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;

*   **re-identifiable data**, from which identifiers have been removed and replaced by a code, but it remains possible to re-identify a specific individual by, for example, using the code or linking different data sets;

*   **non-identifiable data**, which have never been labelled with individual identifiers or from which identifiers have been permanently removed, and by means of which no specific individual can be identified. A subset of non-identifiable data are those that can be linked with other data so it can be known that they are about the same data subject, although the person's identity remains unknown.

This National Statement avoids the term 'de-identified data', as its meaning is unclear. While it is sometimes used to refer to a record that cannot be linked to an individual ('non-identifiable'), it is also used to refer to a record in which identifying information has been removed but the means still exist to re-identify the individual. When the term 'de-identified data' is used, researchers and those reviewing research need to establish precisely which of these possible meanings is intended.

Tissue and data

With advances in genetic knowledge and data linkage, and the proliferation of tissue banks of identified material, human tissue samples should always be regarded as, in principle, re-identifiable.

The increased ability to link data has greatly enhanced the contribution that collections of data can make to research, as it enables researchers to match individuals in different data sets without being able to identify the person. For example, in epidemiological research (concerned with the study of populations), information about individuals and groups may be collected so that features of groups of people can be investigated. These data may or may not have originally been obtained for research purposes.

Banking

While most data are collected, aggregated and stored for a single purpose or activity. Permission may sometimes be sought from participants to 'bank' their data for possible use in future research projects.

'Banked' data may be deposited in a warehouse, similar to an archive or library, and aggregated over time. The Australian Social Science Data Archive, for example, collects computer-readable data on social, political and economic affairs and makes them available for further analysis. Archived data can usually be made available for secondary analysis, unless access is constrained by restrictions imposed by the depositor/s.

Use of the National Statement 's values and principles

The values and principles of this National Statement apply to data collection by researchers, and by others whom they authorise to collect data or to whom they outsource the collection.

These ethical principles for the use of databanks should be applied in the guidelines and procedures established by institutions for the setting up of data collections.

Values, principles and themes that must inform the design, ethical review and conduct of all human research are set out in Sections 1 and 2 of this National Statement. The guidelines and headings below show how those values, principles and themes apply specifically in research that is the subject of this chapter.

GUIDELINES

Research merit and integrity

3.2.1   When planning a databank, researchers should clearly describe how their research data will be collected, stored, used and disclosed, and outline how that process conforms to this National Statement, particularly the requirements for consent set out in paragraphs 2.2.14 to 2.2.18.

3.2.2   To promote access to the benefits of research, such data should be collected, stored and accessible in such a way that they can be used in future research projects.

Data usage

3.2.3   Researchers' use of data from databanks must comply with conditions specified by the providers of the data; in particular, any conditions on the identifiability of the data (see paragraphs 2.2.14 to 2.2.18).

3.2.4   Where research involves linkage of data sets, approval may be given to the use of identifiable data to ensure that the linkage is accurate, even if consent has not been given for the use of identifiable data in research. Once linkage has been completed, identifiers should be removed from the data to be used in the research unless consent has been given for its identifiable use.

3.2.5   It is the duty of the custodian to ensure that the data are used responsibly and respectfully, and that the privacy of participants is safeguarded.

3.2.6   Whenever research using re-identifiable data reveals information that bears on the wellbeing of participants, researchers have an obligation to consider how to make that information available to the participants. Where individual notification is warranted, the custodian of the data will need to take all reasonable steps to re-identify those data.

3.2.7   In most situations, the custodian of data will be the individual researcher or agency who collected the information, or an intermediary such as a data warehouse that manages data coming from a number of sources. In some cases, an independent custodian may be necessary. For example, when coded data are stored in a databank, a custodian independent of both the data collectors and the researchers may be appointed, to maintain the data in coded form while enabling individual participants to access their own identified results or data.

3.2.8   Some uses of data in a databank may be detrimental to people to whom the data relate. Researchers and/or custodians should consider denying or restricting access to some or all of the data for those uses.

Consent

3.2.9   When collecting data for deposit in a databank, researchers should provide clear and comprehensive information about:

(a)   the form in which the data will be stored (identifiable, re-identifiable, non-identifiable);

(b)   the purposes for which the data will be used and/or disclosed; and

(c)   whether they will seek:

(i)   specific, extended or unspecified consent for future research (see paragraphs 2.2.14 to 2.2.16); or

(ii)   permission from a review body to waive the need for consent (see paragraphs 2.3.5 and 2.3.6).

3.2.10 Researchers should recognise that data stored in an identifiable form cannot be used in research that is exempt from ethical review.

3.2.11 Any restrictions on the use of participants' data should be recorded and the record kept with the collected data so that it is always accessible to researchers who want to access those data for research.

3.2.12 Researchers and custodians of the databank should observe any confidentiality agreement about stored data with the participant, and custodians should take every precaution to prevent the data becoming available for uses to which participants did not consent.