Data Management in the APS

Canberra ACT
22 June 1998

Dear Madam President
Dear Mr Speaker

The Australian National Audit Office has undertaken a
performance audit of selected Departments in
accordance with the authority contained in the *Auditor-
General Act 1997.* I present this report of this audit, and
the accompanying brochure, to the Parliament. The report
is titled *Data Management in the APS.*

Yours sincerely

P. J. Barrett
Auditor-General

The Honourable the President of the Senate
The Honourable the Speaker of the House of Representatives
Parliament House
Canberra ACT

Audit Team

Malisa Golightly

Ann Thurley

Alistair Sands

Bill Danaher

Joe Eliasaf

Contents

Abbreviations

ABS	Australian Bureau of Statistics
ACS	Australian Customs Service
AED	Analysis and Evaluation Division of DEETYA
ANAO	Australian National Audit Office
APS	Australian Public Service
AUSSMED	Australian Small and Medium Enterprise Database
AVCC	Australian Vice-Chancellors Committee
BLS	Business Longitudinal Study
BTR	Bureau of Tourism Research
CATI	Computer aided telephone interview
CDD	Corporate Data Dictionary
CDSSU	Corporate Data and Statistical Services Unit of ACS
CELTA	Contracted Entry Level Training Agencies
CEO	Chief Executive Officer
CIMP	Corporate Information Management Plan
CLEAR	Customs Control and Clearance of Lodged Entries
CNCC	Customs National Consultative Committee
DEETYA	Department of Employment, Education, Training and Youth Affairs
DIMA	Department of Immigration and Multicultural Affairs
DIST	Department of Industry, Science and Tourism
DTRD	Department of Transport and Regional Development
DWRSB	Department of Workplace Relations and Small Business
HECS	Higher Education Contribution Scheme
ICSE	Integrated Client Service Environment
IMSC	Information Management Steering Committee
KRA	Key result area
LSIA	Longitudinal survey of immigrants to Australia
NSP	National Survey Program
OGIT	Office of Government Information Technology
OSB	Office of Small Business
PC	Personal computer
QA	Quality assurance
RSU	Research and Statistics Unit
SAS	Statistical Analysis System
SDB	Settlement Database
SDMS	Standards and Data Management Section of DIMA
SME	Small and medium sized enterprises
URD	Unit record data
VET	Vocational, Education and Training of DEETYA

Part One

Summary and Recommendations

Audit Summary

1.   Data management encompasses all stages of the data cycle from defining data needs, planning for the collection and analysis of data in line with defined needs, through assuring data quality and its availability for use, including storage and retrieval, and the sale of data.

2.   Data management is an important component of sound information management. The latter is a broad term that encompasses the management of all information maintained by an agency. It includes, for example, the collection and use of performance information, records management and data confidentiality and privacy practices. Both the broader notion of information management and the specific discipline of data management are important to all Australian Public Service (APS) agencies. Already, the APS is a significant part of the 'information industry'.

3.   The ANAO focussed on the specific area of data management because of recent proposals to reform information management across the APS[1]; concerns from business and non-government organisations about the burden imposed on the providers of data by government agencies; and the possible duplication and overlap in data collections. As well, considerable resources are devoted to the management of data within agencies and accurate and reliable data are essential to meet agencies' identified business needs including decision-making, program administration, policy development and accountability.

4.   Data management generally occurs at two levels within APS agencies, that is, for the overall agency (corporate) databases and for individual data collections. In a devolved environment (with less central control) the focus has been largely on individual collections. However, even in a more devolved environment attention should still be given to data management at the strategic or agency level to ensure that all aspects of data management are driven by identified business needs. As well, priorities need to be established and appropriately resourced. Issues such as overlap, duplication, provider cost and workload levels should be considered across agencies.

---

[1]   See *Management of Government Information as a National Strategic Resource*, Preliminary Report of the Information Management Steering Committee (IMSC) on Information Management in the Commonwealth Government, September 1997. The ANAO notes that the coverage of this report includes all three levels of government, that is, the Commonwealth, States and Territories and local government.

5.   Agencies need to take a strategic approach to data management to encourage credible accountability for resources provided and the effective performance of various functions so that they contribute to the efficient fulfilment of business needs. Such an approach would also allow Chief Executive Officers (CEOs) to better fulfil their responsibilities under the *Financial Management and Accountability Act 1997*, that is, to manage resources in an efficient, effective and ethical manner.

Audit objective, focus and scope

6.   The objective of the audit was to assess the efficiency, effectiveness and accountability of data management by government agencies.

7.   The audit focused mainly on data collected by departments and agencies from non-departmental organisations and institutions. This sort of data is commonly referred to as 'structured data' as it is collected in a predefined form for a specific purpose and is stored in a database. It can be distinguished from 'unstructured data' which relates to file notes, records of meetings, correspondence, reports, spreadsheets and similar material. The audit's scope included the following two forms of structured data:

- statistical data which has been collected for a defined set of 'research' objectives according to accepted statistical methods; and

- operational data which has been derived for a particular transaction or administrative process (such as to check eligibility for a government funded program or compliance with laws and regulations) but is used for a variety of other research and reporting purposes. Such data commonly is maintained in computer-based business applications.

8.   The ANAO established key criteria which were used to test the efficiency, effectiveness and accountability of data management. The main areas covered were:

- the management of data at the agency level including the use of a data management plan or other guidelines, the use of performance information and the role of central data management units;

- data availability including its visibility, accessibility, dissemination and sale; and

- the management of individual data collections including defining data needs and provider cost and workload, data specifications, data quality assurance measures, measures for promoting data use, performance assessment, managing resources and the use of consultants.

9.   The ANAO examined data management in six agencies. The range of agencies examined included both policy-making agencies and client-focused service providers. While data management requirements may vary

depending, for example, on the nature of the business, the general principles applicable to data management remain relevant to all agencies with data holdings. The following six agencies were included in the audit:

- Australian Customs Service (ACS);

- Department of Employment, Education, Training and Youth Affairs (DEETYA);

- Department of Immigration and Multicultural Affairs (DIMA);

- Department of Industry, Science and Tourism (DIST);

- Department of Transport and Regional Development (DTRD); and

- the Office of Small Business (OSB) which formerly was part of DIST but was incorporated into the Department of Workplace Relations and Small Business (DWRSB) after the audit commenced. The audit examined only individual collections managed by the OSB as part of DWRSB.

10.   The ANAO conducted interviews with key agency staff and examined relevant documentation both at the agency level and that associated with individual data collections in each of the six agencies. Twenty-three data collections were examined in detail.

11.   Assistance to the audit was provided by the Australian Bureau of Statistics (ABS) which advised the ANAO on an effective methodology for the identification of data needs, data collection and dissemination methods. The ANAO established an Audit Reference Group to provide input into the audit including in relation to the development of relevant criteria, providing specialist advice on aspects of data management and on the future direction of data management in the APS.

Audit conclusion

12.   Overall, the ANAO found that, while individual collections were generally well managed, there was scope for improving aspects of data management at the agency (corporate) level.

13.   At the agency (corporate) level, mechanisms generally were not in place to ensure efficiency, effectiveness and accountability in relation to data management. Of the mechanisms which should be in place to assist appropriate data management, the ANAO concluded that:

- only two agencies had established strategic data management plans and guidelines and a third agency was in the process of developing a strategic plan;

- only two agencies had allocated responsibility for developing strategic approaches to data management for organisational units. These areas

generally had been effective in developing and implementing agency-wide data measures;

- broadly speaking, the agencies had not progressed to the stage of considering performance assessment for data management; although one agency had developed performance indicators for data quality but the actual assessment of performance was limited to one indicator which did not provide a sufficient basis upon which to assess overall performance;

- while two agencies had mechanisms in place to ensure collections were transparent to those who needed to use them, the ANAO found that data collections at the agency level in the other three agencies were difficult to identify because corporate directories had not been developed. As well, there was no designated area or individual with agency-wide responsibility for the data held in each agency;

- three agencies had in place, or were in the process of considering and developing measures to facilitate data access and sharing at the agency level; but although there was a need in the other two agencies for such an approach, suitable strategies had not yet been developed; and

- agency-wide policies and guidelines did not include a costing methodology to underpin pricing decisions. As a consequence there were inconsistent pricing practices at the individual collection level within two of the agencies.

14.   In relation to individual collections, the ANAO found data collections at the local level generally were well managed and, in particular, data were used for their intended purpose, indicating that data were transparent and easily accessible. However, in some cases there was scope for improving data specifications or data management guidelines and using additional measures to strengthen the integrity of quality assurance measures. Performance assessment of individual collections was the major area in need of improvement. It is likely that the above results reflect the situation in many other agencies and entities.

Recommendations and better practice principles

15.   The ANAO has made a number of specific recommendations which will contribute to improved data management. While this audit has focussed on six agencies and their related individual collections, other APS agencies and data collection bodies would also benefit from examining the findings and recommendations of this report to ensure that their approach is appropriate. To assist in making the report useful more widely in the public sector the ANAO has included a section on better practice

principles at the end of each chapter. The better practice principles draw on both the findings of the audit and the work on an effective methodology for data management undertaken by the ABS.

Agencies' responses

16.   The five agencies, for which the ANAO has made specific recommendations, have agreed with all recommendations made in this report. The OSB has indicated that the report was welcomed.

17.   DEETYA noted in its response to the draft audit report that the overall focus of the report and the recommendations on agency-wide arrangements has been very useful in supporting the direction of certain initiatives introduced last year in DEETYA. In the development of a new culture for DEETYA, the formulation of the Corporate Plan released in December 1997, and in restructuring corporate committees, the Executive has put considerable effort into breaking down the 'chimney stack' approach to management of the department's business. As a result of these developments the Executive has decided to develop an agency-wide data management strategy and to push ahead with development of the Corporate Information Management system project to provide an agency-wide platform for data management.

Key Findings

Agency-wide approaches to managing data collection and dissemination

18. In assessing the strategic agency-wide approach to managing data the ANAO focussed on the:

- development of a data management plan and/or guidelines;
- identification of a person or the establishment of a specific area, steering committee or working group with overall responsibility for data management at the agency level; and
- existence of a performance assessment framework for data management.

Development of a data management plan and/or guidelines

19. The ANAO found that of the five agencies examined only ACS and DIMA had established strategic plans and guidelines to provide corporate direction and an agency-wide approach for managing data. These plans and guidelines are generally sound and better practice is evident. However, ACS's plans and guidelines can be improved by referring to the Information Privacy Principles in the ACS Information Processing Security Policy. DIMAs' plans can also be improved by taking into consideration developments stemming from initiatives sponsored by the Office of Government Information Technology (such as the promotion of a more consistent approach to information and data management across government agencies) in the DIMA Information Management Strategy.

20. DTRD, while not yet having a corporate data management plan, is currently developing one. The process it has adopted, to define needs and engage a consultant to prepare a plan, is sound. Neither DEETYA nor DIST had adopted strategic plans or guidelines in relation to data, although the need for addressing known data management deficiencies on an agency-wide basis had been recognised in DIST.

21. Better practices that were identified include: linking strategic plans for data management to the agency corporate plan and IT plan; the identification in agency plans and/or guidelines statutory requirements and whole-of-government policies which relate to data; and the definition at the agency level of the principles and rules that govern agency data and the distribution of these in guidelines to staff.

Overall responsibility for data management at the agency level

22.   The ANAO found that ACS and DIMA have designated areas with overall responsibility for managing, implementing and monitoring each agency's strategic approach to data management. These areas are an important vehicle for progressing data management strategies as they serve as a common point of advice on data standards and issues within agencies. Such areas are also in a position to monitor and report to the executive on the performance of individual collections and overall progress against strategic plans. While the respective areas in ACS and DIMA largely have been effective in the implementation of agency strategies and monitoring progress against plans, these areas have played a lesser role in relation to monitoring the performance of individual collections. Data accuracy is, however, monitored in DIMA.

23.   DIST and DTRD have designated central areas with limited roles for data management but DEETYA has no area, group or individual with agency-wide responsibilities for data management. The ANAO considers that such areas, groups or individuals perform a valuable role in managing and monitoring agency-wide data activity and performance.

24.   Better practices that were identified include: publishing regular bulletins of statistical information on performance and trends; educational campaigns on agency strategies and measures for data; and establishing cross-program committees, comprising data producers, data users and agency data management staff.

Performance assessment framework for data management

25.   The ANAO found that DIMA was the only agency of the five examined which had established a performance monitoring framework for its data, although DIMA had not progressed beyond collecting performance information for one element of data, that is, dealing with quality. The ANAO recognises that performance assessment of data management is in the early stages of development. Consequently, developing a performance assessment system for data management will be a gradual process involving testing and refinement. The performance indicators developed by DIMA reflected sound principles in that the indicators provide direct links between data management issues and key departmental business outcomes, but there is scope to refine and develop a better balanced set of indicators. The first steps taken by DIMA in this process should be of interest to many other APS agencies.

26.   Better practices that were identified include: the establishment of an agency-wide performance assessment framework to provide the executive with information on progress against agency strategic plans, compliance with departmental standards and the cost/benefit of agency

data operations; and the development of a balanced set of performance indicators which measure efficiency, effectiveness, outputs and outcomes for local-level data management and provide information which can be aggregated at the agency level.

Data availability: visibility, access and dissemination

27. In relation to data available at the agency level for, say, decision-making and research purposes, the ANAO examined:

- data transparency/visibility;
- data accessibility; and
- data dissemination.

Data visibility

28. The ANAO found that ACS data were highly visible as a result of agency wide measures regarding the documentation of data and the creation of useful directories. ACS has directories that identify major databases, statistical collections and smaller end-user computing systems. While the databases which cover transactional data in DIMA could be readily identified, the visibility of statistical collections was limited as there was no directory or area which maintained a record of these collections. The latter situation would be improved significantly by the creation of an agency level directory comparable to the one that exists for DIMAs' business systems.

29. The ANAO found that DEETYA, DIST and DTRD had not addressed data transparency at the agency level and, as a result, it was difficult to identify corporate data collections. In DEETYA it was found that even within one program area there was uncertainty about what the data collections were used *across* the program area. While the poor visibility of a number of databases is acknowledged in DIST, the ANAO did not find any current measures to address these concerns.

30. A better practice that was identified was the recording in a directory or catalogue of all data collections and databases to promote the transparency of agency data holdings.

Data accessibility

31. The ANAO found that ACS and DIMA had established measures that underpinned data access and sharing, such as data warehousing and advanced reporting tools which allow users to perform complex data extractions and manipulations. Issues of data access and sharing are also addressed in the proposed DTRD information management plan. These actions represent better practices.

32. DIST has recognised the importance of adopting a consistent department-wide approach to access and sharing problems but has not yet addressed the problems. In its response to the draft audit report, DIST advised that 'since January 1998, the Department has commenced a major upgrade of its IT infrastructure. An essential component of this upgrade has been the identification and registration of electronically maintained applications and data collections across the Department. The Department believes that this is a significant step towards the agreed goal of agency-wide data transparency and in addressing the access and sharing problem'.

33. The ANAO was advised by DEETYA staff that, because of the department's disparate data needs, there was little reason to access data across the different program areas. However, the ANAO found examples where different program areas are increasingly reliant on sharing data across program boundaries and this highlights the need for an agency-wide approach.

Data dissemination

34. The ANAO found that none of the agencies examined had established at the agency level policies and/or guidelines for data dissemination. The policy and guidelines should also address the legal boundaries relating to data dissemination, including whether there are any legal impediments to the sale of data. Where an agency's policy provides for the sale of data, the principles to be applied when determining the price to be charged should be articulated. DIMA was the only agency to advise the ANAO that it was proposing to examine policies related to the dissemination of information. In this case, concerns about inconsistent dissemination practices had led to the recognition that there was a need to review the department's statistical collections and pricing regimes. In DEETYA, the ANAO found two separate areas which were selling data on the same basis of cost recovery for extraction and dissemination but using different charging methods to price data.

35. Better practices that were identified include: the establishment of an agency-wide policy for data dissemination; and the incorporation in the agency dissemination policy general guidelines on the pricing of data.

Requirements for individual collections

36. To assess the requirements for individual data collections the ANAO examined the following aspects of the data cycle:

- definition of data needs;
- data specification;
- quality assurance;

- data use;
- performance assessment; and
- resources for data collection.

Data needs

37.   The ANAO found that, for all the collections examined, data needs had been defined adequately; collections were authorised appropriately; the authorisation process involved measures to avoid unnecessary duplication and to address the data needs of potential users; and provider workload had been addressed. It was also found that the approaches adopted by the OSB for the Australian Small and Medium Enterprise Database (AUSSMED) provided a useful illustration of different methods that other agencies could use to define and review data needs. OSB established advisory and technical committees to provide expert advice on data needs and collection methodology. OSB had also consulted stakeholders extensively during the design and the ongoing development of the collection.

Data specification

38.   The ANAO found that for all the collections examined specifications and/or guidelines existed which included conceptual statements, classifications and rules for maintaining data confidentiality and privacy. However, in some cases conceptual statements were found to be inadequate and needed to be improved by the inclusion of sufficient information for stakeholders to understand the purpose of a collection, the data contained in it and its potential applications.

39.   A better practice that was identified was a comprehensive manual for each collection that details data management procedures. For large and complex collections it should be available in an electronic form and have indexing/search facilities that allow users to move quickly and easily through a considerable volume of material.

Quality assurance

40.   The ANAO found that quality assurance measures were in place for all the collections examined and that generally these were adequate and being enhanced. For example, ACS is piloting a quality control system which involves cross-checking random samples of import data, undertaking a physical examination of the related consignments and an audit of related commercial documentation to address concerns about data accuracy. For the non-government schools census and financial collections, there would be benefits in DEETYA monitoring the results of the new internet reporting system to assess whether it is improving data accuracy and introducing

tighter targeting and a larger sample in the follow-up checks of the census collection.

41. Better practices that were identified include the: development of software to assist staff to enter, edit and transmit the data electronically to a central site; development of coding manuals that aid the consistent interpretation of complex data; and the assessing and monitoring of data accuracy.

Data use

42. The ANAO found that all the collections examined were used for their intended purposes and that mechanisms existed for monitoring user requirements. The ANAO also found that in some cases in ACS and DIMA it could be difficult to undertake complex data analyses because of design and functional constraints with the systems on which data are maintained. However, appropriate measures had been or were being taken to resolve these problems in both agencies.

43. Most collections had adequate measures to promote data transparency, although there were indications that some collections in the Vocational Education and Training (VET) program area in DEETYA were not highly visible to the area's staff. The ANAO considers that establishing a directory of the collections and data used at program area-level in VET would improve data visibility and serve as a key element in a whole-of-agency data directory in DEETYA.

Performance assessment

44. The ANAO found that the performance assessment framework for collections at the local level was not sufficiently developed to allow adequate monitoring. Where performance was monitored, collections did not use a balanced set of performance indicators to measure all relevant aspects of data management, particularly in relation to efficiency. The ANAO has identified a number of good practices which should provide a useful basis for agencies to determine performance indicators and appropriate sources of performance information. As well, the development of performance assessment for individual collections should be an integral part of an overall agency framework for assessing performance in relation to data.

Resources for data collection

45.   To assess the resources used for data collection the ANAO examined:

- staff skills availability;
- the use of consultants, including the adequacy of the contracting procedures; and
- the use of information technology.

Data management skills

46.   The ANAO found that for the collections examined a variety of methods were used by management to ensure that the necessary skills were available to meet data management requirements. Better practices observed to provide the necessary staff skills were the recruitment of staff with appropriate technical skills; and the development of specific skills and multi-skilling by training in relevant technical skills and, in some cases, staff rotation across a range of areas.

Use of consultants

47.   The ANAO found that for all cases examined the processes used to engage consultants were appropriate. Contracting procedures were transparent and consistent with accepted Commonwealth practices and consultants were managed appropriately. However, for two BTR collections, the contract did not specify adequately the outputs to be delivered by the consultants which resulted in the BTR having to reprocess the consultant's work. To address this shortcoming the BTR needs to re-examine its output requirements and revise the contract with the consultant accordingly.

48.   A better practice identified was the engagement of additional consultants to assist in the review and management of the work of the main consultant. This can be of benefit to agencies developing a complex data collection and/or lacking in-house expertise. Another better practice involved the use of a performance-based, as opposed to a cost-based, approach to managing consultants. Managing consultants on the basis of performance against specified targets and milestones will assist agencies to keep projects on time, to focus consultants on agency requirements and to identify and address problems at an early stage.

Information technology

49.   The ANAO found that all the collections examined relied on information technology although the information technology needs varied significantly from collection to collection and no single system would meet all needs. However, while information technology improved data management and generally met the needs for agencies manipulating

statistical data collections, limitations in information technology restricted the access and use of transactional databases. The two agencies having transactional databases that were examined by the ANAO (ACS and DIMA) were appropriately developing information technology-based solutions such as data warehousing to improve data access and retrieval. Other better practices identified also related to the use of appropriate information technology for direct data entry, automatic editing and direct data transfer from the provider to the collecting agency, data analysis through the use of IT-based research, statistical analysis and reporting tools.

Recommendations

Agency-wide approaches to managing data collection and dissemination

Recommendation No. 1 Para. 2.44

The ANAO recommends that ACS, in revising its Information Processing Security Policy, refers explicitly to Information Privacy Principles of the *Privacy Act 1988* and section 16 of the *Customs Administration Act* to highlight to staff the statutory requirements relating to data and information privacy and confidentiality.

ACS's response: Agreed

Recommendation No. 2 Para. 2.46

The ANAO recommends that to ensure that data collections meet core business needs and can be integrated where necessary, DEETYA and DIST develop strategic plans for data issues with supporting guidelines which establish corporate direction, principles and rules and a common basis for local level data management.

DEETYAs' response: Agreed

DIST's response: Agreed

Recommendation No. 3 Para. 2.75

The ANAO recommends that ACS improves staff awareness of ACS's strategic approaches to data management, such as the Corporate Data Management Strategy and role of the Corporate Data and Statistical Services Section, to assist in ensuring that there is a consistent approach to managing data within the agency.

ACS's response: Agreed

Recommendation No. 4 Para. 2.77

The ANAO recommends that, to assist with the development, implementation and monitoring of strategic corporate approaches to data management, DEETYA, DIST and DTRD assign a specific area or individual with agency responsibility for data management including:

- developing corporate strategic plans, standards and guidelines for data;
- providing advice to local-level areas on data management issues;
- promoting staff awareness of agency data management strategies and measures; and
- monitoring the performance of local-level databases and data collections.

DEETYAs' response: Agreed

DIST's response: Agreed

DTRD's response: Agreed

Recommendation No. 5 Para. 2.99

The ANAO recommends that to enable performance to be measured and improve accountability ACS, DEETYA, DIST and DTRD each should establish an agency-wide performance assessment framework for databases and data collections. The performance assessment framework should include:

- a balanced set of performance indicators which addresses inputs, process, outputs, outcomes and client satisfaction and the collection of information which allows their measurement; and
- sestablishing a system of monitoring, reporting and feedback between the area responsible for performance monitoring and local-level data managers.

ACS's response: Agreed

DEETYAs' response: Agreed

DIST's response: Agreed

DTRD's response: Agreed

	The ANAO recommends that, to measure all relevant aspects of performance, DIMA further develops its data quality performance information project. This should include:

- developing a balanced set of performance indicators which addresses inputs, process, outputs, outcomes and client satisfaction;

- collecting performance information for these indicators from local-level databases and data collections; and

- establishing a system of monitoring, reporting and feedback between the area responsible for performance monitoring and local-level data managers.

DIMAs' response: Agreed

Data availability: visibility, access and dissemination

The ANAO recommends that, to promote data visibility, DEETYA, DIMA, DIST and DTRD each create directories of the data collections and statistical collections held in each agency and that for each collection/database these directories include:

- the title of the collection;

- a contact name and/or phone number of the person responsible for managing the collection;

- a description of the information needs that the collection was designed or purchased to address; and

- other relevant details such as data quality, classifications and links to other data, where necessary.

DEETYAs' response: Agreed

DIMAs' response: Agreed

DIST's response: Agreed

DTRD's response: Agreed

Recommendation No. 8 Para. 3.43	The ANAO recommends that, in order to maximise the benefits of data collected, DEETYA and DIST should determine the extent of the need to access and share data on a department-wide basis and put in place data access and sharing mechanisms which support the identified need.

DEETYAs' response: Agreed

DIST's response: Agreed

Recommendation No. 9 Para. 3.57	The ANAO recommends that, to ensure data are disseminated in line with agency policy and data pricing decisions are equitable and underpinned by an approved costing methodology, ACS, DEETYA, DIMA, DIST and DTRD establish general policy guidelines for the dissemination and sale of data which draw the attention of local-level data managers specifically to:

- whether data can be disseminated;
- relevant statutory requirements applying to data dissemination;
- relevant privacy and security restrictions;
- procedures for obtaining approval to release data (if applicable); and
- the basis for pricing data including the tasks and methods for calculating the price of data.

ACS's response: Agreed

DEETYAs' response: Agreed

DIMAs' response: Agreed

DIST's response: Agreed

DTRD's response: Agreed

Resources for data collection

Recommendation No. 10 Para. 5.29	The ANAO recommends that BTR reassesses its output requirements for the International Visitors Survey and incorporates the revised requirements in the contract conditions at the earliest possible time to ensure that BTR obtains data in a suitable form and of a satisfactory quality.

DIST's response: Agreed

Audit Findings and Conclusions

1. Introduction

This chapter defines the management of data collection and dissemination, describes the background to the audit and sets out its objective, scope, focus, methodology and criteria.

Background

1.1 Data management encompasses all stages of the data cycle from defining data needs, planning for the collection and analysis of data in line with defined needs, through assuring data quality and its availability for use, including storage and retrieval, and the sale of data. The Australian Bureau of Statistics[2] (ABS) has stated that:

> *the aim of data management is to ensure that data holdings are readily visible, easily accessible, and mutually coherent, and that analyses and other outputs based on these holdings are well presented and can be explained.*

1.2 Data management is an important component of sound information management. The latter is a broad term that encompasses the management of all information maintained by an agency. It includes the collection and use of performance information, records management and data confidentiality and privacy practices. Both the broader issue of information management and the specific area of data management are important to all Australian Public Service (APS) agencies. Already, the APS is a significant part of the 'information industry'.

1.3 While this audit focussed on the specific area of data management, a number of audits has been conducted, are underway or are proposed that cover other aspects of information management. These include:

- performance information in the Department of Employment, Education, Training and Youth Affairs (DEETYA);
- a cross-portfolio audit of records management;
- a follow-up audit relating to the protection of confidential client information from unauthorised disclosure; and
- a cross-portfolio audit of data privacy.

1.4 The ANAO focussed on the specific area of data management because of recent proposals to reform information management across the

---

[2] "Consultancy on design attributes for data collection and dissemination", A Report for the ANAO prepared by ABS Statistical Consultancy, August 1997, p. 7.

APS[3], concerns about the burden imposed on the providers of data by government agencies and the possible duplication and overlap in data collections. As well, considerable resources are devoted to the management of data within agencies and accurate and reliable data are important to meet agencies' identified business needs. Such needs include decision-making (including decisions in relation to funding), program administration and policy development. Sound data can also contribute significantly to measuring performance.

1.5 Data management generally occurs at two levels within APS agencies, that is for the agency (corporate) databases and for individual data collections. In a devolved environment (with less central control) the focus has been largely on individual collections. However, even in a devolved environment attention should be given to data management at the strategic or overall agency level. This should ensure that all aspects of data management are driven by identified business needs. As well, priorities need to be established and appropriately resourced. Issues such as overlap, duplication, provider cost and workload should be considered across agencies.

1.6 Agencies need to take a strategic approach to data management to encourage credible accountability for resources provided and the effective performance of various functions so that they contribute to the efficient fulfilment of business needs. Such an approach would also allow Chief Executive Officers (CEOs) to better fulfil their responsibilities under the *Financial Management and Accountability Act 1997*, that is, to manage resources in an efficient, effective and ethical manner.

1.7 In the longer term the strategic view of data management will assist in examining data management issues across the APS. This whole-of-government view could improve efficiency and effectiveness as, by sharing and integrating data, it could become more accessible. As well it may serve to reduce provider load in some cases. The ANAO notes that the report of the Information Management Steering Committee (IMSC), *Management of Government Information as a National Strategic Resource*, is an important reference document which outlines a whole-of-government information management model and includes matters relating to data accessibility, visibility and interoperability (that is, the ability of two or more systems to exchange information for mutual use).[4]

---

[3]   See *Management of Government Information as a National Strategic Resource*, Preliminary Report of the Information Management Steering Committee (IMSC) on Information Management in the Commonwealth Government, September 1997. The ANAO notes that the coverage of this report includes all three levels of government, that is, the Commonwealth, States and Territories and local government.

[4]   Ibid.

1.8   More detailed discussions of various aspects of data management are contained in the relevant chapters of this report.

The audit

Audit objective

1.9   The objective of the audit was to assess the efficiency, effectiveness and accountability of data management by government agencies.

Scope and focus

1.10   The audit focused mainly on data collected by departments and agencies from non-departmental organisations and institutions. This sort of data commonly is referred to as 'structured data' as it is collected in a predefined form for a specific purpose and is stored in a database. It is distinct from 'unstructured data' which relates to file notes, records of meetings, correspondence, reports, spreadsheets and so on.

1.11   The audit's scope included the following two forms of structured data:

- statistical data which has been collected for a defined set of research objectives according to accepted statistical methods; and

- operational data which has been derived for a particular transaction or administrative process (such as to check eligibility for a government funded program or compliance with laws and regulations) but is used for a variety of other research and reporting purposes. Such data is maintained in computer-based business applications.

Audit criteria

1.12   The ANAO established key criteria which were used to test the efficiency, effectiveness and accountability of data management. The main areas covered were:

- the management of data at the agency level including the use of a data management plan or other guidelines, the role of central data management units, and the use of performance information;

- data availability including its visibility, accessibility, dissemination and the commercial sale of data; and

- the management of individual data collections, including defining data needs and provider load, data specifications, data quality assurance measures, measures for promoting data use, performance assessment, managing resources and the use of consultants.

1.13   Relevant criteria specific to the area being discussed are outlined at the beginning of each chapter/section.

Audit methodology

1.14 The ANAO examined data management in six agencies. The range of agencies examined included both policy-making agencies and client-focused service providers. While data management requirements may vary depending, for example, on the nature of the business, the general principles applicable to data management remain relevant to all agencies with data holdings. The following six agencies were included in the audit:

- Australian Customs Service (ACS);
- Department of Employment, Education, Training and Youth Affairs (DEETYA);
- Department of Immigration and Multicultural Affairs (DIMA);
- Department of Industry, Science and Tourism (DIST);
- Department of Transport and Regional Development (DTRD); and
- the Office of Small Business (OSB) which formerly was part of DIST but was incorporated into the Department of Workplace Relations and Small Business (DWRSB) after the audit commenced.

1.15 These agencies were chosen as they provide an understanding of the range of data collection and dissemination activities and approaches to managing those activities in the APS. The OSB was selected in particular because it provided a case study in which consultants undertook all elements of data management under the direction of the OSB.

1.16 Fieldwork was undertaken between August to November 1997. The ANAO conducted interviews with key agency staff involved in or with specific responsibilities for data management. As well, we examined relevant documentation held by each of the six agencies.

Assistance to the audit

1.17 The ANAO contracted the Australian Bureau of Statistics (ABS) to provide advice on an effective methodology for the identification of data needs, data collection and dissemination methods. The framework encompassed recognised industry standards and better practice principles. The material developed by the ABS has been referred to throughout this report. As well, the ABS material and the findings of this audit have been used to develop better practice principles which are included at the end of each chapter of this report.

Audit reference group

1.18 The ANAO established an Audit Reference Group to provide input into the audit including in relation to the development of relevant criteria, providing specialist advice on aspects of data management and on the

future direction of data management in the APS. The Audit Reference Group met regularly throughout the audit. Its members are listed at Appendix 2.

Previous reviews of data management

1.19 Coverage of data management to date has been limited in that previous reviews have concentrated on aspects of government agency data activity rather than data management as an overall function.

1.20 The impact of government statistical collections on small businesses was addressed in the November 1996 report by the Small Business Deregulation Task Force, *Time for Business*. The report included recommendations relating to establishing a central clearance system (or Clearing House) for Commonwealth statistical collections, improving data collection instruments and procedures and measures to promote data dissemination and use.

1.21 Better integration of data management in the context of remodelling information management in the Commonwealth Government was addressed in the IMSC report, *Management of Government Information as a National Resource*. The report addressed, among other things, the hidden costs, particularly for data users, associated with deficient information management systems.

1.22 The ANAO also reviewed an overseas audit report directly relevant to this audit. The United States General Accounting Office report, *Hazardous Waste: Data Management Problems Delay EPA's Assessment of Minimisation Efforts* (1991), addressed the importance of data management planning for the effective and efficient conduct of data collection.

1.23 In addition, the Organisation for Economic Cooperation and Development reported in 1995 on the impact of the United States *Paperwork Reduction Act 1980* on reducing the provider load of Federal government data collections.

1.24 The ANAO used the findings of these reviews to assist in identifying key concerns and refine the audit criteria. For example, the report *Time for Business* raised the issue of provider load. This issue was therefore addressed in the criteria developed by the ANAO.

This report

1.25 The report addresses data management at two levels:

- agency-wide data management practices; and
- the management of individual data collections.

1.26   Chapter 2 examines the management of data at the agency level. The ANAO reviewed agency data management planning, areas with corporate responsibility for agency data matters and the use of performance information for monitoring data management. Chapter 3 addresses data availability issues at the agency collection level, including the visibility of data collections both internally and externally, data accessibility and the commercial sale of data.

1.27   The management of individual data collections at the local level within agencies is discussed in Chapters 4 and 5. The methods used to define and monitor data needs, the adequacy of the quality assurance measures used by data operation areas and the extent to which individual collections employ performance information and monitoring systems are discussed in Chapter 4. Resourcing for data management, including the availability of staff skills, the use of consultants for one or more of the elements of data management (data collection, processing and analysis) and the information technology support provided are discussed in Chapter 5.

1.28   The ANAO has made recommendations which will contribute to improved data management. While this audit has focussed on six agencies and their related individual collections, other APS agencies and data collection bodies are likely to benefit from examining the findings and recommendations of this report to ensure that their approach is appropriate. To assist in making the report useful more widely the ANAO has included a section on better practice principles at the end of each chapter. As discussed above these drew on both the findings of the audit and the work undertaken by the ABS.

1.29   The audit was conducted in accordance with ANAO Auditing Standards and cost $388 000.

2. Agency-wide Approaches to Managing Data Collection and Dissemination

This chapter discusses agency-wide approaches to managing data, including the extent to which agencies have used overall plans, established areas with a central responsibility and employed performance measures for monitoring data operations. While the importance of adopting agency-wide approaches for managing data is recognised, only a few agencies have issued plans and established central data management units. None of the agencies had adopted performance information measures for their data operations. The ANAO has made eight recommendations aimed at improving various aspects of agency-wide data management.

Introduction

2.1   Data management is a dispersed and widespread activity within agencies and across the APS. In a devolved environment the data management function is often undertaken at the local level for each collection. Given the nature of data collections and data use this will, and must, continue to occur. However, as highlighted in Chapter 1 a strategic approach to data management also is important, even in a devolved environment, because it enables an agency to:

- collect and use data to meet overall business needs;

- ensure appropriate data are available to underpin decision-making, policy development and so on;

- reduce unnecessary duplication and/or identify gaps in data needs for the business as a whole;

- facilitate data exchange/sharing within and between agencies to maximise use and minimise costs;

- identify costs of data management; and

- allow for appropriate information technology support, shared systems and data compatibility.

2.2   Without an agency-wide approach effective data management, including the appropriate level of accountability for resources used for the data management function, will be difficult. This is particularly important at a time when the accountability of chief executive officers (CEOs) has been made more explicit, for example through the *Financial Management and Accountability Act 1997 (FMA)*. The FMA Act states, among other things, that CEOs should manage resources in an efficient and effective manner.

2.3 As well as the importance of the strategic approach to CEOs, the environment in which the APS operates also requires an examination of improving methods of service delivery on a continuing basis. One of the identified ways of improving performance in service delivery is through subjecting activities to contestability. Contestability is defined as 'the prospect of competition in public sector activities to improve … efficiency and effectiveness'.[5] In this way CEOs and agencies can test whether the activity is core business or non-core business and determine the most appropriate method to provide the service. As well, subjecting a function to contestability should assist in identifying more cost effective ways of delivering that function.

2.4 The drive to deliver services more cost effectively has focussed attention, to a greater degree than ever before, on outcomes. Like other public sector activities data management should not only be about collecting the data (an input) and analysing the results (an output) but should also be about outcomes, that is, using data. While in this regard developing an appropriate performance assessment framework is important to ensure that the data management function is achieving its objectives, the implementation of the Commonwealth's accrual-based outcomes and outputs framework which is linked to accrual-based budgeting[6] is also important. The accrual budgeting framework which builds on a range of recent Commonwealth reforms focuses on:

- what is being produced by agencies (outputs);
- what resources are administered on behalf of the Commonwealth (administered items);
- what are the purposes of the administered items (outcomes); and
- at what cost (accrual measurement).

2.5 The intent of the new framework is, among other things, to provide consistent information to the Parliament and the community on the use of public resources and improve the awareness of the costs of providing goods and services. Data management is a widespread activity for which costs are generally not known but will need to be considered within the accrual budget context.

---

[5] Department of Finance, *Examining Contestability within the APS: Initial Information*, Management Improvement Discussion Series Paper No.3, November 1995.

[6] See 'Implementing the Commonwealth's Accrual-based Outcomes and Outputs Framework', Exposure Draft, Department of Finance and Administration, October 1997. The first full accrual budget is due in 1999-2000.

2.6 The ANAO recognises that establishing an agency-wide approach in relation to data management is not an easy task and that it is at a relatively early stage in the APS. There are, however, some mechanisms which will facilitate such an approach including:

- developing of a data management plan and/or guidelines;
- allocating the overall responsibility for an agency's data management to a person or a specific functional area; and
- existence of a performance assessment framework for data management.

2.7 The ANAO sought to establish whether agencies examined during the audit had an understanding of these approaches and the extent to which they were being implemented.

2.8 The key criteria are set out at the beginning of each section and are followed by a discussion of the ANAO's findings.

Agency-wide strategic data management plans and guidelines

2.9 A strategic data management plan should set the overall direction for managing data, articulate the principles upon which the plan is based and provide a framework for establishing a coordinated approach to managing data. In a devolved environment, it also serves as a link between the agency's business objectives and individual data collections and databases by providing a central focus for activities at the local level.

2.10 Guidelines on collecting, using and disseminating data should complement the general direction of the strategic plan by providing particular details on key elements of managing data such as the identification of data needs, the need to develop specifications for each collection and quality assurance measures. Guidelines need to provide staff with advice on the rules, standards and procedures adopted at the agency level so that separate data collections and databases within agencies are developed and managed on a consistent basis. This should ensure that the data collected and produced are available in an appropriate form to all those who need to use it.

2.11 In examining agencies' data management, the ANAO sought to determine whether agencies had strategic plans and/or guidelines. These should be linked with corporate and IT plans and address key issues such as:

- compliance with statutory requirements and whole-of-government policies;
- essential data principles and rules; and
- risk management.

2.12   As well, such plans should address matters related to data visibility, accessibility and dissemination, including the pricing of data, which are important determinants of data availability and use. These issues are discussed in Chapter 3.

2.13   Findings against each of the key criteria are discussed under separate headings below.

ANAO findings

Strategic plans/guidelines

2.14   The ANAO found that while the importance of having agency-wide approaches to collecting and using data is recognised, implementation of such approaches is limited. Four of the five agencies[7] examined by the ANAO had identified the need for data management strategies, but only two agencies (ACS and DIMA) had strategic plans and guidelines while a third agency (DTRD) has engaged a consultant to develop an information management plan.

2.15   Given that only ACS and DIMA had plans and guidelines, a brief outline of these documents is provided below. The subsequent discussion examines whether these plans and guidelines provide adequate direction and a framework for data management at the agency level. As well, DTRD's proposed information management plan is discussed briefly in relation to its adequacy in addressing data management.

ACS

2.16   ACS has established a framework of complementary strategic plans and guidelines to promote agency-wide measures for its data. The framework comprises the following:

- *Customs Information Technology Strategic Plan 1996-99* which includes requirements and strategies for more effective use of agency-wide data, data capture and dissemination and information management and security;

- *1995 Review of Customs Statistical Requirements and Capability* which identified deficiencies in ACS's data management and recommended a range of measures to improve data collection and the production and use of statistical data, particularly for departmental purposes;[8]

---

[7]   The original scope of the audit included five agencies but was increased to six with the move of the OSB from DIST to DWRSB after the audit had commenced. The audit examined data management only at the local level in OSB, rather than agency-level data management in DWRSB.

[8]   The Review was undertaken jointly by Customs and the ABS.

- a Statement of Statistical Principles which identifies the rules and quality controls governing the statistical data collected and used by ACS; and

- a Statistical Framework which, based on the Statistical Principles, identifies the key activities (in terms of Key Result Areas) of ACS business and the statistics needed to monitor those activities.

2.17 As well, ACS advised the ANAO that work is being undertaken to develop a broad-based Corporate Information Management Plan (CIMP) which would integrate a number of separate elements including a Corporate Data Management Strategy. The Statistical Principles and Statistical Framework discussed above are to be included in the CIMP in order to highlight their role in the planning and management framework for ACS data.

DIMA

2.18 DIMAs' agency-wide approach to managing data is addressed in the following planning documents:

- DIMAs' Data Management Strategy, released in September 1992, which articulates objectives, key principles and required action to achieve DIMAs' data management aims; and

- DIMAs' Information Management Strategy, released in November 1996 as part of DIMAs' IT Strategy, which defines the principles, objectives and general approaches to be adopted in support of DIMAs' data management. In particular, it states that the Data Management Strategy will be extended to include the management of all information and subsumed into the new Information Management Strategy.

2.19 Some elements of the Information Management Strategy have been implemented but it has not been completed because the area responsible for implementation was disbanded. DIMA advised that responsibility for the Information Management Strategy is to be established in the context of the Strategic Alliance roles which are currently being developed between DIMA and the private contractor engaged by that department to develop applications for business solutions.

Links with other agency plans

2.20 Linking strategic plans and/or guidelines for data with other agency-wide plans, in particular the corporate plan and IT plan, is important because it helps staff to consider data as a departmental asset which should meet business needs. It also underpins links between strategic approaches to data and other agency strategies.

2.21 Both ACS's and DIMAs' strategic planning documents for data management are linked appropriately to each agency's respective corporate

and IT plan. In ACS, data management strategies are included in *Customs IT Strategic Plan 1996-99* which, in turn, is linked to the agency's corporate plan. The IT Strategic Plan clearly establishes the relevance that good data management has to the achievement of ACS's current goals and operational strategies. As stated in the plan, 'many areas of ACS now need extensive access to data from diverse departmental systems - and from some external sources - for the purposes of risk management, analysis, auditing and other compliance activities, profiling and reporting'.

2.22   These links are also made in ACS's Statistical Framework which provides a tool for identifying the key result areas established in the ACS corporate plan and developing the statistics needed to monitor those activities. Details of the Statistical Framework are shown in figure 1.

2.23   DIMA has incorporated its data management strategy in its information management strategy which is part of the department's overall IT Strategic Plan. As with ACS, DIMAs' IT plan highlights that data management is essential for business results, for example, by providing timely access to relevant and accurate data. As part of the overall IT plan, the Information Management Strategy is also linked to DIMAs' Applications Strategy which is intended to establish the IT applications required to support departmental business processes.

Compliance with statutory requirements and whole-of-government policies

2.24   Agencies must manage their data in line with various statutory requirements relating to the information held by government bodies. The ABS has identified the following legislation which may apply to the data collection, dissemination and retention:

- the *Privacy Act 1988*, in particular the Information Privacy Principles contained in sections 14-16 are applicable to the large volume of sensitive data (both personal and commercial) held by government agencies;

- the *Freedom of Information Act 1982*;

- the *Archives Act 1983*; and

- other agency-specific legislation or agreements.

2.25   The ANAO found that DIMA made adequate reference to the Information Privacy Principles in its IT Strategic Plan. In ACS, however, the strategic planning documents did not refer to the *Privacy Act* and section 16 of the *Customs Administration Act*, both of which prevent ACS from releasing certain categories of information. Since ensuring the confidentiality of its data holdings is one of the ACS agency-wide requirements, these two Acts should be mentioned explicitly when ACS revises its information processing security policy. This would highlight

for staff the statutory requirements attached to data privacy and confidentiality and indicate the legal basis of the ACS approach to data security.

2.26 The ANAO considers that agencies, in developing strategic approaches to data management, need to also take into account whole-of-government policies which relate to data and information. There are two areas in particular which require agencies to consider developments and obligations beyond their own immediate requirements:

- the initiatives sponsored by the Office of Government Technology (OGIT) flowing from the Information Management Steering Committee (IMSC) report, *Management of Government Information as a National Strategic Resource*, such as those relating to metadata standards for the Australian Government[9] and a whole-of-Australian-Government search facility[10]; and

- the Commonwealth Statistical Clearing House (the Clearing House), coordinated by the ABS, was established recently with the aims of reducing duplication, improving quality, eliminating statistical collections which do not meet quality standards and minimising provider load (that is, the reporting burden imposed on information providers) for all government collections involving 50 or more businesses. Existing and new statistical collections need to conform with the quality standards and review criteria established by the Clearing House.

2.27 In ACS, the ANAO found that both its IT Strategic Plan and proposed CIMP have been developed with consideration to the initiatives sponsored by OGIT. In DIMA, the IT Strategic Plan refers to the Commonwealth Government's IT reform agenda and the importance of integrating the Information Management Strategy with legislation and broader government policy. In this regard, DIMA should, in implementing its Information Management Strategy, take account of the broad measures recommended by OGIT to ensure that the further development of data management measures in DIMA are in line with wider trends across the APS.

2.28 The Clearing House does not relate to DIMA but does affect some of ACS's data collections. Although most of ACS's data is captured by transaction systems which are outside the scope of the Clearing House, surveys of ACS's business clients have to be reviewed by the Clearing House and meet its quality standards.

---

[9]   *Australian Government Locator Service Implementation Plan*, A Report by the Australian Government Locator Service Working Group, December 1997.

[10]   *Functional Requirements for a Whole-Of-Australian-Government Search Architecture*, A Report by the Search Engine Working Group, January 1998.

2.29 The ANAO found that the staff of the ACS Corporate Data and Statistical Services Unit (CDSSU) were aware of the Clearing House requirements. The ANAO understands that the during the next stage of the ACS National Survey Program the CDSSU will be arranging for surveys of businesses to be reviewed by the Clearing House.

Essential data principles and rules

2.30 Data principles and rules need to be defined at the agency level so that all data collected at the local level are consistent. If the data collected by separate areas within an agency are to become a departmental asset it must conform with departmental standards and rules. According to the ABS, data standards and classifications are a fundamental tool of data management:

> *Application of standards to data acquisition and to new data collections promotes internal, inter-agency and international relatability of data, and provides a framework to satisfy the demand for statistics which can be related across several areas of concern. Relatability enables comparisons or combinations of data from different sources which produces a synergistic gain in data content - the content of data sets which have been related exceeds that of the component data sets.*

2.31 As noted above, ACS intends to include the agency's Statistical Principles and Statistical Framework in its corporate information management plan which integrates data management with other information-related policies. The ANAO considers that these two complementary tools provide a sound basis for the management of ACS's statistical data. The Principles state that all ACS statistical data must be:

- related to corporate objectives, as expressed in the Key Result Areas of the ACS Corporate Plan;
- accurate and reliable;
- based on a common set of key terms used across ACS;
- available in any required breakdown by location or time; and
- as accessible as possible to data producers and potential users.

2.32 Based on the first Statistical Principle that all statistical data should be linked to corporate objectives, the Statistical Framework provides a matrix which separate areas of ACS can use for defining the statistical indicators required to monitor and report on Key Result Areas (KRAs). An example of how the Statistical Framework defines statistical indicators which are linked to KRAs is provided in figure 1.

Figure 1
Using ACS's Statistical Framework to define statistical data requirements: air passenger clearance

The first Statistical Principle is that statistics should be linked to a KRA. The Statistical Framework gives effect to this by mapping out in detail the full range of statistics that is desirable for the monitoring of ACS performance against each KRA.

The Framework has two components.

The first component involves identification of core functions and classifying them according to the relevant KRA, for example:

Key Result Area	Core customer function
KRA 2 - Travel facilitation	Air passenger clearance Sea passenger clearance

The second component involves seeking relevant statistics for each function by reference to a matrix identifying key elements. A comprehensive statistical view of a core ACS function involves measures of activity, cost, efficiency and effectiveness. These measures are related to the following categories:

- a business driver (eg a basic industry unit, like a passenger or a consignment);
- ACS facilitation (eg processing an entry); or
- ACS compliance (eg performing an audit or search).

By way of example, the following shows how the matrix may be applied to the air passenger clearance function:

	Key industry statistic/ business driver of ACS activities	Facilitation statistic	Compliance statistic
Activity	Volume of passenger traffic	Number of people who make declarations Number self-assessing as red/green	Number of people who make declarations checked by type (baggage exams, personal checks)
Cost	Not applicable	Total cost of activity Average cost per passenger	
Efficiency	Not applicable	% through primary line within government standards	Breaches detected by type Goods seized Number of referrals
Effectiveness	Not applicable	Number of complaints/ compliments	Indicator of trafficking activity

2.33 The ANAO considers ACSs' Statistical Principles and Statistical Framework to be examples of better practice as these tools:

- are easy to follow and therefore more likely to used by staff; and

- provide a common method by which the agency's business areas can develop statistical indicators for local and departmental data needs.

2.34 DIMAs' Data Management Strategy included measures for extending the department's existing data standards and establishing rules for data quality. These standards and rules cover, among other things, data change controls, a single common codes system and data validation, accuracy and access. The strategy included, among its goals, the need for DIMA to base its data on established international and national standards in order to enhance data sharing with external bodies and reduce the ongoing maintenance of departmental data.

2.35 A set of departmental statistical standards and principles is also to be developed by the Research and Statistics Unit (RSU) in DIMA to address concerns regarding the coordination and consistency of statistical practices within the department. Common standards should not only improve the consistency and therefore the compatibility of statistics produced by different areas in DIMA but also raise the quality and reliability of statistical data produced in the department. The ANAO considers that the RSU has taken a strategic approach by examining the ACS model (discussed above) for developing statistical principles and standards for DIMAs' own information requirements.

Risk management

2.36 Given the increasing recognition of the importance of data and information to the effectiveness of agency operations, identification of the risks to agency data holdings and the implications of exposure to risk for business results should be considered at the strategic level.

2.37 The ANAO found that ACS has identified in its strategic planning a number of risks to the departmental value of its data and information stemming from inadequate data and information management. These include: data loss and corruption; the loss of control over critical agency data due to the devolution of responsibility for office information; and 'end-user computing' (that is, user-developed applications or databases which are inconsistent with departmental standards, poorly documented, have limited data relatability and are dependent on the original developer).

2.38 In DIMA, the risks stemming from poor data management in terms of costs to business effectiveness and service delivery were recognised in the Data Management Strategy. While DIMA has been able to quantify the costs of data accuracy problems, and the strategic planning for the

Integrated Client Service Environment (ICSE)[11] project has identified risks in the data migration to ICSE, risk management for data and information is omitted in the latest Information Management Strategy. In light of these identified risks, consideration should be given to including data risk management as an element of DIMAs' strategic planning.

2.39   While DTRD has not yet established an agency-wide plan, the ANAO noted that the department had commenced a process to acquire consultancy services for the development of an information management plan and advice on records management software. The requirements and objectives DTRD has specified for its information management reflect a sound understanding of the key principles and elements also needed for an effective plan for data management and provide an appropriate base upon which to establish an agency-wide framework for data and information management.

Conclusion

2.40   The ANAO found that of the five agencies examined only ACS and DIMA had established strategic plans and guidelines to provide corporate direction and an agency-wide approach for managing data. These plans and guidelines are generally sound and better practice is evident. However, ACS's plans and guidelines can be improved by referring to the Information Privacy Principles in the ACS Information Processing Security Policy. DIMAs' plans can also be improved by taking into consideration developments stemming from initiatives sponsored by the Office and Government Information Technology (such as the promotion of a more consistent approach to information and data management across government agencies) in the DIMA Information Management Strategy. The improvements suggested should lead to DIMA better identifying costs associated with data accuracy problems and, as well, DIMA and ACS more easily identifying and addressing problems such as data loss and corruption.

2.41   The ANAO considers that the requirements specified for a corporate information management plan in DTRD indicate that its planning is heading in the right direction based on the key elements required for effective agency data management and information management.

2.42   Neither DEETYA nor DIST had adopted strategic plans or guidelines in relation to data, even though the need for addressing known data management deficiencies on an agency-wide basis had been recognised in DIST. While a strategic plan or guidelines alone will not be the answer

---

[11]   ICSE has the objective of integrating DIMAs' major data applications.

to all the data management requirements and challenges facing agencies, they are an important means of establishing a framework in which agencies can manage their data as a departmental asset.

2.43   The way in which agencies develop strategic plans and guidelines will depend, in large measure, on the executive's assessment of the degree of risk that weak data management poses to the agency. Developing strategic approaches to data management should be in line with the approaches to risk management used by an agency as part of its corporate governance framework. Basing agency-wide data management on a systematic assessment of the key elements of risk will enhance the strategic planning for data and, by assigning specific responsibilities to particular people or organisational functions, improve the accountability of data management in an agency.

Recommendation No.1

2.44   The ANAO recommends that ACS, in revising its Information Processing Security Policy, refer explicitly to Information Privacy Principles of the *Privacy Act 1988* and section 16 of the *Customs Administration Act* to highlight to staff the statutory requirements relating to data and information privacy and confidentiality.

ACS's response

2.45   Agreed. ACS already refers to the responsibilities of the Privacy Officer in the draft IT Security Policy document.

Recommendation No.2

2.46   The ANAO recommends that to ensure that data collections meet core business needs and can be integrated where necessary, DEETYA and DIST develop strategic plans for data issues with supporting guidelines which establish corporate direction, principles and rules and a common basis for local level data management.

DEETYAs' response

2.47   Agreed.

DIST's response

2.48   Agreed. DIST's strategic approach to data management is contingent upon its data risk management framework which, in turn, will draw on the work currently underway in the department to identify and register all electronically maintained applications and data collections.

Overall responsibility for agency data

2.49 Areas with an overall responsibility for agency data can play an important role in assisting with the development, implementation and review of strategic approaches to data management. In a devolved environment, an area with an agency-wide role provides an overall departmental perspective and direction for measures implemented at the program and local levels. As well, such a body can monitor local-level data performance within the agency.

2.50 The nature of the role undertaken by areas with overall responsibilities will differ between agencies according to a number of factors. These include the extent of an agency's data requirements, whether an agency is a primary collector of data or relies more on external data sources (for example, from the ABS or contractors) and the size of an agency. Agencies less involved with collecting data directly or with fewer data requirements and holdings may not face the same need to establish areas with overall data responsibility as agencies with significant data collections. Alternative mechanisms for initiating, coordinating and monitoring agency-wide approaches that can be employed include working groups and committees with specific tasks, for example, promoting data quality.

2.51 To assess the agency approach to data management the ANAO sought to determine whether agencies had an appropriate area, group or individual with overall responsibility for agency data. In particular, the arrangements should underpin the implementation of agency-wide strategies and standards, and monitoring the performance of local-level areas with databases and data collections.

ANAO findings

2.52 The ANAO found that two agencies (ACS and DIMA) had designated areas with departmental responsibilities for data although the nature and extent of their roles varied considerably. While the ANAO found in DIST and DTRD areas with some degree of involvement in data management issues from an overall agency perspective, DEETYA did not have an area responsible for agency data.

ACS

2.53 The ANAO found that ACS had established the Corporate Data and Statistical Services Unit (CDSSU) which has three key functions:

- to implement or facilitate the recommendations of the 1995 joint ACS-ABS Review of ACS Statistical Requirements and Capability;
- to produce corporate statistical data; and
- to provide statistical advice to other areas of ACS.

2.54   The ANAO was also advised by ACS that, in the context of establishing a corporate information management plan and the outsourcing of ACS's IT, consideration was being given to establishing a Corporate Data Management Section to implement the proposed Corporate Data Management Strategy.

2.55   The ANAO found that the CDSSU had played a role in improving the quality of departmental data and promoting agency-wide data standards, particularly through its publication of *Customs Figures: Australian Customs Service Quarterly Statistical Bulletin*. The bulletin is a key vehicle for communicating agency measures for data, such as ACS's Statistical Principles and Statistical Framework (see paragraph 2.16) and raising the visibility of major data sources.

2.56   The regular publication of *Customs Figures*, plus the critical review of data by the CDSSU, had helped to focus attention at the local level within ACS on departmental principles and standards in relation to data. The CDSSU advised the ANAO that it had observed a steady improvement in both the quality and departmental relevance of data provided for *Customs Figures*. As well, the CDSSU advised that data were being produced more efficiently. An internal staff survey, based on Australian Quality Council principles and conducted in late 1996-early 1997, also indicated a general perception that ACS was collecting and producing data which was more useful and aligned better to business needs.

2.57   While the CDSSU has been effective in implementing several corporate data strategies, the ANAO found evidence of problems in relation to the level of staff awareness of these strategies. The internal staff survey in ACS, discussed in the previous paragraph, revealed that there were pockets of improved data management but also areas where staff were critical of the relevance of data collected, the lack of data review mechanisms and difficulties in accessing data. These critical findings reflected several of the criticisms made in the 1995 Statistical Review, indicating that ACS had still to implement all the recommendations of that earlier review such as the development of statistical research tools for improving data access and undertaking detailed reviews of the agency's major data collections.

2.58   Some of these problems are being addressed by ACS as a high priority through the provision data warehouse technology. A data warehouse is a strategic IT application which integrates data from separate databases and provides advanced research and reporting tools for extracting and manipulating data. A corporate data warehouse is being introduced in ACS for priority areas, while 'data marts' (which are small data warehouses) and off-the-shelf reporting tools are being provided to

other areas. These technical measures should overcome concerns about the presentation of, and access to, data.

2.59   The ANAO also noted that the CDSSU has commenced a review of management reporting data for ACS's Border Operations, an area recognised in ACS as having more data quality problems than other areas in the agency. This review is intended as a pilot test for developing a regular procedure for auditing ACS's statistical collection processes which was recommended in the 1995 statistical review of ACS. Undertaking statistical audits of ACS's major data collections should address concerns in some areas of ACS about the absence of review mechanisms for data.

2.60   In addition to these measures, the ANAO considers that educating staff on strategic measures and departmental guidelines for agency data would improve the general awareness of the business purposes of data and its role as a corporate asset throughout ACS. ACS's recent internal review of staff indicated that while the agency executive had a good understanding of agency-wide approaches to data, lower level staff were not as aware of these measures. Incorporating the Statistical Principles and Statistical Framework in ACS new information management plan, as discussed above, will assist in improving the visibility of these tools; but to be effective these measures need to be promoted actively within the agency.

2.61   To this end, the ANAO considers that the CDSSU should undertake a data awareness campaign providing briefings and training on data roles and responsibilities for staff involved in local-level data collection, production and use.

2.62   In terms of the accountability and performance assessment of local-level collections in ACS, the CDSSU has an informal role through its review of data quality in the production of *Customs Figures*. This provides the CDSSU with an indication of some elements of data performance, mainly data accuracy, availability and timeliness. However, the CDSSU does not have formal responsibility for the accountability and performance assessment of agency collections.

2.63   The ANAO considers that responsibility for managing a performance assessment system should be assigned to a specific area or individual to provide the executive with an overall perspective on data management performance across the agency and to assist local areas with implementing such a system. This is discussed further in relation to establishing a performance assessment framework for data management at the agency level (paragraph 2.81).

DIMA

2.64   The ANAO found that DIMA had established several bodies with overall responsibilities for different aspects of the department's data:

- Standards and Data Management Section (SDMS) which has prime responsibility for, among other things, implementing DIMAs' Data Management Strategy, developing and providing standardised data tools, documenting data systems and monitoring compliance with departmental standards and rules. The SDMS is also involved in the development of the new ICSE system;

- Data Accuracy Working Group, comprising both data custodians (systems staff) and data users (business area staff), which has improved the department's Data Accuracy Project and maintains a Data Accuracy Register which identifies tasks designed to improve data quality and address operational inefficiencies which affect data; and

- Data Quality Committee, which is a cross-program body comprising business owners of systems, regional officers and SDMS staff, with a mandate to promote DIMAs' data quality agenda. For example, the DQC initiated the development of performance indicators for data systems (see paragraph 2.85) and undertook a data quality awareness campaign for regional and state offices.

2.65   The SDMS has been the central point in implementing, and monitoring compliance with, the Data Management Strategy. As cross-program bodies, the Data Accuracy Working Group and Data Quality Committee have helped to promote strategic approaches to data management across the department and provide a channel of communication between DIMAs' corporate area (the SDMS) and operational areas.

2.66   The ANAO found that although there were a number of areas with responsibility for different elements of data management, DIMA had been effective in implementing agency-wide approaches to some elements of managing data, improving data quality and operational efficiency. The areas have been able to represent the views of stakeholders effectively and improve communication in the agency. Therefore, data management proposals have high credibility across the department.

2.67   One measure of the effectiveness of these bodies is the outcomes resulting from the Data Accuracy Project which include:

- a reduction of client records for merging from 150 000 at the start of the merge project to 992 in February 1997 (in other words, a reduction in duplication). This has achieved an estimated annual saving of $300 000;

- a reduction of records in the visa error file from an average of 1200 per month in 1994 to less than 100 in early 1997. This represents an annual saving of $400 000;

- improvements in data accuracy have made a significant contribution to reducing the number of referrals of clients to DIMA officers by 63 per cent since 1994. This represents an annual saving of $441 000; and

- the overstayer data error rate of 41.9 per cent following the Migration Reform Act of 1994 dropped to 27.5 per cent in 1997, amounting to an annual saving of $69 120.

2.68 As well, consideration is being given to assigning overall responsibility for visa data, DIMAs' major data category, to the Program and Performance Reporting Section; and overall responsibility for departmental statistical principles and practices to the Research and Statistics Unit.

2.69 However, while the above approach has been effective in managing data quality and monitoring operational efficiency, there are problems in relation to monitoring other aspects of performance. The SDMS collects information on data accuracy and documents the compliance of data systems with departmental standards. As part of the Data Quality Committee, the SDMS also developed a set of performance indicators for key data characteristics which were intended to be used in a data quality performance information project.

2.70 However, this project has not progressed due to the lapsing of the Committee. Consequently, because the SDMS monitors only some of the key characteristics of data quality, DIMA does not possess a comprehensive and balanced picture of data quality performance across its major data applications. While DIMA can point to improvements in data accuracy, it can not say whether such data are available to all those who need to use it (which is the core principle of the Information Management Strategy). This is discussed below in relation to assessing performance (paragraph 2.81).

DEETYA, DIST, DTRD

2.71 The ANAO found that DEETYA, DIST and DTRD did not have areas, groups or a person with overall responsibility for agency data. However, DIST and DTRD have areas with departmental responsibilities in relation to agency information. In DIST, the Information Management Section has undertaken a data flow mapping project in selected branches which was intended to, among other things, report on current data usage, determine whether branch data requirements are being met and identify problems. In DTRD, the Information Services Section (which is mainly responsible for departmental IT matters), together with the Central Records

Management Area have carriage for the consultancy to develop a departmental information management plan which encompasses all areas of agency data. However, neither of these sections has any agency-wide responsibility for the management or accountability of agency data performance. Both of these functions occur within program or sub-programs.

2.72 There is no area in DEETYA with an overall data management role. This reflects the devolution of data management to local areas and limited consideration of agency approaches to data in the department. However, the ANAO noted that the cross-program Performance and Evaluation Committee established recently could provide a mechanism which would fulfil, if required, an organisational role in relation to data management.

Conclusion

2.73 The ANAO found that ACS and DIMA have designated areas with overall responsibility for managing, implementing and monitoring each agency's strategic approach to data management. These areas are an important vehicle for progressing data management strategies as they serve as a common point of advice on data standards and issues within agencies. Such areas are also in a position to monitor and report to the executive on the performance of individual collections and overall progress against strategic plans. While the respective areas in ACS and DIMA largely have been effective in the implementation of agency strategies and monitoring progress against plans, these areas have played a lesser role in relation to monitoring the performance of individual collections. Data accuracy is, however, monitored in DIMA.

2.74 DIST and DTRD have central areas with limited roles for data management but DEETYA has no area, group or individual with agency-wide responsibilities for data management. While the need for establishing areas with overall responsibility for agency data will differ between agencies depending on several factors, including an agency's data requirements, the ANAO considers that such areas perform a valuable role in managing and monitoring agency data activity and performance.

Recommendation No.3

2.75 The ANAO recommends that ACS improves staff awareness of ACS's strategic approaches to data management, such as the Corporate Data Management Strategy and role of the Corporate Data and Statistical Services Section, to assist in ensuring that there is a consistent approach to managing data within the agency.

ACS's response

2.76   Agreed. The ACS has now established areas with responsibility for Information Management and Data Warehousing within the Office of Business Systems Division. This reflects the corporate priority given to strategic approach to data management and associated roles. This approach will continue to be developed and disseminated throughout the organisation in the near future.

Recommendation No.4

2.77   The ANAO recommends that, to assist with the development, implementation and monitoring of strategic corporate approaches to data management, DEETYA, DIST and DTRD assign a specific area or individual with agency responsibility for data management including:

- developing corporate strategic plans, standards and guidelines for data;
- providing advice to local-level areas on data management issues;
- promoting staff awareness of agency data management strategies and measures; and
- monitoring the performance of local-level databases and data collections.

DEETYAs' response

2.78   Agreed.

DIST's response

2.79   Agreed. The need for a specific area or individual to be assigned with agency responsibility for data management is contingent upon the department's data risk management framework.

DTRD's response

2.80   DTRD agrees with this recommendation and notes that the Draft Information Management Plan being developed is expected to address this issue.

Performance assessment at the agency level

2.81   Performance information[12] for key data characteristics should allow the agency executive to gauge the extent to which local-level data collections and databases comply with agency-wide standards, rules and objectives. In this way performance assessment can assist in monitoring an agency's

---

[12]   For more on performance information, see the joint ANAO-Department of Finance and Administration publication, *Performance Information Principles: Better Practice Guide*, November 1996.

progress against objectives defined in strategic planning for agency data and provide feedback to improve performance.

2.82 The ANAO recognises that developing and establishing a performance assessment framework for agency data collections is complex. The ABS has pointed out that the APS is at an early stage in developing performance criteria and methodology for data collections. In view of the different ways in which data is collected and produced, it is also unlikely that there will be a single model by which data performance can be assessed. However, the ANAO considers that agency standards and rules for data provide a starting point for identifying the elements of data performance which might be measured. There are a number of essential characteristics which most data should possess which also provide a set of criteria against which performance indicators can be defined. These are discussed below at paragraphs 2.84 to 2.90.

2.83 In relation to performance assessment at the agency level the ANAO sought to determine whether a performance assessment framework was used to monitor and provide accountability for local-level data operations.

ANAO findings

2.84 The ANAO found that none of the five agencies examined had established at the agency level formal performance information and reporting for their data holdings, although one agency (DIMA) had developed a set of performance indicators to assess data quality and monitor measures to improve it. The discussion below examines this set of performance indicators.

DIMA

2.85 Performance indicators were developed by DIMAs' Data Quality Committee to assist in raising awareness of data quality in the department, identify critical areas needing attention and assess the effect of DIMAs' data quality campaign over time. The results from assessments against these indicators were also intended to be provided to individual data system owners to assist them assess their own data quality priorities.

2.86 The performance indicators included quantitative and qualitative measures addressing eight key data characteristics. Details of the performance indicators for each characteristic are provided in figure 2.

Figure 2
DIMAs' Data quality performance indicators

Data characteristics	Performance indicators
Redundancy Redundancy refers to the duplication of the same information in different systems, databases and files. A goal of the data quality program is to minimise the effect of data redundancy through data sharing and data re-use	• number of copies of data • integration of data • re-use of data
Accuracy Accuracy refers to the data being correct, consistent, compatible, complete and up to date.	• user and client surveys (qualitative) • number of unnecessary referrals resulting from data inaccuracies • number of errors • cost of fixing the errors
Accessibility Accessibility refers to the timely and optimal availability of data for decision makers, that is, data available to the right person, at the right time, producing the right decision.	• number of unnecessary referrals resulting from data inaccuracies • number of onshore clients unable to be processed due to data unavailability • number of errors
Flexibility Flexibility refers to the adaptability of data structures to meet business needs.	• level of data re-use • level of integration • level of data modelling • ease of data reporting
Integrity Integrity refers to the consistent application of business rules and compliance with data standards, for example, data naming, definitions and formats.	• compliance with data standards through applications development and enhancement activities • number of merge cases • number of error holding file records
Security Security refers to protection from unauthorised access; and confidentiality and privacy considerations.	• data is correctly classified • access is controlled • knowledge of DIMA security policy • audit records of update and delete transactions • audit records identifying access to sensitive data

2.87 The ANAO considers that these indicators provide a basis on which DIMA, and other agencies, can build a performance assessment framework for departmental data systems. In particular, the performance indicators for data accuracy and accessibility link aspects of data quality to a key business outcome: DIMAs' processing of clients. This has enabled DIMA to estimate the benefit to its business processes and outcomes of improvements in data quality. For example, performance information on the data accuracy of visa forms has indicated a reduction in errors which DIMA has traced to a corresponding decline in client referrals, with estimated savings of $400 000 per annum.

2.88 Defining performance indicators is one step in the performance assessment process which also requires that information for those indicators is collected, assessed, fed back to managers and used to assist in improving performance. However, the ANAO found that DIMA can only measure performance for data accuracy as information was not collected for the other performance indicators with the lapsing of the Data Quality Committee in mid-1996. DIMA is therefore unable to assess data performance comprehensively nor the effectiveness of the data quality project, except in relation to data accuracy.

2.89 The ANAO also considers that there is scope for DIMA to improve the existing performance indicators for data quality and define indicators for other aspects of data performance such as efficiency and effectiveness. For data accessibility, qualitative information from a range of data users other than staff processing onshore clients would provide a wider perspective on the timely availability of data for those who need to use it. Data visibility is another critical characteristic which could be measured by checking if data holdings are listed in agency or program directories and collecting feedback from users on the ease with which they are able to identify data necessary to perform their jobs.

2.90 It is important that the performance assessment framework uses a balanced set of indicators which address different elements of performance, not simply one aspect of it. This is to ensure that all the key factors which influence outcomes can be examined by managers, rather than only those factors which are easy to measure.

2.91 DIMAs' performance indicators focus mainly on data quality outputs. Although part of the outcome from improvements in data accuracy can be estimated in terms of savings, the ANAO considers that DIMA should also seek to measure the cost/benefit of its data systems by developing performance indicators for inputs (costs) and outcomes (benefits). This would require a mix of quantitative and qualitative measures, particularly for outcome indicators. Quantitative measures for

inputs could be derived from techniques used in activity based costing. While DIMA has been able to devise indicators which quantify some outcomes (for example, savings, numbers of client affected by data errors), user feedback is also important in arriving at an estimate for data benefits.

2.92 The ANAO recognises that developing a performance assessment system for data management will be a gradual process involving testing and refinement. DIMA has therefore taken some important first steps in this process which should be of wider benefit to other APS agencies. However, the DIMA example also points to the need, when establishing a performance assessment framework, to consider not only performance indicators but also the collection of information and procedures for monitoring, reporting and providing feedback on performance results.

Conclusion

2.93 The ANAO found that DIMA was the only agency of the five examined which had established a performance monitoring framework for its data, although it had not progressed beyond collecting performance information for one element of data quality. The performance indicators developed by DIMA reflected sound principles in relation to how the indicators provide direct links between data management issues and their implications for key departmental business outcomes. However, the ANAO considers that there is scope for refining existing indicators and, overall, developing a better balanced set of indicators which address inputs, outputs and outcomes in order for DIMA to be able to assess the cost/benefit of the data it collects, produces and uses.

2.94 The performance indicators developed by DIMA provide a useful starting point for other agencies in developing performance assessment frameworks for their own data requirements. While the DIMA indicators are tailored towards the data needs and business objectives of that agency, the data characteristics which underlie the indicators have broad application which many agencies could adapt to meet their own requirements. In ACS's case, for example, performance indicators for data accuracy, accessibility and integrity (see figure 2) could be adapted to establish a performance information regime based on the ACS's Statement of Statistical Principles and Statistical Framework.

2.95 As well, alternative sources of performance-related information and tools should be used to assist agencies establish a balanced set of performance indicators. Benchmarking and activity-based costing are two methods which could provide information and measures for developing a performance information framework. The use of existing sources of performance-related information for collections at the local level is discussed in Chapter 4.

2.96 Advice should also be sought from program performance monitoring and evaluation areas within agencies on developing performance monitoring and reporting frameworks.

2.97 Moreover, agencies should consider assigning responsibility for implementing and managing a performance assessment system to a specific area or individual. Such a position would provide a contact point and source of advice to assist local-level data management with the implementation of performance assessment. It would also assist the executive by providing an overall perspective on data management performance across the agency.

2.98 The ANAO has made two recommendations below in relation to establishing performance assessment frameworks at the agency level. These recommendations should be considered in conjunction with the ANAO's recommendation in Chapter 4 in relation to establishing performance assessment for collections at the local level.

Recommendation No.5

2.99 The ANAO recommends that to enable performance to be measured and improve accountability ACS, DEETYA, DIST and DTRD each should establish an agency-wide performance assessment framework for databases and data collections. The performance assessment framework should include:

- a balanced set of performance indicators which addresses inputs, process, outputs, outcomes and client satisfaction and the collection of information which allows their measurement; and

- establishing a system of monitoring, reporting and feedback between the area responsible for performance monitoring and local-level data managers.

ACS's response

2.100 ACS agrees that formal performance assessment of databases and data collections is needed. However, it is considered that an agency-wide performance framework is not the most appropriate and cost efficient way of managing system performance and accountability. The ACS has identified the data needed for essential corporate management purposes and will apply appropriate quality controls to these data collections. Other 'local data' collections will be subject to a general statement of data quality management principles to be applied at the local level. At present, system performance and accountability is reviewed by Internal Audit, in accordance with the Annual Internal Audit Plan, developed by the Audit Committee.

DEETYAs' response

2.101   Agreed.

DIST's response

2.102   Agreed. The need for a specific area or individual to be assigned with agency responsibility for data management is contingent upon the department's data risk management framework.

DTRD's response

2.103   Agreed. The Draft Information Management Plan being developed by DTRD is expected to include recommendations to establish an information management planning and review cycle. The planning cycle will be linked into the department's overall corporate planning and performance assessment activities. Information management activities (including data management) will be assessed under this plan according to their contribution to business goals. At a lower or database level, DTRD has already established the data performance measures. Assessment against these measures will in future be carried out within the overall information management framework.

Recommendation No.6

2.104   The ANAO recommends that, to measure all relevant aspects of performance, DIMA further develops its data quality performance information project. This should include:

- developing a balanced set of performance indicators which addresses inputs, process, outputs, outcomes and client satisfaction;
- collecting performance information for these indicators from local-level databases and data collections; and
- establishing a system of monitoring, reporting and feedback between the area responsible for performance monitoring and local-level data managers.

DIMAs' response

2.105   Agreed. DIMA intends to develop a balanced set of performance indicators during the development of its new client application processing system known as ICSE which will progressively replace many of DIMAs' existing systems over the next three years.

Better Practice Principles

From the work undertaken in this audit, the ANAO has identified the following practices which would enhance data management:

Strategic plans and agency guidelines

- Link strategic plans for data management to the corporate plan and IT plan to help staff focus on data as a corporate asset, to meet business needs and coordinate agency-wide approaches to data with agency IT strategies.

- Identify in agency plans and/or guidelines statutory requirements and whole-of-government policies which relate to data to help ensure staff manage data in accordance with legal obligations and to assist the agency align its strategies for data management with wider developments and trends in the Commonwealth Government.

- At the agency-level, define principles and rules to govern agency data and issue these in guidelines to staff. This will provide a common basis and method for the collection and production of data throughout the agency and therefore will enhance the agency's overall ability to share and compare data. The Statistical Principles and Statistical Framework developed by ACS are an example of better practice as they provide a common method by which the ACS's business areas can develop statistical indicators for local and agency data needs, as well as being easy to follow and therefore more likely to be used by staff.

Promoting agency-wide strategies and measures
- Publish a regular bulletin of statistical information on performance and trends for the agency's major business which includes agency strategies and guidelines for data. The regular production of a bulletin will enhance overall data management by:

  - reminding local areas of agency requirements and standards in relation to data;

  - focusing local areas on the requirement to produce data which supports business needs and corporate goals rather than producing data which only meets local needs; and

  - promoting data visibility by identifying data sources.

- Undertake educational campaigns about agency strategies and measures for data throughout the agency to promote the agency's data management framework, raise staff awareness of their role and responsibility for sound data management and seek feedback from

staff on their views and concerns regarding data. Educating staff directly will not only reinforce the agency's data management framework but also enable the agency executive to stay in touch with staff views on how data management affects their work performance.

- Establish cross-program committees, comprising data producers, data users and agency data management staff, to develop and promote agency-wide initiatives. Including representatives from different points in the data life-cycle allows these bodies to identify and consider the range of issues affecting data management, and lends legitimacy to agency-wide measures as these bodies are highly representative of the different interests within the agency.

Performance assessment of data management

- Establish an agency-wide performance assessment framework to provide the executive with information on progress against agency strategic plans, compliance with departmental standards and the cost/benefit of agency data operations; this information should also serve accountability requirements.

- Develop a balanced set of performance indicators which measure efficiency, effectiveness, outputs and outcomes for local-level data management and provide information which can be aggregated at the agency level. The performance indicators developed in DIMA provide an indicative example of output performance indicators upon which agencies could build to meet their own requirements.

3. Data Availability: Visibility, Access and Dissemination

This chapter discusses agency-wide approaches to data availability, particularly in relation to data visibility, access and dissemination including pricing. The ANAO found that only a few agencies had taken an agency-wide approach to data visibility and access. In other agencies without agency-wide measures, the ANAO found it difficult to identify data collections and that there were data access problems. None of the agencies examined had addressed data dissemination or pricing on an agency-wide basis, resulting in inconsistencies among local-level data collections. The ANAO has made five recommendations aimed at improving data visibility, access and dissemination.

Introduction

3.1 The importance of adopting a strategic agency-wide approach to data management has been discussed in Chapter 2. In discussing the development of a plan and/or guidelines to address overall data management, the ANAO highlighted the need to address data availability at the strategic level. This involves policies and measures designed to make data available for various purposes, within and outside agencies. The ANAO notes that some data collected under legislation or specific authority may have restricted availability for privacy or security reasons. However, the planning for, and collection and analysis of, data is aimed generally at making information available for decision-making, research purposes and so on.

3.2 There are a number of elements to be considered in relation to data availability at the agency level and which should then flow through to local collections, particularly data visibility and accessibility. Another issue which requires agency-wide consideration is data dissemination and pricing. The key criteria against which these issues were examined are set out at the beginning of each section and followed by a discussion of the ANAO's findings in relation to the agencies included in the audit.

Data Visibility

3.3 A prerequisite for data use is that those who need to use data must be able to find it. That is, there must be transparency throughout the agency that allows decentralised data collections to be visible across the agency.

The ABS has recommended that within agencies every data collection should be registered or documented and that it:

would seem essential that there is at least some directory or register of data holdings within an agency, either at the agency level or at an intermediate level (division, branch or program level).

3.4   Information which explains the data and its purpose should also be provided. The ABS has identified the following as information (in order of importance) about a data collection which could be recorded in a directory or register:

- the title of the collection;

- a contact name and/or phone number of the person responsible for managing the collection;

- a description of the information needs that the collection was designed or purchased to address;

- the time reference period to which the data refers;

- some key words that describe the contents of the collection;

- measures regarding the quality of the data;

- classifications, data items, populations or other statistical terminology used;

- a description of the way in which the collection links to agency objectives; and

- any other details of interest, for example who undertook the collection, any related previous or planned collections and the cost of the collection.

3.5   The ABS considers that the first three items, at least, should be included in a central directory, while the others could be held in a directory at a lower level of the organisation.

3.6   Recently, the Australian Government Locator Service Working Group (AGLSWG) has developed a metadata[13] standard to promote online access to government data and information. The attributes of the AGLSWG metadata standard are similar to the key information for data visibility defined by the ABS in paragraph 3.4. OGIT advised that all three tiers of Australian government, including the Commonwealth, have endorsed the AGLSWG metadata standard. Further information on this metadata standard is at Appendix 3.

---

[13]   Paragraph 4.30 and 4.31 define and discuss metadata; Appendix 3 defines the Commonwealth Metadata Standard.

3.7 To assess whether data were visible, the ANAO examined measures adopted to promote visibility, including establishing directories or catalogues of agency data holdings.

ANAO findings

3.8 The ANAO found that only two (ACS and DIMA) of the five agencies examined had agency-wide measures for promoting data transparency including a directory of agency databases. However, in DIMAs' case the directory was limited to data held in business systems and did not cover statistical collections whereas ACS had directories for both its business systems and statistical collections. The other three agencies (DEETYA, DIST and DTRD) did not have measures to promote data visibility or directories of data holdings.

ACS

3.9 The ANAO found two sources which served as directories of ACS data holdings. The agency's regular statistical publication, *Customs Figures: Australian Customs Service Quarterly Statistical Bulletin*, identifies the statistical sources currently available within ACS. The bulletin groups the data according to the key result areas defined in the ACS Corporate Plan and includes notes on the purpose of the data and any limitations which might apply. The 1995 *Review of ACS Statistical Requirements and Capacity* also includes an inventory of ACS's statistical collections. The inventory provides details on each statistical collection's title and purpose, the area in ACS where it is located and whether it is a centralised or regional collection.

3.10 These two directories provide a reasonable level of visibility for ACS data collections and their purpose but could be improved by including contact details for each collection.

3.11 ACS is also addressing, on an agency-wide basis, the visibility of data stored in small-scale local databases. These databases hold important data which is not particularly visible outside the immediate area that generates and uses it. To improve the visibility of the data stored in these systems, the *Customs IT Strategic Plan* stipulates that a register of all end-user systems will be maintained and made available to staff and that these systems are to be adequately documented.

DIMA

3.12 DIMAs' Data Management Strategy identified the need to document data adequately through establishing an inventory of existing databases and files. The aims of establishing an inventory were to identify what data is available, where it is located and how it is used. The SDMS developed a

Corporate Data Dictionary (CDD) which is a catalogue of DIMAs' information systems and defines these systems, their holdings and data items. As well, the catalogue identifies the custodians of data collections, business rules applying to data items and links to standard data names.

3.13 The SDMS has also compiled a *DIMA Directory of Information Systems*. Based primarily on the CDD, the directory includes information for four categories of systems: core business systems; non-core business systems; minor departmental systems; and historical systems. This includes the following information for each system:

- title;

- starting date;

- mainframe system code;

- functional description; and

- a diagram with explanatory description of where and how the database fits into the business process which it supports.

3.14 The directory also includes a system ownership table which identifies the area in DIMA with custody for the system and the area with responsibility for the system's development and maintenance. In addition to the information provided for each system a diagram, called the 'DIMA Web', has been prepared which shows all the major departmental systems and the information flows between them. As with the directories in ACS, inclusion of contact details would allow users to identify quickly the relevance of databases to their needs.

3.15 Although the CDD and the directory provide visibility to the data stored in DIMAs' business systems, they do not include any information on data collections which are not mounted on departmental mainframes. For example, major statistical collections managed by the Research and Statistics Unit (RSU) are not included with the CDD or directory and therefore are not visible. While the RSU, the department's major statistical analysis and research area, is aware of significant statistical collections in other areas of DIMA, there is no register or directory of the department's statistical collections. Without a directory of these collections, there is a risk that users outside program areas or the agency itself will be unable to identify data relevant to their needs, that data will be under used and, with staff movement, critical information about statistical collections may be lost.

DEETYA, DIST and DTRD

3.16 As noted above, agency-wide measures to promote data transparency and data directories do not exist in DEETYA, DIST and DTRD, although the consultancy for developing an Information Management Plan

for DTRD will involve documenting departmental information holdings and flows, including data.

3.17 The ABS has recommended that, if no agency-wide directory of data collections exists, there should be directories at an intermediate level such as the branch or program level. In the case of DEETYA, for instance, it may be appropriate to have separate directories of data for the employment and education areas of the department. However, the ANAO did not find intermediate-level directories of data collections in DEETYA, DIST or DTRD. Knowledge of data holdings at this level was based largely on the personal experience of key staff, although this tended to be incomplete. In the Vocational, Education and Training (VET) program area of DEETYA, for instance, the ANAO spoke with all branch heads but was unable to identify any officer or area with a comprehensive knowledge of the data collections used in the program area. This was because the different branches in the VET program area rely on different data collections, some of which were external to the department or are maintained by other program areas within the department.

3.18 The lack of data directories within agencies means that internal and external users are likely to face difficulties and delays in identifying critical data, particularly if it is not easy to find the appropriate officer to answer a request. This is inefficient for both the user and staff involved with answering requests. There is also the risk that incomplete or wrong data will be used. In DEETYA, the ANAO found an instance where separate branches provided two different sets of data, which were mutually inconsistent, in response to the same enquiry.

3.19 As well as impacting on data visibility, the lack of a data directory also increases the risk that data will be lost from the corporate memory when staff move or agencies are reorganised. In DIST, the IT Strategic Plan states that:

> *there has long been a concern that … databases with valuable industry information are not documented, not managed as a whole, are not generally available to officers across the Department and, being managed by individual officers, are in danger of being lost when the responsible staff move to other positions.*

3.20 However, DIST does not have a strategic plan that addresses these concerns. The problems in DIST appear to be similar to those associated with end-user systems[14] in ACS, and other agencies, which have been

---

[14] End-user systems is a term applied to user developed applications which serve as small databases to meet local data needs.

addressed partly by establishing agency-wide control measures, including the creation of a directory of end-user systems and rules for documenting those systems. DIST could therefore learn from those agencies which have already established agency-wide approaches to data management.

Conclusion

3.21 The ANAO found that ACS data were highly visible as a result of agency-wide measures regarding the documentation of data and the creation of useful directories. ACS has directories that identify major databases, statistical collections and smaller end-user computing systems. While the databases which cover transactional data in DIMA were readily apparent, the transparency of statistical collections was limited as there was no directory or area which maintained a record of these collections. The latter situation would be improved significantly by the creation of an agency-wide directory comparable to the one that exists for DIMAs' business systems.

3.22 The ANAO found it difficult to identify data collections at the agency level in DEETYA, DIST and DTRD, none of which has measures for addressing data transparency. In DEETYA this difficulty was compounded in one program area where there was uncertainty about the data collections used across that program. While the poor visibility of a number of databases is acknowledged in DIST, the ANAO did not find any current measures to address these concerns.

3.23 The ANAO considers there would be significant benefits for these agencies in creating directories of agency data holdings similar to those established in ACS and DIMA. Consideration should also be given to including in agency directories the details for each data collection as suggested by the ABS to ensure that a quick reference to key information about individual data is available.

3.24 The ANAO also notes that there are a number of options for establishing directories which agencies should examine in terms of their own requirements and cost effectiveness. These options range from using advanced data warehousing technology through to directories maintained on personal computers (PCs) and published in internal bulletins. In this regard, the ANAO notes that small agencies are not disadvantaged by IT or budgetary constraints in establishing directories. The smaller the number of databases the less the requirement to base a directory on high-cost IT systems.

Recommendation No.7

3.25   The ANAO recommends that, to promote data visibility, DEETYA, DIMA, DIST and DTRD each create directories of the data collections and statistical collections held in each agency and that for each collection/ database these directories include:

- the title of the collection;
- a contact name and/or phone number of the person responsible for managing the collection;
- a description of the information needs that the collection was designed or purchased to address; and
- other relevant details such as data quality, classifications and links to other data, where necessary.

DEETYAs' response

3.26   Agreed.

DIMAs' response

3.27   Agreed. DIMA intends to expand its current visible repository of meta data on data holdings to include data on statistical collections.

DIST's response

3.28   Agreed. The Department, in its 1998 upgrade to its IT infrastructure, is locating and registering all electronically maintained applications and data collections in the Department. This will provide key input to its risk management framework which will, in turn, provide a basis for determining the department's strategic approach to data management in accordance with the ANAO's audit findings.

DTRD's response

3.29   Agreed. A full list of database holdings is being developed as part of the Information Management Plan.

Data access

3.30   To maximise benefits in return for the cost of collecting and storing data, it is essential that all those who need to use data can access it. Access to some data may be restricted appropriately for security or privacy reasons but, in general, data should be available in a format which allows the user to convert or manipulate it for their own requirements, such as in spreadsheets or diagrams.

3.31   As mentioned previously (paragraph 2.26), agencies are discovering that they need to share and integrate data which may be, for example, operational, statistical or financial, across several business functions and

with other agencies. However, a considerable volume of data in agencies is locked up in older IT applications which have limited and inflexible reporting capabilities. These are referred to as 'islands of data' which reduce significantly data access and sharing. There are many technological solutions to this problem including data warehouses, search engines and data research and reporting tools, which provide agencies with the means to enhance data accessibility. These technical measures may need to be matched with strategies which overcome the view that data is the property of the branch or area which collected it by educating staff on the importance of treating data as a corporate asset.

3.32 When examining data access the ANAO sought to establish whether agencies were addressing data access issues at the agency level and, if so, the measures being adopted to enhance data access and sharing within agencies.

ANAO findings

ACS

3.33 The ANAO found that ACS is addressing these issues at the agency level through developing a corporate data management strategy and the introduction of technical solutions as follows:

- data warehousing and data marts (which were discussed in Chapter 2). These systems provide complex research, analysis and reporting functions which allow users to access data integrated from different and separate databases;

- research and reporting tools to enable staff to extract and manipulate data to meet their own data and information needs, including transferring data from applications to PCs; and

- on a selective basis, implementing knowledge engineering systems which apply rules to the processing of business transactions for the purpose of assessing client risk and eligibility.

3.34 The Corporate Data and Statistical Services Unit (CDSSU) of ACS has also promoted the conversion of manually stored data into an electronic format in different areas of ACS in order to produce statistics for management reporting purposes, although this process has been constrained to some extent by delays in implementing a standard PC environment which would allow data storage in a consistent electronic form.

DEETYA

3.35 Due to the absence of an agency-wide approach to data management in DEETYA, the ANAO was unable to form an opinion on

whether data access and sharing was being addressed at the agency level. However, the ANAO was advised that, as the department had disparate data needs, there was a limited requirement to access or share data across program areas.

DIMA

3.36   The Information Management Strategy identifies data (and information) availability to all those who need it and are authorised to use it as a core principle which should guide DIMAs' information and data management.

3.37   The main vehicle which DIMA is developing to promote data access and sharing (or reuse as it is defined in DIMA) is the Integrated Client System Environment (ICSE) project. ICSE is intended to provide DIMA with integrated databases, and therefore will assist in overcoming many of the current technical obstacles involved with sharing and reusing data from separate and disparate applications. Data on clients are to be structured and stored in an integrated manner which will allow data to be accessible for almost all activities involving clients: DIMAs' core business.

DIST

3.38   The DIST *IT Strategic Plan 1995-97* identified as a priority the provision of systems 'to identify, manage and make widely accessible information of agency-wide interest generated in program areas'. However, it was also recognised in this plan that barriers existed to accessing industry client data held in disparate databases within the department and that these would need to be addressed by a consistent department wide approach. The ANAO also found evidence that data cannot be transferred easily between different databases and that staff were unable to compare data across departmental programs because databases were not linked. While approval had been given to develop a consistent departmental approach to these problems, the ANAO was advised that the project had lapsed in mid-1996.

DTRD

3.39   In DTRD, the objectives for the department's information management plan include improving the accessibility of information by providing a framework for the standard and orderly processing of departmental information and promoting the sharing of information from different sources. Such measures are not as yet in place or being developed by DTRD. As noted above, DTRD is engaging a consultant to develop an information plan which will address data (and information) accessibility.

3.40 The ANAO considers that DEETYAs' data access and sharing requirements should be determined by an examination of this issue at the agency level and not by default at lower levels within DEETYA. Although it may not be necessary to provide extensive access and sharing to some data across program areas, there could be a significant requirement to share and integrate some departmental databases for example, for the purpose of providing performance information. Decisions on these matters need to be determined at the agency level by the departmental executive.

Conclusion

3.41 The ANAO found that two of the five agencies examined (ACS and DIMA) had established measures for addressing data access and sharing at the agency level. The ANAO also noted that these issues are to be addressed in the DTRD information management plan which is to be developed by a consultant. In DIST, while the importance of adopting a consistent Department-wide approach to identified access and sharing problems has been recognised, measures to address problems have not eventuated. In its response to the draft audit report, DIST advised that 'since January 1998, the department has commenced a major upgrade of its IT infrastructure. An essential component of this upgrade has been the identification and registration of electronically maintained applications and data collections across the Department. The Department believes that this is a significant step towards the agreed goal of agency-wide data transparency and in addressing the access and sharing problem'.

3.42 The ANAO was advised by DEETYA staff that the department's disparate data needs meant there was little requirement to share or access data across program area boundaries. However, in the course of the discussions, staff also indicated that data is shared between program areas, for example the VET and Schools program areas are increasingly reliant on each other's data. This highlights the need for an agency-wide approach.

Recommendation No.8

3.43 The ANAO recommends that, in order to maximise the benefits of data collected, DEETYA and DIST should determine the extent of the need to access and share data on a department-wide basis and put in place data access and sharing mechanisms which support the identified need.

DEETYAs' response

3.44 Agreed.

DIST's response

3.45 Agreed.

Data dissemination

3.46 Data dissemination in this discussion refers to the external distribution of data; that is, to government and non-government bodies, and the public.

3.47 The ABS states two reasons for establishing agency policies and/or guidelines for data dissemination:

- to avoid individual areas within agencies 'reinventing' policies when applicable policies exist elsewhere in the agency; and

- related to this point, to ensure consistency of dissemination throughout the agency.

3.48 A key issue which should be determined by departments' executives is the pricing or charging policy for data which is disseminated. As with dissemination in general, data pricing policies developed for local-level data collections should be in accordance with agency guidelines to ensure consistency and be based on an approved method of calculating data prices.

3.49 The ANAO sought to establish whether agencies had:

- procedures and guidelines at the agency level to provide a framework and check list to assist local-level areas with disseminating data; and

- developed an agency-wide basis for pricing data which was expressed in a policy or set of guidelines. The ANAO did not examine the agency's processes leading to a decision to 'sell' data but, if agencies were doing so, examined what pricing principles had been applied.

ANAO findings

3.50 The ANAO did not examine if data should be sold but did examine dissemination policies including, where relevant, data pricing. At the agency level and the local level the ANAO found that:

- none of the five agencies examined had established agency-wide policies for the dissemination of data that addressed whether there are any legal impediments to the sale of data and, if data can be sold, the principles to be used to determine the price;

- in DIMA, concerns that data is being disseminated inconsistently and without adequate directions by different areas in the agency have led to the proposal for a review of the department's statistical collections, including the pricing regimes used by different areas;

- in all agencies at the local level, all of the data collections which disseminated data had policies addressing the release of data which had been determined locally; and

- all of the local-level data collections which sold data had guidelines addressing the pricing of data. As well, pricing for all these was based on recovering the costs for extracting and disseminating the data, rather than complete cost recovery (which would include collection and other costs).

3.51 The ANAO noted that, while the practice of charging only for extraction and dissemination costs was the common basis for all those data collections which sold data, there was considerable variation in how extraction/dissemination costs were calculated. To some extent this reflected differences in the types of databases and data products sold, and the complexity of data extraction, between different areas.

3.52 The ANAO found within DEETYA that two separate areas used divergent methods for calculating staff costs in determining the price for data sold, although the extraction and dissemination methods were similar. On the one hand, the University Statistics Section took into account the level of the officer assembling the data, and used the full hourly cost rates (that is, direct salary costs plus on-costs) for ASO1 to SOGB officers as determined by the Department, as the basis for pricing data extractions requiring more than half an hour to assemble. On the other hand, the Data Services Section in the Schools Program area of DEETYA had a set minimum charge of $100 for a data request. This was based on flat hourly rates of $35 for the decision time involved in considering a data request and $45 for the programmer's time, regardless of the level of the officer performing the task, and $20 of computer time for a short report.

3.53 The ANAO considers that the divergence in pricing methods between these two areas for similar data extraction and dissemination tasks reflects the inconsistencies which can arise when data pricing is devolved to the local level in agencies without adequate guidance on the basis for pricing, including agreement on what costs areas can be charged for and how charges are to be calculated. Devolving the decision making for specific pricing schedules to local-level data managers is realistic given the diversity of data products sold by agencies. However, the development of local level pricing schedules should occur within an agency-wide framework which addresses such matters as guidance in relation to:

- which agencies/individuals will actually be charged for data;

- what elements of the process are to be included when calculating the cost, for example, the cost of staff time, overheads and full or partial costs; and

- the methodology to be used to determine the price, that is the hourly rate for staff time.

3.54   This would assist local-level areas to calculate data prices and establish a common basis for the sale of data within agencies. Establishing such a basis is important particularly in relation to ensuring that data services are provided in an equitable manner to all users.

Conclusion

3.55   The ANAO found that none of the agencies examined had established at the agency level policies and/or guidelines for data dissemination. The policies and guidelines need to address whether there are any legal impediments to the sale of data. If an agency decides to sell its data, the principles to be applied to determine the price should be articulated. Unless general guidelines for dissemination and pricing are established for the whole of an agency, data are likely to be disseminated, and prices set, inconsistently. In DIMA concerns about inconsistent dissemination practices had raised the need for a review of the department's different statistical collections, including the pricing regimes used. In DEETYA, the ANAO found two separate areas which were selling data on the same basis of cost recovery for extraction and dissemination but using different charging methods to price data.

3.56   The ANAO considers that general dissemination procedures and guidelines should be established at the agency level to provide a framework in which appropriate dissemination policies can be determined by local-level data managers. Agencies should also provide guidelines that establish the basis for data pricing, including who is to be charged for data services, what costs can be charged for and an approved method or formula for calculating costs and prices.

Recommendation No.9

3.57   The ANAO recommends that, to ensure data are disseminated in line with agency policy and data pricing decisions are equitable and underpinned by an approved costing methodology, ACS, DEETYA, DIMA, DIST and DTRD establish general policy guidelines for the dissemination and sale of data which draw the attention of local-level data managers specifically to:

- whether data are allowed to be disseminated;
- relevant statutory requirements applying to data dissemination;
- relevant privacy and security restrictions;
- procedures for obtaining approval to release data (if applicable); and
- the basis for pricing data including the tasks and methods for calculating the price of data.

ACS's response

3.58   Agreed. The data warehouse currently provides a metadata facility to assist users to understand the meaning and context of data holdings. A Corporate Data Repository has been included for development in the ACS Business Information Plan to promote the visibility of corporate data holdings. This is the first step in managing and standardising the use of data in the organisation.

DEETYAs' response

3.59   Agreed.

DIMAs' response

3.60   Agreed.

DIST's response

3.61   Agreed. The need for general policy guidelines for the dissemination and sale of data is contingent upon the department's data risk management framework.

DTRD's response

3.62   Agreed. Currently only information from one database is available for sale. The matters referred to by the Recommendation are addressed at this level. The draft Information Management Plan is expected to recommend that all aspects of policy relating to the provision of public information be developed at an Agency level as the need arises.

Better Practice Principles

From the work undertaken in this audit, the ANAO has identified the following practices which would enhance data management:

Data visibility

- Agencies should record in a directory or catalogue all data collections and databases to promote the visibility of agency data holdings. To assist users find the data they need, the directory should include (for each collection) information on:

  - the title of the collection;

  - a contact name and/or phone number of the person responsible for managing the collection;

  - a description of the information needs that the collection was designed or purchased to address;

  - the time reference period to which the data refers;

  - some key words that describe the contents of the collection;

  - measures regarding the quality of the data;

  - classifications, data items, populations or other statistical terminology used;

  - a description of the way in which the collection links to agency objectives; and

  - any other details of interest, for example, who undertook the collection, any related collections previous or planned and the cost of the collection.

Data access

- Identify at the agency level data access and sharing requirements and, accordingly, develop a set of measures which promote a common method and approach to accessing and sharing data across program areas to maximise the benefits of existing data and minimise unnecessary duplication.

Data dissemination

- Establish an agency-wide policy for data dissemination to ensure that all data are disseminated in accordance with security, privacy and statistical requirements.

- Each agency should include in its dissemination policy specific guidelines on the pricing of data. These will assist local areas to achieve pricing consistency in accordance with methods and rates set by the executive.

4. Requirements for Individual Data Collections

This chapter discusses the requirements for individual data collections. For each of the collections examined the ANAO assessed the agencies' identification of data needs, development of specifications, quality assurance, data use and performance assessment. The ANAO found that, generally, agencies had addressed data needs, specifications, quality assurance and data use adequately. However, performance assessment of individual collections is not highly developed. The ANAO has made one recommendation in relation to developing balanced performance indicators and monitoring performance for individual collections.

Introduction

4.1　Chapters 2 and 3 of this report have discussed issues in relation to data management from the perspective of the whole of the agency. This chapter examines data management at the local level with the focus on individual collections. Throughout the Australian Public Service there is a great number and variety of data collections which provide essential information to underpin the administration of programs, to assist with decision-making and in the development of policies. As well, a range of data is collected which is widely used by industry, academics and non-governmental organisations.

4.2　The specific processes relating to individual data collections are not governed by any mandatory requirements. However, as with data management at the strategic level, there are a number of principles which, if followed, will help ensure that data collections provide accurate and timely data which fulfil identified business needs.

4.3　The first requirement of local data collections is that they are consistent with agency-wide strategies designed to promote the effective management of, access to and use of data collections. Local level data collections should be designed to ensure compliance with agency-wide strategies in the most efficient and effective manner. Also, they should have valid, reliable and accurate data which are available to all who need it. From a management perspective this means that there must be a reconciliation between the local identification of data needs and the design of data collections that are consistent with agency policies. It also implies the development and application of a performance assessment framework for each data collection so that both local and corporate management have

some assurance that the individual data collections are efficiently and effectively contributing to the achievement of corporate goals.

4.4 The ANAO examined 23 individual collections from six agencies (listed in Appendix 1) against criteria developed by the ANAO in conjunction with the ABS. These data collections fell into two categories: data collections that were designed to provide statistical data for a specific purpose (for example, DTRD road crash data collections); and transactional or operational data collections from which statistical data are extracted (for example, DIMA Settlement Database).

4.5 In discussing the data cycle at the local level we are not suggesting that it is a linear process. That is, figure 3 does not describe a sequence of steps that must be followed. Rather, the diagram highlights the issues that should be considered and their relationships. For example, when defining data needs the issue of provider load also needs to be taken into consideration.

Figure 3

Data management at the local level
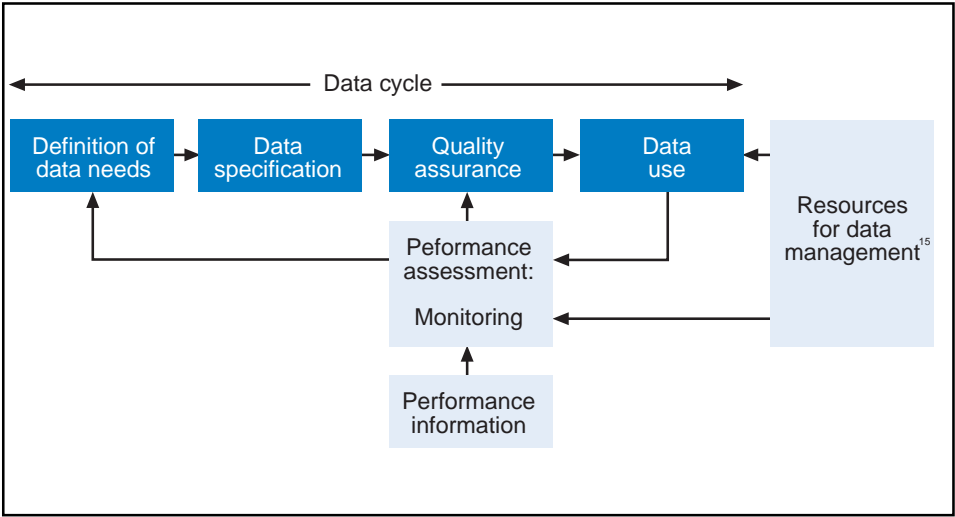
4.6 The areas relating to the data cycle set out in figure 3 are examined below. In Chapter 5 staff skills, the role of consultants and IT systems used in support of the data management cycle are examined.

---

[15]   Discussed in Chapter 5.

4.7   At the beginning of each section below, the ANAO has outlined key points and criteria for each issue and following this is a discussion of the findings.

Data needs

Definition of data needs

4.8   Several factors need to be considered when defining data needs. It is critical that they occur for each individual collection and in light of the overall business needs of the agency. The ABS indicated that:

> *The first and most important step in satisfying any data need is to define the research objectives clearly. From this basis, it is then possible to translate these objectives into specific data requirements.*

> *The objectives should define what the data will be used for and may include such uses as the basis for decision-making, the allocation of funds, analysis of the outcome of policies or programs or determining the direction of future operations.*

4.9   In this section the ANAO sought to assess whether data needs had been defined to enable the data collection to support business needs and user requirements.

ANAO findings

4.10   The ANAO found that data needs had been defined appropriately for all collections examined. In several cases, data needs flowed from legislation, regulations and policy/program requirements. For example, the import data collected by ACS is required by regulations, international trade agreements and the import policies administered under the Trade Facilitation Program in ACS. Much of the import data is also required for the balance of trade statistics compiled by the ABS.

4.11   Where data needs were not outlined in legislation or policy, collection managers used advisory bodies to assist in the identification and definition of data requirements. For example, an interdepartmental advisory group was established to provide advice on the detailed data requirements for the Longitudinal Survey of Immigrants to Australia which is administered by DIMA but produces data for a number of policy areas in DIMA and other Commonwealth departments.

4.12   Figure 4 illustrates the steps taken by the Office of Small Business (OSB) to define the data needs for the Australian Small and Medium Enterprise Database (AUSSMED). It is a good example of several approaches which can be used to identify data needs, particularly for collections which are intended to support a wide range of user requirements. The ANAO considers there would be benefits in data managers using some, all or a mix of the approaches used by OSB to define data needs for new data collections.

<c></>**Figure 4**

Steps in the identification of data needs for AUSSMED

Case study

Identification of the objectives

- The broad objectives of AUSSMED were identified in the 1994 *Working Nation* White Paper which signalled an expansion in official research on small and medium sized enterprises (SMEs).

Obtain appropriate relevant skills and advice

To identify user data needs for the design of the Business Longitudinal Study (BLS), the main survey upon which AUSSMED is based, the OSB:

- engaged staff with expertise in small and medium sized enterprises to assist in the design of the BLS;
- established an advisory committee to provide broad level policy and research advice and to review the development of the database; and
- established a technical committee to provide advice on design and research issues.[16]

Document strategies and seek stakeholders' opinions

- to obtain feedback on the design of AUSSMED, a strategy paper was circulated to government departments, small business, academics and conferences during the development phase of the project.

Documentation and analysis of information needs

- information on data users, research priorities, policy issues and the data required to address these needs was matched in a matrix. This was a useful tool for examining whether AUSSMED was meeting the requirements of users, particularly as the database was intended to support a spectrum of diverse needs extending across several government departments and ranging from export policy through managerial and technical innovation issues to training policy.

Documentation of data applications

- a range of applications for AUSSMED data was identified and detailed in a publication to assist users to recognise how AUSSMED could support their data needs. Applications range from benchmarking and performance information for SMEs, through micro reform issues and the effect of government programs on business, to identifying the characteristics of SMEs which grow or fail.

Maintain the currency of the definition of data needs

Information needs often change over the life of the data collection. To identify emerging information needs and obtain feedback from users on new topics the OSB:

- maintained the technical committee as the main advisory body for the AUSSMED project;
- engaged a consultant with expertise in surveys and SMEs to provide, among other things, advice on new research topics and data outputs;
- presented reports on the results and future surveys of AUSSMED at conferences; and
- visited the States to consult government agencies, business groups and academics.

---

[16] After initial establishment the OSB found that both committees tended to address the same issues. The advisory committee has been discontinued with some of its members coopted to the technical committee.

4.13   In addition to defining needs for new collections, agencies may need to refine existing collections to maintain their currency and to meet changing data needs. This can occur through informal processes where changes are likely to be small and can be based on an understanding of changing needs by managers. If significant changes to existing collections are required, then a formal approach is preferable. Examples of formal approaches identified during the audit include:

- the Bureau of Tourism Research (BTR) redesigned a tourism survey to better meet user needs after a formal evaluation by consultants of its surveys identified gaps in the data collections; and

- reviews of changes to legislation to determine if they change data needs. For example, legislation relevant to the Higher Education Contribution Scheme is reviewed on a regular basis to determine the effect on higher education data collections when changes occur.

Authority and approval

4.14   The conduct of individual data collections should be authorised appropriately. Such authorisation may be derived from legislation, other obligations such as international treaties or given by ministerial or executive level approval. The ABS indicated that the approval process should address, among other things, the availability of other sources of data and the needs of potential users. Without such consideration there is an increased risk that unnecessary duplication of collections will occur with consequent inefficient use of resources.

4.15   The ANAO sought to establish whether appropriate authorisation existed for individual collections which was underpinned by consideration of the factors outlined above.

ANAO findings

4.16   The authority for 13 of the 23 collections came from legislation or international treaty obligations. For example, the legislative basis for DEETYAs' data collections for non-government schools is the *State Grants (Primary and Secondary Education Assistance) Act 1996*. DTRD aviation data are required to meet International Air Transport Association reporting obligations, while some ACS import data are required by international trade agreements.

4.17   For the three BTR collections approval came from a joint Commonwealth-State body of which the federal minister is a member. Four collections were approved by the agency executive, while three short-term collections were approved by program managers.

4.18 The ANAO found that the approval process was adequate in that it addressed unnecessary duplication and the needs of potential users. In two cases, for instance, the approval for new collections included a requirement to reuse data from existing sources. In DIMA, the executive approved the establishment of the Settlement Database (SDB) on the basis that it reused data captured by other DIMA collections. Similarly, approval for AUSSMED included a requirement to link the new Business Longitudinal Survey (BLS) with existing ABS collections.

4.19 In both these cases, the data needs of potential users were a major consideration in the approval process. The rationale for the SDB was the recognised need to develop a specific database to assist settlement service providers in Commonwealth, State and local government and the community sector. Approval for AUSSMED was also based on the recognition of the need for specific data on the performance of small and medium sized enterprises (SMEs) to assist government, business and the research community to understand and develop policies for the SME sector.

Provider load

4.20 Provider load at the agency level was discussed in Chapter 2. Provider load must be considered also in the design of individual data collections because high provider load can:

- impact on data quality;
- reduce survey response rates; and
- increase costs for both the provider and data collection.

4.21 The ANAO assessed whether provider load was considered for individual collections, particularly in relation to the definition of data needs.

ANAO findings

4.22 The ANAO found that for all data collections consideration had been given to provider load when determining data needs. In several cases, the ANAO found that pilot surveys were undertaken during the design phase of collections to, among other things, assess provider load. For example, pilot testing occurred during the development of the BTR National Visitors Survey. BTR also engaged a consultant with expertise in this area to provide an opinion on whether the provider load was acceptable. Similarly, the LSIA involved extensive pilot testing which concluded that a questionnaire taking 90 minutes to complete could be administered appropriately to the survey population.

4.23 The ANAO found that where lengthy questionnaires are used, it is important that the questionnaire design seeks to develop straight forward

questions which avoid confusing the respondent. For instance, while the questionnaire for the AusIndustry Innovation survey is long (62 questions on a 24 page questionnaire), it has been designed so that:

- most questions can be answered by one person without reference to company records; and

- the questions are easy to answer. Multiple choice answers are provided for most questions.

4.24 Another factor which affects provider load is the timing of collection surveys. The timetable for collections needs to take into account the work cycle of the survey population or when data can be most easily and efficiently provided. For example, the OSB changed the timing of the AUSSMED survey in response to advice from data providers that if the survey were conducted at a different time, the data would be more readily available and the workload on providers would be reduced.

4.25 As well as considering provider load, agencies had recognised that assessing provider load is a complex issue involving a number of factors. For example, in relation to the provider load for the LSIA, DIMA indicated that consideration had to be given to the following:

- the total number of people exposed to the burden imposed by any collection;

- the number of collections or administrative processes any individual is involved with; and

- the extent of effort or resources required by any collection.

4.26 DIMA also pointed to the subjective nature of provider load:

where a person is satisfied that a project has relevance and is of benefit to them … then they will be more willing to participate in a lengthy and personal interview, whereas if a collection is seen as irrelevant, unnecessary or of no benefit even a smaller number of questions may be seen as burdensome and intrusive.

4.27 These factors should be considered in any examination of provider load.

Conclusion

4.28 The ANAO found that, for all the collections examined: data needs had been defined adequately; collections were authorised appropriately; the authorisation process involved measures to avoid unnecessary duplication and to address the data needs of potential users; and provider workload had been addressed. The ANAO also identified better practice examples from which agencies could benefit when defining data needs for new collections or redefining existing collections. To enable collections to

support a diverse range of user needs, agencies should examine the mix of approaches adopted by the OSB for the AUSSMED project, listed in figure 4, as these illustrate the different methods by which data needs can be defined at the design stage of a collection and on an ongoing basis.

Data specifications

4.29   Data specifications for individual collections are essential documents that assist managers, collectors and users to understand the purpose of the collection and how to use the data from it. A specification should include information on the following key areas:

- a conceptual statement explaining the purpose of the collection;

- details on the classifications used to define the data;

- whether confidentiality and privacy issues are relevant; and

- the collection methodology, editing and validation procedures.

4.30   The information included in a specification is seen by the ABS as the key to effective data management. This information is referred to as 'metadata' which is described by the ABS as being:

definitions, descriptions of procedures, systems parameters and operational results which characterise and summarise statistical programs. It includes information about the records and fields (variables) that make up the data, together with information about the collection, storage and manipulation of the data.

4.31   Metadata explaining data classifications are particularly important as classifications provide the basis for interpreting and analysing data in a collection. Classifications should be based on established national or international standards so that agency data can be compared and related with other data collections.

4.32   As discussed in Chapter 2, there are statutory, security and commercial requirements in relation to data privacy and confidentiality and these need to be highlighted in any specifications developed to alert collectors and users to the obligations and restrictions attached to the data. As well, an assurance to providers that data will be protected is likely to improve the response rate for collections.

4.33   The ANAO sought to determine if specifications had been developed for each data collection and this is discussed below. We have also discussed, under separate headings, whether specifications addressed conceptual statements, classifications and data privacy and confidentiality.

ANAO findings

4.34   The ANAO found that all of the data collections reviewed had specifications. In general, specifications were incorporated in user guides for collections. The features of an appropriate specification are illustrated in figure 5 of the DEETYA higher education statistical collection. This collection is complex and it is therefore important that the specification not only addresses key issues but is also understood easily by the DEETYA and university staff involved in collecting and processing data for it.

Figure 5
Guide for the DEETYA Higher Education Statistical Collection

Case study

The Higher Education statistical collection provides data on university students, staff and resources. The specification, or Online Manual, is designed to assist university and DEETYA staff involved in the data collection and processing[17]. University staff enter data which, after passing initial edit checks, are transmitted electronically to DEETYA. DEETYA staff undertake further edit checks and, after reconciling differences between their checks and those undertaken by the universities, update the database.

Ease of understanding

• the guide is clearly written and, therefore, easily understood.

Completeness

The guide is complete in addressing important issues, including:

• the purpose of the guide and the reasons for the data collections;

• mandate and type of data sought for the collection;

• changes to data since the last collection. To enable accurate interpretation of the data, it is important that providers be aware of any changes;

• information privacy. Data such as student identification numbers are sensitive and it is important that information providers be explicitly reminded of this;

• data characteristics, such as the classifications;

• technical details relating to file structures;

• data validation rules and checks to assist university staff to process data; and

• edit reports which assist university and DEETYA staff to edit data correctly.

Use of appropriate technology

It is important that the agencies keep the guides up to date and promote their use. Otherwise, inconsistent practices may develop and data quality could be reduced, particularly, as is the case here, if it is a complex data collection with a voluminous guide. Appropriate use of technology allowed DEETYA to:

• make the guide available in an electronic form that can be easily updated; and

• include in the guide a comprehensive index and a computer-based indexing system that allows readers to electronically search for the desired subject and move automatically to the associated point in the guide.

---

[17]   The Online Manual is an element of a software package, DEETYAPAC, that is sent to the universities. It is designed to underpin cost-effective data collection and editing. DEETYAPAC is also referred to in the subsequent section on information technology.

4.35 The adequacy of the specifications in terms of their conceptual statements, data classifications and data privacy and confidentiality is discussed in the sections below.

Conceptual statement

4.36 The ANAO found that all the data collections examined had a conceptual statement. These were often not specifically labelled as a conceptual statement but were included in an introduction, under objectives and so on. It is appropriate to include conceptual statements in an introduction or as part of the collection's objectives as it is under these headings that the purpose of a collection is explained. The DIST Scoreboard Survey provides an example of a sound conceptual statement as it describes clearly the purpose of the collection, the target audience, the data collected and how it can be used:

> *as a reference document for investors, researchers, strategic planners and policy makers … (to) promote international competitiveness, through innovation … to raise awareness of current R&D activity in Australia … Business expenditure on R&D provides a ready indicator of business activity in both incremental and innovative research, and can be used to compare our business R&D performance with that of other countries.*

4.37 However, the ANAO found examples of collections which were poorly defined in that statements were limited to saying 'the purpose of the collection is to provide data'. This is of limited value as it does not provide the agency itself or stakeholders with an explanation of the purpose of a collection, its audience and so on. Where such statements exist agencies need to consider the objectives of the collection and describe these in a conceptual statement to ensure that collection staff and users can understand easily the purpose of a collection, the data contained in it and its potential applications. Providing a statement on the purpose of a collection is particularly important as it serves to focus collection staff on the outcomes of the collection and to inform potential users of the collection's relevance to their needs.

Classifications

4.38 The ANAO found that adequate explanations of data classifications were included in the specifications for all collections reviewed. In terms of the classifications used by individual collections, 13 of 23 collections were based on established international classifications and standards. DTRD aviation statistical collections, for example, use the International Air Transport Association classifications which enables comparisons between Australian data and that from other countries. Four other data collections were based on recognised national standards such as the ABS Australian and New Zealand Industrial Classification.

4.39   The ANAO noted, however, that there were cases where existing classifications do not meet the requirements of a collection and collection designers must develop new classifications or refine existing ones. In such cases, the ABS has indicated that while departing from recognised classifications is appropriate, it is important that existing classifications are examined and every effort made to use them to assist with data relatability.

4.40   The ANAO found one example where existing classifications were revised. In DIMA, the Longitudinal survey of immigrants to Australia (LSIA) collection uses modified ethnicity and language classifications. While existing ABS classifications were used initially, these were revised to better meet user needs. The ANAO considers this to be a case where the collection designers refined existing classifications appropriately, having first tested whether existing classifications met the requirements of the collection.

Data privacy and confidentiality

4.41   The ANAO found that the specifications for separate data collections included rules and guidelines in relation to data privacy and confidentiality. Restrictions relating to staff access and data dissemination were detailed in the specifications. In three cases, training programs for collection staff included modules on the privacy and confidentiality provisions and obligations attached to the data.

4.42   The ANAO noted that in DIST, the tourism surveys do not involve data which is sensitive or requires confidentiality measures, while the Scoreboard survey collects data on the basis that it will be released.

4.43   For all collections involving data which could be used to identify an individual or organisation, there were guidelines referring to the measures in place for protecting the confidentiality of such information. These measures included restricting access to identifiable unit record data (URD) within agencies. In the four cases where URD was disseminated externally, only unidentifiable data was disseminated with sensitive data deleted or aggregated to a point where it is not possible to identify individuals or organisations. The dissemination of URD for the LSIA collection also required the third party receiving the data to sign a contract obliging them to observe several restrictions and which provided for legal penalties if those restrictions were breached.

4.44   For the Settlement Database, the computer program for printing data reports which are to be disseminated externally includes automatic checks to detect and delete small cell sizes (small cell sizes refer to data based on a small number of respondents, for instance, less than four, such that there is a risk that individuals or organisations could be identified).

4.45 Where consultants were used for data collections, the ANAO found that there were contractual obligations relating to maintaining data privacy and confidentiality. For example, in ACS the contract for Client Satisfaction Surveys, which involves a consultant collecting and analysing data, stipulates that the data remains the property of the department and that URD are not to be released to a third party.

Conclusion

4.46 The ANAO found that for all the collections examined specifications and/or guidelines existed which included conceptual statements, classifications and the rules for maintaining data confidentiality and privacy. However, the ANAO considers that in some cases conceptual statements should be revised to explain more clearly to collection staff and data users the purpose of the collection, the data collected and its potential applications.

Quality assurance

4.47 Quality assurance (QA) measures are an important means for ensuring data accuracy and reliability. QA measures find data errors and enable their correction. There are many reasons why errors can occur in data, the most common being human error in entering or recording data (for example, transposition errors such as coding 19 instead of 91). QA measures can involve automatic checks as data is entered into databases, cross-checking data with other sources of related data to identify if large discrepancies occur to follow-up reviews or audits to validate whether the data provided is correct.

4.48 The ANAO sought to determine if the data collections examined had QA measures to identify errors and promote data accuracy in place.

ANAO findings

4.49 The ANAO found that all data collections reviewed had QA measures. The following examples of better practice to improve the quality an reliability of data were noted by the ANAO:

- to assist university staff supplying data for the Higher Education statistical collection, DEETYA provides universities with software which:
    - includes checks that detect input errors automatically and enable corrections to be keyed in immediately. This is an important measure as it is easier and more efficient to identify and correct errors at the source than at a later stage when incorrect data is merged with correct data;
    - has edit reports that aid error detection and correction. For example, there is an edit check which compares DEETYAs' calculation of

student HECS liabilities with university estimates. Significant differences between the two estimates are identified in an exception report and the universities are required to reconcile differences; and

- allows the universities to transmit the data electronically to DEETYA. By reducing the need for DEETYA staff to re-enter data manually, electronic transmission helps reduce the risk of entry errors occurring.

• DTRD prepared a comprehensive manual to assist people coding data to interpret coroners' reports on fatal road crashes. The manual was designed to address problems stemming from inconsistent interpretations of coroners' reports. A supplementary manual provides further instructions on identified problem areas; and

• quality assurance checks were undertaken on aggregated data (as well as on single records) as this is the only way to detect some incorrect records.

4.50 Figure 6 provides examples of the type of quality assurance mechanisms used for the collections examined by the ANAO.

4.51 In ACS, the ANAO found that while ACS's import data (98 per cent of which is sent electronically) is subject to adequate edit checks for separate data entries, ACS is concerned about the internal consistency of some of the data provided. To address this issue ACS is piloting a quality control system which involves cross-checking random samples of import data, undertaking a physical examination of the related consignments and an audit of related commercial documentation. Discrepancies between actual versus reported details are to be recorded and classified by type of error for corrective and follow-up action. The ANAO considers that this quality control system is an appropriate measure to assist ACS assess the reliability of import data, identify causes of error and develop remedies.

4.52 In DEETYA, the ANAO noted that consultants undertaking follow-up checks of schools data had found that most of the responses for the non-government school census and financial collections had data errors. Since the data provided by schools is used for determining funding, the ANAO is concerned that these data errors could result in incorrect payments to schools.

Figure 6
Stages of data quality assurance

Data management stage	Quality assurance processes	Examples and comments
Pre data collection	sample design	The BTR survey of international visitors relies on rigorous sampling procedures to locate interviewers with appropriate language skills.
	design of collection methodology	As appropriate, agencies developed survey methodologies to minimise any likely survey bias that may occur, for example, because an insufficient proportion of the survey population responds. Examples are: DEETYA VET, DIST tourism and the DIMA Longitudinal Survey of Immigrants data collections.
	design of survey instruments	Inappropriate questionnaires place unacceptable demands on information providers and lower the quality of the information. For this reason, for example, BTR sought expert advice on the design of the questionnaire for the National Visitors Survey and pilot tested the questionnaires.
	pilot testing of questionnaires	To identify and correct data collection problems and refine sample size.
Data collection	on-site checks	DEETYA staff involved in the review of the demonstration projects for the Contracted Entry Level Training Agencies (CELTAs) and DIST staff involved in tourism surveys undertook on-site reviews of data collection by consultants to assess procedures and data quality.
Post data collection	single record edits: • range checks • logic checks • internal data consistency	As appropriate, all agencies had range, logic and internal consistency checks. Examples of these are: • a date in April must have a value that lies between 1 and 30; • an accident coded as 'fatal' must have a fatality (DTRD road crash data collection); and • the entered flight distance must be consistent with an estimate based on the flight time and speed (DTRD aviation data collections - check that these fields are consistent).
	aggregate data edits • consistency with previous data • distribution of data collected	DEETYA compares census information supplied by (non-government) schools with the equivalent information from the previous period. Significant differences need to be explained. DEETYA examines information supplied by universities on, for example, the gender ratios for students undertaking university courses. Unusual distributions (for example, all students coded 'male' or 'female') require explanations.
	audit of data collected	Each year DEETYA engages consultants to audit census and financial information supplied by schools. DTRD cross checks national road fatality statistics with State/Territory figures.

4.53   DEETYA advised that it is addressing this issue in two ways. An interactive internet system is being introduced which will enable schools to transmit data directly to DEETYA for the next census. This should improve data accuracy as the data will be subject to checks and edits at the point of data entry. However, the ANAO notes that because this system is to be introduced gradually over several years on a voluntary basis, not all schools will be using it. The ANAO considers that to assess the relative strengths of the new and current system of reporting, DEETYA should monitor and compare data accuracy results between these systems. This would enable DEETYA to assess whether the new system has improved data accuracy and whether there is a need for additional measures to improve data accuracy for schools using the current system.

4.54   The second measure involves a review of the follow-up checks performed by consultants which is to be conducted over the next six months. Issues under review include sample size and targeting. The ANAO considers that tighter targeting of the schools subject to checks and a larger sample would strengthen this QA measure. A stronger system should identify the schools most likely to enter incorrect data and enable corrective action to be taken. Stronger checks will also encourage schools to take more care in providing accurate data. If the schools believe that errors are more likely to be detected, it is expected that they will take steps to improve data accuracy.

Conclusion

4.55   The ANAO found that a range of QA measures was in place for all the collections examined. These measures were generally adequate, although in ACS and DEETYA additional measures were being introduced to address concerns about data accuracy. The ANAO considers that the new QA measures being introduced in ACS are appropriate. In DEETYA, the ANAO considers that the non-government schools census and financial collections could be improved if DEETYA monitored the results of the new internet reporting system to determine whether it has improved data accuracy; and reviewed follow-up checks for the census collection by assessing the cost/benefits of a larger, more tightly targeted sample.

Data use

4.56   Data are collected primarily to be used in decision making, policy development, the determination of funding and so on. There is a significant amount of data collected by government agencies which is also disseminated as a public good to business, academic researchers and the public generally. However, efficient and effective data use depends on the

measures taken by an agency to make the data visible and accessible. These issues were discussed at the agency level in Chapter 3.

4.57 This section focuses on mechanisms in place to promote data visibility and accessibility at the local level, thereby encouraging data use. It also looks at whether the data were used for their intended purposes.

ANAO findings

Data visibility

4.58 The ANAO found that a range of measures was used to promote transparency so that data collections were visible both within agencies and externally. For internal users, the most common mechanism was to report recent collections or reports in staff bulletins. As discussed in Chapter 3, in ACS and DIMA directories of data collections are used to promote data visibility. As well, the ANAO noted that executive summaries of recent reports based on LSIA data are sent electronically to the executive and program managers in DIMA and to the minister's office. This is an effective way of raising the profile of a data collection as it targets key decision makers who are in a position to promote data visibility.

4.59 The main measures to make data visible externally were through publications of data and making data available on the internet. In seven cases, mailing lists were maintained so that data were made visible to target audiences. In ACS, the results from the National Survey Program (NSP) are reported to the Customs National Consultative Committee (CNCC). As the CNCC comprises the ACS executive and peak industry representatives, this forum allows ACS to promote the visibility of the NSP at a high level to key client groups.

4.60 An example of good practice at the local level is illustrated in figure 7 of the AUSSMED collection. This provides an example of measures to raise the visibility of the collection nationally and with an identified target audience.

Figure 7
Measures to promote the visibility, access and dissemination of the AUSSMED data collection

> *Case study*
> - AUSSMED data is first released by the ABS which helps to raise the visibility of the data nationally;
> - reports on AUSSMED data are presented to conferences attended by interested parties, including small to medium industry organisations, academic researchers and policy makers;
> - AUSSMED data products are disseminated free of charge to the target audience in line with the objective to promote the maximum use of AUSSMED data for business and policy-related research;
> - major data products are sent to key media outlets and advertised in targeted publications in order to publicise the AUSSMED collection, particularly to the business and research community;
> - the OSB has arranged for Ausinfo to sell publications through its outlets; and
> - consideration is being given to using the internet to allow direct ordering of AUSSMED data products.

4.61   Discussions with VET program area staff in DEETYA indicated that collections being undertaken within the area were not highly visible. It is likely that data collections within a specific program area will be important and therefore at least need to be visible at that level. The ANAO has recommended in Chapter 3 (paragraph 3.25) that, where agency-wide directories of data do not exist, agencies should consider establishing directories to promote data visibility both within agencies and externally. The ANAO also notes that, as the ABS has stated that it may be appropriate to establish directories of data holdings at the program level, that the VET should examine this option. Developing a directory of program-level collections would increase the visibility of the VET program area's data and could be a building block in establishing a directory of collections for DEETYA as a whole.

Data accessibility

4.62   The ANAO found that local users were able to access data to meet their requirements in most cases. As well, where data were made available externally, they were provided on computer disk or in a machine readable form to enable the user to manipulate the data for their own purposes, or agency staff performed data manipulations in response to requests. For the AUSSMED collection, the OSB and ABS were examining methods which would enable external researchers to access and manipulate unidentifiable unit record data under strictly controlled conditions.

4.63   However, the ANAO found two cases where technical difficulties limited the extent to which data could be accessed and manipulated to meet user requirements. In ACS, data users with a requirement to access client data from several databases face constraints because the databases were not designed to share data or to perform complex statistical analysis. As discussed in Chapter 3, ACS is addressing this issue by developing data warehouse technology. A data warehouse integrates data from separate databases and provides advanced research and reporting tools to enable users to perform complex data manipulation. The introduction of data warehouses and data marts (which are smaller warehouses) are expected to overcome the accessibility problems currently facing users in ACS.

4.64   In DIMA, the ANAO found that some of the design features of the Settlement Database restricted the use of the data for complex statistical tabulations. As well, users indicated that for even relatively straight forward reports the system was not user friendly. To overcome these access difficulties additional SAS[18] tables have been added to the Settlement Database and advanced end-user inquiry tools are also to be provided to enable users to extract and manipulate data more easily. Moreover, the ANAO noted that a Help Desk with an experienced operator was established to assist users with the production of statistical tables and resolving data problems. The Help Desk function is an important key for users seeking to access the data. The ANAO also found that the Help Desk is a channel of communication which enables users to relay access problems and suggested improvements to the administrators of the database.

Data use

4.65   The ANAO found that all of the individual collections examined were used for their intended purpose, indicating that the dissemination and accessibility measures in place were adequate. In the two cases discussed above where problems with data accessibility existed, the ANAO noted that these problems arose when attempts were made to use the data for purposes other than those for which the data was intended, such as using transactional data for statistical purposes rather than processing purposes.

4.66   Eighteen of the 23 data collections examined provided data to external groups; agencies also undertook specific analyses for external users for 12 of these collections. The other five collections examined were designed and used solely to meet internal data needs.

---

[18]   SAS refers to Statistical Analysis System, a software package which is used widely for tabulations and statistical analyses.

4.67   To assist in meeting user requirements, mechanisms had been established to monitor user satisfaction with data outputs in eight of the collections reviewed. As with the definition of data needs discussed above, advisory bodies comprised of data users and technical experts, client surveys and regular liaison with users were the means by which collection designers kept track of user satisfaction and changing needs. For the other 15 collections, collection designers and managers liaised on a regular basis with users to obtain feedback on whether the collection met user requirements.

4.68   The ANAO found that monitoring user needs on an ongoing basis was important not only for identifying new data requirements but also because data needs can change over time. For example, the management of LSIA found, when the actual implementation of the survey was examined, that despite extensive trialing and consultations with users, client needs continued to emerge after the questionnaire was in the field. This necessitated some minor design changes to take account of those needs.

4.69   The LSIA example indicates the importance of ongoing monitoring of user needs for enabling collections to fulfil their purpose and support user requirements. Advisory bodies or regular liaison are important channels of communication between data users and collection designers which allow both parties to understand the requirements and constraints facing each other. These mechanisms have been established and used by LSIA to ensure this occurs.

Conclusion

4.70   The ANAO found that most collections had adequate measures to promote data transparency, although there were indications that some collections in the VET program area in DEETYA were not highly visible to the area's staff. The ANAO considers that establishing a directory of the collections and data used at program area-level in VET would improve data visibility and serve as a key element in a whole-of-agency data directory in DEETYA.

4.71   The ANAO found that all of the collections examined were used for their intended purposes and that mechanisms existed for monitoring user requirements. The ANAO also found that in some cases in ACS and DIMA it could be difficult to undertake complex data analyses because of design and functional constraints with the systems on which data are maintained. However, appropriate measures had been or were being taken to resolve these problems in both agencies.

Data collection performance assessment

4.72   The reasons for establishing a performance assessment framework at the agency level were outlined in Chapter 2.

4.73   Performance assessment for individual data collections is important because in most agencies local level collections generate data which is often critical to agency operations. The information flowing from performance assessment is therefore a key means by which the executive can determine whether local collections are supporting agency objectives in an efficient and effective manner. The devolved nature of data management in the agencies reviewed also means it is particularly important that performance information is available to promote accountability and transparency for individual data collections and contributes to agency level data management.

4.74   Performance information on key data indicators assists locallevel managers of data collections. Particularly, it provides the basis for monitoring key areas such as data quality, helps to identify problems and enables managers to take corrective action.

4.75   As discussed in Chapter 2, the ANAO recognises that performance information in relation to data management is at an early stage in its development. The Audit Reference Group indicated that the development of a full range of performance measures will be complex and its implementation will need to be gradual.

4.76   The ANAO sought to determine for each data collection whether there was:

- a balanced set of performance indicators that addressed such issues as data accuracy, client satisfaction, data timeliness, cost-effectiveness and efficiency; and

- appropriate arrangements for monitoring performance.

ANAO findings

4.77   The ANAO found that all collections had established plans which included milestones and targets for the various stages of the data cycle. Most collections also measured outputs insofar as user logs were maintained to assess the demand for data and identify principal client groups.

4.78   As discussed in Chapter 2, a range of performance indicators for data quality was developed in DIMA but only one of the indicators (data accuracy) was measured. The ANAO considers that the indicators developed by DIMA provide a basis upon which other agencies can develop indicators to meet their own requirements.

4.79 The ANAO also found that several collections used a range of performance measures or indicators which could be combined to provide a basis for developing a balanced set of indicators. DIMA and DTRD have developed the performance measures identified in figure 8.

Figure 8
DIMA and DTRD performance measures

	DIMA	DTRD
Input		
Process	Data accuracy (also an output measure)	Data accuracy (also an output measure)
	Data timeliness	Data timeliness
Output	Number of analytical publications and reports	
Outcome	Requests for data	Citations and use of data
		Data relevance
		Results of associated policies and relevant education campaigns
Quality of service	Satisfaction of branch head	Client satisfaction
	Client satisfaction	Comments by committee members

4.80 While the ANAO acknowledges that there needs to be a focus on outcomes, in a balanced set of measures, inputs should not be ignored. Input measures include the level and cost of resources, such as staff, IT expenditure and overheads. Without knowing the level of inputs and outputs, managers will not be able to determine the cost effectiveness of collections.

4.81 In addition to milestones and targets, the ANAO found that there are other sources of information which, while not currently used for performance purposes, could be used as part of a formal performance assessment framework. In DEETYA, for example, the data collected from non-government schools are checked against school records to determine the accuracy of the data provided. This information could be used to measure the performance of the collections in terms of process and output, to identify problems and to improve performance.

4.82 In ACS, the pilot quality control system for import data (discussed above), which checks the validity of the data provided by import brokers, could also be used as an output measure of data quality. As well, ACS has commenced a pilot test of data quality in the Border Operations program

area and intends to undertake detailed reviews of other major collections. The results of these reviews should provide valuable performance information, while the methodology on which they are based should also identify potential performance indicators and a method by which performance can be assessed on a regular basis.

4.83 The ANAO considers that the information discussed above provides the basis for establishing a sound and balanced performance assessment framework.

Monitoring

4.84 The ANAO found that, where milestones existed in plans, progress was monitored formally for statistical collections. Milestones are not applicable in the case of transactional databases as these systems capture data on an ongoing basis.

4.85 As discussed above, the extent to which collections meet the data needs of users is also monitored and the information generated is used to refine collections. In DEETYA, DTRD and BTR, collections were reviewed formally on a periodic basis, while ACS and DIMA are intending to also undertake detailed reviews of major collections. Periodic reviews of collections are appropriate because determining whether collections have met their objectives and are still relevant to current data needs requires assessment of outcomes over time. As well, systemic problems are likely to emerge over the longer-term.

4.86 However, because the collections examined do not yet have a full balanced set of performance indicators, it is not currently possible to measure all aspects of their performance. In order to monitor performance effectively, it is necessary to develop a full set of performance indicators as part of a formal performance assessment framework.

4.87 The ANAO has made recommendations at paragraphs 2.99 and 2.104 that agencies should establish formal performance assessment frameworks, including developing a balanced set of performance indicators and establishing monitoring mechanisms. This is aimed at establishing a performance assessment framework the agency level but the recommendation is also applicable at the local level of individual collections.

Conclusion

4.88 The ANAO found that the performance assessment framework was not sufficiently developed to allow adequate monitoring. A balanced set of performance indicators needs to be developed and performance against these indicators monitored and used to identify deficiencies and improve

performance. In order to develop a balanced set of indicators, the ANAO has identified above a number of good practices which should provide a useful basis for agencies to determine performance indicators and appropriate sources of performance information. As well, the development of performance assessment for individual collections should be an integral part of an overall agency framework for assessing performance in relation to data, as recommended at paragraphs 2.99 and 2.104.

Better practice principles

From the work undertaken in this audit, the ANAO has identified the following practices which would enhance data management:

- Provider load

  Provider load should be considered when determining information needs. Three ways of doing this that were noted are:
  - seeking information on survey forms on how long it took to complete the survey;
  - monitoring refusal rates for surveys; a high refusal rate can be indicative of a high provider load; and
  - obtaining an opinion from consultants.

- Identification and monitoring of information needs

  The process followed to develop and monitor information needs should have all the necessary steps and include arrangements for the feedback the results of the monitoring process into information needs, including:
  - identifying objectives;
  - obtaining appropriate skills;
  - analysing needs, documenting of strategies (to meet needs) and obtaining stakeholder opinions;
  - documenting data applications;
  - monitoring and maintaining currency of data needs.

- Specifications and guides

  A comprehensive manual should exist for a collection that details data management procedures including, for example, editing rules. Also, particularly for large and complex collections, it should be available in an electronic form and have indexing/search facilities to allow users to move quickly and easily through a considerable volume of material.

- Quality assurance procedures

  A number of good practices relevant to quality assurance were identified, including:
  - software was to assist staff responsible for the supply of data. If required, it should allow staff to enter, edit and transmit the data electronically to a central site;
  - as required, coding manuals should be developed and maintained to help consistent interpretation of complex data; and
  - data accuracy should be assessed and monitored.

- Performance assessment

  A performance assessment framework should be developed for data collections. There should be a range of performance indicators, benchmarks and targets.

5. Resources for Data Collection

This chapter discusses the resources needed to support data management, namely staff skills, the use of consultants and information technology. The ANAO found that data management skills and the use of consultants and information technology was generally appropriate, although it has made a recommendation in relation to the use of consultants by one agency.

Introduction

5.1 The previous chapter discussed key steps in a data management cycle which need to be considered at the local level. This chapter discusses the resources needed to support data management, namely staff skills, the use of consultants and information technology. Data management requires skilled staff because of the specialised nature of most data collections. In general, staff need technical skills related to statistics and information technology. Where technical skills are not available or adequate for the particular requirements of a collection, it may be appropriate to engage consultants to undertake some or all aspects of data management. As well, consultants/contractors might be used to deal with peak workloads so that client service can be maintained in an efficient and effective manner.

5.2 Most data collections are stored using information technology systems. For large complex data collections in particular, many of the activities related to data collection or capture, storage and analysis would not be possible or would take an excessive amount of time without the use of appropriate information technology. However, it is a widely accepted principle of effective data management that information technology should support, rather than drive, the development and management of data collections. That is, data needs should drive data collection, not the availability of technology.

5.3 The criteria against which these issues were examined are set out at the beginning of each section and followed by a discussion of ANAO's findings.

Data management skills and resources

5.4 Given that all elements of data management require specific skills, it is important that agencies develop and implement approaches to ensure that the appropriate range of skills is available. These approaches can include recruiting staff with required skills, training staff, engaging

consultants and establishing advisory bodies to provide technical and other assistance.

5.5 The ANAO sought to determine whether, for each data collection, agencies had taken the appropriate steps to ensure that the required skills were available for effective data management.

ANAO findings

5.6 The ANAO found that agencies had used the range of approaches discussed above (paragraph 5.4) to ensure that appropriate skills were available to meet data management requirements. These approaches are discussed below.

5.7 During the course of the field work the ANAO interviewed the officers responsible for the data management of all the collections examined. In 20 of the 23 collections, staff advised the ANAO that they sought specific technical skills and experience in data management when recruiting staff. Frequently, this was done by recruiting staff with ABS experience. For all of the collections, the ANAO was advised that where necessary, training was provided to equip staff with the required skills.

5.8 The ANAO found that training was provided in a number of ways. The area responsible for Higher Education data collections in DEETYA used a consultant to design and run a SAS training course that met its particular needs. Tailored training courses such as these are more likely to be effective than generic courses because, as the participants are familiar with the case studies, the course material is more easily learnt. As well, training time is more likely to be devoted to the SAS modules which are most relevant to the staff responsible for a particular collection.

5.9 The ANAO found that in two agencies consultants were engaged to provide specialised training because in-house expertise was not available. In DIMA, the Settlement Database employed a consultant to provide on-the-job assistance in the use of a statistical computer language. In Office of Small Business (OSB), a technical consultant has been engaged on an ongoing basis to provide, among other things, training to the team managing the AUSSMED project. This approach enabled staff to receive ongoing support to reinforce initial training and skills transfer.

5.10 Training should not be limited to technical skills training. As discussed previously (paragraph 4.41), the ANAO found that at least three areas provided training relevant to data privacy and confidentiality.

5.11 Given the range of skills needed for data management, it is important that individuals' training needs are identified and arrangements put in place to meet these needs. For example, staff responsible for the

LSIA in DIMA are provided training in accordance with individual learning agreements which are reviewed and updated on a six-monthly basis. In ACS, an ABS training course in statistical skills was provided to a member of staff responsible for the National Survey Program, who had no statistical experience.

5.12 As well as providing training, the ANAO found that in the DEETYA Higher Education and DTRD Aviation data collections there was a practice of rotating staff through all data management functions such as data input, analysis and dissemination. The objective of this is to broaden staff skills and improve their overall knowledge of all elements of the area's operations. By providing staff with experience and skills in all functions, these units are able to deploy staff to areas of highest priority. This also ensures that these areas have back-up staff to fill critical positions in case a member of the area is absent or leaves the section.

5.13 The ANAO found in the area responsible for the DEETYA Schools data collection that only one member of staff had the skills necessary to extract the data. This could result in delays in providing data to clients if this person were absent. The ANAO considers a rotation policy to be a sound means to ensure that data management areas have staff with a full range of required skills and this reduces the risk of a section being dependent on a particular staff member. The practice of rotating staff may overcome this problem and should be considered by the Schools program area.

5.14 Figure 9 demonstrates the type of assistance and/or training provided by consultants within the agencies examined. The figure shows that consultants were not only used for their specific skills but also to assist in handling peak work loads.

5.15 As noted in the above table, DTRD used a consultant to respond to requests for road crash data analyses. This option was taken if departmental personnel were occupied on other matters with a higher priority or the skills were unavailable. The benefits of this approach are that:

- it allows departmental staff to concentrate on core business while maintaining client service levels without the costs associated with additional permanent staff; and

- people with the skill to undertake what may be quite complex statistical analyses are difficult to recruit and, even when they are available, it may be difficult to provide them with the quantity and type of work necessary to maintain their skills.

Figure 9
Agency use of consultancies in the data management cycle

Data management strategy	Agency/program	Consultancy services provided and comments
in-house resources complemented by the use of consultants	DEETYA (Schools)	Data collection, data entry and editing is undertaken annually. Therefore, it is appropriate to use consultants to handle peak work loads. There is a steady, ongoing demand for data and the department undertakes this function.
	DTRD (road crash statistics)	Data are collected from every state/territory for each of the three data collections. One data collection requires the interpretation of coroners' reports and manual data entry. Other collections require the integration of computer files which are in different formats. The analysis of data may require a knowledge of complex statistical techniques. DTRD has found that the diverse range of skills needed can be best provided by consultants. DTRD concentrates on core activities such as data dissemination and data management.
	DIST (BTR tourism statistics)	Data collection for the two surveys requires: the use of computer aided telephone survey techniques; and the location of interviewers with specific multilingual skills at certain airports. BTR could not efficiently and effectively provide these services. BTR concentrates on core activities such as data dissemination and analysis.
consultants (primarily)	DEETYA (VET)	The three data collections reviewed are 'one-off' studies and therefore it would not be efficient for VET to undertake these in-house.
	DIST (AusIndustry) Innovation and Scoreboard data collections	The Innovation survey was an ABS initiative and, given ABS' experience and expertise, the most efficient and effective means for AusIndustry to get information was for the agency to (partially) fund the collection. The Scoreboard collection is undertaken once each year and, therefore, it is appropriate to use consultants to handle peak work loads. AusIndustry concentrates on core activities such as policy development and advice.
	OSB longitudinal study of small business	Longitudinal studies such as this one present management with administrative and staffing problems. They are: complexity; work loads are not spread evenly throughout the year; certain skills are required to develop the survey methodology and other skills are needed to manage it; and data are collected for a specific period and then the resources used are no longer required. OSB has found that the most efficient and effective way to address these problems is to use consultants.

5.16 Another approach to ensure relevant skills are available is to establish specialist units which can provide 'consulting' services to data management areas within the agency. In ACS, the Corporate Data and Statistical Services Unit (CDSSU) was established to, among other things, provide specialised statistical skills. Rather than recruiting staff for separate data management areas, the CDSSU provides methodological and analytical assistance on an as-required basis. A similar approach was evident in DEETYA where the Analysis and Evaluation Division (AED) is available to provide technical advice on data management issues to areas within the department. AED assisted the VET program area in the design and management of three research data collections, providing advice on issues such as research methodology and sampling. The ANAO considers that these examples demonstrate how agencies can benefit through the establishment of an area which provides in-house technical assistance on data management issues.

5.17 As well as the above mechanisms, technical advisory bodies were used to provide expert assistance which otherwise would not be available within an agency. For example:

- BTR received advice on tourism data from technical committees that included people with technical and industry experience. For example, an advisory committee provides advice on matters related to tourism forecasts;

- DEETYA Higher Education data collections are reviewed by a working group that has members representing the Australian Vice-Chancellors Committee (AVCC) DEETYA and ABS. This provides the department with an opportunity to receive advice from those providing data (AVCC) and those using the data (ABS); and

- OSB uses a technical committee to provide high level policy, survey design and research advice and to review the ongoing development of the AUSSMED project. This provides OSB with, among other things, an means of identifying current small business issues.

5.18 Consultants need not be restricted to only undertaking data management functions and skills training. In two instances, a consultant was engaged to, among other things, monitor the performance of a consultant undertaking data management work for the agency. As illustrated in figure 10, BTR engaged a consultant to evaluate, and provide advice on, technical proposals of the main consultant. Similarly, the OSB used a consultant as a technical manager to assist and assess the work of the main consultant for the AUSSMED project.

5.19   The ANAO considers that this approach was appropriate in both cases as the additional consultant provided skills which were not available within the agencies. As both collections were complex, expensive and long-term projects, the use of a second consultant was a sound way of ensuring that the work of the primary consultant met each agency's requirements and that problems were identified and remedied at an early stage.

Figure 10
Development of the National Visitor Survey

> *Case study*
>
> 1. The primary consultant provided:
>   - specialist skills that BTR does not have, for example, in the development and management of large surveys;
>   - computer aided telephone interview facilities for the conduct of tourism surveys; and
>   - staff needed during the initial stages of the consultancy who could develop and field test the data collection techniques.
> 2. Two complementary consultancies provided a better range of skills than could be obtained had all the work been given to one consultant. In particular, BTR were able to obtain independent advice on technical solutions proposed by another consultant, for example, on maximum provider load.
> 3. The project was split into stages (survey development; initial and full pilot testing; and survey conduct) and progress was monitored at the end of each stage. If a decision was made that the project should not proceed to the next stage, the survey could be halted without further cost to the agency.

Conclusion

5.20   The ANAO found that the collections examined used a variety of methods to ensure that the necessary skills were available to address data management requirements. Staff with appropriate technical skills were recruited and the development of specific skills and multi-skilling was promoted by training and, in some cases, staff rotation. The ANAO considers that there could be benefit for other agencies with significant data collections in examining these different approaches to address their own staffing requirements.

Use of consultants

5.21   As discussed previously in this chapter, consultants are one means by which technical expertise and assistance can be obtained to meet data management requirements and priorities. It is essential that consultants

are engaged on a value-for-money basis[19] and are managed appropriately to ensure the agency's data needs are met.

5.22   The use of consultants should support effective data management. In this section the ANAO assessed whether contracting procedures used to obtain consultants were transparent, assisted areas to obtain value-for-money and provided a sound basis for contract management. As well the ANAO assessed whether documentation existed that adequately described the agency's requirements and the criteria used to select the consultants.

ANAO findings

5.23   Consultants were used extensively to undertake all or part of the data management process for 14 of the 23 collections examined. In all 14 cases, the contracting procedures used to engage consultants were appropriate in that:

- the consultant selection process was transparent in terms of the tendering process with all applicants being made aware of departmental needs, timetable and selection criteria; and

- the selection procedure was designed to obtain the best value-for-money in that: appropriate selection criteria were developed; careful consideration was given to how prospective consultants should be identified; and competitive tenders were sought. For example, BTR used a consultant with expertise in purchasing to develop and guide tendering for the National Visitors Survey; and

- the contracts used were in the approved agency form and included time lines which could be used to monitor the consultants' performance (this issue was discussed in paragraph 4.77).

5.24   The documentation associated with the engagement of consultants was satisfactory for 12 of the 14 relevant data collections, specifying agency needs, editing requirements and the outputs required by the agency. However, two data collections maintained by BTR did not fully specify the outputs that the consultant should supply the agency. As a consequence, the information supplied by the consultant was in a format which had to be reprocessed by BTR. This inefficiency could have been avoided by BTR stating clearly in the contract with the consultant that the data should be provided in a format compatible with BTR's specifications. Recommendation 12 at the end of this section addresses the ANAO's concerns in relation the data supplied by consultants involved in the

---

[19]   An assessment as to whether agencies actually achieved the best value-for-money from the engagement of the preferred consultant was outside of the scope of this audit.

International Visitors Survey (ISV) and the need for BTR to be able to use the data without having to reprocess it. The recommendation relates only to the IVS as the other collection which used the consultant for has been superseded by the National Visitors Survey.

5.25   The ANAO also noted that the OSB had adopted a performance basis to manage its contract with the ABS, the primary consultant involved in the AUSSMED project. Originally the consultancy was on a cost basis with the ABS charging OSB when work was completed. A performance basis for managing the consultancy, which the ANAO considers to be more appropriate, involves payment according to an agreed work schedule. It was adopted to assist in keeping the project on time and ensure that ABS delivered the outcomes required by the OSB.

5.26   The ANAO considers the performance-based framework used by OSB to be a sound measure as it clearly defined performance requirements (time frames, milestones, outcomes and the payment schedule) and therefore reduces the scope for misunderstanding between the agency and consultant. A performance approach enables the OSB to monitor and review the progress and quality of the consultant's work, and helps to identify any difficulties or slippages in the delivery of services by the consultant.

Conclusion

5.27   The ANAO found that in 12 of the 14 cases examined the processes used to engage consultants were appropriate. Contracting procedures were transparent and consistent with accepted Commonwealth practices and consultants were managed appropriately. Competitive tendering and associated procedures would assist the agencies in obtaining value for money. However, for two BTR collections the contract did not adequately specify the outputs to be delivered by the consultants which resulted in the BTR having to reprocess the consultants' work. To address this shortcoming BTR needs to re-examine its output requirements and revise the contract with the consultants accordingly.

5.28   In two cases, the ANAO identified better practices where agencies had used additional consultants to assist in reviewing and managing the work of the main consultant. There are benefits in this approach for other agencies where, given to the technical complexity of the collection and/or the lack of in-house expertise, it is necessary to obtain a second opinion on whether a consultant's approach and services meet agency requirements. The ANAO also considers there are advantages in adopting a performance-based, as opposed to a cost-based, approach to managing consultants. Managing consultants on the basis of performance against specified targets and milestones will assist agencies to keep projects on time, to focus

consultants on agency requirements and to identify and address problems at an early stage.

Recommendation No.10

5.29 The ANAO recommends that BTR reassesses its output requirements for the International Visitors Survey and incorporates the revised requirements in the contract conditions at the earliest possible time to ensure that BTR obtains data in a suitable form and of a satisfactory quality.

DIST's response

5.30 Agreed. The International Visitors Survey (IVS) contract for 1998, which is currently being finalised, includes specifications for tables to be received in electronic format as well as in hardcopy format. Previously, these tables were received only as hardcopy tables, leading to additional processing by the BTR staff to prepare the data for publication. The National Visitors Survey (NVS) contract similarly incorporates specifications for electronic tables. This requirement will be a feature of future IVS and NVS contracts. In addition, the BTR is planning a review of its use of information technology which will analyse (among other things) the way that survey data is handled within the BTR.

Information technology

5.31 The scope of the audit did not extend to the ANAO examining in detail the appropriateness of the information technology used to support data collections. The audit instead sought to identify examples where information technology had been used to improve data management and the measures adopted in agencies to address information technology-related problems.

ANAO findings

5.32 The ANAO noted that information technology requirements varied significantly from collection to collection. This reflected differences between statistical collections and transactional data collections and the different purposes for which data were used. For example, user requirements varied from simple tabulations of data to complex statistical analyses requiring highly specialised software to integrate data from separate databases.

5.33 The ANAO found examples of information technology contributing to improved data management and these are outlined in figure 11.

Figure 11

Features of information technology used for data collections

Data management stage	Information technology	Examples and comments
data collection and editing	• computer aided telephone interview (CATI) facilities (provided by consultants)	• A DIST tourism survey uses a telephone survey technique to collect travel information. CATI facilities contribute to data consistency and reliability by ensuring that the same procedures are followed for each interview.
	• use of internet to transmit data	• DEETYA higher education data collections require the universities to transmit the data to it via the internet. This reduces the lag between data collection and information availability.
	• software is provided to data suppliers to enable direct data entry	• DEETYA developed DEETYAPAC to facilitate direct data entry and editing by the information providers. Besides reducing the time to obtain accurate data it leads to increased data accuracy. While DEETYAPAC is 'purpose-made', proprietary software is also available that can be used to provide the same facilities for other collections.
data analysis	• software for information analysis • statistical analyses and preparation of tables • time series analyses (DIST tourism data collections)	• all the data collections reviewed had software that allowed analysts to select the most appropriate data analysis technique. Agencies used the statistical computer languages SAS and SPSS for tabulations and data analysis.
	• data warehousing - that is the integration of different databases	• ACS is introducing data warehousing to allow analysts to make full use of agency data resources by combining disparate databases.
information dissemination	• Internet	• all the agencies reviewed were using the internet to facilitate data access and dissemination.
	• CDROM technology	• data from DIST tourism collections are distributed by placing it on CDROM. This allows a large volume of data to be easily sent to users.

5.34 As discussed previously, the ANAO found that ACS and DIMA face data access problems due to the design features of the systems in which data is stored. In ACS, because the import databases were designed for transaction processing to support business systems rather than statistical purposes, difficulties arise when users attempt to perform complex data manipulations and integrate data from disparate databases. ACS is developing data warehousing and the 'data mart' (small data warehouses) which provide the means to more simply access data in several systems, including the Customs Control and Clearance of Lodged Entries (CLEAR) system.

5.35 In DIMA, the ANAO found that the Settlement Database is seen by users as cumbersome. DIMA is intending to adopt more advanced 'user-friendly' software to enable users to extract and manipulate data to meet their own requirements better.

Conclusion

5.36 The ANAO found that all of the collections examined made use of information technology. The information technology needs varied significantly from collection to collection and no single system would meet all needs. The ANAO found that information technology was used to improve data management although limitations of the technology restricted data access and use if the data sought was from transactional databases. The two agencies having transactional databases that were examined by the ANAO (ACS and DIMA) were appropriately developing information technology-based solutions such as data warehousing to improve data access and retrieval.

Better Practice Principles

From the work undertaken in this audit, the ANAO has identified the following practices which would enhance data management:

Staff skills

- Strategies need to be in place to provide the skills necessary for effective data management. Following an assessment of staff skill needs, agencies can:
  - recruit appropriate staff;
  - provide training and development opportunities;
  - seek assistance from consultants; and
  - obtain expert advice from other areas in agencies with specialist skills.

- Training and development opportunities to improve staff skills can be provided through:
  - staff rotations;
  - formal training programs that are tailored to meet the specific needs of the area and of individual staff; and
  - arrangements that enable staff to receive ongoing support while they develop their skills.

Use of consultants

- Appropriate processes should be in place for the engagement and use of consultants. The processes should ensure that contracting procedures are transparent and assist the agency to obtain value-for-money. If the agency does not have all the necessary skills and/or a data collection is particularly complex, it may be necessary to use a consultant to obtain a second opinion on whether the first consultant's approach and services meet agency requirements. As well, there are advantages in adopting a performance-based, as opposed to a cost-based, approach to managing consultants.

Information technology

- Agencies should examine options available so that they make best use of information technology. Information technology promotes better data management in relation to:
  - data accuracy and cost-effectiveness through systems which enable direct data entry, automatic editing and direct data transfer from the provider to the collecting agency;

– data integration (that is, relatability) and access through data warehousing; and

– data analysis through the use of IT-based research, statistical analysis and reporting tools.

Canberra ACT

22 June 1998

P.J. Barrett

Auditor-General

Appendices

Appendix 1

Data collections examined

Agency	Program area	Collection		Comment
DEETYA	Schools	Recurrent Grants	Census	Census of non-government school students and staff
			Financial Questionnaire	Information on the income and expenditures for non-government schools
	Higher Education	University Statistics	Student, staff and resources	Data collections with information relevant to university students, staff and resources
	Vocational Education and Training	Research studies	Non completing trainees; Contracted Entry Level	A telephone survey of a sample of non-completing trainees and apprentices, and their employers to determine why they did not complete their training
			Training Agencies (CELTAs)	A review of the demonstration projects for Contracted Entry Level Training Agencies
			Take up of New Apprenticeships	A strategic analysis by industry sector of the potential for take-up of the New Apprenticeships during 1997-1998
DTRD	FORS	Road Crash Statistics	Fatal file	Comprehensive data on all road fatalities extracted from Coroners' reports collated annually
			Fatality database (monthly)	Data from State/Territory records for road crash fatalities
			Casualties	As for the above fatality database but also includes accidents which resulted in an injury
	Aviation	Aviation Statistics	International	International flights (regular public transport services between Australian and foreign airports)
			Domestic	Domestic air traffic (regular public transport services in aircraft with over 38 seats, or with a payload of more than 4200 kilograms)
			Regional	Regional air traffic (regular public transport services in aircraft of 38 seats or less, or with a payload of less than 4200 kilograms).
			General aviation	General aviation (all flying except that performed by regular public transport services).

Appendix 1 (continued)

Agency	Program area	Collection	Comment
DIST	BTR	National Visitor Survey	Survey of domestic and international travel undertaken by Australian residents
		International Visitors Survey	Survey of Australian travel by short-term international visitors
		Domestic Tourism Monitor	Survey of domestic and international travel undertaken by Australian residents, superseded by the National Visitor Survey
	AusIndustry	Innovation Survey	Explores the relationship between science and technology and financial prosperity. It was first undertaken as an ABS initiative in 1994. DIST now pays about half the cost through a grant to ABS
		Scoreboard	Details on industry research expenditures to provides benchmarks that industrial groups can use to assess their research performance
DWPRSB	OSB	Australian Small and Medium Enterprise Database	A study over five years to provide data on the characteristics of small to medium sized businesses to determine the factors which influence business performance, including the births, growth and deaths of firms.
DIMA		Longitudinal Survey of Immigrants to Australia	A survey of 5000 immigrants and their accompanying family over three and-a-half years.
		Settlement Database	A secondary data collection that amalgamates data from seven different primary databases. It includes data from on-shore and off-shore (overseas) posts.
ACS		CLEAR (Customs Control and Clearance of Lodged Entries)	Holds import data and is used by ACS to detect imports that may be erroneous, fraudulent or place the community at risk (for example, drugs and firearms).
		Customs client satisfaction survey	Collects data on client satisfaction with services provided by ACS.

Appendix 2

Audit Reference Group

The Group included a range of people from different agencies with expertise in data management, including a representative from the ABS. The Group comprised:

- Ms Malisa Golightly, Executive Director, Performance Audit Business Unit, Australian National Audit Office;

- Dr Ian Heath, First Assistant Secretary, Information Services Division, Department of Health and Family Services;

- Ms Carole Palmer, Director, Information Management, Office of Government Information Technology;

- Mr Warren Richter, Assistant Statistician, Data Management, Australian Bureau of Statistics;

- Ms Ann Thurley, Senior Director, Performance Audit Business Unit, Australian National Audit Office; and

- Mr Ralph Watzlaf, General Manager, Professional Review, Health Insurance Commission.

Appendix 3

Commonwealth Metadata Standard: Australian Government Locator Service[20]

	Metadata element	Comment
1	Title	Name of the resource
2	Creator	The organisation(s) or person(s) responsible for the creation of the resource and included to place the resource in context of government
3	Subject	Specifies the subject or topic of the resource using key words that describe its content
4	Description	Provides a contextual description of the content of the resource. This element describes why the resource is offered and identifies other programs, projects and legislative action wholly or partly responsible for provision of access to the resource
5	Publisher	Names the entity responsible for making the resource available in its current form such as an agency or person
6	Contributor	Identifies significant contributors to the intellectual content of the resource other than the creator
7	Date	Important date(s) relevant to the resource
8	Type	The genre of the resource, eg. homepage, database, working paper etc.
9	Format	The physical format and data representation of the resource
10	Identifier	A string or number that uniquely identifies the resource within its domain
11	Source	Other principal information resources with the same intellectual content from which this resource was derived or generated
12	Language	The language of the intellectual content of the resource eg. English, French
13	Relation	Related information resources that assist discovery of the resource
14	Coverage	Spatial locations and temporal durations of the resource
15	Rights	The terms and conditions relating to access, security classification, privacy, copyright, intellectual property, use and reuse, and pricing. Included to provide provenance and to provide retention information
16	Functional descriptor	Specifies the business function of the agency to which the resource is related
17	Availability	Specifies how the resource can be made available/obtained by providing contact information

---

[20] Australian Government Locator Service Implementation Plan. A report by the Australian Government Service Working Group, December 1997.

Appendix 4

Glossary

Accrual accounting	The accrual basis of accounting recognises the financial effects of transactions or other events in the periods in which they occur, to the extent that those financial effects can be recognised and measured, irrespective of whether cash has been received or paid. APS agencies will be required to prepare accrual-based budgets in 1999-2000. Their budgeting arrangements will need to take into account expenses such as: the consumption of capital items (for example, depreciation of computers) and employee expenses (for example, entitlements such as long service leave).
Cell size	The number of data items in a cell of a tabulation is important because, if a cell count is too small (generally under five), fluctuations in the cell count (produced by data errors and random variations) can result in the data being misinterpreted.
Clearing House (or Commonwealth Statistical Clearing House)	A central clearance system for Commonwealth statistical collections, involving 50 or more businesses, coordinated by the ABS. It was established to reduce duplication, improve quality, eliminate statistical collections which do not meet quality standards and minimise provider load.
Conceptual statement	A statement of the fundamental basis for a collection. It describes the purpose of a collection and can include information on the target audience, the data collected and how it can be used.
Contestability	The prospect of competition in public sector activities to improve efficiency and effectiveness
Data classifications	The coding groups for each data element. For example, gender is generally coded as 'male', 'female' or 'unknown'.
Data dictionary	Provides details of the holdings in a system, for example: • data items (name, type (eg. integer, alphabetic, date), length); • indexes (used to speed up access to computer-based lists); and • access privileges. Data dictionaries are fundamental information technology records. Normally they are only maintained for mainframe or network systems.

Data management	Encompasses all stages of the data cycle from defining data needs, planning for the collection and analysis of data in line with defined needs, through to assuring data quality and its availability for use.
Data warehouses and data marts	A strategic information technology application which integrates data from separate databases and provides advanced research and reporting tools for extracting and manipulating data. Data marts are small data warehouses.
End-user computing or systems	User-developed applications which are designed to meet local data needs. Often they are inconsistent with departmental standards, poorly documented and have limited data relatability.
Information management	A broad term that refers to the management of all information maintained by an agency; for example, the collection and use of statistical and performance information, records management and data confidentiality and privacy practices.
Islands of data	Bodies of data stored on separate IT systems which have limited capacity to share and relate data between different systems. This significantly reduces data access for users outside the area which owns the system.
Key Result Areas	Key Result Areas (KRAs) are the core business functions or programs of an agency.
Longitudinal surveys	Longitudinal surveys involve collecting data from the same group of respondents several times over the life of the survey. Typically, a group is followed to determine what (and possibly why) changes have occurred.
Metadata	Definitions, descriptions of procedures, systems parameters and operational results which characterise and summarise statistical programs. It includes information about the records and fields (variables) that make up the data, together with information about the collection, storage and manipulation of the data. Metadata differs from data dictionaries in that it is more extensive in scope and is designed to underpin data analysis rather than the use information technology.
Operational data	Data captured for a particular transaction or administrative process (such as to check eligibility for a government funded program or compliance with laws and regulations). In this report, the term is used to differentiate it from *statistical* data which relates to data collected for a defined set of research objectives according to statistical methods.

Performance information	Performance information is a tool for management and performance improvement. It assists managers to determine where they are heading, how they will get there, whether they are heading in the right direction and whether they are using resources in the most cost effective manner. It is developed within a framework which includes the objectives and strategies. The framework also includes effectiveness, efficiency and the quality of service provision which relates to how the objectives are being achieved.
Provider load	The reporting burden imposed on information providers. It includes several factors such as the time taken to provide data and (if used) questionnaire length, the total number of people required to provide data for any collection and the total number of collections or administrative processes with which any person is involved.
Relatability	Refers to the ability to combine or integrate data from different sources to produce a synergistic gain in data content and use; that is, the content (and uses) of data sets which have been related exceeds that of the component data sets.
SAS	A software package used for tabulations and statistical analyses.
Structured data	Data that are collected in a predefined form for a specific purpose and are stored in a database. Structured data are distinct from 'unstructured data' which relates to file notes, records of meetings, correspondence, reports, spreadsheets and so on.

Index

Series Titles

Titles published in the financial year 1997-98

Audit Report No.1
Audit Activity Report: Jan-Jun 1997
Summary of Audit Outcomes

Audit Report No.2 Performance Audit
Government Business Enterprise
Monitoring Practices
Selected Agencies

Audit Report No.3 Performance Audit
Program Evaluation in the Australian
Public Service

Audit Report No.4 Performance Audit
Service Delivery in Radio and
Telecommunications
Australian Telecommunications
Authority and Spectrum Management
Agency

Audit Report No.5 Performance Audit
Performance Management of Defence
Inventory
Defence Quality Assurance (preliminary
study)

Audit Report No.6 Performance Audit
Risk Management in Commercial
Compliance
Australian Customs Service

Audit Report No.7 Performance Audit
Immigration Compliance Function
Department of Immigration and
Multicultural Affairs

Audit Report No.8 Performance Audit
The Management of Occupational Stress in
Commonwealth Employment

Audit Report No.9 Performance Audit
Management of Telecommunications
Services in Selected Agencies

Audit Report No.10 Performance Audit
Aspects of Corporate Governance
The Australian Tourist Commission

Audit Report No.11 Performance Audit
AUSTUDY
Department of Employment, Education,
Training and Youth Affairs

Audit Report No.12 Performance Audit
The Pharmaceutical Benefits Scheme
Department of Health and Family
Services

Audit Report No.13 Performance Audit
Third Tranche Sale of the Commonwealth
Bank of Australia

Audit Report No.14 Financial Control
and Administration Audit
Official Travel by Public Sector Employees

Audit Report No.15 Financial Control
and Administration Audit
Internet Security Management

Audit Report No.16 Performance Audit
Equity in Employment in the Australian
Public Service
PSMPC and other agencies

Audit Report No.17 Performance Audit
Sydney Airport Noise Amelioration
Program
Department of Transport and Regional
Development

Audit Report No.18 Performance Audit
Management of the Implementation of the
New Commonwealth Services Delivery
Arrangements
Centrelink

Audit Report No.19 Performance Audit
Risk Management in ATO Small Business
Income
Australian Taxation Office

Audit Report No.20 Performance Audit
Sales Tax
Australian Taxation Office

Audit Report No.21 Financial Control and Administration Audit
Protective Security

Audit Report No.22 Financial Control and Administration Audit
Audits of the Financial Statements of Commonwealth Entities for 1996-97
Summary of Results and Outcomes

Audit Report No.23 Performance Audit
Ministerial Travel Claims

Audit Report No.24 Performance Audit
Matters Relevant to a Contract with South Pacific Cruise Lines Ltd
Department of Employment, Education, Training and Youth Affairs

Audit Report No.25 Performance Audit
Gun Buy-Back Scheme
Attorney-General's Department

Audit Report No.26 Performance Audit
Strategic and Operational Management
National Registration Authority for Agricultural and Veterinary Chemicals

Audit Report No.27 Performance Audit
Managing the Year 2000 Problem
Risk Assessment and Management in Commonwealth Agencies

Audit Report No.28 Performance Audit
Contracting Arrangements for Agencies Air Travel

Audit Report No.29 Financial Control and Administration Audit
Management of Accounts Receivable

Audit Report No.30 Performance Audit
Evaluation Processes for the Selection of
– Records Management Systems
– Internet Access Services
for the Commonwealth
Office of Government Information Technology

Audit Report No.31 Financial Statement Audit
Aggregate Financial Statement prepared by the Minister for Finance and Administration
Year ended 30 June 1997

Audit Report No.32 Performance Audit
The Management of Boat People
Department of Immigration and Multicultural Affairs
Australian Protective Service
Australian Customs Service - Coastwatch

Audit Report No.33 Performance Audit
Commonwealth Management of the Great Barrier Reef
Great Barrier Reef Marine Park Authority

Audit Report No.34 Performance Audit
New Submarine Project
Department of Defence

Audit Report No.35 Performance Audit
DEETYA International Services
Department of Employment, Education, Training and Youth Affairs

Audit Report No.36 Performance Audit
Audit Activity Report
July to December 1997
Summary of Outcomes

Audit Report No.37 Performance Audit
Protection of Confidential Client Data from unauthorised Disclosure
Department of Social Security
Centrelink

Audit Report No.38 Performance Audit
Sale of Brisbane Melbourne and Perth Airports

Audit Report No.39 Performance Audit
Management of Selected Functions of the Child Support Agency
Australian Taxation Office

Audit Report No.40 Performance Audit
Purchase of Hospital Services from State Governments
Department of Veterans' Affairs

Audit Report No.41 Financial Control and Administration Audit
Asset Management

Audit Report No.42 Preliminary Inquiry
Preliminary Inquiries into the Natural Heritage Trust

Audit Report No.43 Performance Audit
Life-cycle Costing in Department of Defence
Department of Defence

Audit Report No.44 Performance Audit
*The Australian Diplomatic
Communications Network – Project
Management*
Department of Foreign Affairs and
Trade

Audit Report No.45 Performance Audit
Planning for Rural Health
Department of Health and Family
Services

Audit Report No.46 Financial Control
and Administration Audit
Internal Audit