The Auditor-General
Audit Report No.29 2005–06
Performance Audit

Integrity of Electronic Customer Records

Centrelink

Australian National Audit Office

ISSN 1036–7632

ISBN 0 642 80889 9

COPYRIGHT INFORMATION

Australian National
Audit Office

Canberra ACT
15 February 2006

Dear Mr President
Dear Mr Speaker

The Australian National Audit Office has undertaken a performance audit in Centrelink in accordance with the authority contained in the *Auditor-General Act 1997*. I present the report of this audit and the accompanying brochure to the Parliament. The report is titled *Integrity of Electronic Customer Records.*

Following its tabling in Parliament, the report will be placed on the Australian National Audit Office's Homepage—http://www.anao.gov.au.

Yours sincerely

Ian McPhee
Auditor-General

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

AUDITING FOR AUSTRALIA

The Auditor-General is head of the Australian National Audit Office. The ANAO assists the Auditor-General to carry out his duties under the *Auditor-General Act 1997* to undertake performance audits and financial statement audits of Commonwealth public sector bodies and to provide independent reports and advice for the Parliament, the Government and the community. The aim is to improve Commonwealth public sector administration and accountability.

For further information contact:
The Publications Manager
Australian National Audit Office
GPO Box 707
Canberra ACT 2601

Telephone: **(02) 6203 7505**
Fax: **(02) 6203 7519**
Email: **webmaster@anao.gov.au**

ANAO audit reports and information about the ANAO are available at our internet address:

http://www.anao.gov.au

Audit Team

Michael McFarlane
Kylie Jackson
Fran Holbert
Eric Turner

Contents

Abbreviations

ANAO	Australian National Audit Office
ATO	Australian Taxation Office
CRN	Centrelink Reference Number
CSC	Customer Service Centre
CSO	Customer Service Officer
DEST	Department of Education Science and Training
DEWR	Department of Employment and Workplace Relations
DI	Data Integrity
DIE	Data Integrity Enquiry (system)
DIMIA	Department of Immigration and Multicultural and Indigenous Affairs
DSSDD	Department of Social Security Data Dictionary
FaCS	Department of Family and Community Services
FAO	Family Assistance Office
GiR	Getting it Right Strategy
ISIS	Income Security Integrated System
IT	Information Technology
POI	Proof of Identity
QOL	Quality On-Line
RSS	Random Sample Survey Programme
TFN	Tax File Number

Glossary

Data integrity

1. The Australian and New Zealand Standard, AS/NZS 7799.2:2003, *Information Security Management*, defines data integrity as:

Safeguarding the accuracy and completeness of information and processing methods.

2. The Webopedia encyclopaedia of computer technology,[1] states that data integrity:

Refers to the validity of data. Data integrity can be compromised in a number of ways:

- *human errors when data is entered;*
- *errors that occur when data is transmitted from one computer to another;*
- *software bugs or viruses;*
- *hardware malfunctions, such as disk crashes; and*
- *natural disasters, such as fires and floods.*

3. The USA Wikipedia Internet site suggests data integrity has the following meanings:

- *the condition that exists when data is unchanged from its source and has not been accidentally or maliciously modified, altered or destroyed;*
- *the condition in which data are identically maintained during any operation, such as transfer, storage or retrieval;*
- *the preservation of data for their intended use; and/or*
- *relative to specified operations, the* a priori *expectation of data quality.* [2]

---

[1] Most definitions above are based on material sourced from the JupiterWeb network's Internet site: <www.webopedia.com> — an Internet based encyclopaedia, dedicated to computer technology.

[2] Sources quoted for these definitions include US Federal Standard 1037C and the National Information Systems Security Glossary (USA).

	Drawing on the above definitions, in this audit the term 'data integrity' is used to refer to the accuracy, completeness and logical consistency of records. Where information is stored in more than one place, the consistency of that information is also important, as well as its 'fitness for purpose'. It is a measure of, and sometimes used synonymously with, 'data quality'.
Program vs Programme	Throughout this report, the word 'program' refers to 'a series of coded software instructions to control the operation of a computer or other machine'. The word 'programme' refers to 'a set of related measures or activities with a long-term aim.' For example, social security programmes. Source: Compact Oxford English Dictionary (2005).
Batch204 (error checks)	Batch204 processing is used to check Centrelink's customer records for structural integrity errors. Structural integrity relates to the specifications of a particular field or a relationship that should exist between two or more fields. For example, a field designed to hold a date, such as a person's date of birth, should only contain a valid date in a particular format.
XDI (error checks)	The XDI checks are based on business rules. Essentially the 'X' stands for one of the business areas or clusters—so that, collectively XDI checks cover, Pension DI checks, Parenting Payment DI checks, Summary Data DI checks and so on.
ISIS record type	Customer records are assigned a record type. For example: PER (for person), ORG (for organisation) and CHI (for child).
Random Sample Survey Programme	The Random Sample Survey (RSS) Programme is part of Centrelink's Business Assurance Framework. The RSS programme involves sampling a number of Centrelink customers from each of the main payment streams to verify the accuracy of information provided by those customers. This process occurs each year and also checks the accuracy

	of calculated payment rates, based on the customer information.
Service reason code	ISIS uses a set of three-letter codes to identify the service reason (essentially the social security programme under which a benefit determination has been made) for customers. For example, JSR refers to JobSeekers, AGE to Age Pension recipients and NSA to NewStart Allowance.

Summary and Recommendations

Summary

Background

1. Like most Australian Government agencies involved in service delivery in the 21st century, Centrelink relies on large and complex information technology (IT) systems to support its extensive business operations. The heart of Centrelink's IT systems is ISIS—the Income Security Integrated System—Centrelink's main customer database.

2. In 2004–05, Centrelink's IT systems performed more than 5.2 billion electronic computations and processed some $63 billion of social security payments to over six million customers. Centrelink grants approximately 2.8 million new claims each year. At September 2005, the ISIS database held information on over 23 million customers—recording details of customers' identity, circumstances and eligibility for benefits under various social security programmes. Approximately 6.2 million of the 23 million records relate to customers with a current benefit determination.[3]

3. In order to distinguish between customer records, a unique identifier is assigned to each record—the Centrelink Reference Number, or CRN. The information in ISIS is organised around the CRN, which links customer information in various parts of the database. For example, the CRN links information on a customer's circumstances and benefit determinations with that in the payments file.

4. Customer information is spread across eleven networked computing environments, with each environment, essentially, servicing a region, state or territory within Australia.[4] Centrelink's data holdings are growing at a rate of approximately 30 per cent each year, and at September 2005, the ISIS database held information in over 440 billion fields, with an average of 21 000 fields of information per customer.

Audit approach

5. The audit examined aspects of the integrity and management of customer data stored on ISIS. In particular, the audit considered measures of

[3] Other records include historical records for customers previously in payment, along with records for organisations and children.

[4] One of the computing environments stores information on Centrelink customers residing outside Australia.

data accuracy, completeness and reliability. The scope of the audit also extended to aspects of Centrelink's IT control environment—in particular, controls over data entry.

6. ANAO considered Centrelink's processes and procedures for entering customer data into ISIS, including the controls surrounding customer registration and the validation of customer data. ANAO also examined Centrelink's existing data integrity error detection and reporting system.

7. ANAO requested, and Centrelink provided, data extracts from all 23 million ISIS records. ANAO tested the contents of a number of mandatory fields to ensure these conformed to Centrelink's business rules and specifications. ANAO's analysis also included a check of logical relationships between various fields.[5] Centrelink customers are required to prove their identity when claiming a pension, benefit, or allowance from Centrelink. ANAO examined details of Proof of Identity (POI) documents recorded on ISIS.

8. A substantial part of ANAO's analysis involved testing the integrity of the primary key[6] of the database—the CRN. ANAO checked for the existence of duplicate CRNs—whether any given value for a CRN was associated with more than one customer—and for multiple CRNs—where an individual customer had been assigned more than one CRN.[7]

9. Fieldwork for the audit was primarily undertaken during April 2005 to October 2005. ANAO acquired over 8 million lines of data, extracted from the agency's data integrity error detection system on 12 July 2005. On 13 September 2005, Centrelink provided ANAO with over 23 million lines of data extracted from the main ISIS database, in accordance with ANAO's specifications.

Overall audit conclusion

10. Centrelink's customer database, ISIS, constitutes one of the largest and most complex Australian Government databases holding information about Australian citizens and residents. With over 23 million records in total, some

---

5 For example, that a customer's recorded date of death did not precede his or her recorded date of birth, or that a customer's marital status (single or partnered) aligned with the payment rate for a benefit that was paid at either a single or partnered rate.

6 The primary key is a means of uniquely identifying each record within the database and a mechanism to link data across various elements of the database.

7 And, therefore, had multiple records in the database.

6.2 million records support a current benefit determination, and in most cases, payment to a customer by Centrelink.

11. This audit found that Centrelink could significantly improve the accuracy and integrity of data stored on ISIS. In particular, Centrelink could improve the integrity of the primary key used in ISIS, and reduce the risks associated with fragmenting customer information across multiple records. Centrelink should also remove training records and obsolete customer records from the production environment of its database. ANAO also found that Centrelink should improve the effectiveness of its existing data integrity checking system.

12. The audit found that up to 30 per cent of customer 'proof of identity' (POI) information recorded on ISIS was insufficient or unreliable in terms of uniquely identifying or substantiating the identity of customers. While much of this information related to historical records, ANAO also found that this information is still relied upon to process new claims associated with those historical records. ANAO noted that Centrelink has tightened some of the controls around POI data entry and that the quality of recently entered POI information appears to be considerably improved.

13. While this audit has highlighted a number of business risks arising from these data integrity issues, including the risk of duplicate or inappropriate payments to customers, the ANAO also found that Centrelink had in place a number of other controls designed to prevent inappropriate payments. Accordingly, the audit found that, while these risks exist, duplicate payments had only occurred in a small number of cases.

14. Therefore, given the scale and complexity of Centrelink's IT operations, and considering the information examined in the scope of this audit, ANAO concluded that Centrelink's electronic customer records are, generally, sufficiently accurate and complete to support the effective administration of the range of social security programmes for which Centrelink is responsible.

15. ANAO also recognises that Centrelink responded promptly to the matters raised during the course of this audit, and commenced a number of initiatives to address specific data integrity issues identified by ANAO, and to generally improve the quality of data in ISIS. Key among these initiatives were projects to analyse and correct the identification of false positive results in the agency's existing data integrity error checking system, the establishment of a Data Quality Team to develop a long term strategy to improve and maintain data quality and work to comprehensively describe the effects of data integrity

errors. Centrelink also undertook to review the operation of the priority rating system for data integrity errors.

16. In addition, Centrelink acted quickly to review cases of potential duplicate payment of customers, and to commit to resolving cases of duplicate and multiple CRNs.

Key Findings

Data entry and exchange (Chapter 2)

17. Having introduced the 'Getting it Right' strategy in 2000, which is founded on the four pillars of: the right person is paid; under the right programme; at the right rate; for the right date(s), Centrelink's intentions of ensuring accurate data and payments were evident to ANAO.

18. Centrelink provides training for all Customer Service Officers (CSOs) in relation to registering new customers on its database. CSOs are also provided with considerable guidance in relation to processing claims and recording customer information. However, decisions about whether customers lodging a new application should be issued with a new CRN rely on the judgement of individual CSOs. The ANAO found that, despite the range of administrative level controls in place, up to 3 per cent of Centrelink customers appear to have been registered more than once on ISIS (for a detailed treatment of customers with multiple registrations, see Chapter 6).

19. Centrelink's IT systems incorporate a number of system level controls, designed to ensure compliance with certain business rules and data entry specifications. However, ANAO's analysis of ISIS data, in particular that detailed in Chapter 4 of this report, indicates that not all data entry business rules have been comprehensively enforced.

20. Centrelink has introduced post-data-entry quality assurance procedures, such as Quality On-Line[8] and a Random Sample Survey

[8] Quality On-Line (QOL) monitors the completeness and correctness of information used in processing customer's claims. QOL was introduced as a quality assurance process to ensure that payments made by Centrelink and the services provided are correct. The QOL system is based on a second person comprehensively checking the correctness of the work of the CSO who initially processed a customer's claim and entered the customer's data into ISIS.

programme[9]. These are designed to detect inaccurate payments or benefit determinations that may have arisen from inaccurate customer data.

Data integrity error detection and reporting system (Chapter 3)

21. Centrelink has in place an extensive data integrity error detection and reporting system, incorporating checks of structural integrity and checks against various programme business rules. However, ANAO noted that the system was not widely used by Centrelink programme managers, nor was the information systematically analysed to reveal trends or identify the cause of particular data integrity failures.

22. Also, ANAO found that the data integrity error detection and reporting system had deteriorated over time and failed to incorporate updated or new business rules, thus producing many false positive results. The system did not discriminate between data integrity errors associated with current Centrelink customers and those associated with historical customer records. Consequently, Centrelink programme managers were not afforded an insight into the true magnitude of particular data integrity errors or the actual level of risk relating to current customers.

23. Centrelink employs a priority rating system to provide a high-level breakdown of data integrity error statistics. However, the ANAO found that the system did not adequately discriminate between errors, nor did it overtly highlight those areas requiring immediate attention by programme managers. This is because approximately two-thirds of all errors were classified as Priority 1 or Priority 2.

24. ANAO also found that, over 50 per cent of the top 87 error definition tables lacked any description of the effect of the error. In this circumstance, programme managers were not presented with sufficient information to recognise the significance, or easily comprehend the likely impact, of particular data integrity problems.

25. ANAO noted that, according to the information contained in Centrelink's data integrity error detection and reporting system, the number of data integrity failures has steadily increased over the past two years. The ANAO acknowledges that a large proportion of reported errors arose from the

[9] The Random Sample Survey (RSS) Programme involves sampling a number of Centrelink customers from each of the main payment streams to verify the accuracy of information provided by those customers. This process occurs each year and also checks the accuracy of calculated payment rates, based on the customer information.

incorrect identification of false positives[10], and over half of all identified errors were associated with historical records. However, Centrelink had made little progress, over the two years preceding this audit, in resolving the errors.

Testing data integrity (Chapter 4)

26. The results of ANAO's analysis of selected database fields, extracted from the 23 million customer records on ISIS, showed that the production environment of ISIS:

- contained at least 10 000 training records—that is, non-genuine customer records created while training Centrelink staff;
- exhibited a degree of inconsistency in the recording of customers' names, with some entries containing a customer's first name, second name and surname, all in the surname field, while leaving the other two fields blank;
- exhibited a degree of inconsistency in recording customers' address details;
- contained entries in particular fields, which were outside the range of legal values defined for those fields; and
- exhibited some anomalies in the recording of customers' dates of birth and death. For example, ANAO found 42 customer records that displayed the same date for the customer's date of birth and date of death and one record indicating that the customer was born two months after his recorded date of death.

27. These findings point to a lack of, or failure of, system level controls, which should enforce conformance with Centrelink's documented business rules and data recording specifications.

28. ANAO's analysis of the data indicated that Centrelink records a false or 'dummy' date of birth, when a customer's true date of birth is not known with certainty. The 'dummy' dates used in the ISIS database are 1 January and 1 July in any given year, although the years 1900 and 1901 are regularly used. ANAO considers that this practice could skew any statistical analysis, based on customer age, although the effect would be most noticeable for age profiles over 100 years. According to ANAO's analysis, approximately 0.5 per cent of

[10] False positive results can arise when a data integrity error checking program tests data against obsolete business rules. These data integrity checking programs report errors where they should not.

recorded dates of birth, for current Centrelink customers, are inaccurate to some extent.

29. ANAO identified that 1.46 million customer records on ISIS had a date of death recorded for the customer—some of which were many decades in the past. ANAO also found that a relatively small number of these records supported a current benefit determination. That is, the data supplied by Centrelink to ANAO, indicated that these customers were current—although not necessarily in payment. Centrelink subsequently advised ANAO that payments had ceased for the majority of these customers, but that the records had been corrupted and continued to display a current benefit determination, when they should no longer do so.

30. Centrelink also advised ANAO that it was required to maintain some records for deceased customers—where there may be an ongoing debt to the Commonwealth, or where the record is associated with a partner record[11]. While recognising Centrelink has a valid business reason for maintaining those categories of deceased customer records, ANAO considers that there is little reason to maintain the large number of records relating to deceased customers, which do not fit into these categories, in the production environment of the database. The existence of these records gives rise to an unnecessary risk to the integrity of Centrelink payments.

31. ANAO found that the data field recording customers' Tax File Number was compromised, in that entries in that field were not unique. Yet, Tax File Numbers are intended to be unique—the one Tax File Number may not be shared by two people. ANAO found that up to 7 000 customer records—3 500 pairs of records—shared the same Tax File Number. ANAO's analysis of Centrelink's data indicated that, in many cases the single Tax File Number was shared in Centrelink's records by a couple, or a parent-child combination, or a sibling combination.

Recording customer identity (Chapter 5)

32. ANAO examined 8.3 million lines of Proof of Identity data to determine the usefulness of the information recorded in ISIS, in substantiating the identity of customers. This involved checking that the POI documents recorded in the database were associated with unique serial numbers or registration numbers.

[11] Where the partner is still alive and in receipt of payment.

33. ANAO's analysis revealed that, as at September 2005, many ISIS records displayed entries inconsistent with Centrelink's policy for recording POI information. Rather than recording valid serial numbers for particular POI documents, thousands of records displayed apparently false serial numbers, such as 99999, 123456, and xxxxx. In addition, many other records displayed a text entry, such as, Citizenship papers, Unknown, and Sighted, rather than a valid serial number.

34. ANAO's analysis showed that only 72.6 per cent of POI records citing Australian Citizenship Certificates contained unique values. In addition, 96.6 per cent of POI records citing Current Australian Passports, and 56.6 per cent of POI records citing Australian Birth Certificates, contained unique values.

35. Overall, ANAO's analysis of four primary POI documents revealed that up to 30 per cent of the recorded details on ISIS were insufficient or unreliable in terms of uniquely identifying or substantiating the identity of customers.

36. ANAO also noted that, since September 2001, Centrelink had introduced a range of system level controls and quality assurance procedures designed to improve the quality of POI information recorded in ISIS. ANAO accepts that Centrelink has made a significant improvement in the quality of POI data entered into ISIS over the past two or three years, and that current procedures are superior to those in place prior to 2001. However, ANAO's analysis included all POI data recorded on ISIS as at September 2005—recent and historical—as historical POI data is still used, in many cases, when processing a new claim for a previous Centrelink customer.

Integrity of the primary key (Chapter 6)

37. Centrelink uses the CRN as the primary key for ISIS. Within any database, the primary key is of great importance. In a well managed database, each customer is allocated one, and only one CRN. In addition, no one CRN is shared by two records within the database.

38. ANAO found that Centrelink's primary key was compromised by the existence of up to 25 000 duplicate CRNs. That is, in 25 000 cases the same CRN had been allocated to two different customers. In addition, ANAO identified up to 500 000 customers with multiple CRNs. That is, those customers had been registered at least twice, under two different CRNs. While the raw numbers appear substantial, they represent approximately 0.2 per cent and 3 per cent, respectively, of all customer records in ISIS.

39. The existence of duplicate CRNs means that the primary key may not be relied upon to uniquely identify Centrelink's customers within ISIS. The effect of multiple CRNs is that customer information may become fragmented across two or more different records. This situation presents a risk of duplicate benefit payments or an inappropriate combination of benefit payments—one on each of the customer's unrelated records.

40. ANAO's analysis indicated that up to 1 000 Centrelink customers possessed a current benefit determination on each of two separate records. In many of these cases, one benefit determination appeared to be linked to a payment while the second did not.[12] In a minority of cases, the data indicated that a customer was current for the same benefit on two records, or that the two records supported incompatible benefit determinations.[13]

41. ANAO provided Centrelink with relevant details and Centrelink investigated the circumstances of these cases. Centrelink then advised ANAO that some of these cases were previously known to exist and that alternative controls were in place to avoid duplicate payments.[14]

42. Therefore, ANAO found that, while the fact that up to 500 000 customers have multiple records presents a risk of overpayment, that risk had been realised in only a very small number of cases. Nevertheless, ANAO considers that Centrelink should address the underlying data integrity issues, rather than rely on an incomplete set of alternative controls to mitigate these risks.

Implications of data integrity issues (Chapter 7)

43. ANAO found that the inconsistent recording of customer's names and addresses creates a number of problems and reduces the integrity of customer data generally. These problems are compounded by the use of dummy values for some date fields, the existence of training records in the production

---

[12] Some benefit determinations are not payment-related. For example, a person may have a current benefit determination to receive a Low Income Health Care Card or for JobSeeker Registration, which allows access to the Job Network.

[13] For example, the recipient of a Carer payment may not be in receipt of another income support payment, such as Age Pension, NewStart Allowance or Parenting Payment. However, such a person may be entitled to receive a Carer Allowance or a Family Tax Benefit payment. [More information is available in Centrelink's publication, *A Guide to Australian Government Payments*].

[14] Centrelink had implemented a duplicate payments filter—a control within the payments system—to stop payment on the second record of a known duplicate pair. Centrelink advised the ANAO that at 18 January 2006, 1 283 of the 2 000 records had been investigated. Centrelink identified one case of overpayment and six cases where a customer had been issued a second Low Income Card.

environment and anomalies in the recording of Tax File Numbers. During the course of this audit, ANAO observed the:

- improper use of data fields—all name elements appearing in the surname field, leaving the first and second name fields blank;
- reversal of first and second names across two records;
- data entry errors and variations in spelling, including the use or non-use of hyphens and/or spaces in two-word name elements;
- inconsistencies in recording addresses; and
- use of values outside those defined as legal values in Centrelink's data dictionary.

44. ANAO considers that inaccurately recording customer details could inhibit Centrelink's ability to effectively analyse its customer data for compliance and fraud detection purposes. Inaccurate data could also reduce the effectiveness of Centrelink's data matching with other agencies and organisations. Obsolete and dummy records could make it difficult for Centrelink to calculate accurate counts of customers or to conduct modelling or data profiling activities that rely on customer age.

45. ANAO found that poor data integrity in Centrelink's electronic POI records could impact on its capacity to effectively detect and prevent fraud, or to engage in data matching activities where a high degree of confidence in the identity of its customers is required.

46. ANAO found that fragmenting customer information across two or more records, through the inadvertent allocation of multiple CRNs to individual customers, presented the greatest risk to maintaining the integrity of Centrelink payments. With two or more—unlinked—customer records, a customer may have two current benefit determinations, be identified as deceased on one record but not the other, or display inconsistencies in personal information across those records.

47. ANAO noted that Centrelink had in place other controls to guard against overpayment, in the cases of multiple CRNs known to Centrelink. This audit found that only a small number of overpayments appeared to be associated with multiple CRN customers. However, ANAO found that, with up to 500 000 customers on ISIS, who have multiple CRNs, Centrelink would be in a stronger position to manage these risks if it were to resolve the underlying data integrity issues.

Centrelink's response

48. Centrelink thanks the Australian National Audit Office for the way in which this audit was conducted. The professionalism of the officers, their evident technical expertise, the working relationship that was fostered and the willingness of all parties to address operational issues throughout the course of this audit has greatly aided Centrelink in quickly implementing continuous improvements to the administration, accuracy, completeness and consistency of its data holdings.

Recommendations

Recommendation No.1
Para 3.40

The ANAO recommends that Centrelink improve the usefulness and effectiveness of its data integrity (DI) reporting system by:

(a) ensuring the timely inclusion of new or revised DI checks whenever new software applications are released, so that the system is always checking data against current business rules; and

(b) enabling the system to clearly identify DI errors associated with current customers.

Centrelink's response: Agreed.

Recommendation No.2
Para 3.54

ANAO recommends that Centrelink, in order to provide programme managers with the capacity to determine the relevant priority of DI issues, including those requiring urgent or immediate attention, revise its priority rating system for DI errors, with a view to:

(a) comprehensively and accurately describing the likely effects of DI errors;

(b) resolving inconsistencies between the stated effects of some errors and the criteria for ascribing particular priority ratings; and

(c) clearly identifying DI errors that pose the greatest risk to the efficient and effective administration of programmes and payments.

Centrelink's response: Agreed.

Recommendation No.3
Para 4.67

ANAO recommends that, in order to address the range of data quality issues identified by this audit, Centrelink conducts a thorough data cleansing exercise within the ISIS database, with a view to:

(a) removing training records and spurious customer records from the production environment;

(b) removing or otherwise inactivating records for deceased customers from the production environment, where there is no continuing business need to retain the records;

(c) improving the accuracy of customers' personal information, particularly in recording the various elements of customers' name and address;

(d) enforcing existing business rules surrounding the use of defined legal values with certain ISIS fields;

(e) resolving possible anomalies in the recorded dates of birth and death for Centrelink customers identified during this audit; and

(f) resolving possible anomalies in the recorded Tax File Numbers for Centrelink customers identified during this audit.

Centrelink's response: Agreed.

Recommendation No.4
Para 5.62

ANAO recommends that Centrelink:

(a) continues to monitor the operation of its Proof of Identity policy and the quality of POI information recorded in ISIS; and

(b) progressively replaces spurious or inaccurate POI information currently recorded in ISIS with accurate information, when processing new claims or undertaking major of reviews of eligibility for existing customers.

Centrelink's response: Agreed.

Recommendation No.5
Para 6.49

ANAO recommends that, in order to improve the integrity of the CRN, the primary key for ISIS, Centrelink takes action to resolve:

(a) all duplicate CRNs — instances where **different** customers have been allocated the same CRN and instances where the **same** customer has a current benefit determination on two or more Centrelink computing environments;

(b) all multiple CRNs — instances where the same customer has been registered under two or more different CRNs; and

(c) all instances of records where a date of death has been recorded against one of a customer's duplicate or multiple records, but not the other(s).

Centrelink's response: Agreed.

Audit Findings and Conclusions

1. Introduction

This chapter provides background information about Centrelink and describes how Centrelink relies on complex information technology systems to support the delivery of a wide range of social security payments and services to its customers. It also provides an introduction to the structure and operation of Centrelink's main customer database. The chapter concludes with an outline of the approach taken in this audit.

Background

1.1 Centrelink is a statutory agency, under the umbrella of the Department of Human Services, within the Finance and Administration portfolio. Centrelink operates under the *Commonwealth Service Delivery Agency Act 1997* to provide Australian Government services in accordance with specified service agreements. The agency administers products and services on behalf of 25 client agencies including the Departments of Family and Community Services (FaCS); Education, Science and Training (DEST); and Employment and Workplace Relations (DEWR). Responsible for the delivery of up to 140 products and services, Centrelink's operations extend to over 1 000 delivery points across Australia. In 2004–05, Centrelink made more than $63.09 billion in social welfare payments to 6.48 million customers.

1.2 Centrelink relies on large and complex information technology (IT) systems to support its various business areas in the delivery of these services. In 2004–05, Centrelink granted 2.77 million new claims and its IT systems performed approximately 5.2 billion electronic computations.

Income Security Integrated System

1.3 A vital component of Centrelink's IT systems is the Income Security Integrated System (ISIS) — Centrelink's main customer database. ISIS constitutes a comprehensive store of information relating to Centrelink's customers. It includes details of customers' identity and eligibility for particular benefits. ISIS also stores historical information about customers' dealings with Centrelink, including a record of payments received, over time.

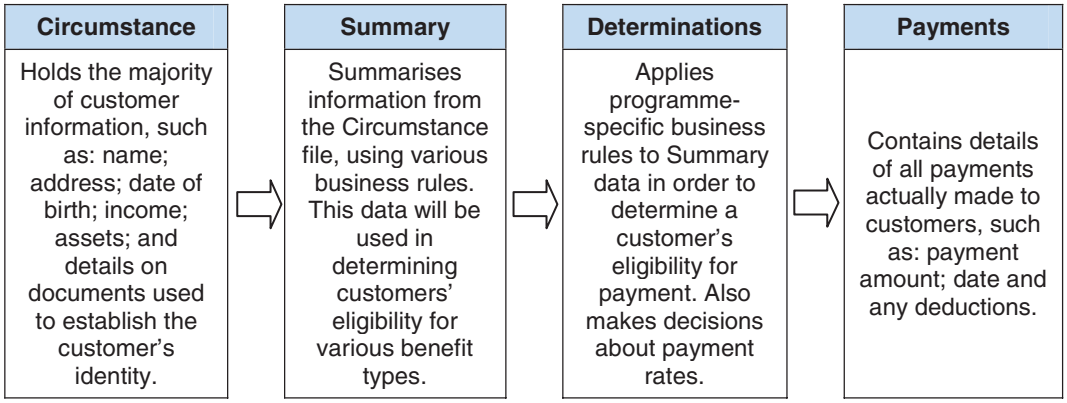
1.4 At July 2005, ISIS contained over 23 million customer records. Different types of customer records are defined within ISIS. Customer records can relate to people or to organisations. A special type of customer record relates to

children.[15] The majority of customer records on ISIS—almost 18 million—relate to people, although only about 6.2 million of these relate to customers currently receiving a Centrelink benefit.

1.5 The information in these records is distributed over a number of different files within ISIS. Key among these are the Customer Circumstance, Summary, Determinations and Payments files. Figure 1.1 illustrates the type of information and functions associated with these files.

Figure 1.1

ISIS file groups



Source: Centrelink

1.6 In order to distinguish between customer records, a unique identifier is assigned to each record—this is referred to as a primary key for the database. The primary key used in the ISIS database is the Centrelink Reference Number (CRN).[16] Each customer is issued with a CRN upon initial registration with Centrelink—for example, when first lodging a claim for Centrelink benefits.

1.7 A customer's CRN is used to link information stored in the various ISIS files. The CRN then provides a means of identifying each customer's data, wherever the data might reside on the ISIS database. For example, the CRN provides a way to connect circumstance data with determinations and payments data for a customer. The CRN drives most data management activities within ISIS.

---

[15] The concept of different record types is described in greater detail in Chapter 4.

[16] Examples of primary keys in other databases might include a person's credit card number, Tax File Number, Drivers Licence number and so on. The CRN is a 9-digit identification number, with a security feature incorporated into a tenth character.

1.8 Because of the size of the ISIS database, customer records are spread across eleven separate computing environments. This architecture balances the load on Centrelink's IT systems, and so helps maintain the efficiency of IT services. These environments are, essentially, geographically ordered. That is, one environment is devoted to storing and processing data relating to customers in a particular state or territory, or part of a state or territory.

1.9 In 2000, Centrelink introduced a series of strategies and actions intended to improve accuracy, correctness, and accountability, leading to quality decision-making throughout Centrelink. This initiative was called 'Getting it Right'[17] (GiR). Centrelink considers GiR an integral and vital strategy for ensuring the accuracy of its data and payments. Managers were expected to make GiR a top priority. GiR is based on four pillars:

- the right person is paid;
- under the right programme;
- at the right rate; and
- for the right date(s).

1.10 To ensure efficient service delivery to customers, Centrelink must ensure that its customer data is accurate, reliable and of high quality. Centrelink must be able to rely upon this data to support the effective and efficient administration of a range of social security programmes.

Audit approach

1.11 The objective of this audit was to examine the integrity of electronic customer records stored on Centrelink's main customer database (ISIS), and to report on Centrelink's management of the data.

1.12 The audit examined Centrelink's customer records and its data management practices, assessing them against the following criteria:

- Centrelink electronic customer records are accurate and complete;
- Centrelink electronic customer records are reliable and internally consistent;

[17] In late 1999, Centrelink's senior management group examined issues of correctness and accuracy, focusing on barriers preventing staff from 'getting it right'. A December 1999 report identified 10 key barriers, and identified actions in train or planned to address them. In April 2000, the Centrelink Board of Management endorsed the 'Getting it Right' strategy. The strategy identified four pillars of correctness: right person, right rate, right date and right programme.

- Centrelink has adequate controls and procedures to ensure high quality customer data; and
- Centrelink effectively manages electronic customer records.

Audit Methodology and Scope

1.13 ANAO approached the audit in three phases, considering aspects of:

- the collection and recording of customer information and data exchange with other agencies;
- Centrelink's existing data integrity error detection and reporting system; and
- data integrity within ISIS, based on a series of analyses of data extracted from the database during September 2005.

1.14 The audit also considered relevant controls, both technical and administrative, surrounding data entry and noted Centrelink's quality assurance processes.

1.15 ANAO included in the scope of the audit, a consideration of the capabilities of Centrelink's data integrity error detection and reporting system. Centrelink supplied the ANAO with data extracts containing information on customer records that were reported to be affected by data integrity errors. The ANAO conducted some testing of these in an attempt to gain greater insight into the potential impact of data integrity errors on Centrelink business.

1.16 A substantial part of the audit revolved around the ANAO's analysis of data integrity within the ISIS database. ANAO requested, and Centrelink provided data extracts containing information, from specific fields, for all 23 million customer records in ISIS. The ANAO tested this data to ensure that selected mandatory fields contained valid data. ANAO also examined aspects of internal consistency in the database—applying these as measures of the accuracy and completeness of customer records.

1.17 ANAO's analysis included an assessment of the integrity of the primary key used in ISIS, that is, the CRN. It also extended to an examination of the electronic records of documents used to establish the identity of Centrelink customers.

1.18 The scope of this audit did not include testing customers' eligibility for any payment/s being received. Nor did it consider the accuracy of payment rates determined by Centrelink. Rather, the audit focused on whether the

customer identification information in ISIS was sufficiently accurate and complete for Centrelink to rely on this information to process customer claims.

Previous audits and reports

1.19 The ANAO has not previously undertaken a detailed performance audit of Centrelink's electronic records.

1.20 The ANAO is currently undertaking an audit of the Rolling Random Sample Survey—a key component in Centrelink's accountability framework. ANAO expects to complete that audit before the end of the 2005—06 financial year.

1.21 Audit Report No. 37 1998–99, *Management of Tax File Numbers*, and Audit Report No. 47 2004–05, *Tax File Number Integrity*, go to a number of issues in connection with Tax File Numbers that are relevant to this audit.

1.22 This audit was conducted in accordance with the ANAO Auditing Standards, at a cost to the ANAO of approximately $415 000.

Structure of report

1.23 This chapter provides a brief introduction to Centrelink and the audit.

1.24 Chapter 2 examines how Centrelink captures and records customer information. It also considers data exchange activities and discusses the controls associated with data entry, generally.

1.25 Chapter 3 provides an analysis of Centrelink's data integrity error detection and reporting system. The chapter explores the effectiveness and usefulness of the system, and presents the results of the ANAO's analysis of the customer records affected by data integrity errors.

1.26 Chapter 4 outlines the methodology used in the ANAO's analysis of the information held in Centrelink's customer records. It presents and discusses the results of this analysis. Considered in the analysis were fields associated with storing customers': name and address; date of birth and date of death; Tax File Number; sex; and marital status.

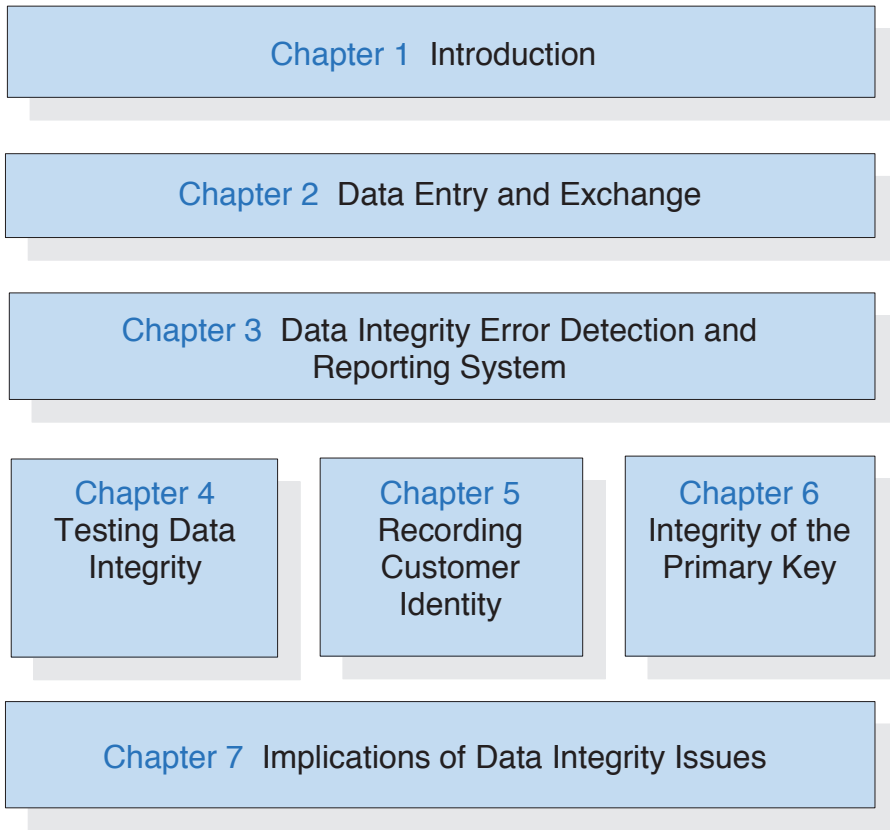
1.27 Chapter 5 presents ANAO's findings in relation to the recording of documents used to establish the identity of Centrelink customers.

1.28 Chapter 6 presents the results of ANAO's assessment of the integrity of the primary key used in the ISIS database (the CRN).

1.29 The final chapter draws on the various audit findings covered in the previous chapters and considers the collective impact of data integrity issues on Centrelink's business.

Figure 1.2

Report structure



Source: ANAO

2. Data Entry and Exchange

This chapter describes how Centrelink captures and stores information about its customers. It outlines procedures for Centrelink staff to directly enter customer data into its database, and how Centrelink exchanges information with other agencies or organisations. The chapter also considers Centrelink's quality assurance procedures, intended to ensure high quality data is maintained in ISIS.

Collecting customer information

2.1 Centrelink needs to collect and record a large amount of information about its customers. Much of this is personal information, and information about the customer's circumstances.[18] Centrelink relies on this information to make decisions about customers' eligibility for various benefits. For these decisions to be correct, the customer information collected and recorded by Centrelink must be accurate and comprehensive.

2.2 Centrelink collects customer information in a number of different ways. Customers can communicate with Centrelink via:

- personal interviews with Customer Service Officers (CSOs) at Centrelink Customer Service Centres (CSCs);
- completed claim forms sent through the mail or lodged at a CSC;
- telephone call to a Centrelink call centre;
- telephone call, using the telephone keypad or interactive voice response system; and
- the Internet.[19]

Data entry

Customer Service Officers

2.3 In order to receive a Centrelink payment, a customer must lodge a completed claim form and be eligible to receive the payment or service they are claiming. People claiming for payment—and their partners—may be interviewed to determine their basic eligibility. An interview can also outline the rates and conditions of payment, advise the customer of their rights and

[18] Centrelink is authorised to collect this information under various provisions in social security and other legislation.

[19] Not all transactions are available over the telephone or through the Internet.

obligations for the continuation of payment, and gather proof of identity, age, residency, and details of customers' income and assets[20].

2.4 During an interview, the CSO either creates a new customer record in the ISIS database, or accesses the customer's existing record. The CSO then directly enters information into ISIS, from the completed claim form and/or information provided by the customer at interview.

Call centres

2.5 Centrelink has 26 call centres in its call centre network. Similar to the procedures used at interview, call centre staff access the customer's record on ISIS and edit existing data or enter new data into the database.

Customer self-service

2.6 Centrelink encourages customers to use a range of self-service options. These include use of interactive telephone services—including an automated voice recognition system—and the Internet. Centrelink customers must first register for the service and be granted a particular level of access. The level of access granted to a customer determines the level of information that the customer can access as well as provide. For example, a higher level of access enables the customer to update personal details such as income received.

Data exchange with other agencies

2.7 Centrelink exchanges electronic data, including customer details, with a number of federal, state and local government agencies, private sector organisations, and foreign governments. The information exchanged falls into three categories:[21]

- a simple yes or no response by Centrelink to enquiries about an individual's eligibility for discounted services (e.g. does the customer have a current Centrelink concession card);
- information about a customer's circumstances is provided to Centrelink by other agencies (e.g. the Department of Immigration and Multicultural and Indigenous Affairs (DIMIA) provides information on

[20] All customers in certain categories must be given a full pre-grant interview. These are cases where: there is a risk of fraud; there are complex qualification issues; there are cultural or language difficulties; the customer wishes to be interviewed; or the customer is aged 15-24 years old and is claiming Youth Allowance as independent (unable to live at home).

[21] The scope of this audit considered the second and third categories—those data exchange activities resulting in Centrelink receiving data from other agencies. It did not consider the first category—activities where Centrelink provided a yes/no response to requests from other agencies.

customers' overseas absences, that may affect eligibility and rates of payment for various programmes); and

- personal information provided to Centrelink to identify new customers (e.g. the Australian Taxation Office (ATO) provides Centrelink with details of individuals claiming the Family Tax Benefit).

2.8 Data exchange varies depending on the requirements of Centrelink and the agency. A list of some of the agencies that Centrelink undertakes data exchange activities with, and the type of information and frequency of exchange, is at Appendix 1.

2.9 In order to exchange information with other agencies, Centrelink and the agencies undertake a matching exercise to create a Mutual Client Index. Once Centrelink and the other agency are confident they have agreed on a person's identity, the customer record is eligible for automatic update of information via the data exchange. Automatic update may lead to consequent changes in benefits or payment rates. Where Centrelink and the other agency are in some doubt as to the identity of the customer, any update of customer information must be processed manually.[22]

Controls associated with customer registration

2.10 Regardless of the manner in which Centrelink receives information, all Centrelink customers must be registered on the ISIS database. In supporting this requirement, Centrelink maintains a National Index—a listing of all customers on ISIS. The system operates such that, before registering a new customer record in ISIS, the CSO must search the index to determine if the customer is already registered. It is important that the search is comprehensive and identifies any existing records before adding a new customer, to avoid duplicate customer records.[23] Centrelink CSOs are provided with the following instructions in relation to registering new customers.

> This is **a Centrelink Must Do**. The instructions below must be followed exactly as they are written. Staff cannot use any discretion when applying this law, policy or procedure, unless clearly stated otherwise.

---

[22] Centrelink requires that people registering for benefits provide sufficient proof of their identity, before processing claims. Centrelink, generally, does not rely on personal details supplied by other agencies in order to create a **new** Centrelink customer record. Two exceptions to this rule are where ATO provides Centrelink with details of persons claiming Family Tax Benefits, and DIMIA provides details of sponsored immigrants. In these cases, where Centrelink cannot match the person concerned to an existing customer record, a new shell (preliminary) record is created.

[23] Duplicate customer records, and a treatment of the potential risks associated with customers having multiple records on ISIS, are examined in Chapter 6 of this report.

> The procedures in this topic detail how to use the search and add (indexing) functions on the Income Security Integrated System (ISIS).[24]

> It is important that thorough searches are completed and the CSO is sure that the correct customer record has been located prior to updating customer details. It is also necessary to complete an exhaustive search before adding a new record for a customer.[25]

2.11 The National Index search uses information such as the customer's surname, first name and/or initial and date of birth. If the CSO is satisfied that the customer is not previously recorded on the system, a new customer record is created and the customer's details are entered. Otherwise, details contained in the existing customer record are confirmed with the customer, updated and the existing record is used.

Audit findings

2.12 ANAO found that Centrelink provides training for all CSOs in relation to indexing customers. CSOs are also provided with considerable guidance in relation to processing claims and recording customer information. Much of the training material is available to CSOs through Centrelink's Intranet. However, ANAO also found that, decisions about registering new customers relied on the judgement of individual CSOs.

2.13 ANAO noted that, despite the range of administrative level controls in place, up to 3 per cent of Centrelink customers appear to have been registered more than once on ISIS (for a detailed discussion of customers with multiple registrations, see Chapter 6).

System level controls associated with data entry

2.14 In any IT system a major control to help ensure that data entered into a database is accurate and correct, is the system level enforcement of business rules. That is, the IT system is programmed so that any data entered, which does not comply with an allowable format or content for a particular field in the database, will be rejected at the time of entry. Immediate rejection allows the CSO entering the data to promptly correct the entry, while the customer or paper form is readily available.

---

[24] Centrelink, *Initial Contact/New Claim Procedures-Overview*, 2005, Centrelink Intranet, [Accessed 29 April 2005].

[25] Ibid.

2.15 For example, an attempt to enter a date that is not a valid date—such as the 30th of February—should be rejected. Also, where a business rule prohibits backdating the commencement of a benefit, an attempt to enter a date prior to the date of the customer lodging an application, should be rejected.

2.16 ANAO considered a Centrelink document titled *Data Integrity checking within Data Modify Routines – A specification*, which outlined many of the business rules associated with entering data into ISIS. These specifications indicate a number of controls that operate whenever data is modified within ISIS, including:

- ensuring only numeric values are entered into 'amount' and 'decimal amount' fields;
- checking for valid dates in all date fields and defining a valid year;[26]
- ensuring 'indicator' fields store only the values 'Y' or an underscore character; and
- special rules associated with the storage of Tax File Numbers and Centrelink Reference Numbers.

2.17 The ISIS database also employs a number of tables, each of which stores a list of valid values for data entered in particular fields. For example, the field storing the customer's sex may only contain one of three legal values—M (male), F (female), U (unknown).

2.18 A large proportion of fields in ISIS permit 'free text entry'—that is, there are no constraints on what a user might enter into a field, although a maximum size limit might be imposed for any particular field. These fields are used to store information such as elements of the customer's name, address, employer's name and other customer circumstance information.

Audit findings

2.19 ANAO found that Centrelink's IT systems incorporated a number of system level controls, designed to ensure compliance with certain business rules. ANAO's analysis of ISIS data, in particular that detailed in Chapter 4 of this report, indicates that not all data entry business rules are comprehensively enforced.

[26] For the entry of a year in any date, the entry must not be greater than 120 years in the past, nor may it be more than 30 years in the future.

Quality assurance processes

2.20 In 2000, as part of the introduction of a broader Business Assurance Framework,[27] Centrelink introduced a system called Quality On-Line (QOL) to monitor the completeness and accuracy of information used in processing customer's claims. QOL was introduced to:

> assure the protection of program[me] outlays for the income support payments and services delivered by Centrelink.[28]

2.21 Over its five years of operation, QOL has evolved from a quality assurance process to a quality control process. The QOL system is based on a second person comprehensively checking the correctness of the work of the CSO who initially processed a customer's claim and entered the customer's data into ISIS.

2.22 Under the QOL system, recently appointed CSOs have 100 per cent of their work checked by a QOL checker, until the CSO is assessed as proficient. CSOs rated as proficient have 5 per cent of their work checked by a QOL checker. A computer program randomly selects the claims of proficient CSOs that are to undergo QOL. The results of all QOL checks are collated and made available to Centrelink managers and team leaders, through a system called QOLStat.

> This information is then fed into QOL's statistical information system, QOLStat, for management information purposes. This process also allows for the checking officer to provide comment to the CSO who submitted the work for checking, within the QOLCheck tool, enabling a real-time feedback loop to assist in learning and improving the skills of the CSO.[29]

---

[27] The Business Assurance Framework was conceived as a comprehensive and integrated mechanism to provide assurance on Centrelink's performance to Government, client agencies, the [Centrelink] Board, stakeholders and customers. The Framework was jointly negotiated and agreed between FaCS and Centrelink. BAF [Business Assurance Framework] has four key principles:

- there should be explicit and binding agreement on what is to be measured;
- there should be explicit binding agreement on how measurement is to be done;
- the same definitions of what is to be measured will be applied at each of three levels—quality control, quality assurance and external assurance;
- the results will be transparent.

Source: Centrelink, *Business Assurance Framework,* Centrelink's Intranet.

[28] Source: 2004, Centrelink, QOL Operating Guidelines, Version 1.4, p4.

[29] Ibid.

2.23 Although primarily focussed on the accuracy of claims processing decisions, and the determination of payments, QOL also provides feedback to CSOs on the correctness and completeness of their data entry activities.

2.24 Centrelink's Business Assurance Framework also incorporates a programme of rolling Random Sample Surveys (RSS). The RSS process involves sampling a number of Centrelink customers each year, to verify the accuracy of information provided by those customers. The process also checks the accuracy of calculated payment rates, based on the customer information. Centrelink describes RSS as:

> Random Sample Surveys are a point in time analysis of customer circumstances designed to establish whether the customer is being correctly paid in accordance with the four pillars of payment correctness under the Business Assurance Framework – right person, right payment, right rate and right date.[30]

2.25 To carry out these RSS reviews each quarter, Centrelink randomly selects a sample of customers from each of the major payment streams and checks their details and payment records. FaCS independently validates the results of this examination.[31]

2.26 The scope of this audit did not extend to a detailed consideration of the RSS programme or QOL. The ANAO has previously discussed aspects of Centrelink's Business Assurance Framework, QOL and the RSS programme in performance audits undertaken during 2002–2003[32]. In addition, the ANAO is undertaking an audit of the RSS programme, concurrently with this audit. The RSS audit report is expected to be tabled in Parliament before the end of the 2005—06 financial year.

Audit findings

2.27 ANAO found that Centrelink has developed a number of systems and procedures to monitor the quality of customer data entered into ISIS. ANAO noted that a range of quality controls—administrative and system level—had been introduced at the point of data entry, helping to ensure that only accurate

---

[30] Centrelink, *Business Assurance Framework, Rolling Random Sample Survey Results, Quarter 4 of 2002–03*, p.3, Centrelink's Intranet, [Accessed 29 April 2005].

[31] Up to October 2004, FaCS had policy responsibility for all major programmes administered by Centrelink. As a result of the machinery of government changes announced in October 2004, the responsibility for RSS was redistributed across four agencies—FaCS, Centrelink, DEWR and DEST.

[32] Audit Report No. 17, 2002–03, *Age Pension Entitlements*, and Audit Report No. 44 2002–03, *Review of the Parenting Payment Single Program*.

and complete customer data was entered into ISIS. In addition, ANAO noted that post-data-entry quality assurance procedures, such as QOL and the RSS programme, were designed to detect inaccurate payments or benefit determinations that may have arisen from inaccurate customer data.

3. Data Integrity Error Detection and Reporting System

This Chapter describes the operation and use of Centrelink's data integrity error detection and reporting system. It also considers the priority rating system used by Centrelink to highlight the importance of particular data integrity errors. The Chapter incorporates the results of ANAO's analysis of data produced by Centrelink's data integrity error detection and reporting system, as at July 2005.

Detecting data integrity errors

3.1 Data accuracy and integrity can be compromised in a number of ways. Centrelink attempts to ensure that all customer information, when initially collected, is recorded accurately on ISIS.[33] Nevertheless, some customer data may be entered incorrectly. For example, a person's year of birth may be accidentally recorded as 1873, rather than 1973.

3.2 A large number of Centrelink computer programs operate on data stored in ISIS—for example, updating fields in customer records, recording the results of determinations and recording payment details. In such a complex IT environment as Centrelink maintains, the operation of some of these computer programs may have unintended consequences, from time to time, in that they can cause data corruption. For example, a program designed to modify data in selected fields may only partially overwrite the previous entry, leaving a combination of the previous entry and the updated entry as the final result.

3.3 Whatever the cause, data integrity failures can represent a risk to the efficient administration of programmes that rely on these data holdings. In 1992, Centrelink introduced a suite of programs to detect data integrity errors within the ISIS database. Part of Centrelink's mainframe computer systems, the Data Integrity Enquiry (DIE) system involves running automated data integrity (DI) error checks against customer records. There are two main DI error detection processes—Batch204 processing and XDI processing. Essentially, the two processes are designed to detect structural errors in the data and data that do not conform to certain business rules.

[33] See the discussion of QOL in the previous Chapter.

Batch204 processing

3.4 Batch204 processing is used to check Centrelink's customer records for structural integrity errors. Structural integrity relates to the specifications of a particular field or a relationship that should exist between two or more fields. For example, a field designed to hold a date, such as a person's date of birth, should only contain a valid date in a particular format. If that field was to become corrupt and contain alphabetic characters, it represents a structural integrity error. Similarly, a field designed to record a person's sex might specify three legal values—M(ale), F(emale) or U(nknown). Any other character stored in such a field would, again, represent a structural integrity error.

3.5 Batch204 error checking programs are run each week. At July 2005, there were 45 discrete Batch204 error checking programs.[34] Thirteen regular Batch204 jobs run every week. These check selected fields in the majority of the ISIS database[35] for DI errors. In addition to running the 13 regular jobs each week, the DIE system incorporates another 32 DI checking programs. A selection of these is run each week, on a rotational basis, so that over a period of approximately two months, all 32 programs are called up.

XDI processing

3.6 The XDI[36] checks are based on business rules. The checking programs are written by the various business areas or clusters within Centrelink. Each week, XDI checks operate on a sample of customer records. Twenty per cent of the database is checked each week so that, over a five-week period the majority of ISIS records are checked.[37]

3.7 XDI checks can also be run on a single customer record when required. Unlike Batch204 DI checks, which essentially operate on a field of data at a time, XDI checks operate on individual customer records. Figure 3.1 illustrates how Batch204 and XDI error checking complement each other.

---

[34] Each program results in a number of individual DI error checks being conducted.

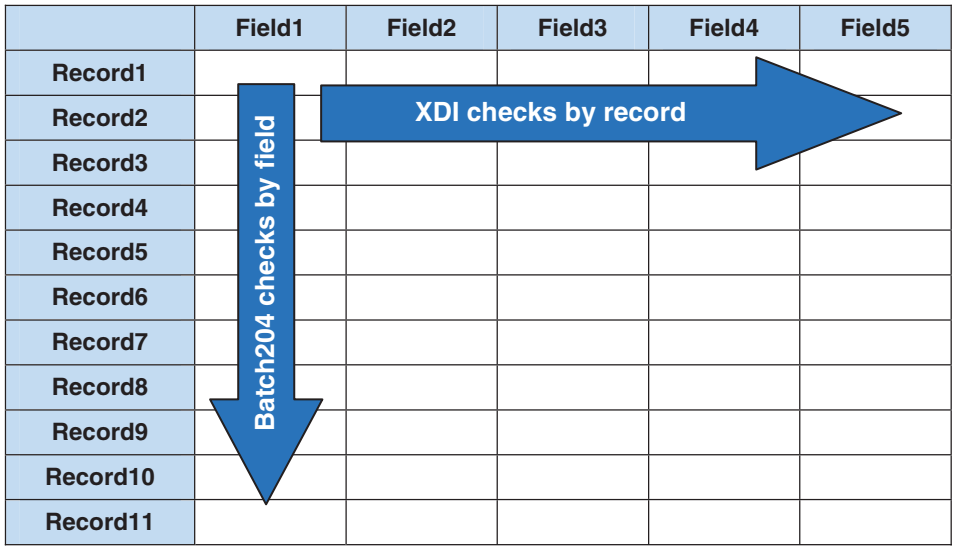
[35] Customer records are assigned a record type. For example: PER (for person), ORG (for organisation) and CHI (for child). Centrelink informed ANAO that 98 per cent of DI checks operate on PER and ORG records only—that is, the checks exclude CHI record types. At July 2005, PER and ORG record types accounted for approximately 17.8 million of the 23.2 million records held on ISIS. This represents approximately 75 per cent of the entire ISIS database.

[36] Essentially the 'X' stands for one of the business areas or clusters—so that, collectively XDI checks cover, Pension DI checks, Parenting Payment DI checks, Summary Data DI checks and so on.

[37] Like Batch204, XDI checks operate on PER and ORG records only.

Figure 3.1

Batch204 and XDI error checking



Source: ANAO, based on information from Centrelink.

3.8 By checking an entire customer record, the XDI process can examine relationships, or compare values across various fields. This permits Centrelink to ensure that a customer's information is consistent with any determinations and payments made to that customer. For example, XDI programs for the Family Assistance Office—or FAO cluster—check that the payment amount in a customer's determination record equals the amount in the payment file. It also checks if a customer is receiving a payment without a corresponding determined amount.

3.9 Together, the Batch204 and XDI processes are designed to provide a comprehensive analysis of the ISIS database and to detect data integrity errors. The results of all weekly DI error checks are stored within the DIE system on Centrelink's mainframe computers. The DIE system stores considerable detail about every DI error detected. A standard set of information is generated for each error check performed. Appendix 2 outlines the 13 pieces of information available for each error—such as error number, description, priority rating, effect and business rule involved.

Priority rating scheme

3.10 Centrelink assigns a priority rating to each DI error type, to reflect the seriousness of the error. Table 3.1 outlines the five-point priority rating scheme, in use at May 2005.

Table 3.1

Priority rating scheme for DI errors

Priority	Description
1	** Critical ** Incorrect payments, staff unable to perform duties
2	High. Adverse publicity, staff workload increased
3	Medium. Data corruption, delays in processing
4	Low. Corrective action required
5	Not allocated. Wish list

Source: Centrelink, DI Reference Information Guidelines, and Production Data Integrity Errors for 3 May 2005, Centrelink Intranet.

Reporting data integrity errors

3.11 While the DIE system holds considerable detail, it is located on the mainframe computer and requires some specialised knowledge to operate to full effect. In order to provide Centrelink programme managers with an overview of data integrity issues relevant to them, in 2000, Centrelink developed a user-friendly, Intranet-based DI error reporting system.

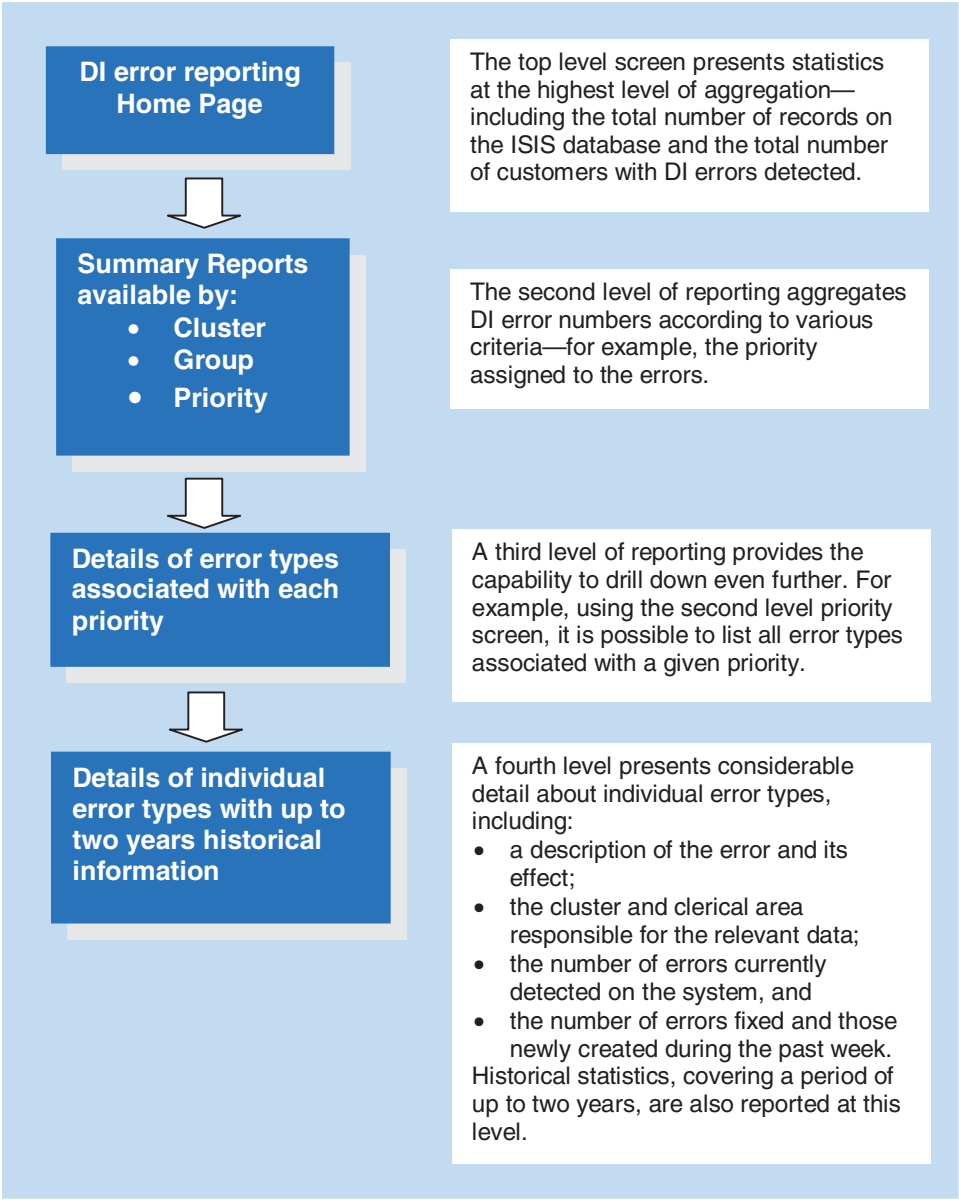
3.12 The DI error reporting system forms part of the Data Management section of Centrelink's Intranet. The reporting system collates information stored in the DIE system and presents it in a series of reports.[38] These reports are updated weekly.

3.13 This system enables users to view information at various levels of aggregation and to drill down to details relating to particular error types. Figure 3.2 illustrates how information on DI errors is presented through the various levels of the DI error reporting system.

[38] Centrelink informed ANAO that the DI error reporting system was not designed specifically as a management information system. It was introduced as a prototype in 2000 and, apart from some minor improvements over time, has not been significantly revised since initial implementation.

Figure 3.2

Overview of DI error reporting system



Source: ANAO, based on information from Centrelink.

3.14 When the ANAO first examined Centrelink's DI error reporting system, in May 2005, the statistics presented in the various reports indicated very high levels of errors. For example, the top level report, titled *Production Data Integrity Errors for 3 May 2005,* sourced from Centrelink's Intranet in the

week beginning 2 May 2005, included the following set of summary statistics for DI errors, at that time.[39]

Table 3.2

Data Integrity Summary Data for 3 May 2005

Summary	Australia Wide
Number of customers (IS1CUST0)	23 242 263
Customers with DI errors	5 126 109
Customers with an XDI error	3 040 096
Customers checked by XDI - no error	14 746 126

Source: Centrelink, Production Data Integrity Errors for 3 May 2005, Centrelink Intranet.

3.15 Upon first inspection, the report informs the reader that over 5 million customer records were affected by DI errors.[40] However, the nature of these DI errors and the status of the customer records must be considered before drawing any conclusion regarding the integrity of ISIS data. These matters are discussed in the next section of this report.

3.16 The statistics reported in Table 3.2 relate to the number of customer records falling into each category—they do not represent the actual number of DI errors detected across the ISIS database. A single customer record may contain one or more XDI errors, one or more Batch204 errors, or a combination of XDI and Batch204 errors. Other screens in the DI error reporting system illustrate the number of DI errors detected. For example, Table 3.3 shows the number of DI errors, by priority.

---

[39] The report also provided a breakdown of the statistics for each of the eleven computing environments in ISIS.

[40] The ANAO confirmed the meaning of each of the descriptors used in Table 3.1 with Centrelink's Data Integrity team. The 'Number of customers' refers to all customer records stored on Centrelink's ISIS database—including old records of people no longer receiving a Centrelink benefit payment. At 3 May 2005, there were 23 242 263 customer records stored on ISIS.

The descriptor 'Customers with DI errors' indicates that 5 126 109 customer records have one or more DI errors associated with them—these errors have been detected through both the Batch204 process and the XDI process.

The third row in the table, 'Customers with an XDI error', shows that 3 040 096 customer records are associated with one or more errors, detected through the XDI process alone. This number is a sub-set of the 5 million customer records at row two of the table. The final row indicates that 14 746 126 customer records have been checked by the XDI process and found to contain no XDI error.

Therefore, with 14 746 126 customer records free of XDI errors and 3 040 096 customer records containing XDI errors, the XDI error checking procedures run over 17 786 222 customer records. ANAO confirmed that this was the case—XDI checks are generally run against customers identified as a 'person' or an 'organisation'. As little as two per cent of XDI checks involve other customer types, such as 'child'.

Table 3.3

DI error counts for 3 May 2005, by priority

Priority	Short description	No. of DI errors as at 3 May 2005
1	** Critical **	624 433
2	High	4 943 798
3	Medium	2 404 802
4	Low	19 708
5	Not allocated	232 731
	Total	8 225 472

Source: Centrelink, Production Data Integrity Errors for 3 May 2005, Centrelink Intranet.

3.17 Accordingly, upon first inspection, combining the information in Tables 3.2 and 3.3 suggests that, at May 2005, the ISIS database contained a total of 8 225 472 data integrity errors, spread across 5 126 109 customer records. However, as previously noted, the nature of the errors and the status of the customer records must be considered prior to drawing any conclusions based on these raw statistics.

Nature of DI errors

3.18 According to the DI error reports extracted by ANAO on 3 May 2005, details were reported for 1 245 different error types. In total, some 5 000 error checks are conducted within the DIE system.

False positives

3.19 ANAO found that some DI checks identify errors where they should not—thereby producing false positive results. ANAO observed that such a situation can arise from a flaw in the error checking program, or as a result of a development elsewhere in the IT system that the DI check does not take into account.

3.20 The following example illustrates how the identification of false positives might occur.

Case study

A particular data integrity check compares values in a data field with a list of valid values stored in the data dictionary. Customer type is a field in ISIS used to identify the type of customer record. Valid values, listed in the data dictionary, include:

- **Person;**
- **Organisation; and**
- **Child.**

A DI error check has been written to compare the value held in each customer's record with the list of valid values. If a customer's record holds a value other than one of the valid values, a DI error is detected and reported.

A new business area within Centrelink might establishe a new type of customer, such as 'Carer'. Although 'Carer' may be a valid Customer type for transactions involving that business area, unless the DI check is updated to include 'Carer' as a valid Customer type, every 'Carer' record will be detected and reported as a DI failure.

3.21 ANAO's discussions with Centrelink staff members revealed that, in some cases, the DI checking programs were not updated to take account of the introduction of new or revised software applications. As a result, the DI checking programs maintained a focus on the previous set of business rules, rather than those employed in the newly introduced software.

3.22 False positives are not necessarily indicative of genuine data integrity failures. Depending on the nature of the false positive, the underlying data may be sufficiently accurate and complete for Centrelink to rely upon the data for any number of electronic transactions. On the other hand, if false positives result from checking an obsolete business rule, then the current business rule will not be checked and some genuine data integrity failures will not be detected or reported.

3.23 Once identified, false positives such as those that would result from the above example, are relatively easy to fix.[41] However, until the cause of false positives is identified, the figures generated contribute to Centrelink's total DI error count.

[41] Centrelink informed ANAO that over 1 million false positive DI errors had been fixed in the week preceding 3 May 2005. These errors were the result of a similar scenario to that outlined in the example.

Audit findings

3.24 There are some 23 million customer records on ISIS. Almost 18 million are regularly checked by the DIE system. Overall, ANAO found that Centrelink's DI error detection and reporting system was designed to provide a detailed assessment of data integrity across the agency's major customer data holdings. However, ANAO also found that an unknown number of false positive DI errors reside within the DIE system. These false positives contribute to the total DI error count, thereby reducing the system's accuracy and effectiveness.

3.25 Put simply, users of the system are unable to determine how many of the 8.2 million reported DI failures result from false positives and how many represent genuine DI errors. Consequently, Centrelink programme managers presented with the DI error reports are not afforded an insight into the true magnitude of any particular DI problem, nor are they easily appraised of the significance of any particular data integrity issue.

3.26 ANAO also found that the use of obsolete or incorrect DI error checking programs meant that Centrelink had not been effectively assessing customer information against an up-to-date set of business rules.

Actions taken by Centrelink

3.27 After receiving ANAO's preliminary findings, in September 2005, Centrelink commenced a project to review the operation of many aspects of the DIE system. In October 2005, Centrelink advised the ANAO that:

> The high level analysis undertaken by Centrelink classified the DI errors into the following categories:
>
> - Ongoing error – current error; or
> - Incorrect error (False positives) – error has been produced incorrectly as the error code has been superseded, or the error has been fixed but is still reporting incorrectly. [42]

3.28 Centrelink's analysis suggested that up to 60 per cent of reported DI errors were false positives. In addition, Centrelink provided the following advice.

> Many errors are dated and have been on the record for a long period….. Centrelink recognises that this reduces the effectiveness of the DI system

[42] Advice from Centrelink's Chief Information Officer, to the ANAO, on 12 October 2005.

reports and will review and explore the possibility of archiving the old errors.[43]

3.29 In addition, Centrelink provided further evidence to ANAO that, as a result of its investigation of false positive errors, the total DI error count had been reduced from 8.2 million—at May 2005—to less than 6 million—at October 2005, and to 5.4 million at 15 November 2005. Centrelink also advised ANAO, as follows:

> Centrelink is in the process of establishing a Data Quality Team. The role of the Data Quality Team is to develop a long term Data Quality strategy and improvement program[me]. This team will also undertake a detailed analysis of the DI reporting and implement improvements.
>
> Centrelink is undertaking a program[me] to reduce the number of errors being reported. Initial fixes for errors were implemented in the September [2005] system release, with further fixes scheduled for December 2005 and March 2006 releases.[44]

Customers status

3.30 The status of a customer's record is also important in determining the significance of DI errors. Customers may be current or non-current. That is, a customer's record may support a current benefit determination, and very often a Centrelink payment,[45] or it may not support a current benefit determination.[46] A DI error on a current customer's record may represent a risk to the accuracy of the related benefit determination and payment, whereas a DI error on a non-current customer's record is less likely to represent a risk to the integrity of Centrelink outlays.

3.31 Centrelink advised ANAO that, at July 2005, approximately 6.5 million customer records supported a current benefit determination. That is, Centrelink had 6.5 million current customers, at that time. The DIE system checks almost 18 million customer records. Therefore, more than 11 million

---

[43] Ibid.

[44] Ibid.

[45] Most Centrelink benefit determinations result in a payment to the customer. However, a small number of benefit determinations, such as JobSeeker Registration and the Low Income Health Care Card, provide the customer with access to non-payment-related benefits or services.

[46] Some customer records that do not support a current benefit determination may be of ongoing business interest to Centrelink. For example, a customer record may hold details of a debt to the Commonwealth. A non-current customer record may constitute a 'partner' record for another customer, who is in receipt of a benefit and whose benefit rate is affected by the income of his or her partner.

non-current customer records are checked, along with the 6.5 million current customer records.

3.32 In terms of a risk to Centrelink outlays, one extreme possibility is that the 5.1 million customers, whose records have associated DI errors, are not currently in payment—that is, these are all non-current customer records. This scenario would represent a very small business risk to Centrelink. However, at the other extreme, if the 5.1 million customer records with DI errors all relate to the 6.5 million customers currently in payment, then the business risk to Centrelink would be much more significant.

3.33 Centrelink confirmed that the DI error detection and reporting system, as it stood at July 2005, did not provide a break down of statistics based on the status of customers, although a customer status indicator was stored in the mainframe DIE system.[47]

ANAO's analysis

3.34 In order to gain further insight into the status of customer records with DI errors detected and reported, ANAO undertook an analysis of Centrelink's DIE data files. ANAO requested, and Centrelink provided, line-by-line data for DI errors detected over the weekend of 9 and 10 July 2005. Each line of data included:

- a Centrelink Reference Number (CRN). The CRN is the primary key for the ISIS database. It is essentially a customer identification number;
- an error code;
- a priority rating for the error;
- a current/non-current indicator; and
- a code representing the computing environment from which the error record was drawn.

3.35 Centrelink provided the data to ANAO on 12 July 2005. ANAO also took copies of a series of standard reports from the DI error reporting system, in order to reconcile the line-by-line data with the aggregated results in the standard error reports. The number of lines of data provided was 8 473 729; the

[47] The raw data, stored in the DI error reporting system on the mainframe computer, does contain a field described as a 'current status indicator'. Centrelink advised ANAO that this field was derived from the ISIS field CUCS.ON.IND, in the customer details file, and that it could distinguish between customers considered current or non-current. However, this field was not used to filter any of the DI error results reported through the Intranet-based Data Management System.

second level report of error count, by priority, indicated a total of 8 473 730 errors detected on that weekend.[48]

3.36 ANAO identified 5 172 848 unique CRNs within the data—indicating that the 8.4 million DI errors existed on 5.1 million customers records.[49] Furthermore, using the current/non-current indicator supplied by Centrelink, ANAO's analysis revealed that 4 856 563, or 94 per cent, of the 5 172 848 CRNs were described as current. Subsequent discussions between ANAO and Centrelink revealed that the current/non-current indicator stored in the DIE system was not a reliable indicator of whether the customer records supported a current benefit determination.[50]

3.37 A more reliable indicator of customer status, sourced directly from the ISIS Determinations File, and mapped against the DI line-by-line data provided by Centrelink, suggested that approximately 2.8 million, or 54 per cent, of the 5.1 million CRNs were associated with customers who held a current benefit determination.[51]

3.38 Therefore, using two different customer status indicators, ANAO's analysis suggested that between 54 per cent and 94 per cent of the 5.1 million customer records with DI errors, related to current Centrelink customers. In ANAO's opinion, neither of these figures is particularly reliable[52]—although the lower figure of 54 per cent is probably the more accurate. In any case, ANAO noted that the DI reporting system did not include a breakdown of DI errors by customer status.

---

48 A discrepancy of one in almost eight and a half million was considered negligible.

49 The issue of a unique CRN representing an individual customer is explored later in Chapter 5 of this report. For the present analysis, ANAO assumed that a unique CRN relates to one customer only.

50 Centrelink advised that, for some particular customer circumstances, if this indicator is turned on it is never turned off. Therefore, some customer records will always be marked as current, despite a benefit status having been cancelled by Centrelink.

51 Chapters 4 to 6 of this report present the results of ANAO's analysis of selected customer identity fields in the ISIS database, for all 23 million customer records. A more reliable indicator of customer status, drawn directly from the Determinations File in ISIS (rather than the Customer Circumstance File) was included in the data extracts used for that analysis. Those data extracts were provided by Centrelink on 13 September 2005. Mapping the current/non-current information from those September files against the DI data files provided in July, resulted in the removal of some customer records from the above analysis—those customers whose status had changed between July and September 2005. ANAO estimated this number to be approximately 319 000, or 6 per cent of customers included in the DIE files.

52 The higher figure is derived from a less reliable customer status indicator that is never switched off in many cases. The lower figure is derived from mapping a more reliable indicator from a September dataset against a July dataset. There was at least a 6 per cent difference between the customer populations in the July and September datasets.

Audit findings

3.39 ANAO found that the DI error reporting system did not adequately distinguish between DI errors associated with current customers and non-current customers. As such, the statistics and information presented through the DI error reporting system did not afford programme managers an insight into the significance of the various DI error counts.

Recommendation No.1

3.40 The ANAO recommends that Centrelink improve the usefulness and effectiveness of its data integrity (DI) reporting system by:

(a) ensuring the timely inclusion of new or revised DI checks whenever new software applications are released, so that the system is always checking data against current business rules; and

(b) enabling the system to clearly identify DI errors associated with current customers.

Centrelink's response

3.41 Agreed.

Priority rating scheme

3.42 Centrelink ascribes a priority rating to each error type—see Table 3.3 for a description of the five point scale in use at July 2005. In addition, the documentation for each error check—see Appendix 2—contains a description of the 'effect' of the error. The effect is defined as an 'indication of the impact of the error on processing'.

3.43 For example, a particular error type[53]—which checks on aspects of a parent-child data link—is classified as a Priority 1 error. Specifications for this error type describe the effect as 'Inconvenience to users, under/overpayments to customers'.

3.44 ANAO examined the alignment between the priority ratings and the descriptions of effect, contained in the specifications for a selection of DI errors. Based on the statistics contained in the DI error reporting system, as at 12 July 2005, ANAO chose: the top 25 Priority 1 errors; top 27 Priority 2 errors; top 25 Priority 3 errors; and top five Priority 4 and Priority 5 errors.

[53] This error type is represented by the code CDU023.

3.45 ANAO's examination revealed that the effect of the error was described as 'TBA' (to be advised) in the specifications for:

- five Priority 1 errors;
- 20 Priority 2 errors;
- 19 Priority 3 errors; and
- three Priority 4 errors.[54]

3.46 Collectively, 47 of 87 error definition tables examined, failed to provide any description of the likely effect of the error. This represents 54 per cent of the sample tested.

3.47 In addition, three of the top 27 Priority 2 errors variously described the effect of the particular error as:

- This may effect benefit payment;
- Incorrect ABY and/or FSL entitlement. Possible overpayment/ underpayment;[55] and
- Potential for under/overpayments to the customer, if the child is a split-custody child, or is a Centrelink customer in their own right (eg on Youth Allowance).[56]

3.48 The effects described above appear inconsistent with the description of a Priority 2 error—they relate more closely to the description of a Priority 1 error.[57] This is because the three descriptions indicate that the errors have the potential to compromise benefit payments.

3.49 ANAO also noted that between the period 3 May 2005 to 12 July 2005, six Priority 1 DI errors, in the FAO cluster were re-prioritised to Priority 4.[58] Centrelink advised the ANAO that each of these six errors related to DI checks on historical data and that they did not impact on a customer's ongoing payment rate. However, ANAO noted that, while the priority rating had changed from 1 to 4, the description of the effect of these six errors remained

---

54 ANAO conveyed to Centrelink a list of the relevant error codes.

55 ABY relates to checks for the ABSTUDY Entitlement Cluster. FSL relates to Financial Supplement Loan.

56 ANAO conveyed to Centrelink a list of the relevant error codes.

57 ANAO reviewed the description of effect for 25 Priority 1 errors, and concluded that those descriptions indicated either a definite over/underpayment, a potential over/underpayment, or receipt of a benefit where no entitlement existed.

58 ANAO conveyed to Centrelink a list of the relevant error codes.

unchanged.[59] Therefore, at 12 July 2005, the stated effect of the errors was inconsistent with the description of a Priority 4 error.

3.50 By definition, Priority 1 and Priority 2 errors represent the greatest risk to Centrelink's business and Government outlays. Priority 1 errors relate to a likely risk of incorrect benefit payments and Priority 2 errors result in adverse publicity and increased staff workload. ANAO's analysis of the figures in Table 3.3 reveals that, collectively, Priority 1 and Priority 2 errors account for 68 per cent of all DI errors. ANAO's analysis of Centrelink's line-by-line data showed that, together, Priority 1 and Priority 2 errors accounted for between 60.9 per cent and 62.7 per cent of errors associated with current Centrelink customers (depending on which customer status indicator is used).

Audit findings

3.51 ANAO found that the priority rating system employed by Centrelink provides a high-level breakdown of DI error statistics. However, with approximately two-thirds of all errors classified as Priority 1 or Priority 2, the system does not adequately discriminate between DI errors, nor does it overtly highlight those areas requiring immediate attention by programme managers.

3.52 ANAO also found that, with over 50 per cent of the top 87 error definition tables lacking any description of the effect of the error, programme managers are not presented with sufficient information to recognise the significance, or easily comprehend the likely impact, of particular DI problems.

Actions taken by Centrelink

3.53 In October 2005, Centrelink advised the ANAO that:

> Following on from the ANAO fieldwork, Centrelink has conducted an analysis and revived an earlier proposal to drastically change the way errors, their descriptions and their priorities are maintained. The improved infrastructure will give the applications teams (the people responsible for maintaining the DI errors) online access to do so through the Centrelink Repository.
>
> [also]
>
> Centrelink has reviewed the priority system and will introduce an updated priority allocation system consistent with problem management processes.[60]

---

[59] One error code states the effect as 'underpayment to customer(s)'. Another error code states the effect as 'potential for under/over payments to the customer'.

[60] Advice from Centrelink's Chief Information Officer, to the ANAO, on 12 October 2005.

Recommendation No.2

3.54 ANAO recommends that Centrelink, in order to provide programme managers with the capacity to determine the relevant priority of DI issues, including those requiring urgent or immediate attention, revise its priority rating system for DI errors, with a view to:

(a) comprehensively and accurately describing the likely effects of DI errors;

(b) resolving inconsistencies between the stated effects of some errors and the criteria for ascribing particular priority ratings; and

(c) clearly identifying DI errors that pose the greatest risk to the efficient and effective administration of programmes and payments.

3.55 This should facilitate the timely resolution of DI errors, in order of priority.

Centrelink's response

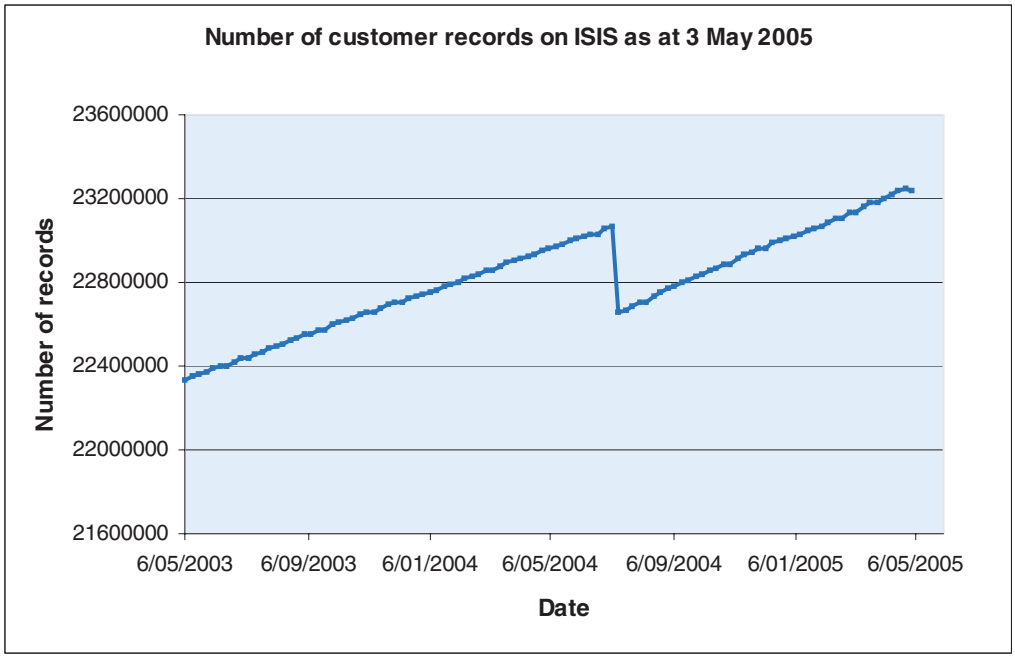
3.56 Agreed.

Distribution of DI errors across computing environments

3.57 Figure 3.3 shows that the number of customer records on ISIS has been increasing at a rate of 2.8 per cent, per annum, over the last two years. Centrelink informed ANAO that the decrease, which appears at July 2004, resulted from an archiving exercise, which removed approximately 412 000 records.

Figure 3.3

Number of customer records over time

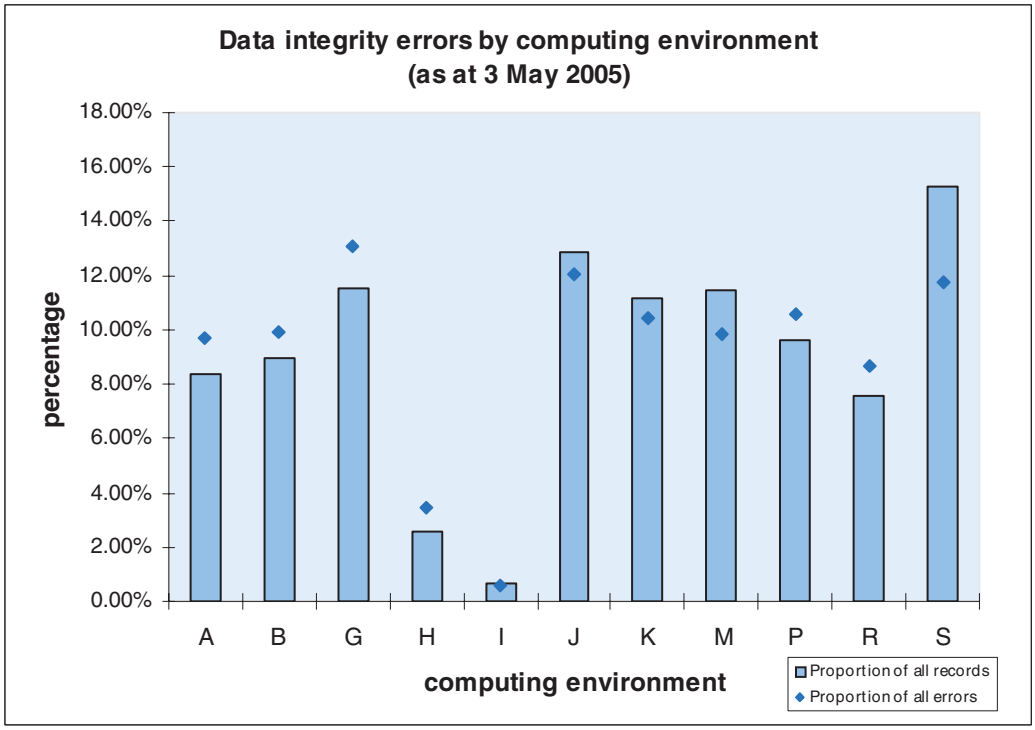


Source: ANAO – using data from Centrelink Intranet – DI error reporting system.

3.58 Figure 3.4 shows the proportion of customer records stored in each computing environment. It also shows the proportion of DI errors associated with each environment.

Figure 3.4

Distribution of customer records and DI errors across environments



Source: ANAO – using data from Centrelink Intranet – DI error reporting system.

3.59 Examination of Figure 3.4 reveals that DI errors are not uniformly distributed across environments. Some environments contain a greater proportion of DI errors, than would be expected, given the proportion of customer records held on those environments. For example, Environment A contains 8.4 per cent of all customer records, but 9.7 per cent of all DI errors.

3.60 Following an examination of the distribution of individual error types across environments, ANAO noted that:

- some error types occurred in one environment only;[61]

[61] ANAO noted that some record types, for example ORG, are all stored in the one environment.

- some error types occurred in significantly greater proportion in one or two environments—that is, the error type was appreciably over-represented in a particular environment, compared to its usual frequency; and
- the count for some error types appeared to be capped at a particular value. For example, ANAO noted that five different error types had the number 10 000 reported in each environment.

3.61 Centrelink confirmed that the reporting of certain errors was capped at a predetermined number—often 10 000. This was done in order to avoid overloading the error reporting system. However, if a count is capped at 10 000 within each environment, only those 10 000 instances of the error are included as part of the total error count.

Audit findings

3.62 ANAO found that capping individual error counts would lead to an inaccurate calculation of the total error count—underestimating the total number of DI errors by an unknown amount.

3.63 ANAO also noted that, as Centrelink's computing environments are essentially geographically based, an analysis of the distribution of particular error types across environments may assist Centrelink to identify quality control issues relevant to particular states or territories.

Trends over time

3.64 Using the historical information contained in the fourth level of the DI error reporting system, ANAO plotted a time series analysis for each of the 87 error types examined.[62] Most of the fourth level reports, for individual error types, contained historical data covering the previous two years. A sample of the graphs, illustrating various trends, is included at Appendix 3.

Audit findings

3.65 ANAO found that, of 87 DI error reports analysed:

- 64 indicated an increase in DI errors over the past two years;[63]
- 11 indicated that DI error counts had not changed significantly; and

[62] ANAO provided Centrelink with a copy of the 87 graphs.

[63] In the case of Priority 1 errors, 17 of the top 25 errors considered, showed an increase in the number of DI failures reported over that period.

- 12 indicated a decrease in the number of DI errors reported.

3.66 Within these reports, ANAO found that five DI error reports indicated a steady increase in DI error count, followed by a plateau. The pattern suggested that an error had been introduced into the ISIS system at some point in time, the numbers steadily increased until the source of the error was identified and fixed, but that the existing errors were left unresolved.

3.67 ANAO observed a number of patterns indicating some type of recurring, or cyclical event, which impacted on error counts. A small number of error types appeared to demonstrate a marked increase in reported DI errors each September. Another cyclical pattern produced an increase in error numbers every 10 or 12 weeks, but in the interim error numbers remained stable. ANAO observed this pattern in five error types.

3.68 ANAO considers that conducting time series analyses of DI error counts could assist Centrelink in identifying the potential cause of some error types, and in targeting data clean-up activities, following rectification of the cause of an error.

Use of DI error reporting by business areas

3.69 ANAO held discussions with staff from a selection of business areas within Centrelink, in order to develop a view as to how the DI error detection and reporting system was used by those business areas. ANAO met with representatives from the following business areas: Seniors, Carers and Means Test; Family and Child Care Services; Education and Training; and Employment Services.

Audit Findings

3.70 Each of these business areas demonstrated a set of procedures for responding to circumstances where a system error caused customers to receive incorrect payments. Most often, the procedures involved collecting intelligence from the Centrelink Network, prompted by customers contacting their Centrelink CSC to report a problem with their payments. Another source of information was the relevant business area Help Desk. Once a problem was identified, the appropriate Business System team set about identifying and addressing the source of the problem, as well as checking to see whether other customers in similar circumstances were, or were likely to be, affected.

3.71 Through discussions with Centrelink staff, ANAO formed the view that these procedures were prompted by, and responded to, errors within the IT applications environment, which caused problems with customers' payments. The clear focus of these activities was to rectify, as quickly as possible, problems with individual customer overpayments or underpayments.

3.72 However, ANAO found that these activities did not represent systems or procedures to actively address data integrity errors, identified through the corporate DI error system. Most business areas informed ANAO that they relied on the IT teams to ensure data quality was maintained. Few staff in the business areas consulted by ANAO, indicated an awareness of the DI error detection and reporting system.

3.73 This view is reinforced by ANAO's analysis of the historical DI error data. The majority of historical records show that each week, approximately the same number of DI errors are fixed as are newly created. However, a 'core' of DI errors is carried over from week to week, with little variation in numbers. ANAO found that, while the business areas' activities might promptly address some new errors, they have made little impact on the core of DI errors identified on ISIS.

Conclusion

3.74 Centrelink has developed an extensive data integrity error detection system, incorporating checks of structural integrity and checks against various programme business rules. This system has evolved over the past ten years and now extends to over 5 000 individual data integrity checks, supplying Centrelink with a wealth of information on instances of data integrity failure in its ISIS database.

3.75 Centrelink's Intranet-based DI error reporting system is currently the only DI management information system, easily accessible to Centrelink programme managers.[64] However, ANAO noted that the system was not widely used by programme managers—nor was the information systematically analysed to reveal trends or identify the cause of particular DI failures. The historical DI error statistics indicate that Centrelink has made

[64] As noted earlier, ANAO recognises that the system was introduced as a prototype and that it has not been substantially reviewed since its introduction in 2000.

relatively little progress in resolving its data integrity issues over the past two years.

3.76 Following our examination of Centrelink's DI error reporting system, ANAO concluded that the system incorporates a number of positive features, such as:

- a comprehensive approach to checking the customer information database for structural integrity and conformance with business rules;
- the production of detailed information about DI errors detected, including historical data, maintained for up to two years;
- a breakdown of DI error statistics by priority, cluster and group; and
- the capacity to drill-down to more detailed information about individual DI errors.

3.77 However, ANAO concluded that the DI error reporting system is deficient in that it:

- does not adequately distinguish between errors associated with customers who are currently in receipt of a benefit determination and those who are not;
- reports error counts that are not necessarily reliable—including an unknown number of false positive results;
- uses some obsolete error checking routines, which result from not always updating the checks following the introduction of new or revised software;
- inadequately defines the probable effect of many DI errors;
- uses a priority rating system, which exhibits a number of inconsistencies; and
- does not assist Centrelink programme managers to readily identify data integrity issues that might require urgent or immediate attention.

3.78 ANAO concluded that Centrelink's DI error detection and reporting system could be further developed, to provide programme managers with more valuable insight into data integrity issues that should be of concern and priority to Centrelink.

4. Testing Data Integrity

This Chapter presents the results of ANAO's analysis of Centrelink's ISIS database. It includes a description of the data and the methodology employed in the analysis, presents the results of specific data integrity tests and discusses the significance of ANAO's findings in relation to Centrelink's administration of a number of social security programmes.

Data description

4.1 In order to conduct the analysis of the integrity of electronic customer records, the ANAO requested Centrelink to provide extracts from the ISIS database. These extracts were to include information stored in a number of fields relating to customer circumstances and identification. ANAO requested the information listed below, for all records held on the ISIS database:[65]

- Centrelink Reference Number;
- record type—for example, person, organisation or child;
- name details—surname, first name, second name, title;
- date of birth;
- date of death (where recorded);
- sex;
- record status—for example, current or non-current;[66]
- address details, including suburb, State and postcode;
- telephone number (where recorded);
- marital status;

---

[65] That is, all 23.5 million records. ANAO requested that the only records excluded were those that Centrelink had logically deleted from the database. Customer records may be logically deleted for a number of different reasons. Logically deleting the record means that the record physically remains on the database, but that Centrelink CSOs and business applications software are unable to access the record. For all practical purposes, such records have been removed from the database.

[66] This is the ISIS field CUCS.ON.IND. The same field was provided with the DIE dataset discussed in Chapter three, as an indicator of customer status. Further discussion with Centrelink suggested that this field was not a reliable indicator of whether a customer was currently in payment or not. Centrelink advised ANAO that a more reliable indication of whether a customer record supported a current benefit determination would be obtained using the information stored in the Determinations File. ANAO agreed to use a separate data extract provided by Centrelink, drawn from the Determination File, to reliably identify those customer records that supported a current benefit determination. Also, see the note at the conclusion of Appendix 4.

- Tax File Number;
- duplicate record indicator (where recorded)[67];
- partner's CRN (where recorded);
- details of documents used to establish proof of customer's identity; and
- details of bank accounts to which benefits were paid (for customers currently in payment).

4.2 Centrelink scheduled the data extracts to run over the weekend of 10 and 11 September 2005, and on 13 September 2005, provided four sets of data extracts to the ANAO. The files contained data in a line-by-line format, organised by CRN, and were comprised of:

- 11 files (one from each computing environment) containing the majority of customer circumstance data listed above—name, address, sex etc.;
- 11 files (one from each computing environment) containing information on proof of identity (POI) documents presented by customers;
- 11 files (one from each computing environment) containing the details of bank accounts into which Centrelink makes payments; and
- 11 files (one from each computing environment) containing information, from the Determinations File, on customers with a current determination for any Centrelink benefit or service.

4.3 The first set of 11 files—customer circumstance data—contained a total of 23 699 220 records. The second set of 11 files—POI data—contained a total of 12 742 853 records. The third set of 11 files—bank account data—contained a total of 6 167 308 records, and the fourth set of 11 files—current determinations data—contained a total of 6 168 030 records. Appendix 4 describes the four datasets in greater detail.

[67] Centrelink is aware of the existence of some duplicate CRNs and has marked these customer records with a duplicate record indicator. When calling up such records, CSOs are alerted to the fact that a duplicate records exists.

Methodology

4.4 The criteria for this audit included an assessment of whether:

- selected mandatory fields in Centrelink's databases contain valid data; and
- Centrelink's electronic records are accurate and complete.

4.5 The audit also considered whether the customer information held in ISIS is sufficiently accurate and comprehensive for Centrelink to rely on, in order to determine customers' eligibility and entitlements. The ANAO assessed the fitness of the information in ISIS by testing a sample of mandatory fields for valid format and content.

4.6 The various fields incorporated into ISIS are defined in a central data dictionary—the Department of Social Security Data Dictionary, or DSSDD, also known as the Data Repository. Centrelink describes the DSSDD as:

> The DSSDD is a tool which holds M204 metadata, that is, data about data. This tool is central to Centrelink's application development environment.
>
> The DSSDD holds the database design of all the M204 files, as well as screen designs and source code.[68]

4.7 The DSSDD describes specifications for particular data fields, including the type of data to be stored in the fields—numeric, date, character, etc.—the field length and, often, a list of legal values associated with a field.[69] ANAO compared the values in the data extracts provided by Centrelink with the legal values defined in the DSSDD. In addition, ANAO conducted a number of logical consistency checks, including an analysis of customers' dates of birth and death.

4.8 All records on ISIS are assigned a record type. In the full dataset provided by Centrelink, ANAO identified nine different record types. These are described in Table 4.1.

---

[68] Centrelink, *Introduction to Business Data Definitions*, 1999, Centrelink's Intranet, [Accessed 6 April 2005].

[69] The legal values are stored in a 'look-up table' or, as known in the M204 environment, a CODE table or INCORE table.

Table 4.1

Record types

OBJECT.TYPE.CODE	Description	Number of records
AOS	Assurance of Support	31 332
CAS	Customer Appointment System	764
CHI	Child	5 193 431
ORG	Organisation	283 555
PER	Person	17 960 774
PRF	Profiling Reference	1
PRU	FICMS Case/Operation	222 458
UOE	Unauthenticated Object Entry	5 148
WLM	Workload Management	1 757
	Total	23 699 220

Source: ANAO's analysis of Centrelink dataset – 13 September 2005.

4.9 The most common type of customer record stored on ISIS is that relating to a person. These records, along with those for organisations, are those one would normally think of as customer records.

4.10 The other record types are used for different purposes. The Customer Appointment System permits CSOs to book appointment times for prospective customers. Centrelink informed ANAO that child records are required where a link between a parent and a child would cross environments.[70] The unauthenticated object entries relate to potential customers who have lodged an intention to claim via the Internet. These are essentially temporary records, awaiting authentication, when the record type will most probably be changed to PER.

4.11 Record types other than PER and ORG are essentially shell records—they contain minimal customer data. Consequently, for much of its analysis, ANAO excluded all record types other than PER and ORG, and for some analysis, concentrated on PER records only.

4.12 Following our analysis, ANAO provided Centrelink with 32 electronic files containing details of records identified as anomalous. These files were

[70] Centrelink informed ANAO that the details of a customer's children are normally stored as circumstance data for the customer and a parent-child link is established, which connects the necessary information. These links may only exist within a single computing environment. In cases where details of the parents/partners and their children are stored on different environments, a CHI record is created in one of those environments and the necessary cross-environment links established. CHI records contain very little data.

provided in order to enable Centrelink to further investigate particular anomalies and to target corrective action for particular records.

Results of field level analyses

Domain integrity checks

4.13 Some fields in the ISIS database are mandatory—they must contain information and not be left blank. ANAO examined a selection of mandatory fields associated with customer circumstance data.[71] These included customers':

- CRN;
- record type;
- name;
- sex;
- marital status; and
- date of birth.

CRN

4.14 ANAO examined the field that holds the customer's CRN. This field is described in the DSSDD as:

> This is the unique identifier for all Centrelink customers (people and orgs).[72]

4.15 ANAO found that all 23 699 220 records in the CRN field conformed to a valid format—a nine-digit number, with a 10th check digit (stored as a character). There were no records containing a blank entry in the CRN field. Chapter 6 of this report explores measures of the integrity of the CRN as a primary key—considering matters such as duplicate use of CRNs and customers with more than one CRN.

Record type

4.16 ANAO examined the record type field and found that all 23 699 220 records contained a valid three-character entry in that field. Table 4.1 provides

[71] Particular fields are defined as mandatory for particular service reasons (or programmes). For example, a customer's marital status is mandatory for age pension, parenting payments and many other programmes that involve payments at either a married or single rate. ANAO used Centrelink's document titled *Service Reasons Codes,* to select a number of fields that were identified as mandatory for the majority of service reasons.

[72] Centrelink, *DSSDD, Element: CNTRLNK.REF.NUM,* Centrelink Intranet, [Accessed on 17 August 2005].

further information on the number of each record type included in Centrelink's dataset.

4.17 ANAO's analysis identified three CHI records that differed significantly from all other CHI records. These three records displayed information similar to that held in PER records. [Centrelink later confirmed that these were, nevertheless, child records.]

Customer's name

4.18 ANAO examined the field that holds the customer's surname. As an unrestricted text field, CSOs are free to enter any computer keyboard characters into this field. ANAO found three blank entries in the surname field of PER records.

4.19 ANAO also found at least 10 000 entries, which do not appear to be genuine, or valid, surnames. These records included such entries as XXXX, YYY, ZZZZ, 'on a selection panel' and a string of location names—for example, West Hobart, West Ryde, Tamworth, Lismore, Canberra, Pipers River—on sequential, or nearly sequential CRNs. ANAO concluded that these entries may represent dummy records created while training CSOs.

4.20 ANAO considers that training records should not exist within the production environment of ISIS. While such records may not be associated with the payment of benefits at this point in time, their existence in the production environment represents an unnecessary contamination of customer data. The records also contribute to the total number of records on ISIS, leading to an inaccurate record count.

4.21 In relation to the recording of customers' given name, ANAO found that 3 228 records contained a blank entry in this field. A person may be known by only one name, and therefore, an entry in the given name field is not mandatory. However, upon further investigation of these 3 228 records ANAO found that many records appeared to contain both a person's surname and first name in the surname field, leaving the first name field blank.

4.22 ANAO also identified a number of spurious entries such as: Unknown; DoNotUse; Testing; and Duplicate Record. In addition, ANAO noted that some entries appeared to contain the names of organisations, although the record type was PER. Other entries for a customer's first name included: ####; 20574; -; ZZZZ; 1AN (ie. Ian, with a numeral 1 replacing the I); and ;YNETTE (ie. LYNETTE with a ; replacing the L). Other entries include zeros rather than the letter 'O' in customer names. These entries constitute illegal values.

4.23 While some of these spurious entries indicate a data entry error on the part of a Centrelink CSO, others are indicative of training records or workarounds—such as the DO NOT USE, or DUPLICATE RECORD entries.

4.24 The ISIS database defines a set of legal values for a customer's title—the preferred courtesy title or title of rank. The system should ensure that only legal values may be entered into the title field. ANAO noted that 123 different legal values for title were included in the look-up table. However, ANAO's analysis revealed that the title field, in the current customer dataset provided by Centrelink, held 231 different values, 193 of which were not included in the table of legal values.

4.25 Within the larger dataset of all PER customer records (that is, current and non-current), ANAO identified 794 different values in the title field, 756 of which were not included in the table of legal values.[73]

4.26 ANAO noted considerable inconsistency in the use of particular values for a customer's title. Table 4.2 presents some examples.

[73] The figures above show that only 38 of the 123 valid values for title, appear to be currently in use. (794 – 756 = 38; 231 – 193 = 38).

Table 4.2

Examples of inconsistent use of values in the field 'Title' in current customer records

Legal value and (defined code)	Codes detected in ISIS	Number of records
Brother (BR)	BR	332
	BRO	336
	BROTH	2
	BROTHR	1
	BRTHER	2
	BRTHR	1
MRS (MRS)	MRS	2 686 871
	MSR	18
	MRZ	3
	MRW	7
	MRSS	344
	MRSSSS	1
	MRSQ	2
	MR S	1
	MMRS	5

Source: ANAO's analysis of Centrelink data – 13 September 2005.

4.27 In total, 13 736 of the PER records on ISIS contained illegal values in the title field—4 332 of these relating to current customers. ANAO concluded that the existence of illegal values in the title field of these records strongly suggests either a lack of, or failure of, the system level control associated with enforcing legal values in this field.

Customer's sex

4.28 The ISIS database defines three legal values for the field recording the customer's sex—these are male, female and unknown. ANAO's analysis confirmed that the data set provided by Centrelink contained only legal values for the customer's sex, or the field was left blank. Sex applies primarily to PER records. Normally, CHI, ORG and other record types do not record sex, and therefore, the field is left blank. However, ANAO identified three PER records where the sex field was blank and three CHI records (out of over five million) where a sex was recorded.

Table 4.3

Customer's Sex

Sex	Number of PER records	Number of current records
MALE	8 639 944	2 171 604
FEMALE	9 302 197	3 995 563
UNKNOWN	18 630	0
BLANK	3	0
Total	17 960 774	6 167 167

Source: ANAO's analysis of Centrelink data – 13 September 2005.

4.29 Within the 17 960 774 PER records tested, 18 630 indicated the person's sex as unknown. This represents 0.1 per cent of all PER records, although it does not include any current records. Although ANAO did not conduct an exhaustive analysis of these 18 630 records, many appeared to be associated with records containing spurious name elements, discussed earlier.

Marital status

4.30 The ISIS database defines eight legal values for a customer's marital status. These are: married; divorced; de facto; widowed; single; separated; not required; and unknown. ANAO analysed data stored in this field for all PER records in the ISIS database. Table 4.4 presents the results of this analysis. The second column of the table presents a count of all PER records on ISIS, while the third column presents figures for current customers only.

Table 4.4

Marital Status Code

Marital Status Code	Number of PER records	Number of current records
(BLANK)	81	0
DEFACTO	868 204	367 629
DIVORCED	459 456	229 999
MARRIED	7 195 204	2 767 426
NOT REQUIRED	506 285	15 462
SEPARATED	1 859 115	769 907
SINGLE	5 405 229	1 412 111
UNKNOWN	251 922	55 176
WIDOWED	1 347 239	549 454
'_' (UNDERSCORE CHARACTER)	68 039	3
Total	17 960 774	6 167 167

Source: ANAO's analysis of Centrelink data – 13 September 2005.

4.31 Neither an underscore character nor a blank entry is included in the list of legal values defined in the DSSDD. While these entries account for only 0.38 per cent of all PER records, they represent an unnecessary contamination of data in the marital status field. In relation to records for current customers, the number of blank and underscore entries is negligible. However, 55 176 current customer records show a marital status code of unknown. This represents 0.9 per cent of current records.

4.32 ANAO identified the following current benefit determinations, each of which involves payment at a married rate or a single rate, associated with 34 857 of these 55 176 records:

- 32 134 Youth Allowance;
- 2 400 Newstart Allowance;
- 322 Austudy; and
- one Parenting Payment Single.

4.33 Many benefit types require a marital status code in order to determine an appropriate rate of payment—some benefit types do not. While still a legal value, 'unknown' may not necessarily be equated with 'not required'. 'Unknown' does not provide useful information for a determination which requires a marital status code.

4.34 After receiving ANAO's initial findings on this matter, Centrelink advised ANAO that:

> .. the unknown marital status code is added when a new customer contacts Centrelink (usually over the phone) with the intention of lodging a new claim. The claim action records some customer details and an unknown marital status is automatically inserted as at the time, the marital status is not known. When the customer lodges a formal claim the correct marital status code is then manually inserted on the customer's record. System error checks are in place, or in the Youth Allowance cases, the customer is paid at the single rate.
>
> Centrelink acknowledges that unknown is not a valid code after the customer has lodged a claim and will investigate corrective action. The business process of recording an unknown status will also be reviewed.
>
> The Parenting Payment Single customer is paid at a manual rate [special legislative provisions apply in this case due to the customer's age], so the marital status code does not impact on that customer's entitlement.[74]

[74] Advice from Centrelink's Chief Information Officer, to the ANAO, on 18 November 2005.

Date of birth and date of death

4.35 ANAO checked the validity of dates stored in the date of birth and, where recorded, date of death fields for all customers.[75] ANAO found that, where a date of birth or date of death was recorded, only valid dates appeared in the dataset.[76]

4.36 ANAO's analysis of all PER records showed that:

- 27 159 records did not contain a date of birth;
- 1 461 533 records contained a date of death;
- the earliest recorded date of birth was 19 July 1873—that record did not have a date of death recorded, nor did the record show a current benefit determination;
- the latest recorded date of birth was 7 September 2005—that record did not show a current benefit determination;
- 33 customers had a date of death, but no date of birth recorded;
- 42 records had the same date recorded for both the customer's date of birth and date of death; and
- one record had a recorded date of birth two months after the recorded date of death.

4.37 ANAO calculated the age of all customers with a recorded date of birth prior to 1 January 1900 and no recorded date of death. Therefore, according to the data in ISIS, these customers are still living—the oldest would now be 132 years of age. Table 4.5 shows the distribution of customers in this category, according to the recorded dates. Figures in the second column are derived from the set of all PER records, while figures in the third column refer to PER records with a current benefit determination.

[75] Examples of invalid dates include: 29 February in a non-leap-year; 30 or 31 February in any year; and the 31st of any month that only contains 30 days.

[76] Only PER records were included in this analysis. ORG and CHI records do not require a date of birth or date of death.

Table 4.5

Distribution of Centrelink customers over 105 years of age, with no date of death recorded.

Age (group)	Number of customers	Number of CURRENT customers	Benefit types
> 108 years	1 239	7	3 x AGE[77], RCA[78]; 2 x RCA, 2 x AGE
108 years	377	9	2 x AGE; 5 x AGE, RCA; 1 x SPL[79]; 1 x RCA
107 years	596	21	12 x AGE, RCA; 4 x AGE; 5 x RCA
106 years	2 615	55	17 x AGE, RCA; 11 x AGE; 5 x RCA; 3 x CCF; 19 x CDA[80]
105 years	962	56	19 x AGE; 26 x AGE, RCA; 10 x RCA; 1 x WID[81]

Source: ANAO's analysis of Centrelink data – 13 September 2005.

4.38 After receiving ANAO's initial findings on this matter, and undertaking supplementary inquiries in relation to the outlying cases, Centrelink advised ANAO that all seven (current) customers, aged greater than 108 years, were still alive and resident in Australia.[82] Centrelink also confirmed that it obtains fact of death information from the Registrars of Births, Deaths and Marriages, and uses this information to update its database.

4.39 ANAO found that the production environment of the ISIS database contains records for almost one and a half million customers, where those records display a date of death—sometimes many decades in the past. ANAO considers there is little reason to maintain such records in the production environment.

4.40 Centrelink informed ANAO that some records for deceased customers need to be kept, as they may attach to an ongoing debt to the Commonwealth,

---

77 AGE = Aged Pension.

78 RCA = Residential Care Allowance.

79 SPL = Special Benefit.

80 CCF = Child Care Benefit for Approved care, CDA = Carer Allowance.

81 WID = Widow Pension.

82 Centrelink advised ANAO that, in five of the seven cases, Centrelink officers telephoned the various nursing homes where the customers lived, and confirmed that each was still a resident. In the remaining two cases, Centrelink provided confirmation of the customers' age and resident status based on the results of a 2004 review, which included a home visit. Centrelink also advised the ANAO that it had confirmed the accuracy of dates of birth recorded on the paper-based files for the seven customers.

or they are associated with 'partner records' for current customers. ANAO accepts that such business reasons may necessarily involve the retention of the records of certain deceased customers within the production environment. However, ANAO questions the need to maintain, in the production environment, records for long deceased customers where no such business reason for their retention exists.

Dummy dates of birth

4.41 Centrelink informed ANAO that, where a date of birth is unknown, or not known with certainty, a false, or 'dummy', date of birth is often entered. The value most often used within Centrelink, in these circumstances, is 1 January. Centrelink also informed ANAO that the years most often used in dummy dates of birth were 1900 and 1901, although a dummy date of birth could be recorded as 1 January in any year. ANAO analysed the day and month of the recorded date of birth, for all current customers and graphed the frequency of each day/month combination. Appendix 5 presents the results of the analysis. The graph shows that, within the current customer dataset, the average number of people born on any given day of the year is approximately 16 800.[83]

4.42 However, the graph also shows that the frequency for customers born on 1 January, in any given year, is 2.26 times greater than the average, at almost 38 000. The graph also shows that 1 July is probably used as a dummy date for recording the date of birth of some customers. According to the graph, the frequency of people born on 1 July, in any given year, is 1.6 times the average figure, at just over 27 000.

4.43 By confining the use of a dummy year of birth to 1900 or 1901, ANAO considers that any age profile analysis, conducted by Centrelink on this data, will be inaccurate. This would especially be the case for an age profile analysis of Centrelink's older customers—say those over 90 years of age—as the 1900/01 year of birth indicates these customers would be 104–105 years old.[84]

4.44 Based on our analysis, ANAO concluded that approximately 0.5 per cent of recorded dates of birth, for current Centrelink customers, are inaccurate to some extent.

---

[83] Figures for the three outlier data points—1 January, 29 February and 1 July—have been excluded for the purposes of calculating an average value.

[84] The use of 1 January 1900 and 1901 accounts for only a small proportion of the total number of records using 1 January as the day and month of birth.

Other fields

4.45 A large proportion of benefit types require an address to be recorded as part of customer circumstance data. Incorporated in a customer's address is the state or territory in which the customer lives. Legal values for the state or territory are defined in a look-up table within the DSSDD.[85]

4.46 ANAO identified 571 668 PER records that did not contain an entry in the state field. The majority of these records appeared to relate to customers with an overseas address.[86] ANAO's analysis revealed that, for those records where an entry existed within the state field,[87] 10 290 records contained illegal values. Of these, 10 191 records contained the entry 'IOB', which relates to customers living overseas. However, ANAO noted that the relevant DSSDD entry stated:

> State Code
>
> This element holds a code that identifies an Australian state or territory. It is intended to be used for addresses.
>
> Do not use when DSS Environment is intended. The codes for IOB and the Sydney split environments are not included.

> [On 18 November 2005, Centrelink advised ANAO that:] Current processing for customers moving overseas is, if a country is recorded that is not Australia, then the state field is to be blank. Previously, overseas customers were identified by using the IOB coding in state. Hence, a number of historical records with a state code of IOB.[88]

4.47 ANAO noted that the other 99 illegal entries in the state field held values such as: S; BN2, HIL; Q; TEX; B'K; YUG and TYN. Fifty-six of these records contained the entry EXT. These often appeared to relate to elements of an overseas address. In addition, ANAO noted that many of the remaining records that contained an illegal value for state, appeared to have the customer's suburb, state and postcode all stored in the field called 'address line 2'. The suburb, state and postcode then appeared to be repeated (in part or in

---

[85] Legal values are: ACT; NSW; NT; QLD; SA; TAS; VIC; and WA.

[86] ANAO did not request Centrelink to provide details of customer's overseas addresses in the data extracts. Overseas address details are stored in different ISIS fields.

[87] CHI records do not include an address. The figures above relate to PER and ORG records only.

[88] Advice from Centrelink's Chief Information Officer, to the ANAO, on 18 November 2005.

full)[89] across the next three fields of data. Some examples of addresses that appear to be corrupted are presented in Table 4.6 below.

Table 4.6

Example of possibly corrupt customer address data

Address line 1	Address line 2	Suburb	State	Postcode
PO BOX [zzz]	WAGGA WAGGA NSW 2650	WAGGA	WAG	NSW
PO BOX [zzz]	SUMMER HILL NSW 2130	SUMMER	HIL	NSW
P O BOX [zzz]	RED HILL QLD 4059	RED	HIL	QLD
1 SMITH ST	ST KILDA VIC 3182	ST	KIL	VIC

Source: ANAO analysis of Centrelink dataset – 13 September 2005.

4.48 After receiving ANAO's initial findings on this matter, Centrelink advised that:

> Some of the issues raised by the ANAO were a result of adding data from external agencies. Centrelink has since made technological improvements in data loading, which has improved the quality of data loaded into the system.[90]

Tax File Number

4.49 For most benefit types, Centrelink requires the customer's Tax File Number (TFN). This is stored on ISIS as an encrypted alphanumeric value.[91] ANAO noted, in Audit Report No. 37 1998–99[92] and Audit Report No. 47 2004–05,[93] that the phrase 'Tax File Number' has a specific legislative meaning. It is a number that is issued to a person by the Commissioner of Taxation.[94] The original and main purpose of the TFN was to be a numeric, unique identifier of clients of the Australian Taxation Office. The TFN is also used by other government agencies when there is a legislative need to verify client identity and establish income levels.[95]

---

89 Depending on the defined length of those fields. For example, the state field accepts the first three characters only.

90 Advice from Centrelink's Chief Information Officer, to the ANAO, on 18 November 2005.

91 CSOs or other Centrelink staff viewing customer records do not, therefore, have access to the customer's actual TFN, yet the IT systems can decrypt and encrypt the TFN for use in data matching or other activities.

92 1999, ANAO, Audit Report No. 37 1998–99, *Management of Tax File Numbers.*

93 2005, ANAO, Audit Report No. 47 2004–05, *Tax File Number Integrity*.

94 Section 202A of the *Income Assessment Act 1936.*

95 1999, ANAO, Audit Report No. 37 1998–99, *Management of Tax File Numbers.* p.27-28.

2005, ANAO, Audit Report No. 47 2004–05, *Tax File Number Integrity*. p.21.

4.50 While many Centrelink publications contain references to customers providing a TFN, the following is typical of the information available to Centrelink customers.

> Centrelink asks people claiming or receiving a payment from Centrelink to provide a Tax File Number (and the Tax File Number of their partner, or parent(s), where applicable). This helps Centrelink check the information people give with information already held by the Australian Taxation Office and some other departments that pay benefits. If a person does not provide a Tax File Number they may not get any payment.[96]

4.51 ANAO analysed the encrypted TFNs provided by Centrelink, as part of the data extracted on 13 September 2005. Of the 17 960 774 PER records provided by Centrelink, 13 777 993 contained TFNs. This represents 76.7 per cent of all PER records.

4.52 ANAO's analysis revealed that 786 092 lines of data indicated the same TFN was associated with two or more CRNs.[97] The fact that TFNs are shared across records with different CRNs, represents a significant weakness in the data integrity of these customer records.

4.53 Table 4.7 shows the distribution of TFNs associated with multiple CRNs.

Table 4.7

TFNs associated with more than one CRN

No. of CRNs associated with a single TFN	Number of TFNs in category
24	1
16	2
11	1
8	1
6	2
5	17
4	342
3	10 696
2	376 232

Source: ANAO analysis of Centrelink dataset – 13 September 2005.

[96] 2004, Centrelink, Centrelink Information — A guide to payments and services 2004–05, pp. 156.

[97] Each CRN on the ISIS database should be associated with a single customer and, consequently, should be associated with a single TFN.

4.54 Where a single TFN was common to two CRNs, ANAO noted that many pairs of records appeared to relate to the same person. The records matched on most elements of the customer's name, date of birth and, in many cases, address details. The matter of customers registered on ISIS under two or more CRNs is discussed in detail in the Chapter 6 of this report, under the sections dealing with the integrity of the primary key and duplicate CRNs.

4.55 In other cases, where a TFN was common to two or more CRNs, ANAO noted that the identity of the customers appeared to be different—either two entirely different people, or members of the same family. For example, ANAO identified approximately 6 756 records—or 3 378 pairs of records sharing a single TFN—which appeared to relate to married couples, parent-child pairs, and sibling pairs.

4.56 ANAO noted 26 680 records—or 13 340 pairs of records sharing a single TFN—which matched exactly on the customers surname, TFN, and address, but not on the customers' first name. Many of these pairs proved to be either a duplicate record for the same person, with their first name spelt differently across the two records, or different family members.

Centrelink's advice

4.57 ANAO provided Centrelink with details of records identified as sharing TFNs. In November 2005, Centrelink advised ANAO that:

> Centrelink has existing controls in place for the TFN through Accelerated Claimant Matching that identifies current records where more than one person uses the same TFN. A sample of the file provided by the ANAO was checked and this showed that the ANAO had identified TFNs associated with 'shell' or 'provisionally deleted' records. These records are not active and do not have a benefit status.
>
> Centrelink has other TFN controls in place that check TFNs after the customer has claimed a payment.[98]
>
> [Centrelink invited ANAO to review the figures in Table 4.7 in light of the customers' status. That is, perform the analysis on current records only.]

Further analysis of Tax File Numbers for current customers

4.58 ANAO examined the 786 092 lines of data used to produce Table 4.7, and found that 184 257 lines of data, or 23.4 per cent, were associated with a current benefit determination. Many of these cases involve a TFN shared

[98] Advice from Centrelink's Chief Information Officer, to the ANAO, on 18 November 2005.

across a current and a non-current record. However, ANAO found that 2 643 lines of data were associated with shared TFNs, where more than one CRN supported a current benefit determination.[99]

4.59 In addition, ANAO found that:

- 345 pairs of records shared a TFN where both records were current for an Age Pension benefit;
- 21 pairs of records shared a TFN where both records were current for a Disability Support Pension benefit;
- 18 pairs of records shared a TFN where both records were current for a Family Tax Benefit; and
- 6 pairs of records shared a TFN where both records were current for a Parenting Payment benefit.

4.60 ANAO provided Centrelink with details of these cases.

Audit findings

4.61 After conducting a range of analyses on the domain integrity of selected fields associated with customer's personal details, ANAO found that several ISIS fields were contaminated with illegal or nonsensical data indicating a lack, or failure, of system level controls, which should enforce conformance with a set of legally defined values. ANAO also found that the field storing customers' Tax File Numbers was compromised, in that TFN entries were not unique. ANAO also found that a number of spurious records, which appear to be training records, exist within the production environment of ISIS.

Analysis of date of death — current customers

4.62 ANAO analysed Centrelink's dataset of customers with a current benefit determination, specifically examining recorded dates of death for customers. Of the 6 167 308 records examined, 617 customer records had a date of death recorded. ANAO noted that some benefit types may validly continue payment for a short period of time, following a customer's death. ANAO used an arbitrary period of grace of six weeks to distinguish between records that

[99] ANAO identified three cases where a TFN was shared by three current records, and 1 317 cases where a TFN was shared by two current records.

could legitimately be included in the current-in-payment file, and those worthy of further consideration.

4.63 ANAO identified 446 records with a recorded date of death within six weeks of the data extraction by Centrelink—13 September 2005—and 171 records where the date of death preceded 1 August 2005. Table 4.8 presents details of these 171 records.

Table 4.8

Customers with a recorded date of death prior to 1 August 2005, with a current benefit determination, as at 13 September 2005.

Benefit type	Number of people in each category
AGE — Age Pension	3
CCF — Child Care Benefit for approved care	1
DSP — Disability Support Pension	1
EIC — Assistance for Isolated Children (pre-2000) student	1
FTB — Family Tax Benefit	48
JSR — Jobseeker	57
LIC — Low Income Health Care Card	51
PPS — Parenting Payment-Single	1
RCA — Residential Care Assistance	7
SKA — Sickness Allowance	1
Total	171

Source: ANAO's analysis of Centrelink dataset — 13 September 2005.

4.64 ANAO noted that the benefit types Jobseeker[100] and Low Income Health Care Card[101] do not attract payments, while the other benefit types are associated with payments. Therefore, according to the data provided by Centrelink, 63 customers who had a recorded date of death prior to 1 August 2005, were paid a benefit on 13 September 2005.

4.65 However, further discussions between ANAO and Centrelink revealed that some of these records had only been 'provisionally deleted' from the

[100] A Jobseeker registration provides the customer with access to the Australian JobSearch network, which is managed by the Commonwealth Department of Employment and Workplace Relations. <Internet address: www.jobsearch.gov.au, accessed 28 October 2005>.

[101] A Low Income Health Care Card provides the customer with access to Pharmaceutical Benefits (PBS) medications at a concessional rate and a lower threshold for the PBS and Medicare Safety Nets. Doctors may also claim incentive payments for bulk-billing LIC holders.

database. Therefore, payment had ceased on these records, yet they continued to display a current benefit determination. Centrelink advised the ANAO that:

> Initial investigation shows that a sample of customers in receipt of payment listed in the [ANAO] report were correctly cancelled as a result of their death.
>
> Investigation showed that there is a problem with the file Centrelink provided to the ANAO with some customers incorrectly reporting a current benefit status when the record was cancelled.[102]

Audit findings

4.66 ANAO found that Centrelink had procedures in place to cancel payments, when informed of a customer's death. However, a small number of records for deceased customers appeared not to have been fully processed.

Recommendation No.3

4.67 ANAO recommends that, in order to address the range of data quality issues identified by this audit, Centrelink conducts a thorough data cleansing exercise within the ISIS database, with a view to:

(a) removing training records and spurious customer records from the production environment;

(b) removing or otherwise inactivating records for deceased customers from the production environment, where there is no continuing business need to retain the records;

(c) improving the accuracy of customers' personal information, particularly in recording the various elements of customers' name and address;

(d) enforcing existing business rules surrounding the use of defined legal values with certain ISIS fields;

(e) resolving possible anomalies in the recorded dates of birth and death for Centrelink customers identified during this audit; and

(f) resolving possible anomalies in the recorded Tax File Numbers for Centrelink customers identified during this audit.

Centrelink's response

4.68 Agreed.

---

[102] Advice from Centrelink's Chief Information Officer, to the ANAO, on 18 November 2005.

5. Recording Customer Identity

This Chapter describes ANAO's analysis of data fields holding information on the documents used by Centrelink customers to establish their identity.

Proof of customer identity

5.1 New Centrelink customers, and existing customers in certain circumstances—for example, customers claiming a new benefit, who were customers before July 1995[103]—must provide sufficient proof of their identity as part of the process to establish their eligibility for a benefit. Appendix 6 illustrates a standard Centrelink form, identified as Form SS231, which outlines the POI requirements. Centrelink employs a system of four levels, or tiers, of POI—depending on the particular benefit sought.

5.2 The four tiers of POI are:

- Tier 0 – no POI required;
- Tier 1 – Proof of birth/arrival in Australia OR approved documents to the value of 50 points (see Appendix 6 for a description of the point value of various documents);
- Tier 2 - Proof of birth/arrival in Australia AND approved documents to the value of 50 points; and
- Tier 3 - Proof of birth/arrival in Australia AND approved documents to the value of 100 points.

5.3 Tier 0—no POI required—is applicable to the following programmes: Family Tax Benefit Part A; Family Tax Benefit Part B; Child Care Benefit; Maternity Allowance; Maternity Immunisation Allowance; Residential Care Assistance; and Assistance for Isolated Children.[104] As can be seen from an examination of Appendix 6, most pensions, allowances and assistance require POI at the Tier 3 level.

5.4 Information from the various documents provided by customers in meeting the POI requirements is entered into the customers' records on ISIS.

---

[103] Revised proof of identity procedures were introduced in July 1995.

[104] For most of these programmes, alternative methods of determining identity are available. For example, in the case of family benefits, a certification of the birth of a child is provided by a hospital, which identifies the mother of the child for benefit purposes. This is not an exhaustive list. Further information on programmes that do not require specific POI is available at Appendix 6 and from Centrelink.

Some of the controls surrounding the recording of this information are discussed in the following sections.

Controls associated with proof of identity

5.5 Included in Centrelink's documented procedures for processing new claims, are instructions for CSOs on the examination and retention of documents, accepted by Centrelink to verify the customer's identity. The procedures state:

> This is a Centrelink Must Do. The instructions below must be followed exactly as written. Staff cannot use any discretion when applying this law, policy or procedure, unless clearly stated otherwise.
>
> This procedure outlines the requirements for sighting, photocopying and certifying POI documents provided by customers.[105]

5.6 Centrelink requires that original documents must be provided for POI purposes. Certified copies may initially be provided in some cases, however, the originals must be provided later—usually within four weeks. Generally Centrelink does not retain original documents. The procedures state that original documents must be copied and certified, by the CSO, as a copy of the original. Documents certified must include a notation by the CSO to the effect that: 'Originals Sighted and Returned' or 'OS&R'.

5.7 The procedures include alternatives for customers without sufficient POI documents and for Aboriginal and Torres Straight Islander people. However, in all cases the CSO must be convinced of the customer's identity.

5.8 The ISIS database records details of the documents, or other means, used by customers to establish their identity. POI information is essentially stored in the database in field pairs—one describing the type of document presented by the customer and another holding the serial number or other identifying number or mark relating to that document. Centrelink has defined a set of legal values for the database field describing the type of document presented.

5.9 ISIS permits information on up to eleven POI references to be stored for each customer. Each line of the data file, provided by Centrelink to ANAO, contained a CRN, environment indicator and 11 fields holding POI details.[106] In

---

[105] Source: Centrelink, *Getting it Right*, *Chapter 2, Initial contact / Sighting, copying, certifying and returning original documents – Overview*, Centrelink Intranet.

[106] Very few records actually contain 11 POI entries. Most often, the customer's record contains three to five POI entries. Where less than 11 POI documents exist for a customer, the remaining fields are left blank.

total, 12 742 853 lines of data were provided by Centrelink. In 4 451 672 cases the entry read:

> NO DOCUMENTS ON RECORD – MAY BE ALT POI OR CONVERSION CUSTOMER WHERE POI ON FILE.

5.10 Therefore, 8 291 181 useable records were included in ANAO's analysis. Of these 4 078 415, or 49.2 per cent, related to customers with a current benefit determination.

5.11 POI documents for which legal values are prescribed include: original Australian birth certificate; current Australian passport; motor vehicle driver's licence; and firearm or shooters licence. Centrelink's guidelines require that an identification number for each of these is stored in the associated field. Within the DSSDD this field is described as:

> The Registration or Serial number is the number which uniquely identifies this document against others of the same type.[107]

5.12 As discussed above, different document types are allocated a point value and various benefit types require the customer to achieve a minimum number of points. For most income support benefits the customer is required to provide at least one document to show a Commencement of Identity in Australia and other documents that add up to 100 points of identification.[108]

5.13 ANAO referred to the Centrelink document titled 'POI Document Coding Guide – Version 9, March 2005', which provides detailed guidance to CSOs regarding the application of the POI policy and recording details on ISIS. The Coding Guide states:

> This Guide will help CSOs enter consistent details from approved identity documents on to the POI system. The SS231 lists the service reasons that require POI and the level of POI that must be attained.
>
> This Guide includes information needed to correctly assess and code POI documents provided, including:
>
> - descriptors of names, address, proof of payment etc.....;
> - details of document specific requirements for the POI screen;

---

[107] Centrelink, *DSSDD entry Field:CLIE.REGO.SER.NUM*, Centrelink Intranet, [Accessed 19 August 2005].

[108] Centrelink, Proof of Identity Requirements (Form SS231). Documents that show Commencement of Identity in Australia include: Australian Birth Certificate; Australian Passport (current); Citizenship Certificate; Australian Visa; Document of Identity (DFAT); Certificate of Evidence of Resident Status (DIMIA); and Certificate of Identity (DIMIA).

- screen shots of the PDS, POI, POIS, POIA and POIH screens and data about the common fields used; and
- a high level workflow diagram.[109]

Current versus historical POI data

5.14 The results of ANAO's analysis of POI data is presented in the following sections of this report. However, the reader should be aware of a distinction between recently recorded POI data and that recorded prior to 2001.[110] Centrelink advised ANAO that more stringent controls for POI data entry had progressively been implemented since September 2001. However, Centrelink also advised ANAO that:

> At the time of the introduction of the Tiered POI Model [September 2001] it was decided not to clean up POI data previously collected.[111]

5.15 Therefore, recently entered POI data is subject to improved controls, yet existing POI data is maintained on ISIS, and used to support the processing of new claims by existing customers so long as it was originally provided after July 1995. In relation to the electronic files of POI data provided to ANAO, Centrelink advised:

> If the customer has had documents input since September 2001 when the POI tiered model was implemented then these documents are displayed. … If the customer had provided POI documents before September 2001 and they were converted as achieving POI tier 3 then these documents are displayed.[112]

5.16 Therefore, the dataset provided to ANAO, by Centrelink, contained some POI records created after September 2001 and some records created prior to September 2001. These two classes of POI records were not able to be differentiated using the data provided to ANAO.

5.17 ANAO decided to include all POI records, provided by Centrelink, in its analysis. Given the descriptions above, and further discussions with Centrelink staff, ANAO formed the opinion that all POI data included in the files, were capable of being used to support claims processing by Centrelink, and therefore, open to examination.

---

[109] 2005, Centrelink, *POI Document Coding Guide – Version 9 – March 2005*, p1.

[110] Centrelink introduced new POI arrangements, including the tiered system, in September 2001.

[111] Centrelink advice of 2 December 2005.

[112] Centrelink advice of 6 September 2005, describing the contents of the POI dataset.

Analysis of POI data

5.18 ANAO examined details for a selection of high-point-value documents on all PER records stored on ISIS, including:

- Australian Citizenship Certificate (70 points);
- Driver's Licence (40 points);
- Original Australian Birth Certificate—Full and Extract (70 points);
- Current Australian Passport (70 points); and
- Non-standard documents (usually 10 or 20 points).

Australian Citizenship Certificates

5.19 ANAO identified 835 889 records that included the POI document type code 'AC'—Australian Citizenship Certificate. Of these, 472 325 records, or 56.5 per cent were associated with current customers. Australian Citizenship Certificates carry a unique identification number. Although the data entered into the serial number field in ISIS should be unique for each record, ANAO's analysis revealed that only 606 451, or 72.6 per cent, of the 835 889 Australian Citizenship POI records contained unique values.

5.20 ANAO observed considerable inconsistency in the recording of Australian Citizenship Certificate data. In many cases ANAO noted that, rather than recording a valid serial number, the database recorded a spurious numerical entry or a text entry. For example, ANAO observed the following entries for Australian Citizenship Certificate:

- 3 046 instances of the entry AC,CITIZENSHIP;
- 2 959 instances of the entry AC,UNKNOWN;
- 2 010 instances of the entry AC,AUST CITIZENSHIP;
- 1 582 instances of the entry AC,1;
- 1 112 instances of the entry AC,99999; and
- 489 instances of the entry AC,123456.

5.21 The Centrelink POI Coding Guide instructs CSOs to enter the following details into the serial number field for Australian Citizenship Certificates:

> Certificate or Register Number. If no number, code the place of issue. If confirmed by DIMIA code "DIMIA confirmed".

5.22 In all, ANAO identified over 300 classes of spurious values where 20 or more records held the particular value, accounting for 41 999 invalid records.

Audit findings

5.23 ANAO concluded that the POI field, holding data for Australian Citizenship Certificates, was compromised to the extent that nearly 30 per cent of the values in that field were unreliable.

Australian Passports

5.24 ANAO identified 2 428 609 records that included the POI document type code 'AP'—Current Australian Passport. Of these, 1 089 409 records, or 44.9 per cent were associated with current customers. Australian Passports carry a unique identification number. Although the data entered into the serial number field in ISIS should, therefore, be unique for each record, ANAO's analysis revealed that 2 345 011, or 96.6 per cent, of the 2 428 609 Australian Passport POI records contained unique values.

5.25 ANAO observed some inconsistency in the recording of Australian Passport data. In many cases ANAO noted that rather than recording a valid serial number, the database recorded a spurious numerical entry or a text entry. For example, ANAO observed the following entries for Australian Passport:

- 4 753 instances of the entry AP,PASSPORT;
- 2 043 instances of the entry AP,1;
- 1 547 instances of the entry AP,UNKNOWN;
- 1 263 instances of the entry AP,DUMMY;
- 868 instances of the entry AP,CURRENT; and
- 651 instances of the entry AP,99999.

5.26 The Centrelink POI Coding Guide instructs CSOs to enter the following details into the serial number field for Australian Passports:

> Passport Number (eg. A9999999) where A=D,E,F,K,L,J,M,X.

5.27 In all, ANAO identified almost 300 classes of spurious values where 10 or more records held the particular value, accounting for 33 549 invalid records. Of the other records, where less than 10 instances of use were identified, a small proportion appeared to contain spurious entries—perhaps 1 to 2 per cent. Some records simply record the values 'IRISH', 'ITALY', 'HOBART' or contained a four digit number that indicated a year—1987, 1984.

Audit findings

5.28 ANAO concluded that the POI field, holding data for current Australian Passports, was compromised to the extent that up to 3 to 4 per cent of the values in that field were unreliable.

Australian Birth Certificate (Full)

5.29 ANAO identified 3 173 236 records that included the POI document type code 'BC'—Original Australian Birth Certificate. Of these, 1 642 359 records, or 51.8 per cent were associated with current customers. Australian Birth Certificates carry an identification number, although the format varies across the states and territories where the certificates are issued. Although the data entered into the serial number field in ISIS should be unique for each record, ANAO's analysis revealed that only 1 795 018, or 56.6 per cent, of the 3 173 236 Australian Birth Certificate POI records contained unique values.

5.30 ANAO observed considerable inconsistency in the recording of Australian Birth Certificate data. In many cases ANAO noted that rather than recording a valid serial number, the database recorded a spurious numerical entry or a text entry. For example, ANAO observed the following entries for Australian Birth Certificate:

- 12 184 instances of the entry BC,BIRTH CERT;
- 3 405 instances of the entry BC,UNKNOWN;
- 2 730 instances of the entry BC,1;
- 2 296 instances of the entry BC,99999;
- 2 227 instances of the entry BC,DUMMY; and
- 1 586 instances of the entry BC,SIGHTED.

5.31 The Centrelink POI Coding Guide instructs CSOs to enter the following details into the serial number field for Australian Birth Certificates:

> Record the birth registration (reg) or entry number (do not confuse this with the serial, application (app) or certificate (cert) number.) AND the four digit year of reg.

5.32 In all, ANAO identified over 40 000 classes of values—some obviously spurious—where 10 or more records held the particular value, accounting for 828 652 invalid records.

Audit findings

5.33 ANAO concluded that the POI field, holding data for Original Australian Birth Certificates, was compromised to the extent that up to 43 per cent of the values in that field were unreliable.

Australian Birth Certificate (Extract)

5.34 ANAO identified 876 310 records that included the POI document type code 'BE'—Australian Birth Certificate Extract. Of these, 390 870 records, or 44.6 per cent were associated with current customers. Australian Birth Certificate Extracts usually carry an identification number, although the format varies across the states and territories where the extracts are issued. Although the data entered into the serial number field in ISIS should be unique for each record, ANAO's analysis revealed that only 603 521, or 68.9 per cent, of the 876 310 Australian Birth Certificate Extract POI records contained unique values.

5.35 ANAO observed considerable inconsistency in the recording of Australian Birth Certificate Extract data. In many cases ANAO noted that rather than recording a valid serial number, the database recorded a spurious numerical entry or a text entry. For example, ANAO observed the following entries for Australian Birth Certificate Extracts:

- 3 644 instances of the entry BE,BIRTH EXTRACT;
- 803 instances of the entry BE,UNKNOWN;
- 552 instances of the entry BE,DUMMY;
- 502 instances of the entry BE,1;
- 396 instances of the entry BE,99999; and
- 367 instances of the entry BE,01.

5.36 The Centrelink POI Coding Guide instructs CSOs to enter the following details into the serial number field for Australian Birth Extracts:

> Record the birth registration (reg) or entry number (do not confuse this with the serial, application (app) or certificate (cert) number.) AND the four digit year of reg.

5.37 In all, ANAO identified over 700 classes of spurious values where 10 or more records held the particular value, accounting for 26 135 invalid records.

Audit findings

5.38 ANAO concluded that the POI field, holding data for Australian Birth Certificate Extracts, was compromised to the extent that up to 31 per cent of the values in that field were unreliable.

Driver's Licence

5.39 ANAO identified 5 079 916 records that included the POI document type code 'DL'—Driver's Licence. Of these, 2 381 771 records, or 46.9 per cent were associated with current customers. Driver's Licences usually carry an identification number, although the format varies across the states and territories where the licences are issued. The data entered into the serial number field in ISIS should be unique for each record. However, ANAO's analysis revealed that only 4 690 394, or 92.3 per cent, of the 5 079 916 Driver's Licence POI records contained unique values.

5.40 ANAO observed considerable inconsistency in the recording of Driver's Licence data. In many cases ANAO noted that rather than recording a valid serial number or combination of numbers and letters, the database recorded a spurious alphanumerical entry or a text entry. For example, ANAO observed the following entries for Driver's Licences:

- 16 070 instances of the entry DL,CURRENT;
- 7 584 instances of the entry DL,DRIVERS LICENCE;
- 4 222 instances of the entry DL,UNKNOWN;
- 2 915 instances of the entry DL,1;
- 2 823 instances of the entry DL,DUMMY; and
- 2 378 instances of the entry DL,99999.

5.41 The Centrelink POI Coding Guide instructs CSOs to enter the following details into the serial number field for Driver's Licences:

> Licence number

5.42 In all, ANAO identified over 1 300 classes of spurious values where five or more records held the particular value, accounting for nearly 130 700 invalid records. ANAO also noted that a proportion of the instances where a value occurred less than five times contained spurious values.

Audit findings

5.43 ANAO concluded that the POI field, holding data for Driver's Licences, was compromised to the extent that up to 8 per cent of the values in that field were unreliable.

Non-standard documents

5.44 ANAO identified 2 763 989 records that included at least one POI document type code 'NS'—Non-Standard documents.[113] Of these, 1 263 109 records, or 45.7 per cent were associated with current customers. Non-Standard document is a code used when the POI documents presented by customers do not fall into one of the categories defined in the legal values table. Nevertheless, CSOs should record sufficient information about the Non-Standard documents to assist in identifying the customer. In a number of cases, ANAO observed that document types, which appeared to fall into one of the legal categories, were recorded as NS. For example, the code NS was followed by an entry such as Driver's Licence No. xxxxx (where xxxxx was sometimes an apparently valid number, and at other times a spurious number, such as 99999.)

5.45 ANAO observed considerable variation in the recording of Non-Standard document data. ANAO observed the following entries for Non-Standard document:

- 227 533 instances of the entry NS,_;
- 143 308 instances of the entry NS,MEDICARE (CARD);
- 57 613 instances of the entry NS,RATES NOTICE;
- 42 069 instances of the entry NS,MARRIAGE CERT;
- 28 882 instances of the entry NS,PAYSLIPS;
- 17 067 instances of the entry NS,STUDENT ID; and
- 14 524 instances of the entry NS,BIRTH CERT.

5.46 The Centrelink POI Coding Guide instructs CSOs to enter the following details into the serial number field for a variety of documents that are non-standard. For example:

> Medicare Card—Code 11 digit number as complete card number. This includes the 10 digit number and the customer's position on the card.

[113] Individual customer POI records may contain more than one NS entry.

> Rates Notice—Account or customer number and (if room) name/initials of Council. Eg, 534573 Brisbane.

5.47 In all, ANAO identified over 200 classes of values where 1 000 or more records held the particular value, accounting for 1 216 347 records, which, while not necessarily invalid, were of little use in contributing to the identity of the customers involved.

5.48 In addition, ANAO noted a number of entries, which appeared frivolous and were of no value in establishing customer POI. These entries included:

- Why not;
- Fictional;
- Why wont it restore;
- Who knows;
- What 4 isn't enough; and
- Fudged for POI;
- Why wont this work;
- Who cares;
- Whatever;
- Why do we need to.

5.49 While many entries contained some useful information in terms of assisting to identify a customer, ANAO formed the view that, on the whole, the data held in the fields associated with the NS code were of little genuine value and only marginally reliable.

Audit findings

5.50 ANAO concluded that the POI field, holding data for Non-Standard documents was compromised and that, generally, values in that field were of limited use in assisting to identify customers.

Further analysis of POI data

5.51 As noted earlier, ANAO analysed all of the POI data held by Centrelink on ISIS and provided to ANAO by Centrelink. Some of this POI data would have been provided to Centrelink and entered into ISIS prior to September 2001 and some will have been provided and entered after new POI arrangements, including the tiered system were introduced in September 2001. Centrelink advised ANAO that:

> Centrelink disagrees with assessing all POI data in the database by the current standards as different coding requirements and edits have been in place in the past, however, they still represent evidence of poor quality recording in some

cases. At the time of the introduction of the Tiered POI Model it was decided not to clean up POI data previously collected.[114]

5.52 ANAO noted that Centrelink had introduced a number of controls to improve the quality of POI data entered into ISIS, and that the most recent POI data should be of a significantly higher quality than that recorded in the past. However, ANAO could not confirm this proposition, as the dataset provided by Centrelink did not contain the date on which the POI information was entered into ISIS.

5.53 At Centrelink's invitation, ANAO undertook further analysis of the POI data, using the date on which the customer's current benefit status was granted. This information was part of the dataset provided by Centrelink in September 2005. ANAO identified 2 752 533 customer records that displayed a current benefit determination, granted on or after 1 September 2001. These account for 44.6 per cent of all current customers, at September 2005.

5.54 ANAO considered records displaying benefits granted between March and September 2005—that is, within the last six months. This amounted to 654 290 customer records. ANAO identified, essentially, the same pattern of spurious entries and multiple entries for serial numbers, in the POI data associated with those customers, as was identified across the whole dataset.

5.55 Further discussion with Centrelink staff revealed that many of these customers had a previous benefit determination associated with their record. For example, a customer may have successfully lodged a claim in 2000, provided POI information, and received benefits for some period of time under that determination. The benefit subsequently ceased, and the customer's record became non-current. Then, sometime between March and September 2005, the customer successfully lodged another claim. The customer's existing POI information remains on the updated record, with a current benefit granted well after the original POI data was entered into ISIS.

5.56 This analysis demonstrates that existing POI records are associated with subsequent benefit determinations for existing customers. Centrelink's revised POI data entry controls apply to new customers or existing customers who are required to provide new POI information. The above analysis does not shed light on the effectiveness of those measures.

---

[114] Centrelink's advice to ANAO of 2 December 2005.

5.57 In December 2005, Centrelink advised ANAO that one of its new procedures for monitoring the quality of POI data, involved the identification of frequently miscoded POI entries. Centrelink's advice was that:

> A dataset is produced each 6 months showing instances where the same serial number was discovered on a number of documents (including Australian Citizenship - AC, Australian Birth Certificate - BC, Drivers Licence - DL, Marriage Certificate - MC and Name Change - NC). The selection routine looks at documents coded in the past 6 months, and is intended to find CSOs that 'frequently' code the same serial/registration number. Frequency is selectable, and is currently set at 3 occurrences by any CSO. The threshold is reduced each month, as the extreme instances are removed by the clean up action on the previous period selections.
>
> When this task was first run (May 04) over 20,000 records were identified. The information was forwarded to Areas, and warning letters were issued.
>
> In the third (most recent) round (Nov 05) 236 records were identified as incorrect. All records were sent to the Areas for correction.[115]

5.58 Centrelink provided ANAO with a copy of the November 2005 report. ANAO accepted this, along with other documents detailing recent system level controls for POI data entry, as evidence that Centrelink's recently introduced controls have significantly improved the quality of *new* POI information being entered into ISIS.

Conclusion

5.59 ANAO's assessment of four primary POI documents used to establish Commencement of Identity in Australia—Australian Citizenship Certificates, Australian Passports, Australian Birth Certificates and Birth Extracts—revealed that, overall, only 73 per cent of values recorded on ISIS, for these four types of documents, constituted unique identification numbers. In addition, ANAO's analysis revealed that other POI data, including those recording details of Australian Driver's Licence numbers, were inaccurate or compromised to some extent.

5.60 Based on our analysis of the data provided by Centrelink, ANAO concluded that up to 30 per cent of the 8.3 million lines of POI information, held on ISIS as at September 2005, was inaccurate, insufficient or unreliable in terms of uniquely identifying or substantiating the identity of customers.

---

[115] Centrelink's advice to ANAO of 2 December 2005.

5.61 ANAO notes the recent improvement in the quality of POI data being entered on ISIS, arising from Centrelink's introduction of more stringent data entry controls and data quality monitoring procedures.

Recommendation No.4

5.62 ANAO recommends that Centrelink:

(a) continues to monitor the operation of its Proof of Identity policy and the quality of POI information recorded in ISIS; and

(b) progressively replaces spurious or inaccurate POI information currently recorded in ISIS with accurate information, when processing new claims or undertaking major of reviews of eligibility for existing customers.

Centrelink's response

5.63 Agreed.

6. Integrity of the Primary Key

This Chapter presents the results of ANAO's analysis of the integrity of the Centrelink Reference Number, which constitutes the primary key for the ISIS database. It includes a description of the data and the methodology employed in the analysis, and presents the results of specific tests designed to detect instances where two or more customers share the same value for their CRN and where a customer might be registered with Centrelink under two or more CRNs. The Chapter also explores the risks and consequences of fragmenting customer information across multiple records.

Methodology

6.1 Centrelink uses customers' CRNs as the primary key in its database. CRNs are used to identify individual customer records, and also, link customers' data stored in different files. Therefore, it is essential that customers have only one CRN and each CRN relates to only one person. If these conditions are breached, the ISIS primary key is compromised. Such a situation would pose a significant risk to the integrity of data held in the database. Consequently, ANAO's analysis involved an assessment of the integrity of the primary key.

6.2 In order to test the integrity of the CRN as a primary key, the ANAO examined whether:

- two or more Centrelink customers shared the same CRN;[116] and
- any Centrelink customers had been issued with more than one CRN.[117]

6.3 For the purpose of this treatment ANAO will use the terms 'duplicate CRNs' and 'multiple CRNs' to refer to the two situations described above. A duplicate CRN is defined as the same nine-digit number existing on two or more computing environments.[118] A case of multiple CRNs is defined as a single customer having two or more records on ISIS, under two or more different CRNs.

---

[116] This also includes the situation where the same customer had two or more customer records, under the same CRN, on different computing environments. (See the following sections for a detailed discussion of this matter).

[117] The scope of this audit does not include searching for cases where customers have fraudulently obtained two or more different CRNs. The scope of the audit is to identify customers that have been mistakenly registered more than once, and as a result, have been issued with more than one CRN.

[118] While Centrelink may employ a system of eleven computing environments for practical or functional purposes, ANAO considered that, collectively, the eleven environments constitute the entire ISIS database. Therefore, if a CRN exists on two environments, it is duplicated within ISIS.

Duplicate Centrelink Reference Numbers

6.4 ANAO analysed all records in the dataset provided by Centrelink—23 699 220 records—to determine if the primary key, the CRN, was constituted of unique values. ANAO found that, within each of the eleven computing environments CRNs were unique. However, ANAO also found instances where the same CRN existed on a number of different environments, and therefore, the CRNs were duplicated within the entire ISIS database.

6.5 The following sections of this Chapter explore the matter of duplicate CRNs. A number of factors bear on this matter. Firstly, the duplicate CRN may be associated with the same customer, who has a record on each of two (or more) environments. Secondly, the duplicate CRN may be associated with two different customers. That is, the same CRN has been issued to one customer, on a particular environment, and to a different customer, on a different environment.

6.6 Where a duplicate CRN is associated with the same customer, the status of each record is important. A CRN on one environment may identify a current record—the customer has a current benefit. The same CRN on another environment may be non-current—the record is not associated with a current benefit determination for the customer.

6.7 As each of Centrelink's computing environments is, essentially, geographically based, when a customer moves from one state or territory to another, Centrelink transfers the customer's record from one environment to another. Centrelink describes the process as:

> The process of transferring records across environments is:
>
> - the customer's record is moved to the new environment and the record in the old environment is deleted;
> - a shadow record is created in the losing environment if required. A 'shadow' record is only a small subset of the customer's full record. The customer and determination records are not included in the shadow, so effectively the customer record is not updatable in the losing environment.
>
> There are occasions when the customer's record is not deleted in the losing environment, due to system constraints. In these cases a flag stops the customer's record from being updated in the losing environment.[119]

[119] Advice from Centrelink's Chief Information Officer, to the ANAO, on 18 November 2005.

Status of records

6.8 Comparing the CRNs of all 23 699 220 lines of data, ANAO identified 239 806 lines of data that pointed to the existence of duplicate CRNs. Further analysis of the 239 806 records revealed 117 886 instances where the same CRN existed on two or more environments. In addition ANAO noted that some 63 700 lines of data were associated with current customer records. The distribution of these 117 886 duplicate CRNs is shown in Table 6.1.

Table 6.1

CRNs duplicated across environments

No. of environments on which CRN exists	Number of CRNs in category	Breakdown of the status of records duplicated across environments
2	114 061	• 1 089 — current on 2 environments • 58 926 — current on 1, non-current on 1 environment • 54 046 — non-current on 2 environments
3	3 629	• 5 — current on 3 environments • 74 — current on 2 environments, non-current on 1 environment • 2 297 — current on 1 environment, non-current on 2 environments • 1 253 — non-current on 3 environments
4	184	• 4 — current on 2, non-current on 2 environments • 119 — current on 1 environment, non-current on 3 environments • 61 — non-current on all 4 environments
5	11	• 9 — current on 1 environment, non-current on 4 environments • 2 — non-current on all 5 environments
6	1	• all 6 records were non-current

Source: ANAO analysis of Centrelink dataset — 13 September 2005.

6.9 ANAO noted that in the majority of cases where a customer appeared to be current on two environments, the data indicated that the customer was current for a non-payment-related benefit type on one environment.[120] While a duplicate CRN, current on two environments, represents a weakness in the integrity of the primary key, if one of the duplicate records does not include a current payment-related-benefit determination, the risk of overpayment is

[120] In the majority of cases, the records indicated that the customer was current for JSR on one record and an income support benefit on the other.

reduced.[121] However, fragmenting information for the same customer over two or more records always attracts some risk to the integrity of customer data.

6.10 ANAO's analysis revealed that 39 CRNs were not only current across two environments, but that they were current for the same benefit type, on each record. ANAO provided details of these records to Centrelink. Centrelink subsequently advised ANAO that:

> Regarding cases highlighted [by ANAO] as potential for double payment: Investigations showed that the records were marked as interstate transfer out and subsequently no payments were issued from the old environment as Centrelink has controls in place. These records have now been corrected.[122]

Duplicate CRNs and record type

6.11 ANAO analysed the duplicate CRNs using the field designating the record type. Table 6.2 shows that, of the CRNs duplicated across two environments, in 88 cases the duplicate CRNs consisted of only Child record types—that is, CHI-CHI. In 727 cases the duplicate CRNs involved one Child record type and one Person record type—that is, CHI-PER.

Table 6.2

Duplicate CRNs by Object Type Code combinations

No. of environments CRN exists on	Object Type Code combinations
2	88 x CHI-CHI, 727 x CHI-PER 1 x PER-ORG 113 245 x PER-PER
3	6 x CHI-PER-PER, 3 623 x PER-PER-PER
4	184 x All PER
5	11 x All PER
6	1 x All PER
Total of	239 806 records, involving 117 886 unique CRNs

Source: ANAO analysis of Centrelink dataset — 13 September 2005.

[121] The risk is not eliminated, as the record is still current and may, at some future point, be amended to include a payment-related benefit.

[122] Advice from Centrelink's Chief Information Officer, to the ANAO, on 18 November 2005.

6.12 Whether these are cases of shadow records for the same customer, or duplicate CRNs for different customers, ANAO considers that fragmenting customer information across different record types represents a data integrity failure.

Same vs different customers with duplicate CRNs

6.13 As mentioned earlier, duplicate CRNs could relate to the same customer or different customers. Commencing with the 239 806 lines of data involving duplicate CRNs, ANAO extracted 238 168 PER records and examined the contents of the name fields for these records.[123]

6.14 ANAO found that 197 167 records resolved into pairs, triplets or quadruplicates that matched exactly on CRN, surname and given name, and therefore, approximately 97 000 of the 117 886 duplicate CRNs involved the same person. That is, one of the records is probably a shadow record.

6.15 Of the remaining 41 001 lines of data that matched on CRN but did not match exactly on surname and first name, ANAO analysed 1 000 records in detail, and found that:

- 11 pairs, or 2.2 per cent of the sample, would have matched on surname but for a spelling error or inconsistency[124] in recording the surname on the two records;
- 58 pairs, or 11.6 per cent of the sample, would have matched on first name but for a spelling error or inconsistency in recording the first name on the two records;
- 381 pairs, or 76.2 per cent of the sample, matched on first name and second name and/or date of birth and/or address details but not on surname—indicating a married surname on one record and a maiden surname on the other;
- 29 pairs, or 5.8 per cent of the sample, matched on surname but not on first name. However, the records did match on address details and the difference in recorded dates of birth indicated that the customers were members of the same family, for example, mother and daughter. If not family members, these records indicated people with the same surname, but different dates of birth. In any case, the two customers sharing the same CRN are, most likely, different people; and

[123] ORG and CHI records do not hold information in the name fields.

[124] Such as the inclusion or exclusion of hyphens, spaces or apostrophes.

- 21 pairs, or 4.2 per cent of the sample, related to clearly different people.

6.16 As a result of the above analysis, ANAO concluded that approximately 10 per cent of duplicate CRNs identified in this exercise, result from the same CRN having been issued to two different people. Centrelink informed ANAO that it was aware of the existence of some duplicate CRNs, and that it had undertaken some work to resolve these. Figure 6.1 sets out how this can occur.

Figure 6.1

How can two people be issued with the same CRN?

A nine-digit CRN is issued for each new customer registration with Centrelink. The CRNs are generated by computer programs using a counter—so that each successive CRN is one greater than the previous.

Consider a situation in which a customer is registered in a particular environment, for example, environment A and assigned a particular CRN according to the value of the counter in that environment.

Sometime later, another customer is registered in a different environment, for example environment B, and is assigned a CRN according to the value of the counter in that environment.

The software controlling the issue of CRNs should ensure that each environment is allocated a particular range of values to be used as CRNs. However, in earlier times, the range available in a number of environments overlapped, and this resulted in some CRNs being created in more than one environment.

In the example above, the two different customers registered in environments A and B, respectively, could share the same value for their CRN.

6.17 As at June 2005, Centrelink estimated that up to 25 000 duplicate CRNs remained unresolved. ANAO's findings—10 per cent of 239 806, or approximately 24 000 duplicate CRNs involving different customers—supports Centrelink's estimate.

Duplicate CRNs and date of death

6.18 ANAO's analysis revealed that over 14 000 records—7 000 pairs of duplicate CRNs—had a date of death associated with one record, but no date of death associated with the duplicate record.[125] Based on an examination of many of the customer identity fields, ANAO estimated that almost 90 per cent of the 14 000 records involved duplicate CRNs for the same customer—that is, involving up to 6 300 customers.

6.19 ANAO's analysis revealed that, of the potential duplicate CRN customers with a date of death on one record, 43 customers had a current benefit determination on one of their records.[126] In most cases the benefit determination related to JSR, which does not involve a Centrelink payment. However, in two cases, customers had a current benefit status for FTB on the same record that also had a date of death recorded. Upon further investigation, Centrelink advised that payment had correctly ceased at the time of the customers' death.[127]

6.20 ANAO also found 404 records—202 pairs—that had different dates of death recorded on each of the two records.[128] Only two of these records displayed a current benefit determination, which was JSR, and therefore, do not appear to represent a risk to payments. Nevertheless, ANAO's analysis highlights the risk associated with fragmenting customer information across duplicate CRNs and the potential for storing inconsistent information about Centrelink customers.

---

[125] The actual number of lines examined was 14 486. Most duplicate CRNs resolved into pairs of records, others into triples or quadruplicates, so the approximate number of customers involved, who have a date of death recorded on one of their duplicate records, is 7 000. In a small number of cases, where a date of death was recorded on both records, and those dates did not match, all other indicators confirmed the same identity for the person.

[126] ANAO checked all available data to ensure that each of the 43 pairs of records related to the same customer. In addition, ANAO identified another 7 pairs of duplicate CRNs in this category, that probably related to the same customer, or two family members. These were excluded from further analysis.

[127] These two records were similar to those records with a date of death recorded, discussed in Chapter 4—provisionally deleted records that still displayed a current benefit status.

[128] Two of the 202 pairs mentioned above actually had records against the same CRN on three different environments, and therefore, were triples. However, in each case, only two records held a date of death.

Audit findings

6.21 In summary, ANAO's analysis of Centrelink's customer database highlighted the following, in relation to the matter of duplicate CRNs.

- The primary key of the ISIS database is compromised by the existence of duplicate CRNs;
- Duplicate CRNs exist in pairs, or multiple combinations, of:
  - non-current records;
  - non-current and current records; and
  - current records.
- Duplicate CRNs exist where the customer is identified as dead on one record and alive on the other—some of these combinations involve at least one current record;
- Duplicate CRNs exist in pairs of incompatible record types; and
- Some duplicate CRN pairs relate to the same customer (with shadow records on different computing environments) others relate to different customers.

6.22 As a result, ANAO concluded that the primary key of the ISIS database could not be relied upon to uniquely identify Centrelink's customers, and that the existence of duplicate CRNs represents a business risk to Centrelink and to the integrity of Government outlays.

Multiple Centrelink Reference Numbers

6.23 As noted previously, each customer should be identified by one and only one, unique CRN. Centrelink's customer registration processes are designed to ensure that, once registered with Centrelink, irrespective of the number of claims for various benefit types made by a customer, all information regarding that customer is stored in a single record under one CRN. In this way, a person may be a Centrelink customer in their youth, such as a Youth Allowance recipient, again in their working years, as a Newstart recipient or FTB recipient, and yet again in older age, as an Age Pensioner—with a full history maintained under one CRN.

6.24 However, if an existing Centrelink customer is registered for a second (or subsequent) time, under a different CRN, the potential exists for the customer's information to be fragmented across multiple records. The IT

systems and applications software operate on the basis that each CRN is associated with a single customer. Therefore, if a customer is assigned two CRNs, the IT environment treats these two records as relating to two discrete customers. Unless the records are linked in some way, the potential exists for multiple benefit payments under two or more CRNs.

Methodology

6.25 ANAO examined the dataset provided by Centrelink in order to identify customers with multiple records involving different CRNs. ANAO's first level of analysis involved identifying records which matched exactly on the customers' first name, surname and date of birth.[129] This generated a list of 1 259 427 PER records, which could potentially involve multiple CRNs.

6.26 Based on previous experience in the analysis of large datasets, ANAO considered that a proportion of these records were, in fact, coincidental matches—that is, different people who happen to share the same first name, surname and date of birth. In order to reduce the number of coincidental matches, ANAO refined the original analysis to include an exact match on the customers' middle initial. This generated a set of potential multiple CRNs numbering 955 224.

6.27 ANAO also noted that 11 536 records matched on customer's first name, surname, middle initial, address and TFN, but did not match on date of birth. This strongly suggested multiple CRNs, in that pairs of records appeared to relate to the same customer, with an error in recording the customer's date of birth on one record. Supporting this proposition was ANAO's observation that the recorded dates of birth for potential pairs often differed in some small respect. For example, same day and month with a difference of one year, or transposing two digits in either the month or day recorded. Adding the 11 536 to the 955 224 above, generated a set of 966 760 records for further analysis.

6.28 ANAO also noted that 18 568 records matched exactly on customer's first name, middle initial, TFN, date of birth and address, but did not match on surname. Most of these records indicated the sex of the customer as female, supporting the proposition that one record held the customer's maiden surname while the other held the customer's married surname. This was further strengthened by the observation that often the customer's marital status code was different on each record. For example, MAR on one, DIV on

[129] Where a customer's surname or first name included a hyphen, apostrophe or spaces between a two-word combination, these characters were ignored.

the other, or SIN on one, MAR on the other. A second group in this dataset exhibited a variation in spelling of the customer's surname—such as O'Connor and OConnor. Adding these 18 568 records to the 966 760 records above, generated a set of 985 328 records.

6.29 Using the above methodology, ANAO was able to develop a reasonable level of confidence that the records identified could relate to customers registered more than once with Centrelink and, therefore, represent multiple CRNs. ANAO notes that the methodology employed will not guarantee the identification of customers with multiple CRNs—there remains the possibility of coincidental matches, and based on previous experience with the analysis of large datasets, these may account for up to 10 per cent of the records identified by ANAO.[130] Nevertheless, ANAO's analysis suggests that up to 1 million records on ISIS may be associated with approximately half a million multiple CRN customers.

6.30 Centrelink informed ANAO that it had conducted a similar analysis of ISIS records in an attempt to identify multiple CRN customers, although using different matching algorithms, some of which relied more heavily on address details than ANAO's methodology outlined above. Through its analysis, Centrelink identified approximately 640 750 possible multiple CRN customers. Treating address details with greater importance, Centrelink's analysis identified approximately 355 900 multiple CRN customers.

6.31 Given the numbers suggested by Centrelink's analysis, ANAO's value of approximately 500 000 appears to constitute a reasonable estimate of the number of multiple CRN customers. It lies between the lower and upper estimates given by Centrelink's analysis. Although still likely to contain some coincidental matches, ANAO employed this test dataset for further analysis.

Multiple CRNs and customer status

6.32 Of the 985 328 records in the test dataset, 211 842 records, or 21.5 per cent of the test set, were records for current customers. Table 6.3 shows the number of multiple CRNs associated with customers and provides detail on the status of customer records.

---

[130] In addition, ANAO identified a relatively small number of customers (approximately 4 275 of the 985 328) who appeared in both the primary data matching group and one of the secondary matching groups. For example, a customer may have records under two CRNs, where those records match on first name, middle initial, surname and date of birth, and a third record under another CRN, but with a variation in the spelling of the surname or the recording of the date of birth.

Table 6.3

Extent of Multiple CRNs

No. of CRNs associated with an individual customer	Number of customers in category	Breakdown of the status of Multiple CRNs for individual customers
15 or more	108	86 customers were non-current on all 15 or more CRNs 22 customers were current on one CRN only
5 - 14	207	124 customers were non-current on all 5 - 14 CRNs 83 customers were current on one CRN only
4	1 219	657 customers were non-current on all 4 CRNs 562 customers were current on one CRN only
3	17 168	9 393 customers were non-current on all 3 CRNs 7 775 customers were current on one CRN only[131]
2	462 835	260 540 customers were non-current on both CRNs 201 202 customers were current on one CRN only 1 093 customers were current on both CRNs

Source: ANAO analysis of Centrelink dataset — 13 September 2005.

6.33 Table 6.3 shows that a total of 481 537 customers may have been registered with Centrelink under two or more CRNs. Furthermore, that 209 644 of these customers have only one CRN displaying a current benefit determination. In addition, the table also shows that 1 093 customers may have two CRNs, displaying a current benefit determination on both records.

6.34 ANAO examined the records for these 1 093 customers. A small number of records matched on all name elements, address and TFN, however, did not match on date of birth. ANAO concluded that, although some record pairs shared the same TFN, the difference in dates of birth suggested that the customers were a parent and child. These records were excluded from further analysis, along with a small number of coincidental matches identified by ANAO.

6.35 Refining the group of 1 093 customers, to remove coincidental matches and false matches due to family members with the same TFNs displayed on their records, resulted in a dataset of 1 080 pairs of customer records. As a customer with a current benefit determination on each of two records presents the greatest risk to the integrity of Centrelink's payments, ANAO considered a

[131] ANAO identified 55 coincidental matches in this group. These records were excluded from further analysis.

further refinement of this group. Therefore, the following analysis involves three sets of data:

- Reasonable level confidence data match — the number of customers in this group was 1 080 — customers matched exactly on first name, second initial, surname and date of birth;[132]
- Higher level confidence data match — a subset of the first dataset, the number of customers in this group was 793 — customers matched exactly on first name, second name, surname, date of birth, and either TFN or address (one of the non-matching pairs of TFN or address may be absent);[133] and
- Highest level confidence data match — a subset of the first and second datasets, the number of customers in this group was 213 — customers matched exactly on first name, second initial, surname, date of birth, TFN and address (both entries must contain exactly the same TFN **and** both entries must contain exactly the same address—none of those four values may be blank).[134]

6.36 ANAO examined details of the current benefit determinations for customers in each of the three groups described above. In particular, ANAO identified those multiple CRN customers, whose records showed the same benefit determination on each record. For example, customers were identified with a current benefit determination for Age Pension on both records. ANAO noted that some record pairs displayed the same commencement date for the benefit determination, while other record pairs displayed a different commencement date. Table 6.4 presents the results of ANAO's analysis.

---

[132] This group also contains the (high confidence) pairs with a different date of birth recorded and the maiden name/married name pairs.

[133] For example, a pair of records matching on the primary criteria AND sharing the same TFN were included. A pair of records matching on the primary criteria AND sharing the same address were included, as long as the pair did not contain different TFN values. That, is one record may contain a TFN value and the other of the pair may contain a blank entry for TFN. The result is a medium confidence data set incorporating either a match on TFN or address, when sufficient TFN and address information was available to reasonably exclude or include the pair.

[134] Essentially, these records matched on all available customer identification data.

Table 6.4

Distribution of possible double benefit determinations

Double benefit determinations	No of customers (from the REASONABLE confidence data match)	No of customers (from the HIGHER confidence data match)	No of customers (from the HIGHEST confidence data match)
Age Pension	102	9	2
Disability Support Pension	13	3	0
Family Tax Benefit	62	2	2
Parent Payment Partnered	2	0	0
Parent Payment Single	14	0	0
NewStart Allowance	5	1	1
Youth Allowance	6	3	1
Carer Allowance	5	2	0
JobSeeker Registration	316	302	99
Low Income Health Care Card	11	9	1

Source: ANAO analysis of Centrelink dataset — 13 September 2005.

6.37 In addition to the double benefit determinations shown above, ANAO identified a number of incompatible combinations of benefit determinations. For example, seven customers displayed a combination of Parent Payment Partnered and Disability Support Pension; six customers displayed a combination of Parent Payment Single and Parent Payment Partnered; and eight customers displayed a combination of Disability Support Pension; and NewStart Allowance.

6.38 In October 2005, ANAO provided Centrelink with full details of the records associated with these potential double payments and incompatible payments.

Centrelink's advice

6.39 In November 2005, Centrelink advised ANAO that it had conducted preliminary investigations of the highest confidence level dataset and a sample of the reasonable confidence level dataset. Some customers were known to have multiple CRNs and notes appeared on each of the records to this effect. Controls prevented double payments on such records. Centrelink advised that a number of potential multiple CRNs were the result of coincidence matching—further investigation showed that they were not the same person.

6.40 In only a small number of cases did there appear to be an inappropriate combination of active payments on both records. Centrelink advised that it would further investigate these cases. In addition, Centrelink advised that:

> Centrelink will investigate the large number of multiple reference number records. This can be caused by both user and system processes that result in adding a multiple record. Centrelink will also investigate the tightening of control for the business processes and system procedures for adding new customers to the database.[135]

Date of death and multiple CRNs

6.41 From the dataset of customers with potential multiple CRNs, ANAO identified 1 473 customers whose data included a date of death recorded on at least one CRN.[136] Further analysis revealed that 1 332 customers had a date of death recorded on one CRN but not the other. In addition, the data revealed that 122 customers had the same date of death recorded against their two CRNs[137] and 19 customers had a different date of death recorded on their two CRNs.

6.42 Of the 1 332 customers with a date of death recorded on one CRN but not the other, five customers also displayed a current benefit status on one of their CRNs. In one case (Age Pension), the recorded date of death was within six weeks of the data extract.[138] In another (Low income Health Care Card), the recorded date of death was just nine days after the date of granting the benefit status.

6.43 In the remaining three cases (two Age Pension, one Disability Support Pension), these customers showed a date of death on one CRN but had a current benefit determination on their other CRN, where no date of death was recorded. Furthermore, according to the recorded commencement dates for the benefits, the determinations were made after the customers' recorded dates of death.

---

[135] Advice from Centrelink's Chief Information Officer, to the ANAO, on 18 November 2005.

[136] These were customers, whose details matched exactly on name elements, address elements, date of birth, TFN and, in many cases, telephone number, but had a different CRN. ANAO actually identified approximately 10 000 customers whose details matched on less strict criteria, but that group contained a large proportion of coincidental matches. ANAO estimates that somewhere between 2 000 and 5 000 multi CRN customers may have a date of death recorded on one record. Although we have foregone some genuine matches, by concentrating on the 1 473 customers referred to above, ANAO is confident that they, at least, are highly probable multiple CRN customers with a date of death recorded on at least one of their CRNs.

[137] One of these customers had the same date of death recorded against three different CRNs.

[138] See the previous discussion on periods of grace following the death of a customer.

Audit findings

6.44 ANAO found that the ISIS database contains up to 1 000 000 customer records involving multiple CRNs—that is, approximately 500 000 customers registered more than once on Centrelink's database. This represents approximately 3 per cent of all (person) customer records on the ISIS database.

6.45 In addition, ANAO found that approximately 210 000 of these 500 000 customers had a current benefit determination on at least one of their records. This represents approximately 3.5 per cent of Centrelink's current customer population. Within this group, up to 1 000 customers had a current benefit determination on each of their two records, although rarely would both determinations result in a payment to the customer.

6.46 ANAO also found that, due to the fragmentation of customer information across multiple records, there is a risk of customers receiving double payments for a particular benefit, or a combination of incompatible payments. For some other customers, where a date of death is recorded on one CRN, ANAO noted a continued current benefit determination on another CRN. Although the majority of these cases involved a non-payment benefit, this situation may also present a risk to the integrity of Centrelink payments.

Conclusion

6.47 ANAO concluded that the primary key of Centrelink's customer database was compromised by the inclusion of duplicate and multiple CRNs. ANAO considers that fragmenting customer information across two or more CRNs presents the risk of duplicate benefit payments or a combination of inappropriate benefit payments. Fragmenting customer information also presents the risk of only closing off one record upon the death of a customer, leaving the other record to support a benefit determination. A similar risk exists if a customer becomes ineligible for a benefit payment and relevant customer circumstance data is only updated on one CRN.

Centrelink's advice

6.48 In November 2005, Centrelink advised ANAO that:

> Centrelink has checked samples of the Highest level Multi CRNs [file]. Generally the results do not involve dual payments of the same benefit, as the sample indicates that both records may be in receipt of different benefits or payments that are compatible. One customer was in receipt of FTB under multiple CRNs for different children. This matter is being investigated. Centrelink acknowledges the issues raised by the ANAO regarding the fragmentation of records and will undertake corrective action for the cases identified.[139]

Recommendation No.5

6.49 ANAO recommends that, in order to improve the integrity of the CRN, the primary key for ISIS, Centrelink takes action to resolve:

(a) all duplicate CRNs — instances where **different** customers have been allocated the same CRN and instances where the **same** customer has a current benefit determination on two or more Centrelink computing environments;

(b) all multiple CRNs — instances where the same customer has been registered under two or more different CRNs; and

(c) all instances of records where a date of death has been recorded against one of a customer's duplicate or multiple records, but not the other(s).

Centrelink's response

6.50 Agreed.

[139] Advice from Centrelink's Chief Information Officer, to the ANAO, on 18 November 2005. In January 2006, Centrelink confirmed that the investigation referred to above found that one customer had received duplicate payments under two different CRNs.

7. Implications of Data Integrity Issues

This Chapter draws on the findings of the three major themes of the audit and discusses the overall impact of the identified data integrity issues on Centrelink's administration of a number of social security programmes.

Inconsistent recording of names and addresses

7.1 ANAO found that the inconsistent recording of customer's names and addresses creates a number of problems and reduces the integrity of customer data generally. In summary, during the conduct of this audit, and in particular in relation to our examination of duplicate and multiple records, ANAO observed instances of:

- improper use of data fields — surname, first name and second name all stored in the surname field, leaving the other fields blank;
- reversal of first name and second name across two records;
- two given names stored in the first name field, leaving the second name field blank;
- data entry errors, or spelling errors in surnames, first names and second names, street names and suburb names;
- use or non-use of hyphens and/or spaces in two-word surnames and/or first names;
- inconsistent recording of house or unit numbers—eg. Unit 1, 26 … ; 1/26 …; or recording a unit number in address line 1 and the remainder of the address in address line 2; and
- recording the State and Postcode in the suburb field rather than the State and Postcode fields.

7.2 ANAO considers that inaccurately recording customer details could inhibit Centrelink's ability to analyse its customer data for compliance and fraud detection purposes. Centrelink exchanges and matches data with many other government agencies. Inaccurate customer data is likely to reduce the effectiveness of these data matching activities.

7.3 Centrelink CSOs rely on the accuracy of customer records on the National Index to identify existing customers. Inaccurate customer data may lead to incorrect identification during this process, and a CSO inappropriately

creating a new record for an existing customer. Finally, inaccurate customer data may lead to a poorer quality of customer service, for example, using inaccurate data when addressing correspondence to customers.

Implications for Centrelink's business

Results of field level analyses

7.4 ANAO's analysis of various ISIS fields that contain basic customer information, such as name, address and proof of identity data highlighted the following issues:

- less than ideal quality control on data entry—in particular the accuracy and completeness of customer's names;
- the use of a wide range of values outside those defined by the data dictionary as legal values for particular fields;
- the use of dummy dates of birth;
- the existence of training records in the production environment; and
- duplicate occurrences of TFNs across different customers, when TFNs are supposed to consist of unique values.

7.5 ANAO considers that these matters may impact on Centrelink's ability to efficiently and effectively:

- uniquely identify its customers, and register each customer once only;
- data match with other Government departments and agencies;
- detect the presence of duplicate and multiple registrations of customers within its database;
- provide accurate counts of Centrelink's customers;
- conduct modelling or data profiling activities that rely on customer age and/or date of birth; and
- conduct compliance and fraud detection activities.

Proof of identity

7.6 ANAO's analysis of the POI data provided by Centrelink found that up to 30 per cent of Centrelink's POI data, stored on ISIS, may be of little use in

actually establishing the identity of its customers.[140] ANAO's analysis revealed that some POI fields appear to be more reliable than others, in that a greater proportion of records in those fields consisted of unique values. For example, 97 per cent of values describing Australian Passports were unique, while only 57 per cent of values relating to Australian Driver's Licences were unique.

7.7 ANAO considers that the lack of integrity in electronic POI records may impact on Centrelink's ability to effectively implement its POI policy. A lack of POI integrity also presents a risk to the efficient and effective operation of Centrelink's fraud detection and prevention activities.

7.8 ANAO noted Centrelink's significant advances in improving and actively monitoring the quality of POI data in ISIS, since the introduction of the Tiered POI system in 2001. However, historical POI information is still relied upon when processing claims for many existing customers. While the historical information remains on the database, the overall integrity of POI data will be reduced.

Integrity of the primary key

7.9 ANAO's analysis highlighted the existence of duplicate CRNs and multiple CRNs. As a result, the primary key may not be relied upon to uniquely identify Centrelink customers, and to serve as the basis for ensuring a comprehensive record of customer information. Consequently, customer information may be fragmented across multiple records, or may be inconsistent across multiple records.

7.10 Irrespective of whether the customer records affected by duplicate or multiple CRNs support a current benefit determination, fragmenting information across customer records always presents a risk to data integrity. It impacts on Centrelink's ability to effectively data match with other agencies and organisations, detect and prevent fraud and accurately report on the number of customer records.

7.11 A significant risk occurs where two or more records support a current benefit determination—the risk being either double payments on the same benefit type, concurrent payments on two or more incompatible benefit types, or the issuing of multiple concession entitlement cards. This audit revealed

[140] This audit did not attempt to reconcile the electronic POI records with the paper-based customer files, which according to Centrelink's POI recording policy should hold photocopies of original documents presented by customers. The paper-based files may hold more useful information—many of the electronic records do not.

that Centrelink had other controls in place to prevent duplicate payments. For example, a duplicate payment filter operates to stop the second payment on a known duplicate customer record. Nevertheless, ANAO concluded that addressing the underlying data integrity issues would provide a better and more durable solution to managing these risks.

Conclusion

7.12 Centrelink's customer database, ISIS, constitutes one of the largest and most complex Australian Government databases holding information about Australian citizens and residents. With over 23 million records in total, ISIS holds just on 18 million records relating to people and organisations. Of these, as at September 2005, some 6.17 million records supported a current benefit determination, and in most cases, payment to a customer by Centrelink.

7.13 This audit has highlighted a number of issues in relation to the accuracy and integrity of data stored on ISIS, in particular that relating to recording documents used to prove a customer's identity, and the effectiveness of the primary key as a unique identifier. ANAO also concluded that many of the approximately 12 million non-current records on ISIS are superfluous and constitute an unnecessary risk to the integrity of the dataset. These records should be removed from the production environment, although Centrelink may wish to retain access to some of these records off-line.

7.14 Nevertheless, given the scale and complexity of Centrelink's IT operations, and considering the information examined in the scope of this audit, ANAO concluded that Centrelink's electronic records are, generally, sufficiently accurate and complete to support the effective administration of the range of social security programmes for which Centrelink is responsible.

7.15 ANAO also recognises that Centrelink responded promptly to the matters raised during the course of this audit, and commenced a number of initiatives to address specific data integrity issues identified by ANAO, and to generally improve the quality of data in ISIS. Key among these initiatives were projects to analyse and correct the identification of false positive results in the agency's existing data integrity error checking system, the establishment of a Data Quality Team to develop a long term strategy to improve and maintain data quality and work to comprehensively describe the effects of data integrity errors. Centrelink also undertook to review the operation of the priority rating system for data integrity errors.

7.16 In addition, Centrelink acted quickly to review cases of potential duplicate payment of customers, and to commit to resolving cases of duplicate and multiple CRNs.

Ian McPhee
Auditor-General

Canberra ACT
15 February 2006

Appendices

Appendix 1: Data exchange with other agencies

Table A1.

Data exchange between Centrelink and other agencies

Agency	Data exchanged	Frequency
Australian Valuation Office (AVO)	Centrelink identifies customers that are potentially exceeding the assets test limit and passes information to the AVO for the AVO to determine the value of the customers' assets. The AVO produces a valuation of particular assets.	Annual request from Centrelink.
Department of Veterans' Affairs (DVA)	DVA requests information from Centrelink on its customers. Centrelink returns details of payments due to those customers. DVA pays the customers.	Daily.
Department of Immigration and Multicultural and Indigenous Affairs	Some migrants have a sponsor who provides an assurance of support for the migrant. In this case a shell record[141] is created on ISIS, with a link to the supporter. Full proof of identity is required before any benefits are paid. DIMIA provides details of overseas absences of Centrelink customers. Certain Centrelink staff (in the international centre in Hobart) have direct access (a real-time link) to DIMIA information, to verify customer information.	Daily. Online—real-time.
Australian Taxation Office	Annual income provided to Centrelink for income test. Where a Family Tax Benefit is applied for through the ATO, Centrelink may create a new customer record for that customer. Centrelink holds ATO supplied income information separate from customer-supplied information.	Daily.
Several agencies (eg Child Support Agency, ComSuper, New Zealand Government)	Changes in benefits paid to customers by other agencies are notified to Centrelink. These may trigger changes to the rate of Centrelink payments to those customers.	Varies from daily to annually.
Medicare Australia	Centrelink request confirmation that customers are eligible for benefits (eg immunisation, child care). Medicare Australia response is yes/no.	Daily.

Source: ANAO discussions with Centrelink staff.

[141] A shell record is a type of customer record in ISIS that contains minimal information on the person concerned. A shell record cannot support the payment of a benefit.

Appendix 2: DI error code definition table

Sample Table Definitions

The following section gives a detailed description of each of the components of the standard DI Table. It includes information which will be stored by Data Management into the DIE reference file. Each error is to be described in the following format as in the following example:

Error Number	The error number assigned to this data integrity check. A 6 digit numeric number indicates this check is completed by a batch program. A number in the format AAAnnn (eg PGA001) indicates that this check is completed by an online (XDI) program.
Error Short Text	A 32 character short description of error text. Shown on summary DI reports and on screens within the DIE system.
Error Long Text	A 154 character long description of the error. 2 lines of 77 characters
Effect	Indicates the impact of the error on processing. . 231 characters maximum. 3 lines of 77 characters.
Priority	Priority assigned to this error. Priorities are allocated using same guidelines as Production processing faults recorded on Quantum. 1 digit number.
Refer To	Section to which queries should be directed. Maximum of 53 characters. E.g. For PGA, this will always be 'PGA National Help Desk'
Keywords	12 character
Action	Indicates what action should be taken to correct this DI error. In many cases, no clerical action will correct the problem; resolution will depend on case by case approach. Maximum of 770 characters. 10 lines of 77 characters.
Report	Indicates what data should be reported if an error is found. Displayed on DE screen in the DIE system. This may be the same as short message text or may consist of one or more of the following: - message text - data group name/field name which is in error - values held in the data field There may be up to 20 occurrences reported for a single customer record.
Contact Name	Name of the area providing the business rule. Maximum of 20 characters.
Cluster	Cluster code for area to which errors should be reported.
Contact Area	Area to which contact name belongs. Maximum of 6 characters.
Business Rule	Gives the business rule for the data groups checked by this module. If the business rule is NOT met, an error should be reported. Note that the business rules are based on the assumption that meta data in Centrelink's data dictionary (DSSDD) is both current and correct.

Source: Centrelink. Retirement, Disabilities, Parenting, Carers and Workflows Service Delivery User Specification, version 4.0, 9 February 2005.

Appendix 3: Sample time series analyses of DI error statistics

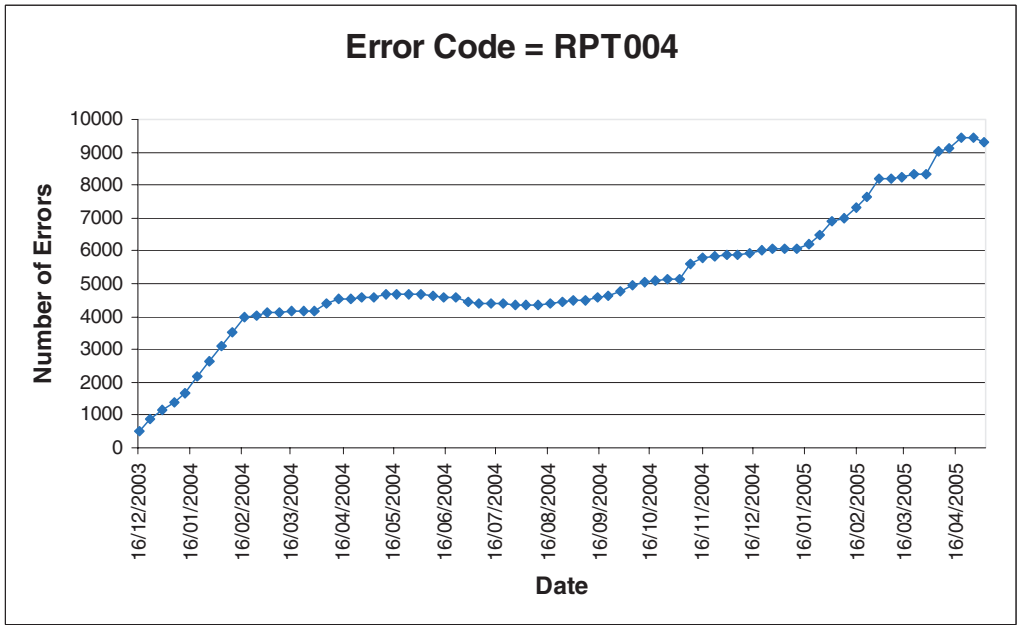
ANAO analysed the DI error statistics, including historical error counts, for the top 87 errors as at May-July 2005—the top 25 Priority 1 errors, top 27 Priority 2 errors, top 25 Priority 3 errors and top 5 Priority 4 and Priority 5 errors.

ANAO provided Centrelink with a copy of each of the 87 graphs produced, and discussed how these time series analyses might provide an insight into the possible causes of some errors.

Chapter 3 of this report outlines ANAO's findings. This appendix provides a sample of some of the trends in error numbers, identified during ANAO's analysis.

Figure A 1

Example of increasing number of errors

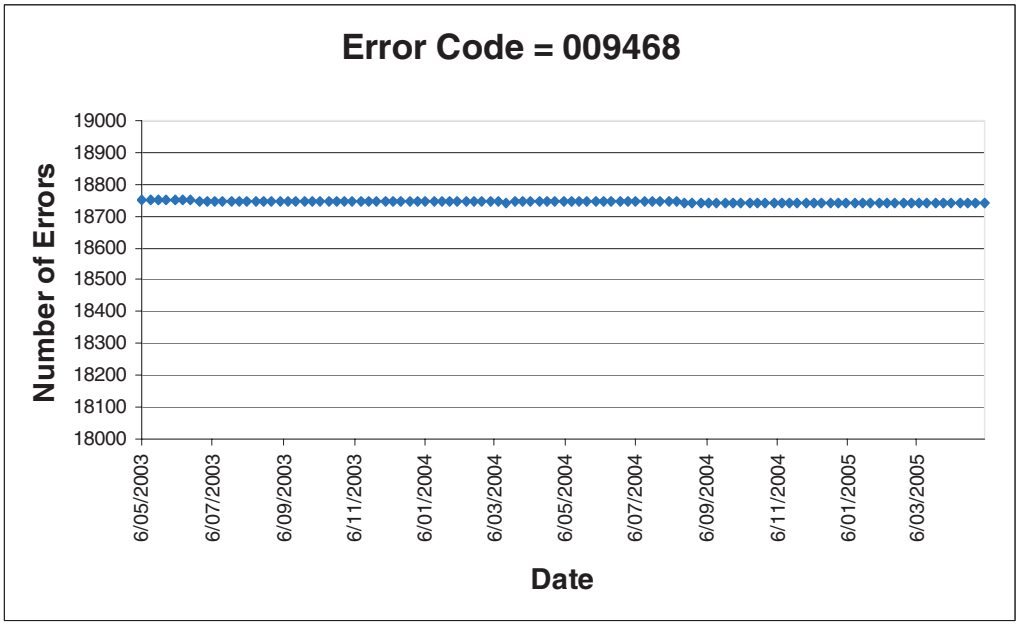


Source: ANAO – using data from Centrelink Intranet – DI error reporting system.

This is an error check which detects the lack of a reporting regime for the earnings of NewStart customers, where there is a stimulus indicator in the matching pay file.

Figure A 2

Example of a stable error count

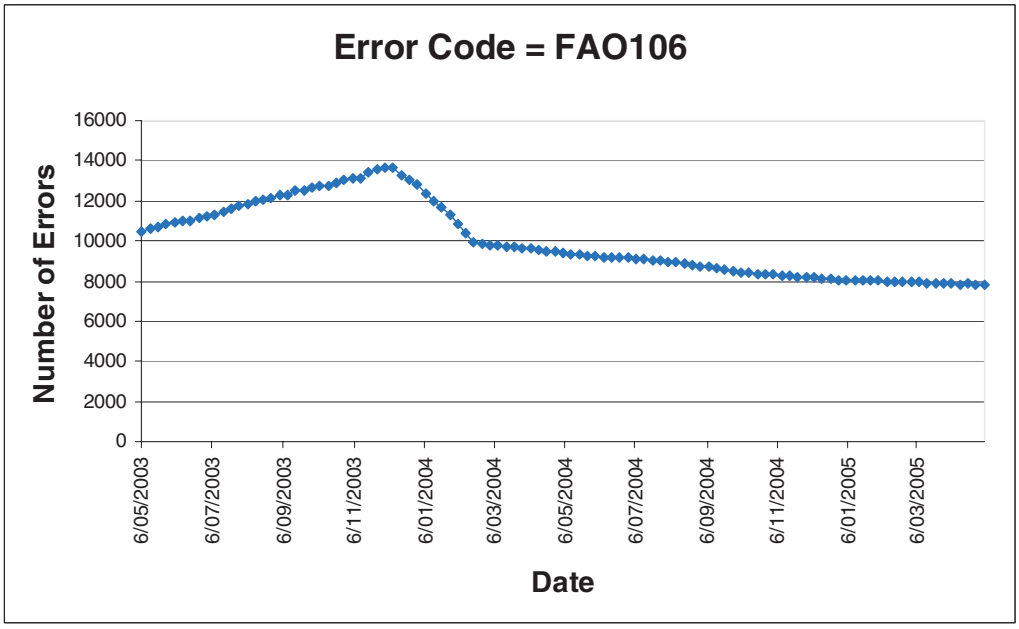


Source: ANAO – using data from Centrelink Intranet – DI error reporting system.

This error check detects the absence of a date of receipt of customer details, where this information is mandatory.

Figure A 3

Example of decreasing number of errors

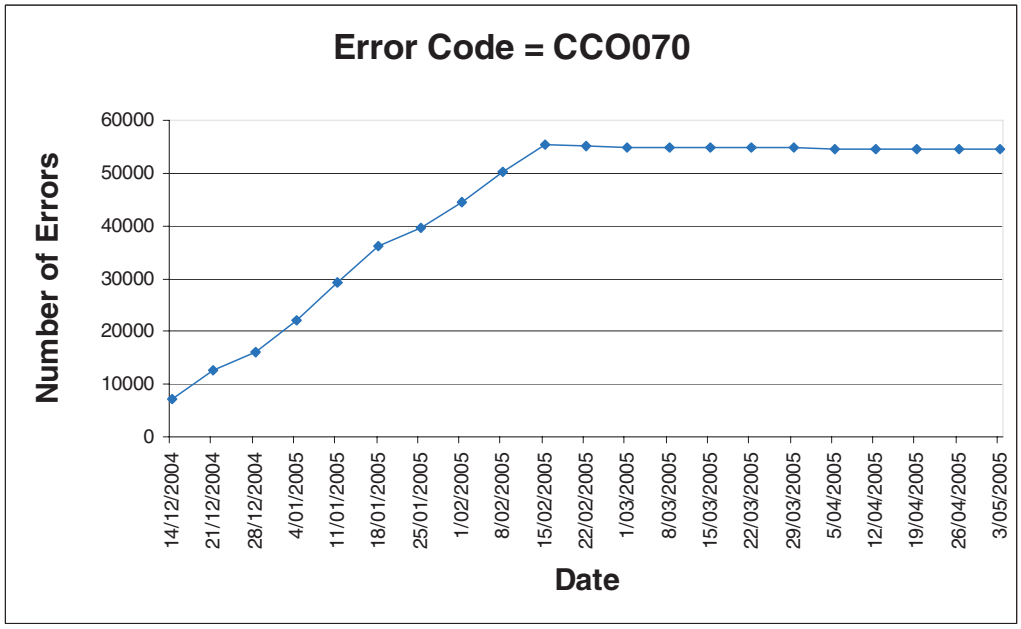


Source: ANAO – using data from Centrelink Intranet – DI error reporting system.

This error check detects missing taxation details for FAO customers.

Figure A 4

Example of a source of error fixed, but no data clean-up

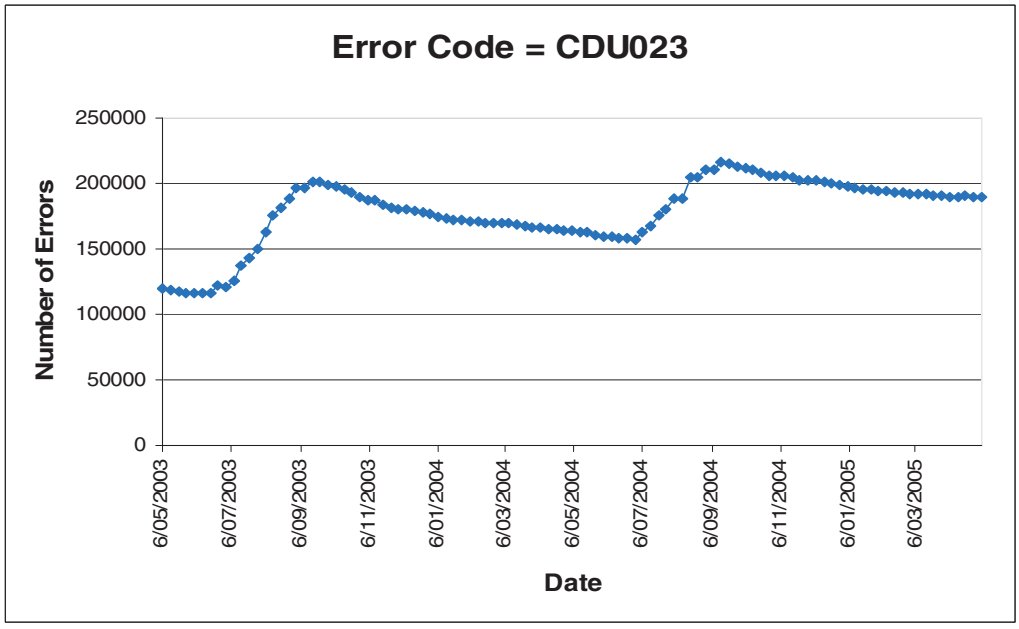


Source: ANAO – using data from Centrelink Intranet – DI error reporting system.

This error check detects a discrepancy between Health Care Concession Card entitlement and delivery of the Health Care Concession Card to a customer.

Figure A 5

Example of a cyclical pattern in error numbers

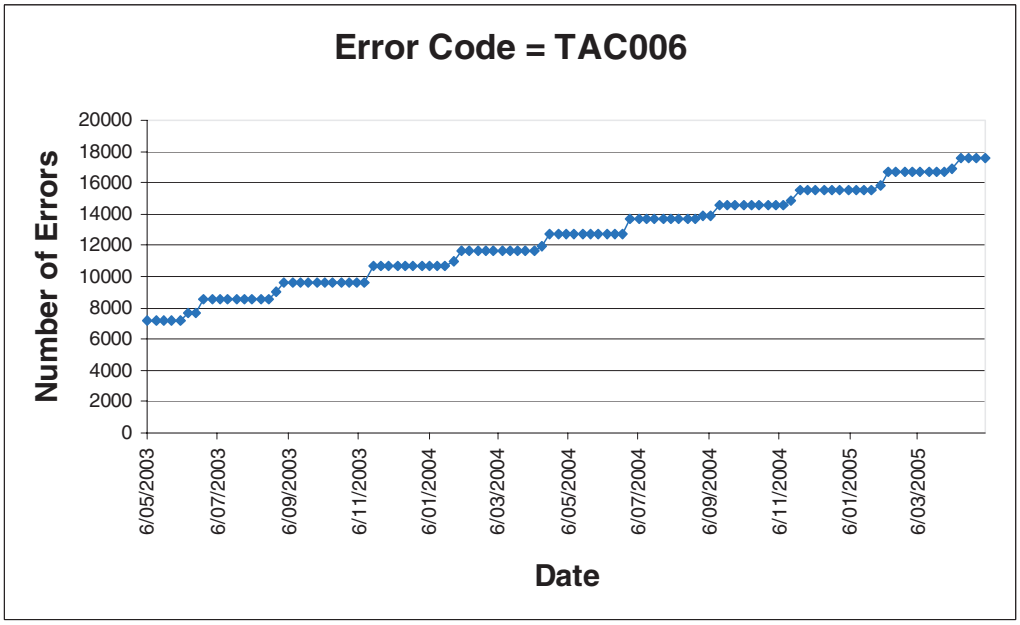


Source: ANAO – using data from Centrelink Intranet – DI error reporting system.

This error check detects a mismatch between information on a National Index and the local KIDS file for a child.

Figure A 6

Example of a 10-12 week cyclical pattern



Source: ANAO – using data from Centrelink Intranet – DI error reporting system.

This error checks customer data related to trusts and companies.

Appendix 4: Description of data provided by Centrelink

Customer circumstance data – 23 699 220 records **containing the following:**

Field name	Description	Data type	Length
CRN	Centrelink Reference Number (nine digit plus a 10th check digit as CHAR)	CHAR	10
OBJECT.TYPE.CODE	Type of customer record PER-person, ORG-organisation, CHI-child,-AOS-assurance of support, etc.	CHAR	3
CUNM.SURNAME	Customer's Surname	CHAR	30
CUNM.1ST.NAME	Customer's first name	CHAR	30
CUNM.2ND.NAME	Customer's second name or middle initial (if recorded)	CHAR	30
CUNM.TITLE	Courtesy title	CHAR	12
CUNM.DOB	Date of birth - eight digit, in format YYYYMMDD	DATE	8
CUNM.SEX.CODE	Sex	CHAR	1
CUCS.ON.IND	'Y' to represent a current customer (see discussion of this field in the body of the report)	CHAR	1
CUAD.ADDR.1ST.LINE	First line of address details	CHAR	50
CUAD.ADDR.2ND.LINE	Second line of address details	CHAR	50
CUAD.SUBURB.LOCALITY.NAME	Suburb	CHAR	50
CUAD.AUS.POSTCODE	Australian postcode	NUM	4
CUAD.STATE.CODE	State	CHAR	3
CUPH.PHONE.NUM	Telephone number	NUM	20
CUDD.DEATH.DATE	Date of death (if recorded, otherwise blank), - eight digit, in format YYYYMMDD	DATE	8
CUPT.MARITAL.STS.CODE	Marital status code	CHAR	3
CUTF.TFN	Customer's Tax File Number – as encrypted string of numerals and alphabetic characters	CHAR	9
CUN.DUP.CRN.IND	An indicator to show that this CRN is a known duplicate (unresolved on ISIS)	CHAR	9
PARTNER.CRN	Partner CRN - if recorded (no check digit)	CHAR	9

Customer POI data – 12 742 853 records containing the following:

Field name	Description	Data type	Length
CRN	Centrelink Reference Number (nine digit plus a 10th check digit as CHAR)	CHAR	10
The dataset provided by Centrelink comprised another 33 fields—11 sets of 3 fields—each set of which was comprised of:			
POI.DOC.UNIQUE.ID (1 to 11)	A number to indicate the POI record for a customer	NUM	1
POI.DOC.CODE (1 to 11)	A code representing the type of POI document presented by the customer—eg. Birth certificate, passport, drivers licence etc.	CHAR	2
POI.REGO.SER.NUM(1 to 11)	A number or combination of letters and numbers identifying the document listed above, such as a birth certificate issue number, passport number, or drivers licence number	CHAR	20
Not all customers had 11 sets of POI information associated with their records. If a particular customer only had four POI documents recorded, the first 12 fields would be populated and the remaining 21 fields would be blank			

Customer bank account data – 6 167 308 records containing the following:

Field name	Description	Data type	Length
CRN	Centrelink Reference Number (nine digit plus a 10th check digit as CHAR)	CHAR	10
Bank account details	A six-digit BSB number, followed by a semicolon, followed by a bank account number or identifier. If more than one active bank account was recorded for a customer, a '/' was inserted following the first record and another set of BSB, ';' and account number followed	CHAR	varied
Institution code	A code indicating payment made to a particular institution or nominee, rather than directly to the customer.	CHAR	3

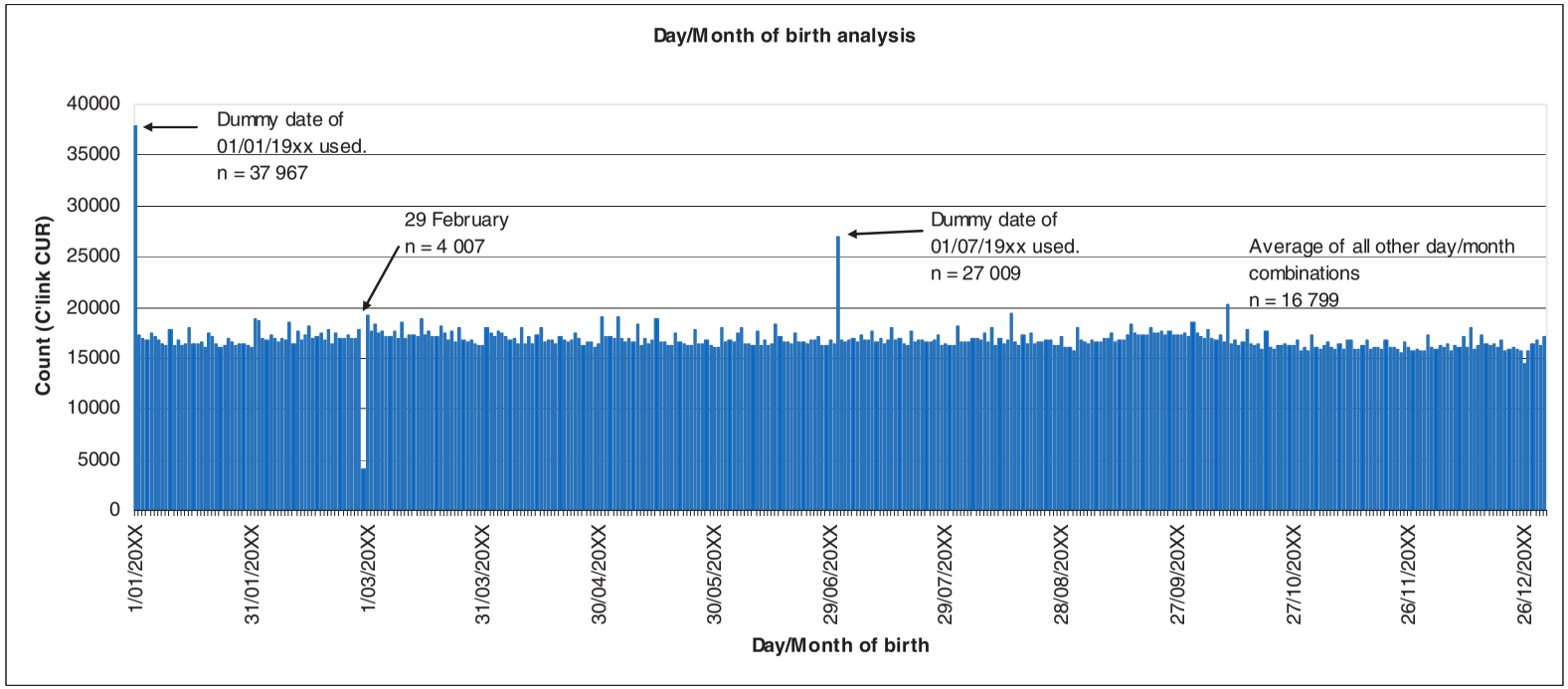
Customer status data – 6 168 030 records **containing the following:**

Field name	Description	Data type	Length
CRN	Centrelink Reference Number (nine digit plus a 10th check digit as CHAR)	CHAR	10
BENEFIT TYPE details Details of benefit determinations, as at 10 September 2005	This field consisted of sets of: • a three letter code indicating a benefit type; • the three letters 'CUR', indicating a current determination exists for the benefit type; and • eight digits representing the commencement date for that benefit determination	CHAR	varied

Centrelink advised ANAO that the customer status dataset only included details for customers whose record showed a current benefit determination. The dataset was constructed from the Determinations File, using ISIS fields known as xxDS.STS.CODE fields (where xx stands for a variety of benefit types—such as PN for pensions, PG for parenting payment NS for Newstart etc.).

By joining this customer status dataset with any of the other three datasets, ANAO was able to reliably identify current Centrelink customers.

Appendix 5: Analysis of day and month of birth



Source: ANAO analysis of recorded dates of birth in Centrelink's current customer dataset — 13 September 2005.

Appendix 6: Proof of identity

Australian Government | Centrelink

SS231.0412 (Page 1 of 2)

Proving your identity to Centrelink

As a Centrelink customer you are required to prove your identity when claiming a pension, benefit, allowance or service. You must establish your identity by providing **original** documents (not photocopies) from this approved list. Centrelink requires you to prove both your:

- Commencement of Identity (proof of your birth or arrival) in Australia; and
- Use of this Identity.

NOTE: The document you use to show your Commencement of Identity cannot also count towards the POINTS required for the payment or service claimed.

If you have any difficulty in obtaining or providing these documents, you should contact Centrelink as soon as possible.

If you have previously met the proof of identity requirements, and you are reclaiming within 52 weeks of receiving a Centrelink payment, fewer proof of identity documents may be required. If you think this applies to you, contact Centrelink as soon as possible.

Where possible, Centrelink will use the documents you provide to prove your age, residence, income and/or assets if that is relevant to the payment or service you are applying for. However, to be eligible for some payments or services, you may need to provide additional documents.

For claims of:

ABSTUDY (living allowance)
Age Pension
Austudy
Bereavement Allowance
Carer Payment – *both claimant and care receiver*
Disability Support Pension
Exceptional Circumstances Relief Payment – *claimant (and partner if applicable)*
Farm Help – *claimant (and partner if applicable)*
Newstart Allowance
Parenting Payment – *claimant (and partner if applicable)*
Pensioner Education Supplement
Sickness Allowance
Special Benefit
Widow Allowance
Youth Allowance

You will need to provide:

- at least one document (listed below) to show Commencement of Identity in Australia; **AND**
- OTHER documents that add up to 100 points from the approved list (see below and overleaf).

For claims of:

Carer Allowance – *both claimant and care receiver*
Mobility Allowance

You will need to provide:

- at least one document (listed below) to show Commencement of Identity in Australia; **AND**
- OTHER documents that add up to 50 points from the approved list (see below and overleaf).

For claims of:

Low Income Health Care Card – *claimant (and partner if applicable)*
Commonwealth Seniors Health Card – *claimant (and partner if applicable)*
Health Care Card for foster children – *for the foster child*

You will need to provide:

Any documents from the approved list (see below and overleaf) that add up to 50 points.
Proof of Residency may also be required. Please see your claim form for details.

Commencement of Identity in Australia

Document	Explanation/description	Points
Australian Birth Certificate	Original Australian birth certificate, extract or birth card in your name/former name.	70
Australian Passport (current)	Australian passport in your name/former name. Expired passports are not acceptable.	70
Citizenship Certificate	Australian citizenship certificate in your name/former name.	70
Australian Visa	Australian visa, current at time of entry to Australia as resident or tourist, showing your name/former name.	70
Document of Identity (DFAT)	Document of Identity issued in your name/former name by the Department of Foreign Affairs and Trade to Australian citizens or persons who possess the nationality of a Commonwealth country, for travel purposes.	70
Certificate of Evidence of Resident Status (DIMIA)	Certificate of Evidence of Resident Status (Form 283) issued by the Department of Immigration and Multicultural and Indigenous Affairs, showing your name/former name.	70
Certificate of Identity (DIMIA)	Certificate of Identity issued by the Department of Immigration and Multicultural and Indigenous Affairs to refugees and non Australian citizens for entry to Australia.	70

NOTE: If more than one of the above documents is provided, the additional documents will count as points.

Use of Identity

Document	Explanation/description	Points
Defence Discharge Papers	Australian Defence Force discharge papers, in your name/former name.	70
Shooter's or Firearm Licence	Current shooter's or firearm licence showing signature and/or photo and same name as claim.	70
Security Licence	Current security protection industry or crowd control licence, showing signature and/or photo and same name as claim.	70
Bank/Financial Institution card, statement or passbook	Current ATM or credit card showing your name and signature. Statement or passbook from current savings or cheque account showing your name and same address (if applicable) as your claim. **Cannot accept**: cards issued by organisations other than banks, credit unions or building societies, overseas accounts or ATM or internet receipts/statements.	40
Child's Birth Certificate	Australian birth certificate for a child showing your name as parent/guardian. **Cannot accept**: sibling's certificate.	40
Australian Driver's Licence - Motor Vehicle	Current state or territory issued driver's licence, learner's permit or provisional licence showing signature and/or photo and same name and same address as claim.	40
Australian Divorce Papers	Australian divorce papers in your name/former name, e.g. Decree Nisi, Decree Absolute.	40
Educational Certificate	Up to 3 school/education qualification certificates for different years in your name/former name (school/TAFE/university/Registered Training Organisation (RTO)).	40
Australian Marriage Certificate	Marriage certificate issued by a state or territory government agency. **Cannot accept**: church or celebrant issued certificates.	40
Mortgage Papers	Legally drawn mortgage papers for an Australian residence in your name/former name.	40
Name Change	Legal change of name certificate or deed poll certificate.	40
Overseas Passport	Current overseas passport with valid entry stamp or visa.	40
Registration Certificate from a Professional Board	Registration certificate from a national or state/territory professional registration board, e.g. doctors, nurses, dentists, physiotherapists, accountants.	40
Trade Certificate	Current Australian trade certificate in your name/former name. Must be signed by issuer or claimant.	40
Veterans' Affairs Gold Card	Current Department of Veterans' Affairs Gold Card issued in your name.	40
Reference from Indigenous Organisation	Reference from an Aboriginal/Torres Strait Islander organisation showing referee's full details and length of time they have known you.	20
Educational Report or Reference	Up to 3 school/education reports or references, including enrolment confirmations for different years or semesters, in your name/former name (school/TAFE/university/RTO).	20
Student ID Card	Current student ID card issued in your name with signature and/or photo (school/TAFE/university/RTO).	20
PAYG Payment Summary	PAYG payment summary, less than 2 years old, with tax file number. **Cannot accept**: Centrelink issued payment summaries.	20
Insurance Renewal	Current insurance renewal for house, contents, vehicle, boat, crop insurance in your name and showing same address as claim.	20
Tenancy Agreement or Lease	Current formal residential tenancy agreement or lease in your name and showing same address as claim.	20
Medicare Card	A current Medicare card showing your name.	20
Motor Vehicle Registration	Current motor vehicle registration showing your name, same address as claim and proof of payment.	20
Other Overseas Documents	Up to 3 overseas documents (equivalent to Australian documents listed of at least 20 points value), includes lapsed overseas passports.	20
Other Licence	Up to 3 current Commonwealth, state or territory licence for coxswain, boat, aircraft etc. Must have your photo and/or signature and same address as claim (if applicable). **Cannot accept**: recreational fishing licences.	20
Proof of Age Card	Current proof of age or photo identity card issued by a government agency in your name with photo and/or signature.	20
Rates Notice	Paid rates notice in your name and showing same address as claim, less than 12 months old.	20
Utility Account	Up to 3 paid utility accounts e.g. gas, water, electricity or phone in your name and showing receipt number and same address as claim, less than 12 months old.	20
Electoral Enrolment	Proof of electoral enrolment card issued in your name and same address as claim.	10
Other Financial Documents	Up to 3 current financial documents, such as superannuation, shares, life insurance, credit card statement or managed investment documents issued in your name. **Cannot accept**: hire or lease agreement.	10
Health Insurance Card	Current health insurance card showing your name.	10
Motoring Association Card	Current membership card or documents issued in your name.	10
Taxation Notice of Assessment	Taxation notice of assessment in your name less than 2 years old.	10
Employment Records	Termination notice, separation certificate, report or reference from employer in your name. **Cannot accept**: payslips.	10

Index

Series Titles

Audit Report No.28 Performance Audit
Management of Net Appropriations

Audit Report No.27 Performance Audit
Reporting of Expenditure on Consultants

Audit Report No.26 Performance Audit
Forms for Individual Service Delivery
Australian Bureau of Statistics
Centrelink
Child Support Agency
Medicare Australia

Audit Report No.25 Performance Audit
ASIC's Implementation of Financial Services Licences

Audit Report No.24 Performance Audit
Acceptance, Maintenance and Support Management of the JORN System
Department of Defence
Defence Materiel Organisation

Audit Report No.23 Protective Security Audit
IT Security Management

Audit Report No.22 Performance Audit
Cross Portfolio Audit of Green Office Procurement

Audit Report No.21 Financial Statement Audit
Audit of the Financial Statements of Australian Government Entities for the Period Ended 30 June 2005

Audit Report No.20 Performance Audit
Regulation of Private Health Insurance by the Private Health Insurance Administration Council
Private Health Insurance Administration Council

Audit Report No.19 Performance Audit
Managing for Quarantine Effectiveness–Follow-up
Department of Agriculture, Fisheries and Forestry
Biosecurity Australia

Audit Report No.18 Performance Audit
Customs Compliance Assurance Strategy for International Cargo
Australian Customs Service

Audit Report No.17 Performance Audit
Administration of the Superannuation Lost Members Register
Australian Taxation Office

Audit Report No.16 Performance Audit
The Management and Processing Leave

Audit Report No.15 Performance Audit
Administration of the R&D Start Program
Department of Industry, Tourism and Resources
Industry Research and Development Board

Audit Report No.14 Performance Audit
Administration of the Commonwealth State Territory Disability Agreement
Department of Family and Community Services

Audit Report No.13 Performance Audit
Administration of Goods and Services Tax Compliance in the Large Business Market Segment
Australian Taxation Office

Audit Report No.12 Performance Audit
Review of the Evaluation Methods and Continuous Improvement Processes for Australia's National Counter-Terrorism Coordination Arrangements
Attorney-General's Department
The Department of the Prime Minister and Cabinet

Audit Report No.11 Business Support Process Audit
The Senate Order for Departmental and Agency Contracts (Calendar Year 2004 Compliance)

Audit Report No.10 Performance Audit
Upgrade of the Orion Maritime Patrol Aircraft Fleet
Department of Defence
Defence Materiel Organisation

Audit Report No.9 Performance Audit
Provision of Export Assistance to Rural and Regional Australia through the TradeStart Program
Australian Trade Commission (Austrade)

Audit Report No.8 Performance Audit
Management of the Personnel Management Key Solution (PMKeyS) Implementation Project
Department of Defence

Audit Report No.7 Performance Audit
Regulation by the Office of the Gene Technology Regulator
Office of the Gene Technology Regulator
Department of Health and Ageing

Audit Report No.6 Performance Audit
Implementation of Job Network Employment Services Contract 3
Department of Employment and Workplace Relations

Audit Report No.5 Performance Audit
A Financial Management Framework to support Managers in the Department of Health and Ageing

Audit Report No.4 Performance Audit
Post Sale Management of Privatised Rail Business Contractual Rights and Obligations

Audit Report No.3 Performance Audit
Management of the M113 Armoured Personnel Carrier Upgrade Project
Department of Defence

Audit Report No.2 Performance Audit
Bank Prudential Supervision Follow-up Audit
Australian Prudential Regulation Authority

Audit Report No.1 Performance Audit
Management of Detention Centre Contracts—Part B
Department of Immigration and Multicultural and Indigenous Affairs

Better Practice Guides

User–Friendly Forms Key Principles and Practices to Effectively Design and Communicate Australian Government Forms	Jan 2006
Public Sector Audit Committees	Feb 2005
Fraud Control in Australian Government Agencies	Aug 2004
Security and Control Update for SAP R/3	June 2004
AMODEL Illustrative Financial Statements 2004	May 2004
Better Practice in Annual Performance Reporting	Apr 2004
Management of Scientific Research and Development Projects in Commonwealth Agencies	Dec 2003
Public Sector Governance	July 2003
Goods and Services Tax (GST) Administration	May 2003
Managing Parliamentary Workflow	Apr 2003
Building Capability—A framework for managing learning and development in the APS	Apr 2003
Internal Budgeting	Feb 2003
Administration of Grants	May 2002
Performance Information in Portfolio Budget Statements	May 2002
Life-Cycle Costing	Dec 2001
Some Better Practice Principles for Developing Policy Advice	Nov 2001
Rehabilitation: Managing Return to Work	June 2001
Internet Delivery Decisions	Apr 2001
Planning for the Workforce of the Future	Mar 2001
Contract Management	Feb 2001
Business Continuity Management	Jan 2000
Building a Better Financial Management Framework	Nov 1999
Building Better Financial Management Support	Nov 1999

Managing APS Staff Reductions (in Audit Report No.49 1998–99)	June 1999
Commonwealth Agency Energy Management	June 1999
Cash Management	Mar 1999
Security and Control for SAP R/3	Oct 1998
Selecting Suppliers: Managing the Risk	Oct 1998
New Directions in Internal Audit	July 1998
Controlling Performance and Outcomes	Dec 1997
Management of Accounts Receivable	Dec 1997
Protective Security Principles (in Audit Report No.21 1997–98)	Dec 1997
Public Sector Travel	Dec 1997
Audit Committees	July 1997
Management of Corporate Sponsorship	Apr 1997
Telephone Call Centres Handbook	Dec 1996
Paying Accounts	Nov 1996
Asset Management Handbook	June 1996