESRCInfoCentre/index.aspx>
- <sup>14</sup> <http://assda.anu.edu.au/>
- <sup>15</sup> <http://www.foi.gov.uk/sharing/index.htm>
- <sup>16</sup> <http://quads.esds.ac.uk>
- <sup>17</sup> de Wolf et al, “Part II: HIPAA and Disclosure Risk Issues”, *IRB: Ethics & Human Research* 28 No. 1 (2006): 6-11.
- <sup>18</sup> Rodgers and Nolte, (2006): 85-98.
- <sup>19</sup> A. Zarate and L. Zarate, “Essentials of the disclosure review process: A federal perspective,” *Journal of Empirical Research on Human Ethics* 1(3), (2006): 51 – 62.
- <sup>20</sup> Ministry of Health. 2006.
- <sup>21</sup> *Tri-Council Policy Statement: Ethical Conduct for Research Involving Humans*, published jointly by the Medical Research Council of Canada, the Natural Sciences and Engineering Research Council of Canada, and the Social Sciences and Humanities Research Council of Canada (1998). [http://www.ncehr-cnerh.org/english/code\\_2/](http://www.ncehr-cnerh.org/english/code_2/)
- <sup>22</sup> *National Statement of Ethical Conduct in Research Involving Humans*. Australian National Health and Medical Research Council (2006) [www.nhmrc.gov.au/ethics/human/ahec/projects/statementsec.htm](http://www.nhmrc.gov.au/ethics/human/ahec/projects/statementsec.htm)

---

<sup>23</sup> <http://www.dpconline.org/docs/events/0607thorley2.pdf>

<sup>24</sup> The ethical dilemma facing many government researchers when data sets are within easy reach may have a historic precedent. For example, the beginning of the eighteenth century saw the emergence of grave robbing, which, with the passage of time, became by far the most significant means of getting bodies. The earliest grave robbers were surgeon-anatomists or their pupils, and there was often a close liaison between them and the body snatchers, with the latter providing several thousand bodies annually. <http://www.asa3.org/ASA/topics/ethics/PSCF3-95Jones.html>

<sup>25</sup> Zarate and Zarate, “Essentials of the disclosure review process: A federal perspective” *Journal of Empirical Research on Human Ethics*. 1(3) (2006): 51 – 62.

<sup>26</sup> <http://www.dpconline.org/docs/events/0607thorley2.pdf>

<sup>27</sup> [http://www.mrc.ac.uk/strategy-data\\_sharing\\_policy.htm](http://www.mrc.ac.uk/strategy-data_sharing_policy.htm)

<sup>28</sup> <http://www.esds.ac.uk/news/eventdocs/bestmewrmar06.ppt>

<sup>29</sup> <http://www.dpconline.org/docs/events/0607thorley2.pdf>

<sup>30</sup> <http://www.foi.gov.uk/sharing/index.htm>

<sup>31</sup> <http://www.icpsr.umich.edu/>

<sup>32</sup> Zarate and Zarate, (2006): 51-62.

<sup>33</sup> Ibid, pp. 51-62.

<sup>34</sup> O’Rourke et al. (2006): 63 – 84.

<sup>35</sup> de Wolf et al. “Part I: What Is the Requirement for Data Sharing?” *IRB: Ethics & Human Research* 27 No. 6 (2005): 12-16.

<sup>36</sup> Ibid, pp. 12-16.

<sup>37</sup> O’Rourke et al. (2006): 63 – 84.

<sup>38</sup> Ibid, pp. 63–84.

<sup>39</sup> de Wolf et al. “Part I: What is the requirement for data sharing?” *IRB: Ethics & Human Research* 27 No. 6 (2005): 12-16.

<sup>40</sup> <http://assda.anu.edu.au/>

<sup>41</sup> [www.stats.govt.nz/products-and-services/table-builder/leed-table-builder.htm](http://www.stats.govt.nz/products-and-services/table-builder/leed-table-builder.htm)

<sup>42</sup> <http://www.nzfvc.org.nz/>

<sup>43</sup> <http://www.e.govt.nz/>

<sup>44</sup> <http://www.brcss.net/>

<sup>45</sup> <http://www.digitalstrategy.govt.nz/>

<sup>46</sup> Ibid.

<sup>47</sup> <http://www.esrc.ac.uk/ESRCInfoCentre/index.aspx>

<sup>48</sup> <http://quads.esds.ac.uk>

---

<sup>49</sup> <http://www.esds.ac.uk/qualidata/about/introduction.asp>

<sup>50</sup> <http://iassistdata.org/publications/iq/iq29/iqvol293.pdf>

<sup>51</sup> [http://www.esrcsocietytoday.ac.uk/ESRCInfoCentre/opportunities/Commissioning\\_updates/index80.aspx?ComponentId=8266&SourcePageId=5940](http://www.esrcsocietytoday.ac.uk/ESRCInfoCentre/opportunities/Commissioning_updates/index80.aspx?ComponentId=8266&SourcePageId=5940)

<sup>52</sup> <http://www.esrcsocietytoday.ac.uk/ESRCInfoCentre/research/resources/QUADS.aspx?ComponentId=13169&SourcePageId=919>

<sup>53</sup> Ministry of Health. 2006.

<sup>54</sup> Paragraph 32 of the Operational Standard for Ethics Committees is presently being updated to reflect the amendment to Right 7(10) of the Code of Health and Disability Services Consumers' Rights.

<sup>55</sup> Ron Patterson, *New Zealand Doctor*, 30 June, 2004.

<sup>56</sup> <http://www.nuffieldtrust.org.uk/publications/detail.asp?id=0&PRid=45>

<sup>57</sup> <http://www.foi.gov.uk/sharing/index.htm>

<sup>58</sup> [http://www.mrc.ac.uk/strategy-data\\_sharing\\_policy.htm](http://www.mrc.ac.uk/strategy-data_sharing_policy.htm)