Australian Government
Australian Institute of Health and Welfare

An AIHW framework for assessing data sources for population health monitoring

Working paper

Australian Government
Australian Institute of Health and Welfare

Authoritative information and statistics to promote better health and wellbeing

An AIHW framework for assessing data sources for population health monitoring

Working paper

Australian Institute of Health and Welfare
Canberra
Cat. no. PHE 180

The Australian Institute of Health and Welfare is a major national agency which provides reliable, regular and relevant information and statistics on Australia's health and welfare. The Institute's mission is *authoritative information and statistics to promote better health and wellbeing*.

© Australian Institute of Health and Welfare 2014

This product, excluding the AIHW logo, Commonwealth Coat of Arms and any material owned by a third party or protected by a trademark, has been released under a Creative Commons BY 3.0 (CC-BY 3.0) licence. Excluded material owned by third parties may include, for example, design and layout, images obtained under licence from third parties and signatures. We have made all reasonable efforts to identify and label material owned by third parties.

You may distribute, remix and build upon this work. However, you must attribute the AIHW as the copyright holder of the work in compliance with our attribution policy available at <www.aihw.gov.au/copyright/>. The full terms and conditions of this licence are available at <http://creativecommons.org/licenses/by/3.0/au/>.

Enquiries relating to copyright should be addressed to the Head of the Digital and Media Communications Unit, Australian Institute of Health and Welfare, GPO Box 570, Canberra ACT 2601.

A complete list of the Institute's publications is available from the Institute's website <www.aihw.gov.au>.

ISBN 978-1-74249-648-1

Suggested citation

Australian Institute of Health and Welfare 2014. An AIHW framework for assessing data sources for population health monitoring: working paper. Cat. no. PHE 180. Canberra: AIHW.

Australian Institute of Health and Welfare

Board Chair
Dr Mukesh Haikerwal AO

Director
David Kalisch

Any enquiries about or comments on this publication should be directed to:
Digital and Media Communications Unit
Australian Institute of Health and Welfare
GPO Box 570
Canberra ACT 2601
Tel: (02) 6244 1032
Email: info@aihw.gov.au

This publication is printed in accordance with ISO 14001 (Environmental Management Systems) and ISO 9001 (Quality Management Systems). The paper is sourced from sustainably managed certified forests.

Please note that there is the potential for minor revisions of data in this working paper. Please check the online version at <www.aihw.gov.au> for any amendments.

Contents

Acknowledgments ..... **iv**
Abbreviations ..... **v**
Summary ..... **vi**
Introduction **1**
The AIHW's assessment framework **2**
Key elements of data quality 3
1. Institutional environment 4
2. Relevance 4
3. Timeliness 5
4. Accessibility 5
5. Interpretability 5
6. Accuracy 6
7. Coherence 6
Data quality assessment template 7
References **9**
List of tables 10
List of figures 10
List of boxes 10

Acknowledgments

This working paper was written by Ann Hunt, Peta Craig and Louise York from the National Centre for Monitoring Arthritis and other Musculoskeletal Conditions at the Australian Institute of Health and Welfare. It was reviewed by Lisa McGlynn, Justine Boland and Pamela Kinnear.

The paper was also prepared under the guidance of the Arthritis and Other Musculoskeletal Conditions Advisory Group whose members are: Lyn March (Chair), Flavia Cicuttini, Robert Cumming, Peter Ebeling, Ann Taylor, Pam Webster, Mellick Chehade, Chris Maher, Paul Hodges and Tania Winzenberg.

The project was funded by the Department of Health.

Abbreviations

ABS	Australian Bureau of Statistics
AIHW	Australian Institute of Health and Welfare
DQS	data quality statement

Summary

This paper outlines the Australian Institute of Health and Welfare's (AIHW) assessment framework for determining the suitability of specific data sources for population health monitoring.

AIHW's Assessment Framework

When identifying potential data sources for population health monitoring, it is important to ensure they are 'fit-for-purpose'. The AIHW has developed a 3-step process to assess potential data sources for population health monitoring:

- **Step 1** collects information about the data source
- **Step 2** identifies the potential to inform key monitoring areas
- **Step 3** assesses the quality of the data, using a modified version of the Australian Bureau of Statistics (ABS) *Data Quality Framework* (ABS 2009), to determine its 'fitness-for-purpose' by establishing its utility, strengths and limitations.

The assessment framework has been designed for use by the AIHW and others with an interest in assessing new data sources for use in population health monitoring. With adaptation, it may also have wider applications in other sectors or subject areas.

For an example of the application of the assessment framework, see the AIHW working paper *Assessment of the Australian Rheumatology Association Database for national population health monitoring* (AIHW 2014a).

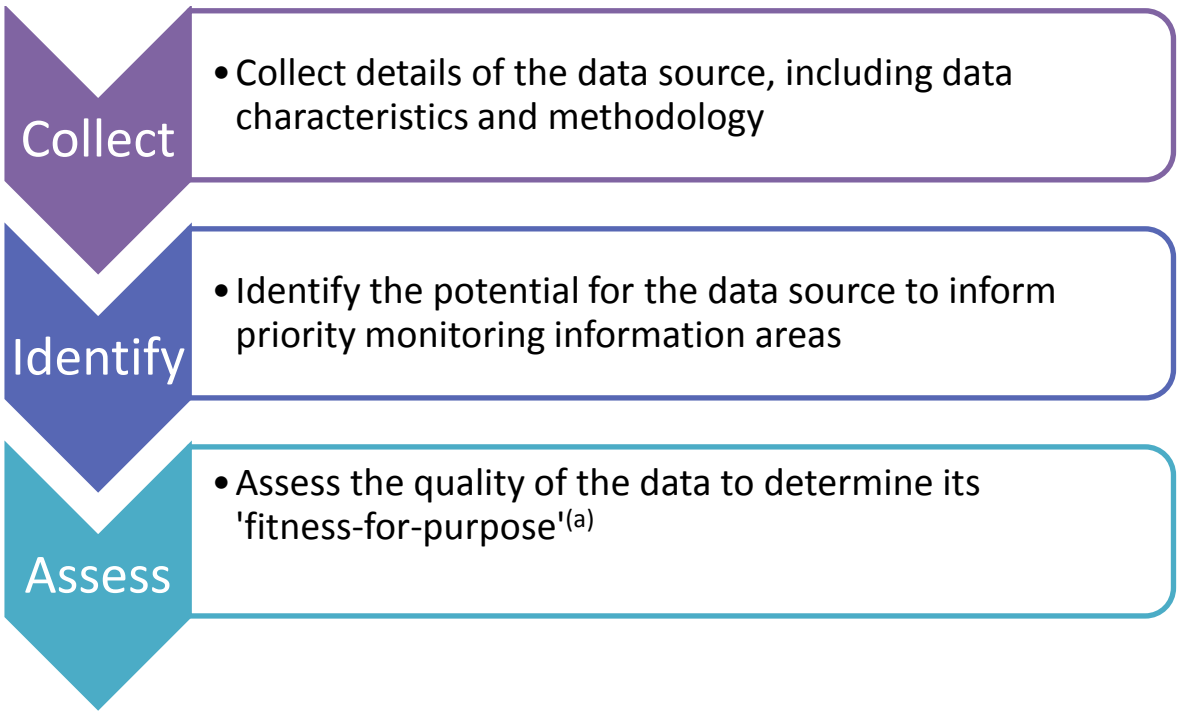
Introduction

Core roles for the AIHW include exploring the usefulness of potential new data sources for national monitoring purposes, driving increased transparency and standardisation of existing national health datasets, and developing standards for new datasets.

There is a growing interest in viewing data as a strategic resource and in making the most of existing data collections (Breunig 2012). This is consistent with government strategies aimed at unlocking government information assets to benefit the broader community (Department of Finance and Deregulation 2009). The broader usage of data collections has the potential to improve efficiencies, lower costs, release stored information and provide a more comprehensive picture of a condition or a disease.

The AIHW has developed a 3-step process to assess the utility of potential data sources for population health monitoring (as outlined in Figure 1). The first step is to collect information about the data source, the second step is to identify the potential to inform key monitoring areas, and the third step is to assess the data quality to determine its 'fitness-for-purpose'.

This framework has been designed to support assessment of potential datasets and to identify their benefits and uses for national chronic disease monitoring.



(a) Adapted from the ABS's *Data Quality Framework* (ABS 2009).

Figure 1: The AIHW's 3-step process for assessing data sources for population health monitoring

The AIHW's assessment framework

The first step of the assessment process collects information about the data source, using categories as described in the template at Table 1. This includes:

- data source characteristics, such as the type, purpose, description, management and contact details for the data collection
- methodology, including the scope, geographic coverage and frequency/timing of the data source.

Table 1: Data source information template (for use in Step 1 of the AIHW's framework for assessing data sources for population health monitoring)

Full name of survey or data collection	
Type of data source	Includes survey type (registry or administrative) and scope (national, state or regional).
Brief description	Brief outline of data source and information relevant for monitoring musculoskeletal conditions.
Purpose(s)	Main stated purpose or purposes of the data source.
Collection methodology	Key features of the collection methodology (administrative or survey) and data collection method (CATI, self-completion, administrative).
Scope (theoretical coverage of relevant population)	Population that is potentially covered, noting the inclusion of an Indigenous identifier (if appropriate).
Coverage (actual)	Actual population covered (response rate and if applicable, the retention rate).
Geographic coverage	National, state or other, noting if there is adequate coverage for remote areas, non-English speakers and so forth.
Frequency/timing	Year(s) in which data have been collected.
Basic collection count	For example: treatment episodes, separations and so forth.
Size	Sample size or number of records in most recent reference period.
Collection management organisation	The organisation chiefly responsible for collecting and managing the data.
Further information	A web link with further information.

The second step determines the potential for the data source to inform priority information areas for monitoring, such as information on risk factors, incidence, prevalence, prevention, management, treatment, quality of life, disability, mortality and health expenditure (see template at Table 2). The monitoring areas identified for different data sources potentially vary, depending on the condition or disease under investigation and the focus of the monitoring work. The information areas included in Table 2 have been selected for monitoring arthritis and other musculoskeletal conditions (AIHW 2014b).

Table 2: Priority monitoring information area template (for use in Step 2 of the AIHW's framework for assessing data sources for population health monitoring)

Full name of survey or data collection		
Priority information area	**Details**	**Data available**
Risk factors	Modifiable and not.	
Prevalence and incidence	Prevalence and incidence, injury and severity.	
Prevention, treatment and management	Prevention, treatment and management (in general practice, other primary health care, specialist and hospital settings and medication use).	
Quality of life	Includes pain, disability, functioning, problems at school, work disability, loss of productivity, social participation, and mental health, carer impacts.	
Disability and death	Disability, death and burden of disease (as a summary measure of these two factors combined).	
Expenditure, costs	To the individual, family members or carers, or the health system.	
Population demographics	For example: age, sex, location, Indigenous status, marital status, personal identifiers for data linkage (if relevant).	

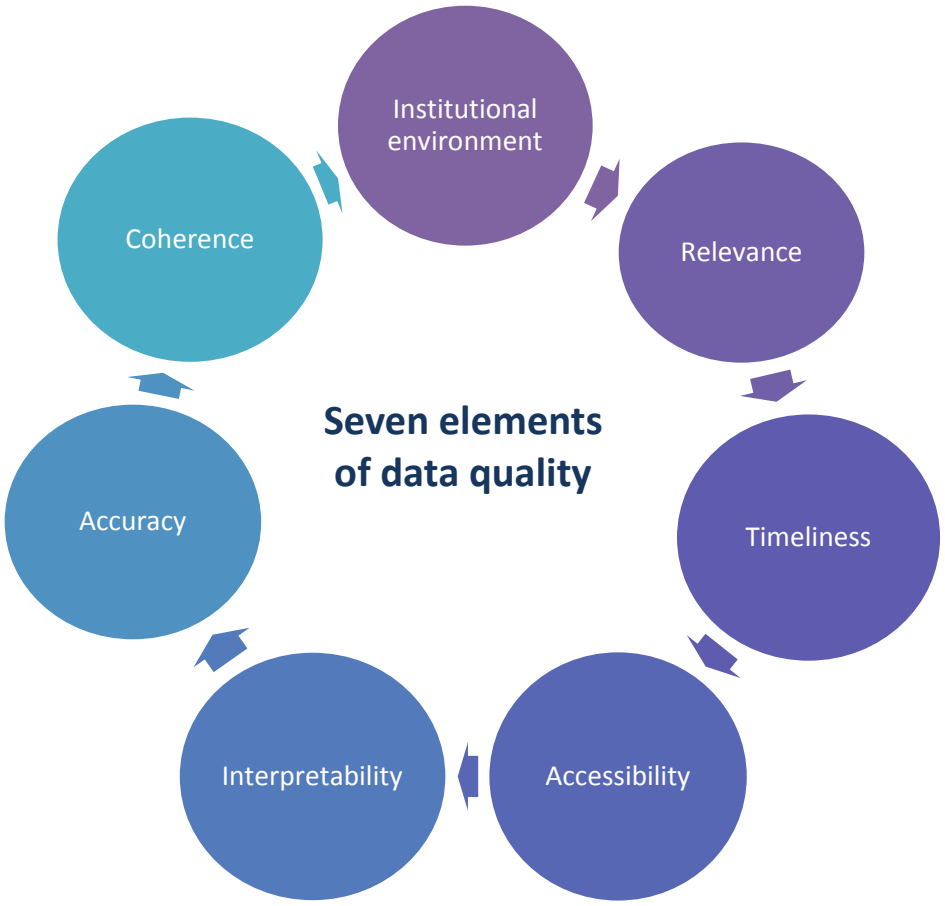
The third step assesses the quality of the data to determine if it is 'fit-for-purpose'. This step determines the utility of the data source, allowing an 'on-balance' assessment considering the relative strengths and limitations of information available across all of the priority information areas. The methodology for assessing data quality is based on the Australian Bureau of Statistics (ABS) *Data Quality Framework* (ABS 2009).

Key elements of data quality

Step 3 of the AIHW's assessment framework consists of seven elements of data quality, as outlined in Figure 2. These elements include the institutional environment, timeliness, accessibility, interpretability, relevance, accuracy and coherence. These elements of data quality are derived from, or consistent with, the following frameworks:

- the ABS *Data Quality Framework* (ABS 2009)
- the AIHW *Data quality statements policy and guidelines* 2011 (internal document)
- the Australian Commission on Safety and Quality in Health Care *Strategic principles for a national approach to Australian clinical quality registries* (ACSQHC 2010)
- the Canadian Institute for Health Information *Data Quality Framework* (CIHI 2009).

All seven elements of data quality are considered as part of the assessment process, but not necessarily weighted equally. The weight given to each element is largely influenced by the intended end use for the data. When assessing data quality, many of the individual elements interrelate with the other elements. For example, the age of a data source (an aspect of timeliness) may impact on its applicability (an aspect of relevance).



Source: Adapted from the ABS's *Data Quality Framework* (ABS 2009).

Figure 2: Seven elements for assessing data quality

1. Institutional environment

The first element of the framework considers the origin of the data collection and the arrangements under which the collection is governed and administered. This enables an assessment of the context, which may influence the validity, reliability or appropriateness of the data. The ABS identified 6 core aspects to consider when assessing the institutional environment (Box 1).

2. Relevance

The second element of the framework considers how well the data source meets the needs of users in terms of the concepts measured, and the populations represented. It is important to identify the purpose of the data collection and determine if the data measures the concept required by the user. For example, if someone wanted to investigate the number of very tall people living in Australia and the intended data source did not include measurement of an individual's height, it would be deemed unsuitable.

It is also important to assess the representativeness of the data source by determining who is included and who is excluded in the data source, and the impact of excluding particular individuals, groups or locations. Using the example above, if the data source only included

children, it would not be representative of all very tall people living in Australia and so would be deemed unsuitable for a study of this group.

Box 1: Core aspects for assessing the institutional environment supporting a data collection

- **Impartiality and objectivity:** whether the production and dissemination of data are undertaken in an objective, professional and transparent manner.
- **Professional independence:** the extent to which the agency producing statistics is independent from other policy, regulatory or administrative departments and bodies, as well as from private sector operators, and potential conflict of interest.
- **Mandate for data collection:** the extent to which administrative organisations, businesses and households, and the public at large, may be compelled by law to allow access to, or to provide data to, the agency producing statistics.
- **Adequacy of resources:** the extent to which the resources available to the agency are sufficient to meet its needs in terms of the production or collection of data.
- **Quality commitment:** the extent to which processes, staff and facilities are in place for ensuring the data produced are commensurate with their quality objectives.
- **Statistical confidentiality:** the extent to which the privacy of data providers (households, enterprises, administrations and other respondents), and the confidentiality of the information they provide, are guaranteed (if relevant).

Source: Modified from the ABS's *Data Quality Framework* (ABS 2009).

3. Timeliness

The third element of the framework considers the reference period to which the data relate, and the frequency with which the data are provided and published. Timeliness reflects the length of time between the availability of the data source and the event or phenomenon it describes. It is an important consideration when assessing the quality and usefulness of data, as in most instances the data need to be 'recent enough' to reflect the current situation.

4. Accessibility

The fourth element of the framework considers users' ease of access to data and supporting materials, including the ability of users to identify relevant information in a convenient and suitable manner. The cost of the data source is also an aspect of accessibility for some users. Other aspects of accessibility include ease of access to relevant publications and reports and data, as well as information on detailed data or unit-record data.

5. Interpretability

The fifth element of the framework refers to the availability of information to help provide insights into the data. This includes information about the variables used; the availability of metadata (data that describes other data); the measures of accuracy; and the concepts and classifications used. Interpretability is an important component of quality as it helps information to be better understood and used appropriately.

6. Accuracy

The sixth element of the framework refers to how well the data correctly describe the condition it was designed to measure. This element of quality relates to how well the data portray reality and has implications for how useful and meaningful the data are for interpretation or further analysis. Accuracy should be assessed in terms of the major sources of errors that potentially cause inaccuracy (see Box 2). Factors that impact on the overall validity of the information for users should be described.

Box 2: Major sources of errors

- **Coverage error:** occurs when a unit in the data is incorrectly excluded, included or duplicated in the data.
- **Response error:** a type of error caused by records being intentionally or accidentally inaccurate or incomplete. This occurs not only in statistical surveys, but also in administrative data collection where forms are not well understood by respondents.
- **Non-response error:** incomplete information for a record (that is, when some data are missing). The use of any imputation strategies should be noted (where values are assigned for missing data).
- **Sample error:** where sampling is used, the impact of sample error can be assessed using information about the sample design, total sample size and the size of the sample in key output levels. For sample surveys, response rates should be provided and, where applicable, retention rates.
- **Other error sources:** includes errors caused by incorrect processing of data; rounding errors involved during collection, processing or dissemination; high variability (or instability) of data due to small numbers; and other quality-assurance processes.

Source: Modified from the ABS's *Data Quality Framework* (ABS 2009).

7. Coherence

The seventh element of the framework refers to the internal consistency of a data collection, as well as how it compares with other sources of information, within a broad analytical framework and over time. The use of standard concepts, classifications and target populations promotes coherence, as does the use of common methodologies across collections. Coherence does not necessarily imply complete numerical consistency, but rather consistency in methods and collection standards. See Box 3 for aspects of coherence to consider when assessing data quality.

Box 3: Aspects of coherence for consideration

- **Changes to data items:** the extent to which particular data items might be available over time, or if significant changes have occurred to the way data are collected.
- **Comparison across data items:** the capacity to make meaningful comparisons across multiple data items within the same collection. The ability to make comparisons may be affected if there have been significant changes in collection, processing or estimation methodology which might have occurred across multiple items within a collection.
- **Comparison with previous releases:** the extent of any significant changes in collection, processing or estimation methodology in this release compared with previous releases, or any 'real world' events which could have impacted on the data since the previous release.
- **Comparison with other products available:** the extent to which other similar data sources are comparable or 'tell the same story'. This aspect may also include identification of any other data sources with which the data cannot be compared, and the reasons for this, such as differences in scope or definitions.

Source: Modified from the ABS's *Data Quality Framework* (ABS 2009).

Data quality assessment template

Completion of the data quality assessment is the third and last step to enable an informed judgement on the utility of a potential data source for population health monitoring. The assessment framework presented here applies the seven data quality elements to meet Step 3 of the framework. These seven data quality elements are consistent with dimensions of the AIHW's *Data Quality Statement* (DQS) and the ABS's *Data Quality Framework* (ABS 2009). The AIHW requires the production of a DQS for every data collection for which it acts as data custodian and a DQS must be included in all releases which draw on these data. Through this, users are informed of data limitations and can make informed judgements about the use of the data.

A series of questions has been developed to address each of the data quality elements noted above (see Table 3). When assessing each data quality element, it is important to work in close cooperation with the key contact officers for the data source to ensure the conclusions drawn are accurate and valid.

To maximise the utility of evidence gathered during this assessment process, and the transparency of the assessment, the template at Table 3 will generally be used in conjunction with additional commentary to describe the assessment rating against a three-point scale ('Yes', 'Partially' or 'No'). Once the individual data quality elements have been analysed, it is then possible to make an overall assessment of the data's quality, considering both the relative strengths and limitations of the data source. The overall assessment is often an 'on balance' assessment considering the relative strength of information across all of the priority information areas, plus any quality issues relating to the data source.

Table 3: Data quality assessment summary template (for use in Step 3 of the AIHW's framework for assessing data sources for population health monitoring)

Data quality element	Key question	Yes	Partially	No
Institutional environment	Is the production and dissemination of data undertaken in an objective, professional and transparent manner?			
	Is the agency producing the data independent and free from potential conflicts of interest?			
	Are there sufficient resources for the collection and production of the data?			
	Are there processes, staff and facilities in place to ensure data quality?			
	Does the data source comply with privacy and legislative requirements for managing data?			
Relevance	Does the data collection measure the concept identified by the end user of the data?			
	Is the data source representative of the target population identified by the end user of the data?			
Timeliness	Are the data up-to-date and current?			
	Are there likely to be subsequent surveys or data collections?			
Accessibility	Are there processes in place to facilitate data access (e.g. Ethics Committee where appropriate; data transmission arrangements)			
	Can the data source be provided in a timeframe suited to the user's requirements?			
	Are the data available in suitable formats?			
	Are data available at costs affordable for the user's project?			
Interpretability	Is metadata available to support correct interpretation of the data?			
Accuracy	Do the data reflect the condition or situation it was designed to measure?			
	Are potential or acknowledged sources of error described?			
Coherence	Does the data source use standard concepts, classifications and target populations?			
	Does the data source use methodologies comparable with other data collections?			

Source: Adapted from the ABS's *Data Quality Framework* (ABS 2009).

References

ABS (Australian Bureau of Statistics) 2009. ABS Data Quality Framework, May 2009. ABS cat. no. 1520.0. Canberra: ABS. Viewed 17 March 2014, <http://www.abs.gov.au/ausstats/abs@.nsf/Latestproducts/1520.0Main%20Features2May%202009?opendocument&tabname=Summary&prodno=1520.0&issue=May%202009&num=&view=>.

ACSQHC (Australian Commission on Safety and Quality in Health Care) 2010. Strategic principles for a national approach to Australian clinical quality registries. Sydney: ACSQHC. Viewed 19 March 2014, <http://www.safetyandquality.gov.au/wp-content/uploads/2012/03/Strategic-and-Operating-Principles-for-Australian-Clinical-Quality-Registries-AHMC-endorsed-Nov-2010.pdf>.

AIHW 2014a. Assessment of the Australian Rheumatology Association Database for national population health monitoring: working paper. Cat. no. PHE 181. Canberra: AIHW.

AIHW 2014b. Data sources for monitoring arthritis and other musculoskeletal conditions. Arthritis series no. 19. Cat. no. PHE 175. Canberra: AIHW.

Breunig R 2012. Small-scale survey data collection: improved coordination of Commonwealth government activities and more efficient data usage. Canberra: Australian National University.

CIHI (Canadian Institute for Health Information) 2009. The CIHI Data Quality Framework 2009. Ottawa, Ontario: CIHI.

Department of Finance and Deregulation 2009. National Government Information Sharing Strategy. Canberra: Department of Finance and Deregulation. Viewed 16 September 2014, <http://www.finance.gov.au/files/2012/04/ngiss.pdf>.

List of tables

Table 1: Data source information template (for use in Step 1 of the AIHW's framework for assessing data sources for population health monitoring) 2

Table 2: Priority monitoring information area template (for use in Step 2 of the AIHW's framework for assessing data sources for population health monitoring) 3

Table 3: Data quality assessment summary template (for use in Step 3 of the AIHW's framework for assessing data sources for population health monitoring)....8

List of figures

Figure 1: The AIHW's 3-step process for assessing data sources for population health monitoring....1

Figure 2: Seven elements for assessing data quality 4

List of boxes

Box 1: Core aspects for assessing the institutional environment supporting a data collection....5

Box 2: Major sources of errors 6

Box 3: Aspects of coherence for consideration....7

When identifying potential data sources for population health monitoring, it is important to ensure they are 'fit-for-purpose'. This working paper outlines the Australian Institute of Health and Welfare's 3-step process used to assess potential data sources for population health monitoring purposes.