The Auditor-General
Audit Report No.16 2007–08
Performance Audit

Data Integrity in the Child Support Agency

Child Support Agency

Department of Human Services

Australian National Audit Office

ISSN 1036–7632

ISBN 0 642 80992 5

COPYRIGHT INFORMATION

Australian National
Audit Office

Canberra ACT
30 November 2007

Dear Mr President
Dear Mr Speaker

The Australian National Audit Office has undertaken a performance audit in the Child Support Agency within the Department of Human Services in accordance with the authority contained in the *Auditor-General Act 1997*. Pursuant to Senate Standing Order 166 relating to the presentation of documents when the Senate is not sitting, I present the report of this audit and the accompanying brochure. The report is titled *Data Integrity in the Child Support Agency.*

Following its presentation and receipt, 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

Fran Holbert
Michael McFarlane
Greg Dare

Contents

Tables

Figures

Abbreviations

ANAO	The Australian National Audit Office
ATO	The Australian Taxation Office
CNVRT	Converted or Conversion records
CRN	Centrelink Reference Number
CSA	The Child Support Agency
CSID	Child Support Identification number
CSO	Child Support Officer
DHS	Department of Human Services
ELP	Entry Level Programme
IT	Information Technology
MCR	Multiple Customer Record (dataset)
PI	Procedural Instruction
RACS	Restricted Access Customers
TFN	Tax File Number
TSO	Technical Support Officer

Glossary

Case status	Each child support case has an associated case status. ACTV—active CAN—cancelled CLSD—closed EWAR—ended with arrears EWLD—ended with liability calculation due INEG—ineligible REGN—registration in progress WDRN—withdrawn
Cuba	Cuba is the main electronic database used by the CSA to support the administration of the Child Support Scheme. It is not an acronym. The database was named after Cuba, in Roman mythology, the goddess of children.
CSA Customers	The CSA uses the term 'customer' to refer to each person, employer or organisation with whom it has dealings. Within the database, customers are divided into two categories—individuals and non-individuals. The respective codes used in Cuba are IND and NIND. Individual customers are the people involved in child support cases—the payers, payees, spouses and children. Non-individual customers are the employers, companies and overseas child support agencies that assist the CSA in the administration of the Child Support Scheme. In this report, we use the terms 'customers' and 'employers' to distinguish between the two customer types.
Role Type Code	Each customer in a CSA case is assigned a role. CHD—child PYE—payee PYR—payer SPS—spouse

Summary and Recommendations

Summary

Introduction

1. The Child Support Agency (CSA) was established in 1988 to administer Australia's Child Support Scheme. The CSA currently forms part of the Department of Human Services (DHS).

2. Through its administration of the *Child Support (Registration and Collection) Act 1988* and the *Child Support (Assessment) Act 1989* the CSA is responsible for the:

- assessment of how much child support is payable, through the application of a formula;
- registration of court orders for child and spousal maintenance, court registered agreements and administrative assessments;
- collection of child support; and
- disbursement of child support.[1]

3. Since the introduction of the Child Support Scheme, the CSA has managed a total of 1.4 million child support cases involving almost 4.6 million people. In 2006–07, the CSA reported transferring some $2.68 billion in child support payments between parents. At June 2007, the CSA managed an active caseload of approximately 800 000 child support cases, involving around 1.4 million parents and approximately 1.2 million children.

4. The primary electronic database supporting the administration of the Child Support Scheme is named Cuba.[2] The data in Cuba are organised around a number of core business functions including: case management; customer relationship management; child support assessments; accounting; and administrative support. The design of Cuba incorporates two significant data constructs—case and customer. Each child support case involves a number of customers—usually a payer, a payee and one or more children. Each customer may be associated, in various roles, with more than one child support case.

---

1 CSA, *What the Child Support Scheme aims to do* [Internet]. Available from <http://www.csa.gov.au/agency/facts.aspx> [accessed 2 August 2007].

2 Cuba is not an acronym—the system was named after Cuba, the goddess of children in Roman mythology.

5. While not in the largest category of Australian Government databases, Cuba is one of the more complex—given the various relationships that can exist between payers, payees, children, third parties, employers, financial institutions and overseas government agencies, across multiple child support cases.

Audit scope and objective

6. The objective of the audit was to examine the integrity of electronic records stored on the CSA's database—Cuba—and to report on the effectiveness of CSA's management of the data.

7. The audit assessed the CSA's electronic case and customer records and data management practices against the following criteria:

- CSA's case and customer records are accurate and complete;
- CSA's case and customer records are reliable and internally consistent;
- CSA has adequate controls and procedures to ensure a high quality of data capture and recording; and
- CSA effectively manages case and customer records.

8. The audit considered aspects of the CSA's data capture and recording practices, including data exchange with other agencies, along with technical and administrative level controls surrounding the CSA's data entry. A substantial part of the audit focussed on the analysis of data integrity within the various tables[3] of Cuba. The data extract from 16 Cuba tables comprised, in total, 142 958 924 lines of data.

9. Data within these tables were tested to ensure that selected mandatory fields contained valid entries. We also addressed aspects of internal consistency in the database, applying these as measures of the accuracy and completeness of customer records. The analysis included an assessment of the integrity of the primary key[4] for both the case and customer tables.

10. A major objective of our analysis was to identify any CSA customers who had been issued with more than one customer identification number and

[3] A table is a component of the database that stores related records. For example, the case table stores certain information about CSA cases; the case indicators table stores information on a selection of indicators that relate to CSA cases.

[4] 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.

to assess the business risk posed by fragmenting customer information across active CSA records.

Conclusion

11. Based on analysis of an extensive extract of CSA records, the ANAO concluded that the majority of records in Cuba are sufficiently accurate, complete and reliable to support the effective administration of the Child Support Scheme. Anomalous records that were detected usually accounted for a relatively small proportion of all records in the database.

12. While relatively few in number compared to the entire record set, the presence of erroneous records in the database indicates a weakness in effective control systems for data entry and recording. The CSA has recently introduced a Data Quality Improvement Programme, including a series of activities designed to test specific aspects of data quality in Cuba. The CSA will benefit from extending this programme to incorporate a comprehensive check of the application of all relevant business rules within the database.

13. Furthermore, the CSA should draw on the findings of this audit, and the information obtained through its Data Quality Improvement Programme, to identify and address weaknesses in data quality control systems. The inclusion of some controls should be relatively straightforward—such as a technical level control to ensure that an 'end date' does not precede a 'start date', within a line of data. Other controls, such as enforcing a standardised approach to recording names and addresses, may present more of a challenge, yet are essential to improving the overall quality of customer data in Cuba.

14. Most of the errors and weaknesses identified in this audit pose a minimal risk to the CSA's overall administration of the Child Support Scheme. However, particular errors or anomalies on individual records can result in an inaccurate calculation of child support liability. For the families involved, the effects can be significant. One of the objectives of the Child Support Scheme is that 'parents share in the cost of supporting their children according to their capacity'. Incorrect child support liability calculations, resulting from errors on customer records, pose a risk to the achievement of this objective.

15. This report highlights a number of areas in which the CSA could significantly improve the quality and reliability of data in Cuba, by addressing:

- inconsistent and inaccurate recording of dates, names and addresses;
- redundant records and training records in the production environment of the database;
- information fragmented across multiple customer records;
- weaknesses in the accuracy and reliability of Tax File Numbers and Centrelink Reference Numbers; and
- corrupt records arising from a data conversion in 2002.

Key findings by chapter

Data Capture and Recording—Chapter 2

16. The Cuba database commenced in 2002 with the conversion of customer and case information from an older mainframe computer system. While many data quality issues associated with the conversion—such as corrupt records—have been resolved by the CSA, a small number still persist.

17. All Child Support Officers (CSOs) undertake a standardised entry level training program and have access to national guidance material to assist them in their work. The training programme contains a module on Cuba navigation and work management. However, it does not attempt to provide staff with specific skills in data entry. These are generally gained through on the job experience. However, data entry quality assurance practices varied across CSA sites. Some team leaders checked 100 per cent of CSOs' work in their first four weeks on the job, others relied on staff to report instances of incorrect processing.

Case Records—Chapter 3

18. Testing found that the primary key for the case table—the case identification number—was sound, and contained no duplicate values.

19. The ANAO's analysis revealed that a relatively small proportion of liability start dates and liability end dates recorded in Cuba may be invalid or unreliable. Some liability start dates precede the commencement of the CSA, while some liability end dates indicate extraordinary apparent case durations.

20. A relatively small number of records demonstrated a weakness in the integrity of start and end dates pertaining to some case indicators. Further inconsistencies in the use of dates were evident in a small proportion of records in the liability calculation table. In particular, several inconsistencies were noted in the CSA's use of low dates and high dates[5] to denote the commencement and end of a period.

21. The data in Cuba include some records that display an inconsistency between case status and their current case liability determination. Some active

[5] The CSA uses 01/01/0001 as a low date and 31/12/4000 as a high date. While not valid dates in themselves, low dates and high dates are used in a variety of circumstances—as a placeholder for a valid date; to indicate that the true date is unknown; or to indicate that a valid date is not able to be recorded at a particular point in processing a record. The high date also indicates that a particular line of data within a table is current.

cases appear not to be associated with a current liability determination, while some non-active cases show a current liability determination.

22. The impact of these anomalies on the overall administration of the scheme is likely to be minimal. However, the potential impact on the individual customers involved may be significant, as these data integrity issues may result in incorrect liability determinations. In September 2007, the CSA advised the ANAO that it had commenced corrective action on a number of anomalous case records identified during this audit.

Customer Records—Chapter 4

23. Testing found that the primary key for the customer table—the child support identification number—was sound and contained no duplicate values.

24. A relatively small number of records in the customer table exhibited weaknesses in the integrity of date fields used to describe particular periods—with end dates preceding start dates for some customer indicators.

25. In relation to the accuracy of recording customers' names in Cuba, analysis revealed:

- 105 records where the name fields were blank;
- the existence of over 200 apparent training and other spurious records in the customer table;
- inconsistent application of business rules and acceptable standards for recording customer title, given name, middle name and surname in up to 1 per cent of records in the customer table; and
- 14 per cent of all customer records—current and historical—display the entry 'UNKNOWN' as the customer's surname.[6]

[6] The majority of these are child records. Prior to the conversion of data from the old CSA system to Cuba, in 2002, many child records did not incorporate a surname. These records were migrated to Cuba with the surname of 'UNKNOWN'. At the time of conducting this audit, there were over 389 000 active child records displaying the surname 'UNKNOWN'.

26. In relation to the accuracy of recording customers' dates of birth and dates of death, analysis revealed a number of weaknesses. According to the data recorded in Cuba:

- the oldest payee is 116 years of age and the youngest payee is 10 years of age;
- the oldest child, associated with a currently active case, is 51 years of age; and
- records for nine customers display the same date for the customers' date of birth and date of death, and 19 customer records, associated with active cases, also display a date of death.

27. Up to 12 per cent of customer records, associated with active cases, do not display a current address. In some instances mandatory address fields were left blank or displayed information other than address details—such as telephone numbers, email addresses or passwords.

28. Tax File Numbers (TFNs) play a central role in identifying CSA customers and facilitating data exchange with the Australian Taxation Office (ATO). Analysis confirmed that no single value for a TFN was recorded more than once in Cuba. However, 58 individual customer records contained invalid TFNs.[7] In addition, 913 pairs of customer records displayed duplicate Centrelink Reference Numbers (CRNs).

29. Within the individual customer table, 118 676 records are not, and have never been, associated with a CSA case. These include records apparently created during staff training. While they present only a slight business risk to the CSA, they are redundant records and should be removed from the production environment of Cuba.

30. In September 2007, the CSA advised the ANAO that it had commenced a process to correct anomalous customer records identified during this audit.

[7] In 56 cases the recorded TFNs consisted of more than nine digits, and in two cases the recorded TFNs failed the ATO check digit algorithm.

Employer Records—Chapter 5

31. The employer table in Cuba contains just under 259 000 records. In approximately 38 000 instances the employer's name appears more than once—constituting multiple records for employers. Almost 21 000 of these have been identified by the CSA as redundant records.

32. Employers play an important role in collecting child support payments by withholding a portion of their employees' wages and remitting this to the CSA. Multiple records for employers create difficulties for CSOs processing these payments and create unnecessary re-work, transferring payments between the multiple accounts.

33. Analysis of the employer table revealed numerous instances of mandatory fields containing blank entries. This situation points to a weakness in the application of business rules within the employer table.

Multiple Customer Records—Chapter 6

34. The CSA uses the child support identification number (CSID) as the primary key for customer records. In a well managed database each individual customer is allocated only one identification number. If a customer has more than one CSID there is a risk of fragmenting that customer's information across two or more unrelated records.[8]

35. The CSA has identified up to 18 000 customers with duplicate records and marked these records accordingly.[9] The ANAO sought to identify whether any other customers had been issued with more than one CSID, which had not been detected by the CSA. Through a series of internal data matching activities, 27 633 customers in this category were identified. Of this group, 360 customers show currently active cases on each of their CSIDs.

[8] Multiple records most often occur when an existing or previous customer of the CSA becomes involved in a subsequent child support case. It is possible for the CSO processing the second child support application to create a new record for the customer, rather than calling up the existing customer record.

[9] The ANAO identified some 12 783 marked in accordance with the Procedural Instruction: *A Guide to Duplicate Payers/Payees Records, Version 1.1,* 2002. In addition, the ANAO identified another 5000 customer records, distinguishable as duplicate records but not marked in strict accordance with the Procedural Instruction.

36. Within the group of multiple record customers, the ANAO observed:

- different case liability determinations across the two records;
- different income details;
- some of the multiple records were linked to the same CSA case, while other records related to different cases;
- 493 multiple record customers displayed two different TFNs, while 99 displayed two different CRNs;
- 136 customers displayed a date of death on one of their records but not on the other; and
- incompatible customer roles—such as a child on one record and a payer or payee on the other.

37. Customers with multiple records account for less than 1 per cent of all customer records and, therefore, pose only a slight risk to the CSA's overall administration of the Child Support Scheme. However, for the families involved, the errors in calculating child support liabilities can be significant. For example, the ANAO identified one payer with two customer records, each associated with an active child support case. Against one CSID, the customer's annual taxable income was recorded as $15 000 and an annual case liability of $5600 was displayed. Against the other CSID, the customer's annual taxable income was shown as $130 000, generating a case liability of $24 500.

38. The results of these analyses were provided to the CSA. The CSA investigated the circumstances surrounding the anomalous records and, in September 2007, advised that, for a relatively small number of customers, child support liability calculations were incorrect. The CSA also advised that it had commenced corrective action on these records.

Implications of Data Integrity Issues—Chapter 7

39. This audit has highlighted a number of specific opportunities for the CSA to improve the quality of data in Cuba. As well as improving the quality of individual customer records, the CSA could also improve the overall management of data in Cuba through a more consistent enforcement of business rules within the database.

40. The CSA would also benefit from the introduction of an active programme to regularly test the accuracy and reliability of customer and case data. The CSA is aware that a number of corrupt records exist in Cuba, arising

from the 2002 data conversion and has commenced work to resolve these issues. The CSA will benefit from further developing its Data Quality Improvement Programme.

41. The value to be gained from these improvements takes on added significance as the CSA moves to implement a new version of Cuba in July 2008.[10] A comprehensive cleansing of current data and the consistent application of business rules, within the database, should ensure that the CSA commences operation of the new version of Cuba, in 2008, with the highest possible quality dataset.

Summary of DHS's response

42. The Department of Human Services (DHS) appreciate the assurance provided by the ANAO's Data Integrity audit report outlining that the majority of records in the Child Support Agency's (CSA) primary database, Cuba, are sufficiently accurate, complete, and reliable to support the effective administration of the Child Support Scheme. However, where anomalies on individual records may potentially affect a customer's child support liability, we will resolve these errors and develop mechanisms and controls to support the continuous improvement of the data holdings in Cuba.

43. Furthermore, DHS values the recognition provided by the ANAO of the work commenced by the Data Quality Improvement Program within CSA and agrees with the benefit that would be gained with incorporating ongoing testing of data quality and the application of business rules, in Cuba, as part of this program.

[10] As part of the introduction of the third stage of reforms of the Child Support Scheme.

Recommendations

Recommendation No.1

Para 3.56

The ANAO recommends that the CSA:

- investigate and correct records in the case, case indicator and case liability tables that display anomalies in the various start and end date fields;
- ensure a consistent use by CSOs of high dates and low dates in case liability start date and case liability end date fields; and
- resolve instances of inconsistencies between case status and current case liability determination.

DHS's response: Agreed.

Recommendation No.2

Para 4.79

The ANAO recommends that the CSA identify and remove redundant records from the individual customer table in Cuba.

DHS's response: Agreed.

Recommendation No.3

Para 4.81

The ANAO recommends that, either separately or as part of its Data Quality Improvement Programme, the CSA:

- cleanse data in the fields describing customers' names and addresses;
- investigate and resolve anomalies in customer TFN and CRN records;
- develop and implement a quality assurance system to ensure a consistent standard of recording names and addresses in the revised version of Cuba to be introduced in 2008; and
- review and improve the effectiveness of data entry controls to ensure that, for individual customer records:
  - only valid, nine-digit TFNs may be entered;
  - only valid values, stored in a Cuba Code Table, are able to be entered;
  - dates of birth are recorded accurately; and
  - all mandatory fields are populated with valid entries.

DHS's response: Agreed.

Recommendation No.4

Para 5.19

The ANAO recommends that the CSA:

- cleanse the employer table of redundant records, including occurrences of multiple records for employers;
- resolve the issues surrounding the use of eight-digit and nine-digit TFNs for employers; and
- ensure that valid data populate all mandatory fields in the employer table.

DHS's response: Agreed.

Recommendation No.5

Para 6.60

The ANAO recommends that the CSA resolve all instances of multiple customer records in Cuba.

DHS's response: Agreed.

Audit Findings and Conclusions

1. Introduction

This chapter provides some background to the operations of the Child Support Agency, and in particular, the main electronic database that supports the CSA's administration of the Child Support Scheme. It also provides information on the objective, scope and conduct of the audit.

The Child Support Agency

1.1 The primary role of the Child Support Agency is to administer the Child Support Scheme. The Scheme was introduced in 1988 to

> 'strike a fairer balance between public and private forms of support [for children] to alleviate the poverty of sole parent families'[11]
>
> and aims to ensure that:
>
> - parents share in the cost of supporting their children according to their capacity;
> - adequate support is available to all children not living with both parents;
> - Commonwealth involvement and expenditure is limited to the minimum necessary for ensuring children's needs are met;
> - incentives for both parents to participate in the labour force are not impaired; and
> - the overall arrangements are non-intrusive to personal privacy and are simple, flexible and efficient.[12]

1.2 The *Child Support (Registration and Collection) Act 1988* established the Child Support Agency. The *Child Support (Assessment) Act 1989* implemented an administrative assessment of child support in accordance with a formula. The legislation provides for:

- assessment of how much child support is payable by administrative assessment;

---

[11] Cabinet Sub-Committee on Maintenance, *Child Support: Discussion Paper on Child Maintenance*, Canberra, 1986, p. 14—quoted in *Child Support Scheme Facts and Figures,* 2005–06, p. 5.

[12] Australian Government, *An examination of the operation and effectiveness of the Child Support Scheme, Government Response to the Report by the Joint Select Committee on Certain Family Law Issues*, Canberra, November 1997, p. 3—quoted in *Child Support Scheme Facts and Figures,* 2005–06, p. 5.

- registration of court orders for child and spousal maintenance, court registered agreements and administrative assessments;
- collection of child support;
- disbursement of child support; and
- changes to the administrative assessment because of special circumstances.[13]

1.3 Since October 2004, the CSA has been part of the Department of Human Services. At June 2007, the CSA employed some 4100 staff at 36 sites around Australia. In 2006–07, the agency administered over 800 000 cases involving approximately 1.4 million parents and 1.2 million children, and reported facilitating the transfer of approximately $2.68 billion in child support payments between parents.[14] The CSA had a departmental appropriation budget for 2006–07 of $370 million.[15]

1.4 The CSA is currently undertaking an extensive change agenda. It is implementing a new Child Support Scheme, following the Government's acceptance of the majority of recommendations of a Ministerial Taskforce on Child Support.[16] Stage One of the reforms was introduced on 1 July 2006 and Stage Two on 1 January 2007. Stage Three of the reforms will introduce a number of significant changes on 1 July 2008. A key feature of the changes encompassed in Stage Three is the introduction of a revised formula for the calculation of child support liabilities.

1.5 The agency is also undertaking a range of internal reforms under the 'Building a Better CSA' initiative, with a focus on customer service, organisational change, communication and stakeholder engagement.[17] The

---

13 CSA, *What the Child Support Scheme aims to do* [Internet].
Available from <http://www.csa.gov.au/agency/facts.aspx> [accessed 2 August 2007].

14 Department of Human Services, *Annual Report 2006–07*, DHS, Canberra, 2007, p. 50.

15 ibid., p. 9.

16 Department of Family and Community Services, *In the Best Interests of Children - Reforming the Child Support Scheme - Summary Report and Recommendations of the Ministerial Taskforce on Child Support,* FaCSIA, Canberra, 2005.

17 The 2006–07 Budget provided $150 million over four years to provide: more personalised and face to face services for parents with complex cases; on-line services; simplified processes; call recording; reduced CSA team sizes; and seamless connection of customers to other service providers. In addition, the initiative aims to improve staff training and development, revise CSA operational guidelines to eliminate systemic biases and increase education and communication programs.

CSA is also in the process of disengaging its Information Technology (IT) infrastructure from the Australian Taxation Office (ATO).[18]

The CSA's primary database—Cuba

1.6 The primary database supporting the CSA's administration of the Child Support Scheme is named Cuba.[19] Cuba incorporates financial and accounting information, customer information and information required by the *Child Support (Registration and Collection) Act 1988.*

1.7 The Relational Database Management System, in which Cuba operates, uses the DB2 proprietary software from the IBM Corporation. At the time of this audit the Cuba database consisted of over 230 DB2 Tables, incorporating over 5000 data fields. The database holds information relating to current child support cases as well as historical information on all previous child support cases. While not in the largest category of Australian Government databases, Cuba is one of the more complex—given the various relationships that can exist between payers, payees, children, third parties, employers, financial institutions and overseas governments, across multiple child support cases.

1.8 The data in Cuba are organised around a number of core business functions. These are:

- Customer relationship management;
- Case (payer/payee relationship);
- Child support assessment;
- Collection;
- Accounting;
- Disbursements; and
- Administrative support.[20]

[18] When CSA was created in 1998 the agency existed within the ATO, with the Commissioner for Taxation also fulfilling the role of Child Support Registrar. Consequently, CSA's IT infrastructure was originally developed within that of the ATO. Although the CSA now falls within the Human Services Portfolio, at the time of this audit CSA's IT systems, including Cuba, resided on ATO's IT infrastructure.

[19] Cuba is not an acronym—the system was named after Cuba, the goddess of children in Roman mythology.

[20] Child Support Agency—Application Architecture Team, *Cuba Functional Model*, CSA, Canberra, November 2006.

1.9 The Cuba system also supports the electronic exchange of information with a number of external bodies, including Centrelink, the ATO, employers and financial institutions. This data exchange is essential to the effective administration of the Child Support Scheme.

Cases and customers

1.10 The design of the CSA's Cuba database incorporates two significant data constructs—case and customer. A straightforward child support case might involve two parents and one or more eligible child of the relationship. A case might also involve a relevant dependant child in the care of either parent,[21] or a third party such as a grandparent or other carer. Each of these individuals is considered a customer of the CSA.

1.11 Any individual customer may be involved in more than one case—for example, as the result of a second or subsequent relationship. Customers may also be assigned different roles within those cases. For example, an individual may be a payee in one case and a payer on another case; or a child might be claimed for on one case and a relevant dependant of the payer or payee on another case. An individual customer may also have been an eligible child in his or her parents' case, and later, a payer or payee in his or her own case.

1.12 Cuba is designed to uniquely identify each child support case, by assigning each a unique case identification number. It is also designed to identify each CSA customer, by assigning each a unique identification number called the Child Support Identification number (CSID). These two numbers are stored in database fields in Cuba. Each represents a primary key[22] for the data constructs case and customer—the case identification number pertaining to records of CSA cases and the CSID pertaining to customer records.

1.13 In this way the managers and users of Cuba can establish relationships between individual customers—associating them with cases, amending or modifying the nature of the association over time and ending the association, while still keeping an historical record of the period of association.

1.14 As an indication of the complexity of relationships in the Cuba database, during the course of this audit the ANAO identified:

---

[21] A relevant dependant is usually a child from another relationship, not the relationship involving the payer and payee associated with a particular case.

[22] 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.

- one case with 21 different customer records attached—a payer, a payee and nineteen children: some claimed for as eligible for child support, at various stages in the case; some claimed as relevant dependants of the payer and some claimed as relevant dependants of the payee[23];
- two children who had each, at various stages and in various roles, been associated with 16 CSA cases;[24]
- 22 130 customers who were, over the course of their dealings with the CSA, associated with five or more cases; and
- almost 4000 cases that, over their course, displayed at least 10 different customer records associated with them.

1.15 Case records include information specific to each case—such as the status of a case, the case liability start and end date, a range of case indicators, audit information and information on the database operators who create and amend case details. Case records are also related to other information stored in various tables throughout Cuba. These include the customers associated with the case, child support periods and registrations relating to the case and liabilities calculated at various stages during a case.

1.16 Customer records include information specific to each customer—such as the customer's name, date of birth, contact telephone numbers, Tax File Number and Centrelink Reference Number. Customer records are also related to other tables in the database holding customer information, such as address, income, employer and the customer's involvement in other CSA cases.

1.17 Case and customer records are linked, or associated, in a table called case_customer_role. This table establishes the combination of case identification number, customer identification numbers and the customers' respective roles, that comprise each child support case.

Audit approach

1.18 The objective of the audit was to examine the integrity of electronic records stored on the CSA's database—Cuba—and to report on the effectiveness of CSA's management of the data.

---

[23] Some of the children were also associated, in various roles, with other cases.

[24] The record for the first of these children displayed a date of birth in 1991, making the child 16 years old at the time of the data extract from Cuba. The 16 cases involved a total of 13 different adults (payers and payees) and numerous other children. The record for the second child also displayed an association with 16 separate cases. This record displayed a date of birth in 2003, making the child four years old at the time of the audit.

1.19 The audit assessed the CSA's electronic case and customer records and data management practices against the following criteria:

- CSA's case and customer records are accurate and complete;
- CSA's case and customer records are reliable and internally consistent;
- CSA has adequate controls and procedures to ensure a high quality of data capture and recording; and
- CSA effectively manages case and customer records.

Audit methodology and scope

1.20 The audit considered aspects of the CSA's data capture and recording practices, including data exchange with other agencies, along with technical and administrative level controls surrounding the CSA's data entry.

1.21 A substantial part of the audit focussed on the analysis of data integrity within the various tables of Cuba. Following a series of preliminary investigations of the nature, structure and extent of the Cuba database, the ANAO requested and the CSA provided a large extract of data from the production environment of Cuba. The data extract from 16 Cuba tables comprised, in total, 142 958 924 lines of data. A list of the tables provided, and the number of records associated with each table, is at Appendix 1.

1.22 Data within these tables were tested to ensure that selected mandatory fields contained valid entries. The testing also addressed aspects of internal consistency in the database—applying these as measures of the accuracy and completeness of customer records. Analysis included an assessment of the integrity of the primary key for both the case and customer constructs.

1.23 A major objective of the analysis was to identify any CSA customers who had been issued with more than one customer identification number and to test whether customers in this situation were associated with active CSA cases on more than one customer identification number. The CSA was provided with a series of electronic files resulting from the analysis of the data extracts from Cuba.

1.24 During the audit the ANAO:

- observed data entry activities and held discussions with Child Support Officers (CSOs) and other key CSA staff at three CSA sites—Adelaide, Melbourne and Townsville;
- interviewed programme specialist and information systems staff at the CSA's corporate headquarters in Canberra;
- reviewed various CSA files, information sheets and publications; and
- examined the CSA's Intranet and Internet sites.

1.25 Fieldwork for the audit was undertaken during the period March 2007 to July 2007. This audit was conducted in accordance with the ANAO Auditing Standards, at a cost to the ANAO of approximately $401 000.

Structure of this report

1.26 This chapter provides some background information on the operation of the CSA and the high level structure of information stored in Cuba. It also outlines the approach taken in the audit.

1.27 Chapter 2 discusses data capture and recording in Cuba. It considers the migration of data from the CSA's previous database to Cuba in 2002, data exchange with other agencies and data entry by CSA staff.

1.28 Chapter 3 presents the results of the field level analysis of data in the case table.

1.29 Chapter 4 presents the results of the field level analysis of data in the individual customer table, while Chapter 5 focuses on the employer table.

1.30 Chapter 6 presents the results of the analysis of undetected multiple records for CSA customers. That is, customers that may have multiple records in Cuba, not previously identified as such by the CSA.

1.31 Chapter 7 considers the overall audit findings arising from the analyses and discusses the possible business risks posed by data integrity weaknesses.

2. Data Capture and Recording

This chapter describes how information is captured and recorded in Cuba. It also examines the training, reference materials and other support that the CSA provides its staff, especially as these relate to the quality of the CSA's data capture and recording.

How data enters Cuba

Data migration and conversion

2.1 Cuba was first implemented in March 2002. Prior to that time the database supporting the administration of the Child Support Scheme resided on the 'old CSA mainframe system'—as it is often referred to in CSA documentation. Data in the old system was migrated to Cuba. The CSA refers to customer records created in this way as 'converted' or 'conversion' records.

2.2 At the time of the conversion, the CSA was aware that some data in the old system was of poor quality. The CSA issued a document called *'Cuba Implementation Data Integrity Guide—A guide to eligibility errors and other data conversion issues'*, dated March to July 2002. The Data Integrity Guide recognised that:

> Due to the poor quality of data recorded in the old CSA mainframe system, inconsistencies, omissions or corruptions mean that Cuba is unable to process any variations until the errors are rectified.[25]

2.3 Although Cuba represented a significant improvement on the old system, its implementation was limited by some persistent data quality issues. The Data Integrity Guide illustrates a particular example:

> Cuba automates many more functions than the previous system and as a consequence sometimes requires more detailed case data to achieve this higher level of automation. In some circumstances the previous CSA system did not reliably and/or accurately store some elements of case data and consequently this information could not be automatically migrated to specific locations in Cuba (e.g. court order details).
>
> Children are also now uniquely identified in Cuba. This means that care and relationship information is also unique for each carer that the child has a relationship with at any given time. Altering the care details on one case will

[25] CSA, *Cuba Implementation Data Integrity Guide—A guide to eligibility errors and other data conversion issues,* 2002, p. 3 and p. 7.

> simultaneously trigger complementary adjustments on any related cases involving that child and/or carer. This is one of the major advantages of Cuba over the previous CSA system.
>
> However, as children were not unique identities in the previous CSA system, many inconsistencies remain undetected. These inconsistencies have migrated to Cuba. When Eligibility is triggered, Cuba is unable to determine which information is correct and is therefore prevented from finalising a variation until the inconsistency is dealt with.[26]

2.4 The Data Integrity Guide provided detailed instruction to CSA staff on how to identify and fix many of the problems associated with converted records. Since 2002, many of the data conversion issues have been resolved and the quality of data in Cuba improved. However, during fieldwork for this audit the ANAO was informed that some data conversion issues remain unresolved.

2.5 One example offered by CSOs interviewed by the ANAO was that of care records for children.[27] These CSOs described a situation in which care arrangements on a converted case, which may have been dormant for some years, are corrupted when that case is reactivated.[28] The CSOs indicated that in such situations they were required to re-build the care arrangements on the converted case in order to fix the error and allow the new case to be properly processed.

Data exchange

2.6 The CSA is authorised to exchange data with a number of other government departments and agencies. Key among these are the ATO and Centrelink. The CSA relies on data exchange with the ATO to gather income information on its customers. Various fields in the Cuba income table are populated through this data exchange. Customers' Tax File Numbers may also be obtained from the ATO's systems.

---

[26] ibid.

[27] That is, data indicating care arrangements—the proportion of time each child spends in the care of each parent, and the periods during which those arrangements applied.

[28] Reactivation of such a case often occurs by association with a new case. That is, if one or more of the customers associated with the converted case later returns to the CSA in a new case, Cuba will cross-reference the association and call up information from the converted case. (See the second paragraph of the example from the DI Guide, quoted in paragraph 2.3). If corrupt information exists on the (until now) dormant converted case, Cuba cannot proceed to process eligibility on the new case.

2.7 Similarly, the CSA exchanges data with Centrelink, including information on the customer's status within the Family Tax Benefit programme. The CSA may also collect child support payments from customers' Centrelink benefit payments, through garnishee.

Data entry by Child Support Officers

2.8 Child Support Officers work with Cuba every day—recording new applications for child support, registering child support cases, updating customer records as their circumstances change and generally managing customer records.

2.9 CSOs undergo an intensive eight week training programme when they join the CSA—called the Entry Level Program (ELP). The curriculum and training modules for the ELP are standardised across Australia, but ELP co-ordinators at each site, in consultation with local staff including Quality Advisers and Technical Support Officers, may choose to emphasise particular topics or supplement the curriculum with additional modules.

2.10 The ELP is presented by experienced CSA staff—rather than by specialist trainers—and focuses on the Child Support Scheme's legislation, the CSA's corporate culture and values, and the service delivery aspects of the agency's business. Although the ELP contains a module on Cuba navigation,[29] it does not attempt to provide staff with specific skills in data entry—these skills are acquired on the job.

2.11 On the job training involves working closely with an experienced CSO, handling actual cases. After satisfying the requirements of the ELP and being allocated to a team, the new CSO will ideally work with a dedicated 'buddy', initially observing the work then progressively performing an increasing amount of work independently.

2.12 New CSOs undergo a probation period, with reviews at 2.5 and 5.5 months—a 'back on track' program is available to assist those whose progress is below the required standard. The CSA uses competency guides to assist Team Leaders in monitoring the progress of new staff.

2.13 During fieldwork, a number of the CSOs interviewed by the ANAO advised that the ELP had assisted them in gaining an understanding of the

[29] CSA, *Cuba Navigation and Work Management – Facilitator Guide, version 5.0 (January 2007)* and *CSA, Cuba Navigation and Work Management – Participant Workbook, version 4.0 (January 2007)*, Canberra, 2007.

CSA's business and of the child support legislation, but provided less comprehensive instruction on aspects of day-to-day work such as using Cuba. They advised that between two and three months hands-on experience is required, following the completion of the ELP, to be reasonably competent at navigating through Cuba and entering data.

2.14 A number of CSA staff advised that the buddying system is extremely effective under ideal conditions, but instances occurred—due to pressure of work, or shortages of staff—where recent ELP graduates were assigned to a succession of buddies, or even left without a buddy for a period of time. These CSOs were of the opinion that when this occurs, the effectiveness of the process is reduced. In some instances buddies might, themselves, have little experience. For example, one group of CSA staff informed the ANAO that ELP graduates had been assigned to buddies who were graduates of the immediately preceding ELP course.

2.15 The CSOs observed during fieldwork, including relatively inexperienced staff, generally performed data entry at a high standard of competence. They were usually able to navigate easily in Cuba and to find customer details readily.

Reference materials for CSOs

2.16 All CSOs have access to Procedural Instructions (PIs), a help function in Cuba, the Cuba help desk, the Guide,[30] and various support tools such as the Electronic White Pages and the CSA's online child support calculator.

2.17 Approximately 100 PIs are available to staff on the CSA's Intranet, covering topics such as payment options, agreements, capacity to pay and change in care arrangements. The ANAO examined a number of PIs and observed that they appeared to be comprehensive. CSA staff interviewed during fieldwork confirmed this impression, but advised that although the PIs are comprehensive, they are not widely used during phone calls with customers, as they are quite long. CSOs indicated that they found it difficult to locate the required information quickly.[31]

2.18 Several CSOs reported that the CSA's Online Law and Policy Guide is very useful and easy to use. However, CSOs said that the Guide does not

[30] CSA's online law and policy Guide, available at: <http://www.csa.gov.au/guide/index.htm>.

[31] For example, *PI–Agreements* is 20 pages long, including more than 13 pages of step-by-step instructions on how to process a child support agreement.

provide guidance on procedures or protocols for data entry in Cuba. CSA staff advised that their preferred, and most often used source of support and reference for data entry matters, was their colleagues. Cuba has an inbuilt Help Function, although most CSOs interviewed—in three CSA sites—described the help function as less than useful.

Quality assurance

2.19 During fieldwork, a variety of quality assurance methodologies were observed in use at different sites. At one site, staff interviewed advised that there is limited checking of the quality of data entered into Cuba, with managers relying on staff to report instances of incorrect processing. Remedial action would then be taken to correct data entry errors, often with the assistance of a Technical Support Officer (TSO). Staff at another site advised that 100 per cent of transactions performed by new staff are checked during their first four weeks in the job. Following this, a sample of the CSO's work would be checked between five and ten times per week. The limited sample of three CSA sites indicates that data entry quality assurance practices vary across sites.

2.20 The CSA staff interviewed advised that team leaders conduct *ad hoc* checks from time to time, involving a sample of case records, in order to examine accuracy in respect of a particular issue—for example the recording of care levels.

2.21 The CSA commissioned a review of data quality and integrity in Cuba, and received a consultant's report in January 2007.[32] One of the recommendations of this review encouraged the CSA to 'establish a Data Quality project to drive forward data cleansing and process improvement.' The CSA responded positively to this recommendation and established a Data Quality Improvement Programme at the beginning of 2007.

2.22 The Data Quality Improvement Team involves representatives from each CSA Division and State Office. The team is actively engaged in identifying data quality and integrity issues and ensuring appropriate remedial work is undertaken either by a team member or the relevant business area within the CSA.

2.23 The CSA's Draft Data Quality Improvement Programme Framework states:

[32] CPT Global, *Data Quality and Integrity Review–Cuba, Final Report,* 2007.

The objective of the Data Quality Improvement Program is the continuous improvement in the quality of the CSA's data and associated business processes through the joint involvement of all stakeholders to:

- detect (potential) data quality issues;
- quantify and analyse data quality issues;
- identify the reasons (root cause) of data quality issues;
- establish and implement improvement strategies (including system enhancements, improvements to business processes and education and training strategies); and
- establish appropriate reporting mechanisms to monitor improvements.

[In addition, the framework envisages]

The successful establishment of the Data Quality Improvement Program will provide continuous improvement in the quality of the CSA's data including:

- increased data integrity;
- reduction in rework and work-arounds;
- provide greater integrity in the delivery of the Child Support Scheme;
- improved transparency of business processes; and
- provide the Executive with a degree of confidence of the quality of the CSA's data.[33]

2.24 Although the Data Quality Improvement Programme is in its early stages, the ANAO encourages the CSA to continue to develop and implement a programme of continuous improvement in data quality. The results of the analysis of Cuba tables presented in this report should assist the CSA in setting priorities for data cleansing and improving data entry controls.

[33] CSA, *Draft Data Quality Improvement Program Framework,* 2007, p.7.

3. Case Records

This chapter presents the results of the ANAO's analysis of three tables supporting data related to CSA cases. These tables were: case, case indicator and case liability.

Case table

3.1 The CSA provided the ANAO with a copy of the Cuba table 'case'. The file contained 1 439 570 lines of data, each line containing information on one case. Of these, 1 033 631 records were identified as having been created during the conversion process, when data in the previous child support database was migrated to Cuba, in March 2002. The remaining 405 939 cases displayed a CSA user identification code, indicating that these records had been created by Child Support Officers directly in Cuba.

Case identification number

3.2 The case table contains the case identification number—the primary key for the data construct 'case'. All 1 439 570 records contained unique values. That is, no values in the primary key, for case, were duplicated.

Case status

3.3 The case table also contains information on the status of each case and the dates upon which liability begins and ends for each case. The case status field holds a four-character code indicating the status of each case. The CSA's Entity/Attribute Definition for the case table states that valid values for these codes are stored in a code table with a code type of 'Case Status'.[34]

3.4 All entries in the case status code field consisted of valid values. The distribution of case status codes is shown in Table 3.1. Using these case status codes, the ANAO calculated that 789 297 cases were either active or had ended with arrears outstanding, at the time of extracting the data from Cuba in April 2007. This number compares well with the figure of 781 026, published by the CSA in *Child Support Scheme—Facts and Figures 2005–06.*[35]

---

[34] CSA, *Entity/Attribute Definition for Entity Type: CASE*. Provided to ANAO 23 March 2007.

[35] CSA, *Child Support Scheme, Facts and Figures, 2005–06*, p. 4. Reported as CSA Caseload (all active cases and cases ended with arrears as at 30 June 2006).

Table 3.1

Distribution of case status codes

No. of records	Case status code	Description
743 992	ACTV	Case Active - Case has one or more Current or Suspended Entitlements
12 406	CAN	Registration Cancelled
516 706	CLSD	Case Closed
45 305	EWAR	Case Ended and Arrears Outstanding
161	EWLD	Case Ended with Liability Calculation Due
111 144	INEG	Case Ineligible
2 313	REGN	Registration in Progress
7 543	WDRN	Case Withdrawn
1 439 570		Total

Source: ANAO analysis.

Case liability start and end dates

3.5 The ANAO examined the fields describing the case liability start date and case liability end date. The CSA defines these fields as:

- LBLTY_STRT_DT — date the registration was accepted by CSA and the case begins; and
- LBLTY_END_DT — date the last liability in the case ends.[36]

3.6 In 76 425 records, the liability start date was recorded as 01/01/0001. This is often referred to in the CSA's database documentation as a 'low date'. While not a valid date itself, the low date is used in a variety of circumstances—as a placeholder for a valid date; to indicate that the true date is unknown; or to indicate that a valid date is not able to be recorded at a particular point in processing a record.

3.7 For example, if an application for child support is received by the CSA and subsequently determined to be ineligible, it is still necessary for the CSA to record the fact that the application was made, and the result of the application. For an ineligible case, there is no genuine liability start or end date. In such an instance, a case record is created with a case status of INEG and liability start date and liability end date are both set to 01/01/0001.

[36] CSA, *Entity/Attribute Definition for Entity Type: CASE*. Provided to ANAO 23 March 2007.

3.8 In 76 263 of the 76 425 cases mentioned above, both the liability start date and liability end date display 01/01/0001, and consistent with the practice outlined in the example given above, all have a case status of:

- INEG (ineligible) n = 74 619;
- CAN (cancelled) n = 203;
- WDRN (withdrawn) n = 1 601; or
- REGN (registration in progress) n = 2.

3.9 However, 162 records were identified where the liability start date was recorded as 01/01/0001 and the liability end date was recorded as a valid date within the range 11/03/2002 to 02/03/2007. The case status codes for these cases were CAN (n=152), INEG (n=3) or WDRN (n=7), and therefore, these records do not involve any active cases.[37] As such, this situation represents a negligible business risk to the CSA's management of active child support cases, but reflects an inconsistency in the practice of using low dates for recording liability start and end dates for inactive cases.

3.10 In addition, 36 406 records were identified where the liability start date was recorded as 31/12/4000. This date is often used as a 'high date' within the Cuba database. A high date usually indicates a current line of data in a table. When a line of data is superseded or 'end-dated' that line becomes part of an historical record. It is replaced by a (new) current line of data, with an end date set to 31/12/4000. A high date may also be used if the true date is unknown, but it is expected to be known at some point in the future. Of the 36 406 cases with a liability start date set as a high date:

- 454 records displayed a liability end date set as a low date (01/01/0001);
- 27 067 records displayed a liability end date set as a high date (31/12/4000); and
- the remaining 8885 records displayed a valid liability end date within the range 29/03/2000 to 31/03/2007.

3.11 The third category, above, represents an anomaly as the dates indicate that liability for those cases ended sometime between the years 2000 to 2007 inclusive, but that liability will not start for those cases until the year 4000. Once again, the case status codes for these records did not include any active

[37] The majority, but not all, of these 162 records were conversion records (CNVRT).

or ended with arrears cases, and therefore, the business risk arising from these data anomalies is negligible.

3.12 Table 3.2 shows how low dates, high dates and valid dates are used in combination. For example, the first row of the table shows that 27 067 cases have a high date set as both the liability start date and liability end date, and that none of those cases is active or has ended with arrears.

Table 3.2

Use of low and high values in liability start and end dates

Liability start date	Liability end date	No. records All records (ACTV & EWAR)	
High	High	27 067	(0)
High	Valid	8 885	(0)
High	Low	454	(0)
Low	Low	76 263	(0)
Low	Valid	162	(0)
Low	High	0	(0)
Valid	Low	3 800	(0)
Valid	Valid	1 319 897 **See discussion below**	(787 393)
Valid	High	3 042 **See discussion below**	(1 904)
Total		1 439 570	(789 297)

Source: ANAO analysis.

3.13 As illustrated in Table 3.2, the analysis revealed an inconsistent use of low dates and high dates to record liability start and end dates within the case table of Cuba. While many of these instances do not represent a significant business risk to the CSA's administration of current cases, the last two rows of Table 3.2 (valid/valid and valid/high) show that some active cases may be affected.

3.14 For the 1 322 939 records displaying a valid liability start date and either a valid or high liability end date,[38] the range of liability end dates extended from 07/07/1988 to 31/12/4000. Within this group, a significant number of liability end dates displayed alternatives to the standard high date

[38] 1 319 897 + 3 042 = 1 322 939.

of 31/12/4000. For example, end dates of 30/12/4000, 01/12/4000, 29/11/4000, 31/12/3999, 01/01/3000, 31/12/2999, 28/07/2103, and 31/12/2099 were observed.

3.15 The ANAO selected an arbitrary cut off value of 01/01/2050 to divide this group into two subgroups for further analysis—essentially, those illustrated in the last two rows of Table 3.2. That is, 1 319 897 records displayed a liability end date between 07/07/1988 and 31/12/2049, while 3042 records displayed a liability end date later than or equal to 01/01/2050.

3.16 Of the 3042 cases with a liability end date beyond 01/01/2050, 1904 are either active or have ended with arrears.[39] These liability end dates appear to be unreasonable. That is, a liability end date beyond the year 2050 is not reasonable for a case commenced during or before 2007. The legislation underpinning the Child Support Scheme provides for child support up until the youngest eligible child turns 18 years.[40] [41] On this basis, the 'true or genuine' liability end date for these 1904 cases is likely to be prior to 01/01/2050. In fact, the genuine liability end date for Stage 2 cases is likely to be prior to 01/04/2025—that is, 18 years after the data extract.

3.17 Of the 1 319 897 cases with a valid liability start and end date, 787 393 are ACTV or EWAR. The ANAO examined the relationship between start and end dates for these cases.

Apparent duration of cases

3.18 The ANAO calculated the apparent length of these cases based on recorded values for case liability start and end dates. According to the values recorded, the longest case will run for 60 years[42] and 99 207 cases will exceed 18 years. Of these, 97 000 display a case status code of ACTV and a further 102 display a case status code of EWAR. Cases displaying an apparent duration in excess of 19 years may reflect an anomaly in the data—specifically, the presence of unreliable liability end dates.

---

[39] That is, the case status code shows that the case has ended with arrears (EWAR), despite the fact that the liability end date is at some point in the future, beyond 01/01/2050.

[40] Stage 1 of the Child Support Scheme was introduced by the *Child Support (Registration and Collection) Act 1988* and relates to court orders and court registered agreements. Stage 2 was introduced by the *Child Support (Assessment) Act 1989* and relates to administrative assessments under a formula.

[41] *Child Support (Assessment) Act 1989* s12 states that 'a child support terminating event happens if:…the child turns 18;..'. In certain cases the period may be extended until the end of the school year in which the child turns 18 years—s151B(1). Some Stage 1 cases, involving court orders and spousal maintenance, may have extended periods of liability.

[42] Liability start date of 13/02/1990 and liability end date of 31/12/2049—this case displays a status of ACTV.

3.19 Table 3.3 below, shows the case status and liability start and end dates recorded for eight cases with an apparent duration in excess of 45 years. The table shows case status codes of ACTV, EWAR, CLSD and INEG associated with these cases. Each of these case records demonstrates an anomaly—such as the ineligible case with a liability start date in the past and a liability end date in the future, and the closed case with a liability end date in the future.

Table 3.3

Apparent duration of cases—exceeding 45 years

Status Code	Liability Start Date	Liability End Date	Apparent case duration (yrs)
ACTV	23/03/1989	4/08/2034	45
ACTV	6/02/1990	15/12/2035	46
INEG	13/12/1966	15/09/2017	51
ACTV	13/05/1994	8/05/2045	51
ACTV	21/05/1992	1/01/2044	52
CLSD	6/02/1995	9/04/2047	52
EWAR	1/05/1989	1/05/2048	59
ACTV	13/02/1990	31/12/2049	60

Source: ANAO analysis.

3.20 Analysis also revealed 8937 cases where the recorded values for liability start date and liability end date were identical.[43] In the majority of these cases—6702—the case status was CLSD. In addition, five cases were WDRN, 480 were INEG and 374 were CAN. However, one ACTV case and 1370 EWAR cases displayed the same date for liability start and end. These records demonstrate an anomaly in that an ACTV or EWAR case is unlikely to have commenced and ended on the same day.

3.21 One EWAR case was identified where the liability end date preceded the liability start date—a logical impossibility. In addition, the ANAO identified five ACTV cases and 16 REGN cases with a recorded liability end date in the past.[44] This situation represents an anomaly as these cases should, logically, not display a liability end date in the past. A case status of REGN indicates that the CSA is processing the application for registration. While it may be possible for such a case to display a liability end date just prior to the

[43] The liability start and end dates for these cases did not include a low date or high date value.

[44] That is, before the date of the data extract.

date of data extraction, some liability end dates for these cases were as early as 1995 and 1997.

Early liability start dates

3.22 Testing of the liability start dates recorded in the case table, identified 26 active cases with recorded start dates before 01/06/1988—the commencement of the Child Support Scheme. The earliest liability start date identified was 06/10/1976. These 26 active cases also display an apparent case duration of between 22 to 42 years.

3.23 Some of these cases may relate to applications for spousal maintenance or to the registration of court orders in existence at 01/06/1988. However, given the apparent duration of these ACTV cases, and the fact that they all display liability end dates in the future, the CSA was requested to investigate the circumstances surrounding these records. In September 2007, the CSA advised that:

> CSA has investigated all the records and has identified that they all refer to circumstances where the registration date, for the case, is prior to the commencement of the Child Support Scheme. The particular registration record is in fact, refused and has never been active - these cases have more than one registration recorded. The active registration record is correct and there is no impact on customers.
>
> CSA agrees that these records appear anomalous, however, data correction is not required.[45]

Audit findings

3.24 Following an examination of selected fields within the case table, it was concluded that:

- the primary key for the case table was sound;
- the use of case status codes conformed to the CSA's business rules and technical specifications;
- the use of low dates and high dates, as case liability start dates and case liability end dates, was inconsistent;
- some case liability start dates—preceding the commencement of the Child Support Scheme—may be invalid or unreliable; and

[45] Advice from the CSA—18 September 2007.

- some case liability end dates may be invalid or unreliable—as they indicate extraordinary apparent case durations.

3.25 The business impact of data anomalies in the case table is minimal. The number of anomalous records is relatively small compared to the total number of records in the case dataset. If some of the cases displaying an active status, but liability end dates well into the future, have actually ended, this will affect the accuracy of the CSA's summary statistics for active cases.

Case indicators table

3.26 While the case table holds the primary key for the data construct 'case', it is supported by a number of additional tables that rely on the case primary key, and hold other relevant information on CSA cases. One such table is called case indicators. This table holds information on one or more indicators for a case. An indicator may show whether a case is private collect or CSA collect,[46] whether the case is currently associated with a low income non-enforcement period, or whether the case is associated with a Social Security Appeals Tribunal review. A case may have more than one indicator.

3.27 The case indicator table contained 1 623 770 lines of data, each line containing information on one instance of a case indicator pertaining to a case.

3.28 The indicator type code field holds a code defining each case indicator. Valid values for these codes are stored in the code table with a code type of 'Case Indicator Type'.[47] All entries in the indicator type code field consisted of valid values. The distribution of indicator type codes is shown in Table 3.4.

[46] A private collect case is one in which the payer makes child support payments directly to the payee, under a private arrangement. A CSA collect (or agency collect) case is one in which the payer makes child support payments to the CSA, either directly or through employer withholdings. The CSA then passes the payments to the payee.

[47] CSA, *Entity/Attribute Definition for Entity Type: CASE_INDICATORS.* Provided to ANAO 23 March 2007.

Table 3.4

Distribution of case indicator type codes

No. of records	Indicator type code	Description
715 683	COLL	CSA Collect
30 308	INTI	International In Case
44 983	INTO	International Out Case
9 340	INTU	International Unknown case
9 674	LINE	Low Income Non-Enforcement period
630 218	NCOL	Private Collect
4 427	ONC	Overall Non Care
174 822	PCC	Private collect candidate
—	PENF	Payee Enforcement in progress
4 114	REQ	CSA Private Coll (Registrar initiated opt out)
201	SSAT	SSAT Review Lodged

Source: ANAO analysis.

Note: Each code represents a specific characteristic of a case. For example, 'International In Case' indicates that the payee resides outside Australia and the payer resides in Australia. The specific meaning of each of these codes is not directly relevant to the matter under discussion, however, further information on the Child Support Scheme may be obtained from the CSA's Internet site, www.csa.gov.au.

3.29 There were 1 119 285 unique case identification numbers identified within the 1 623 770 lines of case indicator data. Given that the case table held 1 439 570 unique case identification numbers, there are 320 285 cases that have no case indicators associated with them. Of these 170 039 cases are ACTV.

3.30 The ANAO anticipated using the case indicators COLL and NCOL to distinguish between cases that were CSA collect or private collect. However, given the number of cases not displaying any case indicators, we sought further information from the CSA on how cases were identified as either CSA collect or private collect. The CSA advised that the field describing payment method exists elsewhere in Cuba and that the case indicator table is only updated when a change in collection method occurs. Therefore, we were not able to use the COLL and NCOL indicators in the case indicator table to reliably distinguish between CSA collect and private collect customers.

Case indicator start dates and end dates

3.31 The CSA defines the fields 'start date' and 'end date' as:

Start date (Mandatory)—Date from which the case indicator becomes valid, usually the change processed date; and

End date (Mandatory)—Date on which the case indicator ends. [48]

3.32 Analysis revealed that there were 2496 instances where the recorded end date was earlier than the recorded start date—1582 of these displayed a case status code of ACTV and 219 displayed a case status code of EWAR. These records demonstrate a weakness in the integrity of dates describing the valid period of case indicators—as indicators should not end before they start.

3.33 While 2240 of these records displayed the start date as being only one day after the end date, others displayed the end date as being well before the start date. In the most extreme instance a case indicator record showed an end date of 13/08/1992 and a start date of 26/10/1999.

Audit finding

3.34 Following an examination of selected fields within the case indicator table, it was concluded that:

- the use of indicator type codes conformed to the CSA's business rules and technical specifications; and
- a relatively small number of records demonstrate a weakness in the integrity of start and end dates pertaining to some case indicators.

Case liability table

3.35 The CSA provided a Cuba table called case liability. The file contained 27 567 612 lines of data, each line containing information on a single calculation of liability for a case.[49] A case may have liability calculations performed a number of times, over the life of the case. Often, these case liability calculations relate to different child support periods. Therefore, many cases display multiple entries in the case liability table.

3.36 The twenty–seven and a half million lines of data in the case liability table represent liability calculations for 1 319 768 unique cases.[50] The average number of liability calculations per case was 21. One particular case displayed 947 separate liability calculations, while two other cases displayed in excess of

[48] CSA, *Entity/Attribute Definition for Entity Type: CASE_INDICATORS.* Provided to ANAO 23 March 2007.

[49] Stage 1 cases contain liability calculations for each child or spouse claimed for. Stage 2 cases contain liability calculations for the liable party only.

[50] Based on a count of unique values for case identification number.

800 separate liability calculations. In all, 8748 cases (or 0.7 per cent of the 1.3 million cases represented in the case liability table) displayed 100 or more case liability determinations. There were 837 671 cases (or 63.5 per cent of the 1.3 million cases represented in the case liability table) that displayed fewer than 22 separate case liability determinations and 42 632 cases that displayed only one case liability determination.

3.37 For those cases displaying more than one case liability calculation, the ANAO sought to determine which particular line of data represented the current liability determination attached to each case. According to the CSA's Entity/Attribute Definition for the entity case liability, two fields identify when a line is created and end-dated.

> START_DATE (mandatory) — Date the row was created; and
>
> END_DATE (mandatory) — Date the row was end-dated. If current, equals high values.[51]

3.38 Analysis confirmed that all 27 567 612 lines of data held a valid start date, within the range 01/06/1988 to 31/03/2007. The ANAO also examined the end date field and found that 16 691 535 lines (60 per cent) held a valid end date within the range 08/11/1988 to 31/03/2007, while 10 876 077 lines (40 per cent) in the case liability table displayed an end date of 31/12/4000.

3.39 The table contained many instances of an individual case displaying multiple rows of case liability data with an end date of 31/12/4000. Therefore, end date, alone, was insufficient to identify the current line of case liability data pertaining to a particular case. Two other fields in the case liability table are:

> PERIOD_START_DATE (mandatory) — First day of the period on which the calculation is based; and
>
> PERIOD_END_DATE (mandatory) — Last day of the period on which the calculation is based.[52]

3.40 All records in the table contained an entry for period start date and period end date.[53] However, the earliest period start date recorded was 23/07/1972—this date precedes the commencement of the CSA by 16 years. In

---

[51] CSA, *Entity/Attribute Definition for the entity CASE_LIABILITY,* provided to ANAO on 23 March 2007.

[52] ibid.

[53] This means that the mandatory requirement for these fields to contain data is fulfilled.

all, 166 lines of data in the case liability table displayed a period start date prior to 01/06/1988.

3.41 The latest recorded value for period start date was 31/12/2099, with the year 2099 appearing in 45 lines of data—indicating that the calculations relate to a period commencing in 92 years time. In all, 97 830 lines of data in the case liability table showed a period start date later than the data extract date—and therefore, in the future. While some case liability calculations may be appropriate for child support periods commencing in the (near) future, the case liability table contains 2393 entries displaying a period start date at least one year in the future.

3.42 The earliest recorded values for a period end date were 25/02/1948 and 22/11/1949—most likely data entry errors—with 25 lines in total displaying a period end date prior to the commencement of the CSA.[54]

3.43 The latest recorded value for a period end date was the high date of 31/12/4000—this value appeared in 9404 lines of data. Other period end date values spanned various days and months in the years 4000, 3999, 3000, 2999, 2205, 2199, and numerous other years well into the future. In total, 12 497 lines of data in the case liability tables displayed a period end date after the year 2025.

3.44 In many instances, these anomalies related to cases that had been closed, deemed ineligible or cancelled. In order to assess the potential impact of liability start and end date anomalies on current CSA cases, the ANAO examined case liability information for cases with a status of ACTV or EWAR.

3.45 Given the large number of lines of data in the case liability table, the ANAO approached the CSA for advice on a method to reliably select the line of data that represents the current case liability determination for each case. The CSA advised that the line containing the current case liability determination could be identified where the date of the data extract, 01/04/2007, fell between the start date and end date AND also fell between the period start date and period end date.

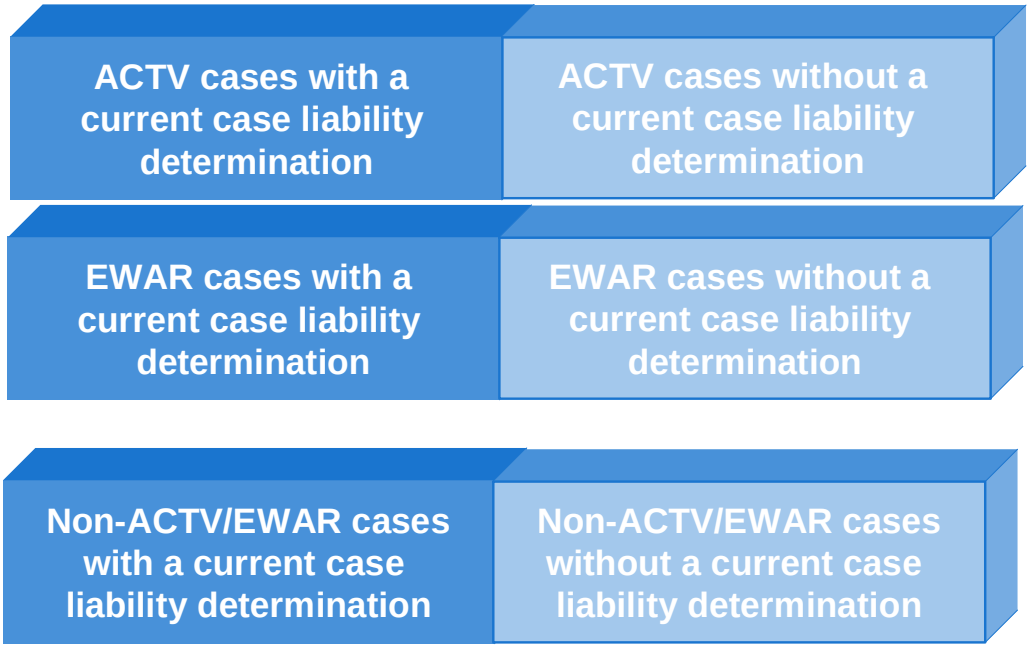
3.46 Employing the CSA's advice, the ANAO examined the 27.5 million lines of case liability data and identified 745 296 current case liability determinations relating to 744 398 individual cases.[55]

[54] Many, but not all of these, displayed a period start date after the period end date—a logical impossibility given the definitions of these two fields.

3.47 This information was divided into three categories—represented in Figure 3.1 below. The following analysis concentrates on these current case liability determinations.

Figure 3.1

CSA cases and case liability determinations



Source: ANAO analysis of case liability and case tables extracted from Cuba.

3.48 The subsequent analysis considered the following combination of circumstances:

1 - ACTV cases *with* a current liability determination;
- ACTV cases *without* a current liability determination;

2 - EWAR cases *with* a current liability determination;
- EWAR cases *without* a current liability determination;

3 - nonACTV/EWAR cases *with* a current liability determination; and
- nonACTV/EWAR cases *without* a current liability determination.

[55] That is, a small number of cases held more than one current case liability determination. The CSA informed the ANAO that Stage 1 cases held a liability assessment for each eligible person (spouse and/or children).

3.49 Within the first grouping, 743 125 ACTV cases showed a current case liability determination,[56] and 867 ACTV cases did not show a current case liability determination.[57] In one of the 867 ACTV cases noted above, no corresponding entries for that case in the case liability table were identified. This means that in one instance, an active case is not associated with any case liability determinations—either historical or current.

3.50 In the remaining 866 cases, 115 displayed an end date one day prior to the date of extracting the data from Cuba. Therefore, these 115 cases were most likely being updated during the data extraction process, and that a new case liability entry was pending. The remaining 751 cases were associated with 12 452 lines of data in the case liability table, however, none of these met the CSA criteria for currency.

3.51 Within the second grouping shown at paragraph 3.48, 97 EWAR cases[58] displayed a current case liability. In 17 cases the current liability was zero. In 79 cases the current liability amount ranged between $52 and $12 000. In one case, there were two current case liability determinations—one for zero dollars, the other for $260. On the face of it, a case that has ended would not normally be expected to be associated with a current liability determination,[59] and therefore, these 79 cases appear to be anomalous.

3.52 Within the third grouping shown at paragraph 3.48, 1181 lines of case liability data were associated with 1167 non-ACTV or EWAR cases. Of these, 698 displayed a current case liability of zero, while 483 displayed a current case liability within the range $1 to $33 916. Once again, one would not normally expect a non-active case to be associated with a current liability determination.

3.53 In September 2007, the CSA advised that it had investigated these cases[60] and that:

> **Analysis:** The cases relate to spousal maintenance and court orders where not all relevant details have been populated - i.e. 'court order made on dates'.

---

[56] At least one current case liability determination. Stage 1 cases may have more than one current case liability determination.

[57] According to the criteria supplied by the CSA regarding the identification of a current line of data in the case liability table.

[58] In 98 lines of case liability data. That is, one Stage 1 case held two current liability entries.

[59] The ANAO examined the CSA's functional specifications relating to the Common Action Block for Liability Determination (CSA Reference: v12.7_SS_28022006) but could not identify any specific processing rule relating to the calculation of case liability determinations for EWAR cases.

[60] That is, the 79 cases referred to in paragraph 3.51 and the 483 cases referred to in paragraph 3.52.

This particular aspect of this issue relates back to the conversion of legacy cases to Cuba. CSA is currently undertaking corrective action in relation to each of these cases, in line with the *Court Variations Procedural Instruction* - specifically relating to the correction of Stage 1 converted cases.

CSA Response: CSA is already aware of these conversion issues and will correct the data accordingly. CSA wishes to note that a Procedural Instruction, *Court Variations*, exists (since March 2002) to specifically address the correction of the cases.

Customer Impact:

– [For the 79 cases referred to at paragraph 3.51]—nil, because liability is still being generated correctly;

– [For the 483 cases referred to at paragraph 3.52]—nil, non active registrations, and as such CSA is not using the court order liability that has been generated.[61]

Audit findings

3.54 Following an examination of selected fields within the case liability table, it was concluded that the table:

- contains some period start dates and period end dates preceding the commencement of the Child Support Scheme;
- contains some period start dates well into the future, compounded by the inconsistent use of high dates; and
- displays a number of inconsistencies between case status and current case liability determination.

3.55 Once again, the number of records identified as anomalous or unreliable is relatively small compared to the entire case liability table dataset. Therefore, the likely impact of these data integrity issues on the effective administration of the Child Support Scheme is slight. However, regardless of the overall impact on the scheme, an error in liability calculations may have a significant impact on the particular customers involved.

[61] Advice from the CSA—18 September 2007.

Recommendation No.1

3.56 The ANAO recommends that the CSA:

- investigate and correct records in the case, case indicator and case liability tables that display anomalies in the various start and end date fields;
- ensure a consistent use by CSOs of high dates and low dates in case liability start date and case liability end date fields; and
- resolve instances of inconsistencies between case status and current case liability determination.

DHS's response:

3.57 DHS agrees that the anomalies identified will need to be corrected. To this end, CSA will review all anomalies identified and develop appropriate corrective action, in line with material impacts to customers. Furthermore, CSA will review the inconsistent high dates and low dates, and active cases without a current case liability determination, outlined in the audit report, and will undertake corrective action, where required, in addition to ongoing monitoring of these fields, via CSA's Data Quality Improvement Program.

4. Customer Records

This chapter presents the results of the ANAO's analysis of selected fields in a number of Cuba tables supporting data related to CSA customers. Among these were tables named: customer; customer indicators; individual customer; non-individual [employer] customer; individual customer address; non-individual customer address; and individual customer name (history).

Customer tables

4.1 Similar to the structure of the data construct 'case', Cuba incorporates a data construct of 'customer'. A series of related tables hold information about the CSA's customers. The CSA provided a copy of the customer table. This is the table holding the primary key for customer records. The file contained 4 829 880 lines of data, each line containing a customer identification number,[62] a code to indicate the type of customer (an individual or a non-individual—such as an employer) and system information on when the record was created.

Customer identification number

4.2 Inspection of the field holding the customers' identification numbers revealed that the lowest value was 26 and the highest value was 999 999 999. All 4 829 880 records contained unique values. That is, no values in the primary key, for customer, were duplicated. There were 3 454 462 records, or 71.5 per cent of all customer records, identified as converted records, created during the data conversion process of 2002.

Customer type code

4.3 Only two legal values are listed in the code table for customer type—individual (IND) and non-individual (NIND). Essentially, an individual is a person and a non-individual is an organisation, such as an employer or overseas child support agency. All entries in the customer type code field conformed to one of the legal values. Examination confirmed 4 571 020 records marked as IND and 258 860 records marked as NIND.[63]

---

[62] Also known as the Child Support Identification Number, or CSID.

[63] One of the CSIDs held the value 999999999. This value did not appear in the non-individual customer table (which held only 258 859 records).

Customer indicators

4.4 Similar to the use of case indicators referred to previously, Cuba employs a number of customer indicators to hold information on customers. The CSA provided a Cuba table called customer indicators. The file contained 11 247 722 lines of data, each line containing an entry relating to one indicator for a customer. A customer may have more than one indicator.

4.5 Some examples of customer indicators, and the number of customer records associated with them are given in Table 4.1. The full set of customer indicator codes is included at Appendix 2.

Table 4.1

Selected customer indicator codes

No. of records	Customer indicator code	Description
2 681 530	CNLK	Customer is also a Centrelink customer
9 098	INTR	Customer requires an interpreter
759 111	TRAC	Tracing activity being undertaken in relation to customer
32 524	DEAD	Customer is deceased
1 694 651	DEFT	Default income used for customer

Source: ANAO analysis.

4.6 All values in the customer indicator type code field were included in the list of legal values in the code table. In addition, nine legal values listed in the code table have not been used in the customer indicator table.

Customer indicator start and end dates

4.7 Analysis identified 10 635 lines of data that displayed an end date earlier than the start date. In 1644 instances the end date displayed a low date value of 01/01/0001. In 80 per cent of cases, the end date displayed preceded the start date by less than 10 days. However, a significant number of lines of data indicated a difference of some years between the start and end dates.

4.8 Once again, these records demonstrate a weakness in the integrity of dates describing the valid period of customer indicators—as indicators should not end before they start.

Customer identification

4.9 A number of fields in the customer table go to the issue of identifying CSA customers. The next phase of the analysis examined fields holding the customer's name—title, first name, middle name and surname—date of birth and date of death, address details, Tax File Number and Centrelink Reference Number.

Customer names

4.10 The CSA provided a copy of the individual customer table from Cuba. This is the table holding the bulk of personal information about individual CSA customers. The file contained 4 571 020 lines of data, each with a unique CSID.

4.11 There were 105 entries that contained blank entries in all name fields—title, first name, middle name and surname.

4.12 In addition, at least 245 name entries in the individual customer table appeared to relate to non-individual customers—that is, organisations, employers or overseas child support offices.

4.13 Among the entries for individual customer names, there were over 200 spurious entries that did not appear to constitute valid names but may represent training records. Some of these names were recorded as: 'XOXOXO', 'DUMMY CASE', 'ZZZZZ', MR X, and a number of single letter entries for the customer's first name and surname.

4.14 The CSA has published a number of Procedural Instructions for staff to follow in relation to identifying customers who have duplicate records on the database. A *Guide to Duplicate Payers/Payees Records, Version 1.1* directs staff to alter the name elements of a duplicate as follows:

> Add **XXX** to the front of the existing first name and surname (no spaces).
>
> Change the middle name of the duplicate customer record to **Duplicate.**
>
> e.g. change Reginald Edward Dwight to
>
> First name: **XXXReginald**
>
> Middle name: **Duplicate**
>
> Surname: **XXXDwight**

This ensures that the original name can be easily discerned from customer lists, involvement lists etc. while at the same time being clearly marked in a consistent manner as to be recognisable as a duplicate record.[64]

4.15 Within the individual customer table 12 783 records conformed to this protocol. However, further analysis revealed considerable variation in the marking of duplicate records. A sample of the various methods observed is presented in Table 4.2 below.

Table 4.2

Sample of various methods used to mark duplicate customer records

Title	First name	Second name	Surname
	XXX	DUPLICATE	XXX
Mr	Andrew		DUPLICATE
Mr	Archie	XXXXXXXXXXXX	DUPLICATE
	AlyshaXXXXXXXX	Kaye	DUPLICATE
DO NOT USE	Jane	DUPLICATE	DUPLICATE Smith
XXMr	Ray		Duplicate
			DUPLICATE CHILD

Source: ANAO analysis.

4.16 It was estimated that in addition to the 12 783 correctly marked duplicate records, up to 5000 other customer records, that are still recognisable as duplicates, have not been marked in accordance with the CSA's Procedural Instruction.

4.17 The individual customer table also includes many entries with a recorded surname of 'UNKNOWN'. In the previous CSA system, that is before Cuba, many child records did not incorporate a surname for the children associated with child support cases. On migrating customer data from the old system to Cuba, many child records continue to display 'UNKNOWN' for the child's surname. Analysis detected 645 957 records—14 per cent of all customer records, current and historical—displaying 'UNKNOWN' as the customer's surname. Of these, over 389 000 were associated with children linked to active cases.[65]

---

64 CSA, *A Guide to Duplicate Payers/Payees Records, Version 1.1,* 2002, p. 5.

65 The CSA advised the ANAO that it was seeking to improve the accuracy of child surname data through its Data Quality Improvement Programme, improved data exchange with Centrelink and as part of its collection of customer information in preparation for Stage 3 of the Child Support Scheme reforms. [Advice from the CSA—18 September 2007].

4.18 The ANAO sought to remove the blank entries, duplicates, spurious and invalid name entries from the table and concentrated on the remaining entries. Within this refined group, considerable variation in the recording of the various elements of customers' names was evident.

4.19 For example, the field 'current title' is described in the CSA's Entity/Attribute Definition for individual customer as 'Text that describes the name title.'[66] The code table lists 21 valid values for current title—such as: Mr, Ms, Br, Prof and Dr. The analysis identified 288 different values in this field, within the individual customer table. Over 200 of these values occur once only, and many appeared to contain one or more of the customer's name elements.

4.20 Other entries in the current title field appeared to arise from data entry errors or inconsistent use of abbreviations for customer title. For example, the titles 'Miss' and 'Mstr' appear in the code table as valid values. However, the following variations in coding for these titles were noted.

Table 4.3

Variation in coding customer title

Code displayed in individual customer table	Count
MISS	417 306
****MISS****	1
MIS	12
MISSS	2
MISS'	2
MISSQ	1
MISA	1
MIISS	1

Code displayed in individual customer table	Count
MASTER	110 808
MAS	324
MASER	1
MAST	166
MSTR	3

Source: ANAO analysis.

4.21 Within the individual customer table, over 900 records were identified where the first name field contained a blank entry, yet the surname field contained text. Often the surname field appeared to contain a first name and a

[66] CSA, *Entity/Attribute Definition for the entity: INDIVIDUAL CUSTOMER*. Provided to ANAO on 23 March 2007.

surname. On other occasions the surname field simply contained an apparent surname.[67]

4.22 Other records appeared to hold the customer's first name in the title field or other combinations of name elements in inappropriate fields. Further insight into inconsistencies in recording name elements became apparent during the analysis of multiple records for the same customer. This topic is covered in a later chapter of this report. However, in relation to the recording of name elements the following variations were noted, among records that have a high probability of representing undetected duplicate customer records.

Table 4.4

Examples of inconsistent recording of name elements

Title	First name	Second name	Surname
	Amanda	Jane	*Smyth*
Ms	Amanda	Jayne	*Smyth*
Miss	Jacqueline	L	*Jones*
	Jacqueline	Louise	*Jones*
	John	Lance Richard	*Brown*
Mr	John	Lance	*Brown*
Ms	Jan	Maree	*Smith Jones*
	Jan	Maree	*Smith – Jones*
	Jade	Ashley	*Smith*
Mr	Jade	Ashley *Smith*	*Smith*

Source: ANAO analysis. The names used are fictitious.

4.23 Further examination highlighted a particular name element anomaly—that is, where the customer's surname appears in the second name field as well as the surname field. There were 26 457 records in the individual customer table that displayed this characteristic. While these records account for less than one per cent of all customer records, they indicate a weakness in the CSA's practices for recording customers' names.

4.24 Based on the sample of recorded names examined in this audit, and excluding those records variously marked as 'unknown' and 'duplicates', it was estimated that up to one per cent of recorded names in the individual customer table may not conform to an accepted standard. The ANAO also

[67] A number of these records appeared to contain names of non-individual customers, although they appeared in the individual customer table.

considered whether these names formed part of the CSA's active customer dataset. That is, not all customer records in the individual customer table are actually associated, nor have they ever been associated, with a CSA case. (See the section on unlinked customer records later in this chapter).

4.25 Many of the spurious entries and probable training records fell into this category. These records represent a minimal risk to the administration of the Child Support Scheme—as they are not linked to a case. However, the records remain part of the CSA customer list—in the individual customer table—as redundant records.

4.26 For those records linked to CSA cases—current and historical—anomalies in recorded names make data matching more difficult. The CSA data matches with the ATO and Centrelink, among others, and may be forgoing positive matches with customer records in these agencies because of these anomalies in recorded names. Recommendation No.2 at paragraph 4.79 recommends that the CSA remove redundant records from the individual customer table.

Customer date of birth data

4.27 The ANAO examined the field that holds the customer's date of birth. Of the 4 571 020 records in the individual customer table, 86 220—or 1.9 per cent of all records in the table—displayed a value of 01/01/0001 for the date of birth. In 20 475 instances these records were not associated with any CSA case, and in 65 745 instances the records were associated with at least one CSA case.[68] Of those associated with a case, 6157 CSIDs were associated with a total of 6552 ACTV cases.

4.28 The CSA's Entity/Attribute Definition for the individual customer table states that date of birth is mandatory for a child. Within the 6552 ACTV cases associated with records displaying a date of birth of 01/01/0001, 1446 were associated with child records.[69] The majority of these child records contained a strong indication that the record was a duplicate entry, created in error, or that the child's surname was unknown. Nevertheless, these 1446 child records that, at some stage, had been associated with ACTV cases, failed the test of a valid entry in the mandatory date of birth field.

[68] In fact, the 65 745 CSIDs were associated with 88 920 cases.

[69] That is, CSIDs with a link to a role type code of CHD in the case_customer_role table.

4.29 One child record, with a date of birth of 01/01/0001, also displayed a current line of data associated with an ACTV case. The name field in that record indicated a duplicate record. However, a second record displaying a current line of data on the same case that shared the same name as the first record, but was not marked as a duplicate. This suggests that the child is claimed for twice on the same case.

4.30 The earliest recorded date of birth was 7 September 1891.[70] This customer's record did not have a date of death recorded. Furthermore, the customer was recorded as a payee on an active case.

4.31 Ten recorded dates of birth were prior to 1 January 1900 and another twenty records displayed a date of birth of 1 January 1900. Only one of these records displayed an entry in the date of death field. In addition, 22 of the 30 records were variously marked as duplicate records or customer UNKNOWN. The ANAO concluded that for the eight apparently valid customer records,[71] either the recorded dates of birth are in error or a date of death should be recorded against these records.

4.32 Continuing the analysis of outlier records in the date of birth field, the ANAO examined 39 customer records[72] that displayed:

- a role type code of PYR, PYE or SPS (Payer, Payee or Spouse);
- a date of birth indicating that the customer's age was 85 years or older;
- no date of death recorded; and
- an association with an ACTV case, with a process end date of 31/12/4000 and a case liability end date in the future—indicating an active line of data.

4.33 According to the data held in Cuba, the oldest payee is 116 years of age with their case liability start date in 2004 and case liability end date in 2020. The data also indicate that five payers and one payee/spouse, who are in their nineties, will be associated with ACTV cases until 31/12/2099.

---

70 There were three records with a recorded date of birth prior to this. However, those three records were marked as duplicates.

71 That is, records not marked as duplicates.

72 The 39 records spanned 46 lines of data as some customers displayed entries with role type SPS and PYE. The ANAO also removed marked duplicate records and records displaying UNKNOWN in any of the name elements.

4.34 There may be CSA customers older than 85 years who are paying and/or receiving child support. However, our analysis of the CSA's data suggests that a combination of the use of high dates—for case liability end dates—and, in some cases at least, a weakness in the accuracy of recorded dates of birth has given rise to these anomalous records.

4.35 In some cases the anomaly may also have arisen due to a lack of information about the death of a customer. Once again, combined with the use of sometimes unrealistically high dates for case liability end dates, in the absence of other information, these active cases and customers will remain in that state on the database. Analysis of the records strongly suggests they are current, and presumably, the CSA has included these records in its published statistics about the number of active cases and customers. If so, those statistics will be in error, although the number of customers/cases involved appears to be relatively small.

Date of death data

4.36 Of the 4 571 020 records in the individual customer table, 32 918 records displayed a date of death. Nine customers had the same date recorded for their date of birth and date of death.

4.37 The ANAO examined records displaying a date of death, where the customer records were also attached to ACTV or EWAR cases. In total, 5847 customer records were identified in this category. However, upon further investigation, 5828 of these records displayed an entitlement end date or process end date in the past.[73] This left 19 customer records that displayed:

- a date of death between 20/03/1993 and 26/02/2007;
- a high date of 31/12/4000 in the process end date field—and therefore represent active lines of data;
- an entitlement end date between 16/07/2007 and 31/12/4000;
- a case status of ACTV or EWAR—(16 ACTV and 3 EWAR); and
- no entries in either the terminating event code field or the terminating event date field.

4.38 The 19 records displayed various entries in fields indicating eligibility or ineligibility for Stage 1 or Stage 2 of the Child Support Scheme. These

[73] That is, prior to 01/04/2007, the date of data extraction. Process end date refers to the date when the line of data in the table was ended.

entries appear inconsistent in that they display a date of death and an entitlement date in the future.

Child records

4.39 Only in exceptional cases does the Child Support Scheme provide for child support entitlements after a child reaches 18 years of age.[74] The ANAO examined the CSA's data for child records (role type code = CHD) that displayed a date of birth indicating the person was 19 years of age or older. In total, 2236 lines of data were identified in the case_customer_role table,[75] that indicated a child record where the age of the child was 19 years or older and the record was associated with an ACTV case.[76] The ages displayed on these records ranged from 19 to 59 years. Linking the CSIDs contained in the 2236 lines of data with associated records for all cases in which these particular customers were involved, produced a dataset of 4355 records.[77]

4.40 However, some of the records related to Stage 1 cases,[78] or indicated that the child's Stage 2 eligibility had ended or the child role code indicated that the child was not claimed for in a particular case (child role code = T). In an attempt to identify a high risk group of records where the recorded date of birth was inconsistent with other characteristics of the customer record, the ANAO undertook further analysis and identified 17 records, where the:

- role type code displayed CHD (child);
- child role code displayed C, R or D (claimed for, dependant of payer, or dependant of payee)—indicating a child role with an impact on the case liability calculation;
- date of birth gave the child's age within the range 19 to 51 years;
- Stage 2 eligibility indicator displayed Y, and the registration type code displayed ASS (assessment)—thereby removing records that may have been associated with a court order;

---

74 Although an entitlement may exist until the end of the school year in which a student turns 18 years.

75 The case_customer_role table shows the association of customers with cases and the various roles attributed to customers within cases.

76 These records also displayed a high date of 31/12/4000 for process_end_date and a future date for entitlement_end_date.

77 In other words, the ANAO mapped the 2 236 CSIDs back to case_customer_role to pick up cases other than ACTV, or cases in which the customer displayed a different role (ie. PYR, PYE or SPS).

78 Court orders and court registered agreements may provide for liability to extend beyond the age of 18 years.

- process end date displayed a high date of 31/12/4000 and the entitlement end date displayed a date in the future,[79] within the range 4/12/2007 to 17/03/2018—indicating a current line of data;
- child liability indicator did not display an N;[80] and
- record was associated with an ACTV case.

4.41 The 17 lines of data related to 16 people—one person (that is, one CSID) fulfilled the above criteria on each of two active cases. These customers are too old to be children currently associated with active cases. In some instances, the records were associated with other lines of data displaying a different role type code—such as payer or payee, suggesting that at least one of the records is in error.[81]

4.42 Analysis also identified 165 lines of data indicating 163 payers and payees, on ACTV cases, whose recorded date of birth indicated an age less than 18 years.[82] While this may be correct for many cases, the youngest payee record displayed a date of birth in 1997, suggesting the payee was 10 years old at the time of the data extraction. The youngest payer, according to the recorded date of birth, was 15 years old. The ANAO sought further information from the CSA in relation to the validity of the outlying records in this dataset. In September 2007, the CSA advised that:

> As part of CSA's Data Quality Improvement Programme, processes to regularly report these potential anomalies, and review the data, to ensure the data entered is reflective of the customer's circumstances, is currently being undertaken.[83]

---

79 As at the date of data extract—1 April 2007.

80 If the child liability indicator displays an N, the payer is not liable for the child in that case.

81 The ANAO is aware that a customer may have been a child in a CSA case at some point in the past, then assume the role of a payer or payee in another case, as an adult. Many of the combinations of role type code mentioned above do not seem to fit this set of circumstances. For example, one customer shows an involvement in a case as a payee and spouse. That case showed an entitlement start date of 28/06/1988 and entitlement end date of 31/12/2099. The registration type was court order. The person then shows as a child (R), in another case: registration type is Assessment. Yet the entitlement start date displays 6/09/1997 and entitlement end date 5/09/2015. Therefore, according to the database records, this person, whose recorded date of birth is in 1956, displays as a child record on a case that started nine years after a previous case started, in which the person was classified as a payee/spouse.

82 The dataset was filtered to include only those lines of data with a high date set in the process end date field.

83 Advice from the CSA—18 September 2007.

Customer address data

4.43 The ANAO examined the data in the customer address table of Cuba. The extract provided by the CSA contained 7 704 095 lines of data, covering current and historical address information for all customers. According to the CSA's Entity/Attribute Definition for the customer address table:

> STATUS CODE is mandatory. This is a code for the status of the address. Valid values are stored in code table with code type of 'Status'.[84]
>
> The Code Table indicates three valid values:
>
> - ACTV—active;
> - INV—invalid; and
> - REIN—reinstated.

4.44 Analysis identified 2 443 486 address entries with a status of ACTV and an end date in the future.[85] Within this group 2934 records[86] displayed blank entries in the field called address line 1—despite showing CSIDs and various entries in other fields of the address table. The CSA's Entity/Attribute Definition states that address line 1 is mandatory. Also identified were 897 records with a blank entry in the field called locality name. The CSA's Entity/Attribute Definition also states that locality name is mandatory.

4.45 At least 100 entries for address line 1 did not appear to constitute valid entries for a customer's address. In many instances the entry appeared to be in the form of a note, such as:

- 'Prefers to be called Cathy';
- 'Authority: Kelly D Smith';
- 'Surname formerly Jones';
- 'Don't tell PYE where PYR is';
- 'Address and Ph No is Mum's House'; and
- 'Please refer to case manager'.[87]

4.46 Other entries in address line 1 included customers' mobile phone numbers, passwords and telephone contact hours. These elements did not constitute valid address entries for CSA customers.

---

[84] Entity/Attribute Definition for Entity Type: CUSTOMER_ADDRESS. Provided to ANAO 23 March 2007.

[85] In order to ensure that lines of data that have been end dated are not included.

[86] Approximately 0.1 per cent of all ACTV address entries.

[87] The surnames used are fictitious.

4.47 A field called customer address type code is also mandatory and draws on codes stored in the Cuba code table. The following entries were observed with in this field.

Table 4.5

Address type codes

Code	Description	Number of entries
BUS	business	5 016
POS	postal	2 260 856
REG	registered	225
RES	residential	173 954
RIO	Not In Code Table	3 435

Source: ANAO analysis of customer address table—ACTV records.

4.48 As noted in the last row of Table 4.5, a code—'RIO'—was recorded for 3435 customer address records. However, that code does not appear as a valid code within the Code Table. The ANAO examined these records, but was not able to identify any unifying characteristic to account for the use of the RIO code.

4.49 In September 2007, the CSA advised that the code, RIO, stands for Registrar Initiated Opt Out and refers to the customer address applying at the time of the customer moving from CSA collect to private collect. The CSA agreed that it would update the Cuba Code Table to include the RIO code.

4.50 The CSA relies to a large extent on address information to communicate with its customers. In order to test the completeness of information stored in the customer address table, the ANAO compared the list of 1 299 948 payers and payees, whose records displayed a current[88] association on an active[89] CSA case, with the list of 2 443 486 ACTV[90] address records.

4.51 That comparison found that 167 992 of these customers did not display an active entry in the address table.[91] Matching the 167 992 customer records

[88] That is, the line of data in the case_customer_role table displayed an end date of 31/12/4000.

[89] That is, case status displayed ACTV.

[90] With an end date of 31/12/4000.

[91] With an end date in the future. Further analysis revealed that some CSIDs were associated with address records that displayed ACTV but had an end date in the past. See the following sections of the report for further detail.

with historical address information revealed that over 90 per cent of these customer records were associated with one or more address records that displayed a status of INV (invalid). In addition, two per cent were associated with UNC (unconfirmed) address records. Five per cent were associated with ACTV address records, however, those records displayed an end date in the past. In 1952 instances, the customer CSIDs did not display any entry in the customer address table at all.

4.52 On the basis of this analysis, it would appear that the CSA does not hold a current address for up to 12 per cent of payers and payees who are currently associated with at least one active CSA case. The ANAO assumes that many of these cases are private collect—although we could not test that assumption using the data provided by the CSA—and that the address details are simply out of date. Nevertheless, the lack of current address information for CSA customers would be expected to make effective communication with those customers more difficult. The ANAO notes that, in preparation for the introduction of Stage Three of the Child Support Scheme reforms, the CSA is attempting to update address details for all customers.

Postcode data

4.53 The ANAO examined all ACTV address entries in the customer address table, for the entries in the postcode field. Postcode is described as optional, in the Entity/Attribute Definition for customer address.

4.54 A zero appeared in the postcode field of 41 132 records. The majority of these related to overseas addresses and many of the records contained a notation 'Do Not Touch', or similar. These overseas address records are treated differently by the CSA and were ignored in the following analysis.

4.55 Comparing the recorded postcodes against the Australian Postcode Database, maintained by Australia Post,[92] revealed that 4201 ACTV entries in the customer address table recorded postcodes did not appear in the Australian Postcode Database.

4.56 The pattern of many of these entries suggests that the postcode for particular localities had been changed. That is, the customer address details in Cuba could well have been correct at the time of recording the postcode,

[92] The full Postcode Database is available from the following Internet site: <www1.auspost.com.au/> postcodes. The ANAO downloaded the file on 31 July 2007. This version of the Postcode Database displayed a message that it was last updated on 24 July 2007.

however, the actual value for the postcode has since changed and the record has not been updated.

4.57 However, some records displayed incorrect postcodes. For example, a number of inappropriate three-digit postcodes were recorded for addresses within the ACT.[93]

Tax File Numbers

4.58 Customers' Tax File Numbers (TFNs) are stored in the individual customer table. Of the 4 571 020 lines of data in the individual customer table, 2 430 680 did not contain a TFN, while 2 140 340—approximately 47 per cent of records in the individual customer table—contained an entry in the TFN field.

4.59 The ANAO sought and was provided with advice from the Australian Taxation Office on the format and structure of TFNs. The ATO advised that TFNs consist of either eight or nine digits—the eight-digit TFN is used by companies, partnerships and trusts, while the nine-digit TFN is used by individuals.[94] The ATO also provided an algorithm to check the validity of TFNs.[95]

4.60 Further analysis of the CSA's data extract identified 58 customer records where the recorded TFNs appeared to be invalid. In 56 cases the recorded TFNs consisted of more than nine digits—47 records displayed a ten-digit TFN and nine records displayed an eleven-digit TFN. In a further two records, the recorded TFNs failed the check digit algorithm used to verify the validity of TFNs.

4.61 Analysis also confirmed that no single value for a TFN appears more than once in the individual customer table. This is not to say that every TFN stored against a customer's record is correct, simply that no duplicate values were detected within the TFN field.

4.62 The ANAO also investigated Centrelink Reference Numbers (CRNs), stored as part of the CSA's customer records. A full treatment of the results of

---

93 The Australian Postcode Database lists only two valid three-digit postcodes for the ACT—200 as the postcode for ANU and 221 as an 'LVR Special Mailing' location for Barton. Entries in the address table in Cuba included 201, 203, 207, 209, 241, 243, 244, 249, 253, 266, and 267.

94 The ATO TFN Integrity Unit confirmed that ATO had not issued any ten-digit TFNs. [Source: Advice from the ATO Software Industry Liaison Unit, 30 May 2007].

95 The ATO employs an algorithm, or mathematical function, to generate TFNs. All TFNs include a 'check digit'. That is, an in-built security feature to check the validity of TFNs. If a number fails the check digit algorithm, it is not a valid TFN.

this analysis is included in the following section of this report. However, the following matter is also relevant to accuracy of recorded TFNs. Of interest was a set of records that displayed duplicate CRNs but different TFNs. Analysis identified 273 pairs of customer records (546 lines of data in the individual customer table) in this category.

4.63 A customer should only be associated with one CSA customer identification number, one TFN and one CRN.[96] Therefore, the case of a pair of records displaying the same CRN but different TFNs presents a logical anomaly. If the pair of customer records relate to different people then at least one of the recorded CRNs is incorrect, as two people should not share the same CRN. If the pair of records relate to the same person then the database contains multiple records for the same customer—which would reflect a weakness in the integrity of the database. Furthermore, in the second case—same CRN but two different TFNs—one or both of the recorded TFNs is incorrect.[97]

4.64 This type of data anomaly is often associated with multiple customer records, and the fragmentation or inconsistent recording of information across those multiple customer records. The matter of potential multiple customer records is examined later in this report. However, given the importance of correct and accurate TFNs to the CSA's business—used to identify customers, match and receive taxable income information and to intercept tax returns, among other functions—the ANAO recommends that the CSA investigate these anomalies. See Recommendation No.3 at paragraph 4.81.

Centrelink Reference Numbers

4.65 The ANAO examined the field containing the customers' Centrelink Reference Number (CRN). In 2 571 718 instances, or 56.3 per cent of all individual customer records, the records displayed an entry consisting of nine digits and one alphabetic character. This format is consistent with Centrelink's CRN. All CRNs recorded in Cuba conformed to the Centrelink check digit algorithm.

4.66 Analysis identified 1829 customer records where a single value for a CRN appeared on two or more different CSA customer records—that is,

[96] ANAO observed, during fieldwork for this audit, that these three identifiers were commonly used to filter and identify in-bound telephone calls to CSOs, and to identify and verify the identity of CSA customers. For example, the customer search screen on Cuba allows a CSO to enter any one of the three numbers and returns a matching record from Cuba.

[97] Or if the customer has been issued with two TFNs, then the integrity of TFNs may be brought into question.

multiple customer records sharing a CRN. In some instances the customer records displayed the same given name, other name, surname and date of birth—suggesting that these were probably multiple records for an individual customer. In other instances the two customer records appeared to relate to two clearly different people. In either case, the fact that two customer records display the same CRN is indicative of a weakness in the integrity of information stored in the CRN field.

4.67 The 1829 lines of data related to 913 customers—910 displaying two records that shared a CRN and three displaying three records that shared a CRN.[98] Within the group of 910 customers:

- 273 displayed a different TFN on each record;
- 615 displayed a TFN on one record and no TFN on the other; and
- 22 did not display a TFN on either record.

Unlinked customer records

4.68 There were 4 571 020 records in the individual customer table—each record identified by a unique CSID. Comparing the CSIDs contained in the individual customer table with those contained in the case_customer_role table revealed that 118 676 CSIDs in the individual customers table did not have a corresponding record in the case_customer_role table. This means that the 118 676 customer records are not, and have never been, associated with any case.[99]

4.69 Some of the characteristics of the 118 676 unlinked customer records include:

- 20 475 records, or 17.3 per cent, displayed a date of birth as 01/01/0001;
- 6489 records, or 5.5 per cent, displayed a value in the TFN field;
- 3128 records, or 2.6 per cent, displayed a value in the CRN field;
- 62 882 records, or 53 per cent, displayed the customer's sex as UNKNOWN;
- 116 309 records, or 98 per cent, displayed the customer's marital status as either UNKNOWN or a blank entry;

[98] Each of these displayed a TFN on only one of their three CSIDs.

[99] If a record had been previously associated with a case, there would at least be a 'history line' in case_customer_role although the line would have been end-dated.

- 23 804 records, or 20 per cent, contained the word UNKNOWN as part of the customer's name—most often in the surname field[100]; and
- 1850 records, or 1.6 per cent, contained the word DUPLICATE as part of the customer's name—most often in the other given name field.[101]

4.70 One particular customer's name appeared on 19 records[102]—15 of these records displaying a date of birth of 25/01/1965 and four records displaying a date of birth of 01/01/0001.

Audit finding

4.71 The ANAO concluded that many of the 118 676 unlinked records were probably training records or arose from data entry errors. Many appeared to involve multiple records for the same customers.[103] Whatever their origin, as these records are not, and have never been, linked to CSA cases, they are redundant.

Child roles

4.72 According to the CSA's Entity/Attribute Definition for the case_customer_role table, the following fields are defined as:

> STAGE_1_ELIGIBILITY_INDICATOR—Text(1), Mandatory—Indicator used to show that case role is eligible for Stage 1. Values : Y / N
>
> STAGE_2_ELIGIBILITY_INDICATOR—Text(1), Mandatory—Indicator used to show that case role is eligible for Stage 2. Values : Y / N
>
> CHILD_LIABILITY_INDICATOR—Text(1), Mandatory—Indicates that the payer is liable for the child in a case. It is the result of the presumption/proof of parentage eligibility test. Values : Y, N.
>
> PROCESS_END_DATE—Date, Mandatory—Date row was end dated, if current equal to high values.
>
> ENTITLEMENT_END_DATE—Date, Optional—Date the liability is to end for the particular customer in the case.[104]

---

[100] Another 13 records displayed blank entries for all name elements (title, first, other and surname).

[101] In addition some 50 records also displayed DUPE or DUP, often with XXX in other fields, indicating a duplicate record marker.

[102] This customer also had a twentieth CSID that was associated with a case.

[103] While none of these records was linked to a case, often another record, under another CSID appeared to be linked to a case.

[104] CSA, *Entity/Attribute Definition for Entity Type: CASE_CUSTOMER_ROLE.* Provided to ANAO 23 March 2007.

4.73 Collectively, these fields describe a child's eligibility for child support and whether the payer is liable for a child in a particular case.

4.74 There were 1 484 904 current lines of data in the case_customer_role table with a role type code of CHD, case status of ACTV and entitlement end date in the future. These records covered 1 216 789 unique CSIDs. They detail the association of these children with active CSA cases—whether as claimed for, not claimed for or claimed as a relevant dependant in a case.

4.75 Examination of the Stage 1 and Stage 2 eligibility indicator fields, showed that 1940 records contained a blank entry in each of these fields—despite the indication, above, that both fields are mandatory. All but eight of these records displayed an 'N' in the child eligibility indicator field, suggesting that the payers were not liable for the children in those cases. The eight records displayed blank entries in each of the child eligibility indicator field and the Stage 1 and Stage 2 eligibility indicator fields.[105]

4.76 The CSA investigated this matter and advised that these eight records are anomalies and, as of 18 September 2007, had been corrected.[106]

4.77 Within the 1 484 904 current lines of data it was also noted that some records in the Stage 1 eligibility indicator field displayed an 'E'—once again, despite the indication, above, that the valid values are 'Y' and 'N'. The CSA advised that the 'E' stood for 'Ended'. Table 4.6 shows the various combinations of entries in Stage 1 and Stage 2 eligibility indicator fields.

Table 4.6

Stage 1 and Stage 2 Eligibility Indicators for child records

	Stage 2 Eligibility Indicator				
Stage 1 Eligibility Indicator	**Blank**	**E**	**N**	**Y**	**Total**
Blank	1 940	3 787	9 994	1 464 747	1 480 468
Y	4 406		13	17	4 436
Total	6 346	3 787	10 007	1 464764	1 484 904

Source: ANAO analysis of current lines of data, for active cases, from case_customer_role table, with role type code = CHD.

4.78 The table indicates that 17 lines of data indicate a 'Y' in both the Stage 1 and Stage 2 eligibility indicator fields. Such a situation appears contradictory. However, the ANAO was not able to investigate further, given the limitations

[105] The eight records also indicated a child role code of D or R—relevant dependants of the payee or payer.

[106] Advice from the CSA—18 September 2007.

of the data extract.[107] The CSA agreed that these 17 records were anomalous and that it would correct the data accordingly.[108]

Recommendation No.2

4.79 The ANAO recommends that the CSA identify and remove redundant records from the individual customer table in Cuba.

DHS's response:

4.80 DHS agrees. CSA will undertake to establish mechanisms to continuously identify redundant records, and remove the records from the production database accordingly.

Recommendation No.3

4.81 The ANAO recommends that, either separately or as part of its Data Quality Improvement Programme, the CSA:

- cleanse data in the fields describing customers' names and addresses;
- investigate and resolve anomalies in customer TFN and CRN records;
- develop and implement a quality assurance system to ensure a consistent standard of recording names and addresses in the revised version of Cuba to be introduced in 2008; and
- review and improve the effectiveness of data entry controls to ensure that, for individual customer records:
  - only valid, nine-digit TFNs may be entered;
  - only valid values, stored in a Cuba Code Table, are able to be entered;
  - dates of birth are recorded accurately; and
  - all mandatory fields are populated with valid entries.

DHS's response:

4.82 It is agreed that CSA should undertake data cleansing of customers' names and addresses, in addition to resolving anomalies in customer TFN and

[107] ANAO is aware that a case may contain both a Stage 1 and Stage 2 component (Source: Functional Specifications – Register Case V14.3_AS_15112006, p.30. Functional Overview, paragraph 5). However, we cannot see how an individual child could have both indicators set to Y on the same line of data in Case_Customer_Role.

[108] Advice from the CSA—18 September 2007.

CRN records. CSA would also like to note that controls currently exist within Cuba where invalid TFNs cannot be entered.

4.83 As noted by the ANAO, CSA's Data Quality Improvement Program is in its early stages, and the development of a quality assurance system and review of data entry controls will be a focus for this program, to help continuously improve the quality of CSA's data.

5. Employer Records

This chapter presents the results of the ANAO's analysis of employer records. These are records associated with employers, government and private sector organisations and overseas child support administration agencies.

Employer table

5.1 The CSA defines an employer as an organisation that has dealings with the CSA[109]—largely, employers that assist the CSA in the collection of child support payments by means of withholding a portion of employees' wages or salary. The employer table contained 258 859 records, each incorporating 39 data fields.

5.2 All 258 859 records contained unique values. That is, no values in the primary key, for the employer table, were duplicated.

5.3 A number of fields are indicated as mandatory in the Entity/Attribute Definition for the employer table. A field called 'status code' is defined in the Entity/Attribute Definition for this table as:

> Status code: Mandatory. Current status of the employer.[110]

However, despite the field being identified as mandatory, 99 382 records displayed a blank entry in this field. Cuba's code table listed four valid values for the code 'organisation status'. These are shown in Table 5.1 below, along with the number of records associated with each of the code values.

[109] CSA, *Entity/Attribute Definition, Entity Type: Non-Individual*. Provided to ANAO on 23 March 2007.

[110] ibid.

Table 5.1

Organisation status code

Code	Description	Number of records
ACTV	ACTIVE	134 485
CEAS	CEASED	2 844
DUPL	DUPLICATE	20 798
INSV	INSOLVENT	1 350
	Blank entry	99 382

Source: ANAO analysis.

5.4 Other mandatory fields held a number of blank entries. Table 5.2 shows the number of blank entries associated with a selection of mandatory fields in the employer table.[111]

Table 5.2

Blank entries observed in mandatory fields

Field name	Description	No. of blank entries
Legal_name	The legal name of the customer	2
Section_72A_indicator	Denotes whether non-individual customer is eligible for section 72A	18 844
Section_72B_indicator	Denotes whether non- individual customer is eligible for section 72B	14 530
Section_120_indicator	Denotes whether non- individual customer is eligible for section 120	18 541
Section_161_indicator	Denotes whether non- individual customer is eligible for section 161	19 188
FTB_section_120_indicator	Denotes whether non- individual customer is eligible for FTB section 120	124 375
FTB_section_72AB_indicator	Denotes whether non- individual customer is eligible for FTB section 72AB	124 389

Source: ANAO analysis.

[111] These fields contain a range of indicators relating to the CSA's legislative power to issue notices and collect monies and information from third parties. For example, the CSA can issue a notice to someone who owes a child support debtor (payer) money requiring that the money be paid to the CSA in satisfaction of child support related debts (section 72A, *Child Support (Registration and Collection) Act 1988.*

5.5 Mandatory fields that held no blank entries included:

- Registration date;
- Registration received date;
- Registration source code; and
- Hold reallocation start (and end) dates.

5.6 As a result of these analyses, the ANAO concluded that the mandatory nature of some fields, in the employer table, was not enforced.

5.7 A number of fields are identified as optional in the Entity/Attribute Definition for the employer table. The fields tested included:

- employer role code;
- employer type code; and
- hold reallocation indicator.

5.8 These fields held only valid values from the code table, or a blank entry.

TFNs in the employer table

5.9 A particular field in the employer table holds the employers' Tax File Numbers. Of the 258 859 records in the table, 63 821 displayed a value for TFN.[112] These consisted of 26 660 nine-digit TFNs and 37 161 eight-digit TFNs.

5.10 As noted earlier, the eight-digit TFN is used by companies, partnerships and trusts, while the nine-digit TFN is used by individuals. Of those records displaying nine-digit TFNs most of the 26 660 records clearly related to companies—not individuals. The fact that an almost equal proportion of eight and nine-digit TFNs are stored in the employer table indicates a possible weakness in the integrity of the TFN field within that table.

5.11 Using the algorithms provided by the ATO, the ANAO checked the validity of both eight and nine-digit TFNs and found that all 63 821 entries passed the check digit test.

[112] And 195 038 displayed a zero as the entry for TFN.

Names stored in the employer table

5.12 The field 'legal_name' stores the name of each employer. Two records in the table displayed blank entries for the legal name. A small number of records—approximately 50—displayed apparently invalid entries for the legal name. Some of the names displayed included: ZZZZZZZZZZZZZZZ; XXXXXXXXXXX; XXXCREATED IN ERROR; USE EPC (followed by a nine-digit number) and DO NOT USE.

Multiple records in the employer table

5.13 Employers play an important role in the collection of child support payments from some CSA customers—through a system of withholding part of an employee's wages or salary, and forwarding the payment to the CSA. The names of many companies were repeated across records in the table. A large number of employers operate multiple sites across Australia, and in many cases, multiple payroll offices. For example, a large banking corporation could have branches in every State and Territory, with each State and Territory operating a payroll office to service the employees in their region. It is, therefore, possible for employers to be listed a number of times within the employer table[113]—each listing associated with a unique CSID. Multiple listings for employers might also arise when large organisations consist of a number of divisions, subsidiaries or in the case of franchises.

5.14 With this in mind, the ANAO examined the legal names in the employer table for multiple occurrences, which may impact on the efficiency of the CSA's business. This analysis identified 231 490 distinct legal names within the 258 859 lines of data in the employer table. In 220 862 cases, the legal name appeared only once in the table. In the remaining cases, the entry appeared between two and 7776 times.[114] In total, these accounted for 37 997 lines of data indicating potential multiple records for companies in the table.

5.15 Many of these multiple records had been identified by the CSA as redundant. Table 5.1 indicated that 20 798 records displayed a status of DUPL (duplicate) and another 2844 displayed a status of CEAS (ceased). In addition, for some thousands of records, a field called 'indicator comments' held information such as 'DO NOT USE', 'DUPLICATE' and 'USE CSID' (followed by a nine-digit number).

[113] As long as each entry is unique and can be distinguished from other entries for the 'parent' company.

[114] The entry appearing 7 776 times was REDUNDANT EPC.

5.16 The ANAO further examined the 134 485 records that displayed an ACTV status. On 262 records, even though the records displayed an ACTV status, the indicator comment field displayed a contradictory message—such as Do Not Use, Duplicate of CSIDxxx, Use CSIDxxx and Ceased trading. The ANAO removed these 262 records from the dataset of 134 485 ACTV lines of data and analysed the remaining 134 223 records for multiple customer records (not so identified or implied by any other data in the employer table).

5.17 This analysis identified 4682 lines of data associated with potential multiple employer records. Within this dataset one employer was associated with 22 separate CSIDs. These 22 records contained nothing in the 'indicator comments' field to distinguish them. Furthermore, all 22 records displayed a zero in the 'group number' field. The 22 records did display a range of values in the 'State code' field, as follows:

- 11 records displayed NSW;
- three displayed VIC;
- three displayed SA;
- four displayed QLD; and
- one record displayed ACT.[115]

5.18 A possible business impact of having such multiple records in the employer table is that it is difficult for a CSO to correctly identify the appropriate employer CSID to which they should link a payer. During visits to four CSA offices,[116] CSA staff advised that multiple employer records often present a problem for the efficient processing of employer withholdings. An example provided by these CSA staff involved a given employer having two CSIDs—the first associated with three payers and the second associated with five other payers. In such an example, it usually transpires that an employer will remit all monies—that is, withholdings for all eight payers—using a single CSID. Reconciling employer withholdings then becomes more difficult for CSA staff, who must manually journal payment entries from one account to another. This unnecessary work affects the efficiency of the CSA's business processes.

[115] The ACT record, therefore, was the only record uniquely distinguishable in the group.

[116] Canberra, Adelaide, Melbourne and Townsville.

Recommendation No.4

5.19 The ANAO recommends that the CSA:

- cleanse the employer table of redundant records, including occurrences of multiple records for employers;
- resolve the issues surrounding the use of eight-digit and nine-digit TFNs for employers; and
- ensure that valid data populate all mandatory fields in the employer table.

DHS's response:

5.20 DHS agrees. The anomalous records and the identified mandatory fields not populated will need to be investigated and corrected. CSA will also undertake to establish mechanisms to continuously identify redundant employer records and multiple records for employers, and remediate accordingly.

6. Multiple Customer Records

This chapter presents the results of the ANAO's analysis of individual customers who may have multiple records on Cuba. That is, individual customers who have been issued with two or more CSIDs that are not linked or connected in any way.

6.1 In order to test the integrity of the primary key for customer identification in Cuba, the ANAO examined whether:

- two or more CSA customers shared the same CSID; and
- any CSA customers had been issued with more than one CSID.

6.2 For the purpose of this treatment we will use the terms 'duplicate records' and 'multiple records' to refer to the two situations described above. A duplicate record is defined as the same CSID appearing more than once in the database—that is, the same CSID shared by two different people. Multiple records refer to a single customer having two or more records on Cuba, under two or more different CSIDs.[117]

6.3 As described in paragraph 4.2, no duplicate CSIDs were identified in the dataset provided by the CSA. However, the ANAO did find instances of multiple customer records.

How multiple records arise

6.4 The CSA's customer registration processes are designed to ensure that, once registered with the CSA, all information regarding a particular customer is stored in a single record under one CSID. In this way, a person may be registered as a claimed-for child in one child support case, be treated as a relevant dependant in another case, later become a payer in their own case, and later still become a payee in another case—with a full history maintained under one CSID.

6.5 If an existing customer is registered for a second (or subsequent) time, under a different CSID, the potential exists for the customer's information to be fragmented across multiple records. At the very least, there is a risk that undetected multiple records will result in inaccurate reporting of customer

[117] The ANAO is aware that CSA uses the term 'duplicate' payer, payee or child records to refer to (what we have defined as) multiple records. In this report, where quoting a CSA document or an actual customer entry that uses the word duplicate, we will use that term. On all other occasions, we will use the term multiple records or multiple customer records to refer to the same customer having been assigned two or more CSIDs.

numbers and other statistics that depend on customer numbers. Further, the existence of multiple records could make the task of reconstructing a comprehensive history of a customer's dealings with the agency extremely difficult or impossible. Multiple records also make it more difficult to identify a particular customer reliably, complicating the task of data exchange with other agencies.

6.6 Where multiple records exist for a payer, and more than one is linked to a case, there is a risk that each of the case assessments will be inaccurate, as the amount of child support payable is modified if a payer is assessed to pay child support to two or more payees.[118] Where multiple records exist for a child, there is a risk that the child might be counted twice in an assessment.

Known multiple records

6.7 The CSA is aware of the existence of a number of multiple customer records in Cuba. A document entitled '*A Guide to Duplicate Payers/Payees Records*', and another entitled '*A Guide to Duplicate Child Records*' appear on the CSA's intranet. These guides provide background on the importance of avoiding the creation of multiple records, together with instructions on how to treat existing multiple records when they are detected.[119] In the data extract provided, 12 783 individual customer records contain the word 'DUPLICATE' in the middle name field, with 'XXX' in the first name and surname fields, and TFNs removed, as mandated by these guides. In addition, the discussion in Chapter 4 suggests that Cuba may contain up to 5000 additional duplicate records, that do not conform with the CSA's instructions, but are marked in some manner as duplicates.

Identifying potential multiple customer records

Methodology

6.8 The ANAO examined the dataset provided by the CSA in order to identify customers with multiple records involving different CSIDs. The first level of analysis involved identifying records which matched exactly on the customers' first name, middle initial, surname and date of birth.[120] This

[118] Examples of Cuba records in this category are presented later in this chapter.

[119] Also see Chapter 4, at paragraphs 4.14 to 4.16.

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

generated a list of 102 891 records, which could potentially involve multiple CSIDs.

6.9 Based on previous experience in the analysis of large datasets, we considered that a proportion of these records were, in fact, coincidental matches—that is, different people who happen to share the same first name, middle initial, surname and date of birth. In order to reduce the number of coincidental matches, we refined the original analysis to exclude records that did not match on the entire entry for middle name (where present).

6.10 Analysis also showed that 18 records matched on customer's first name, middle initial, surname, and some elements of customer address records, but did not match on date of birth. This strongly suggested multiple CSIDs, 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 the 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.

6.11 In addition, 494 records matched exactly on the customer's first name, middle initial, date of birth and some elements of the customer's address records, 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. Mapping these records to the name history file verified that, within a pair of records, the customer displayed both names within their name history file entries. A second group of records in this dataset exhibited a variation in spelling of the customer's surname—such as Bennett and Benett.

6.12 Examining the combined list of potential multiple customer records showed that many displayed 'DUPLICATE' or 'UNKNOWN' as one of the name elements. The ANAO also noted other characteristics that would make it difficult to determine whether particular entries constituted multiple customer records.

6.13 Furthermore, many of these customer records were not, and have never been, associated with a CSA case (refer to the previous discussion in Chapter 4 on unlinked customer records). These extraneous records were identified and removed from the dataset. For more information on the process used to refine the initial list see Appendix 3.

6.14 Removing these records from the combined list of potential multiple customer records resulted in a dataset of 55 266 lines of data (the multiple record customer test set, or MCR test set). This MCR test set represents, with a reasonable level of confidence, customers who have multiple records on Cuba. It represents 27 633 customers who each have two records under two different CSIDs. The ANAO examined this dataset in order to identify any significant business risks associated with fragmenting or duplicating information across multiple customer records.

Association with CSA cases

6.15 Within the MCR test set, at least one CSID (of the pair of CSIDs for an individual) is associated with at least one CSA case. However, each CSID (of the pair of CSIDs for an individual) may also be associated with one or more CSA cases. For example, consider the set of records shown in Table 6.1.

Table 6.1

Example of multiple customer records

CSID	Full Name	Date of birth	Cases associated with the CSID (case status)
2133399y	Ms Jennifer K Smith	4/06/1977	none
55434240x	Ms Jennifer K Smith	4/06/1977	2 (CLSD, ACTV)
194399y	Mr James L Harrison	24/08/1963	1 (ACTV)
2124240y	Mr James L Harrison	24/08/1963	3 (CLSD, EWAR, ACTV)

Source: ANAO analysis of Cuba data extracts. The names and CSIDs used are fictitious.

6.16 In the above example, Ms Smith has two CSIDs. One of Ms Smith's CSIDs has never been associated with a CSA case, however, her other CSID is associated with two cases—one that has closed, and one that is active. Mr Harrison has two CSIDs. One of Mr Harrison's CSIDs is associated with one active case. The other CSID is associated with three cases—one of which is active, another ended with arrears and the third case is closed.

6.17 Within the 27 633 customers (55 266 lines of data) in the MCR test set:

- 8054 customers (16 108 lines of data) had one of their CSIDs associated with at least one case and the other CSID was not associated with a case; and
- 19 579 customers (39 158 lines of data) had both of their CSIDs associated with at least one case.

6.18 The first group—8054 customers who only have one CSID linked to a case—represents a slight business risk to the CSA's administration of the Child Support Scheme, as the unlinked record exists on the database and could, at some point in the future, be (mistakenly) associated with a new CSA case.

6.19 Of greater concern is the second group—those 19 579 customers with two CSIDs, both associated with CSA cases. Within this group many of the child records display an entitlement end date in the past. That is, the child record may still be associated with an active case although that particular child may have turned 18 years of age and the entitlement for that child ended.[121]

6.20 The ANAO identified child records with an entitlement end date in the past and removed these lines of data from the test set.[122] We also removed lines of data relating to payers and payees where those lines of data displayed a process end date other than 31/12/4000[123].

6.21 Refining the test set in this way, revealed that 4672 customers displayed at least one current entry—from the case_customer_role table—on each of their CSIDs. In total, the 4672 customers were associated with 11 487 current entries; an average of approximately 2.5 current entries in case_customer_role per customer. Table 6.2 shows that 3249 customers were associated with two cases—one on each of their CSIDs. Also, 936 customers were associated with three cases—one case on one CSID and two cases on the second CSID.

[121] The case may still be active because younger children of the relationship may still display an entitlement end date in the future.

[122] Some customers in this category displayed a role type code of CHD on one CSID and PYE or PYR on the other CSID. If the CHD record displayed an entitlement end date in the past, only the CHD record was removed from the test datset.

[123] See previous discussion on the meaning of a high date in the process end date field.

Table 6.2

Number of cases associated with multiple record customers[124]

Number of cases	Number of customers
2	3 249
3	936
4	322
5	120
6	28
7	12
8	4
9	1
	Total 4 672

Source: ANAO analysis.

6.22 In the third line of Table 6.2, the entry shows that 322 customers were associated with four cases, however, it does not indicate whether these were arranged as two cases per CSID or three cases on one CSID and one case on another CSID. The eighth line of Table 6.2 shows that one customer displayed an association with nine cases. In this instance further analysis revealed that the customer displayed six cases on one CSID—four of which were active[125]—and three cases on the second CSID—all of which displayed a case status of ineligible.

6.23 The ANAO further examined the records that represent the highest risk group—those customers with both CSIDs currently associated with ACTV cases. That analysis identified 360 customers from the MCR test set that displayed at least one currently ACTV case on each of their CSIDs.

6.24 Of the 360 customers, 40 displayed a connection to the same case number. That is, both CSIDs were linked to the same case. The role type code combinations for these 40 customers were as follows:

- 31 displayed Child—Child;
- 6 displayed Child—Payee; and
- 3 displayed Payer—Payee.

[124] Where the multiple record customers display at least one current association with a case on each of their CSIDs.

[125] The two remaining cases displayed a status of INEG and CLSD.

6.25 The 31 Child—Child combinations represent a situation where the same child is involved twice on the same case, although under two CSIDs. Some pairs of records suggested the child was claimed for twice; others indicated the child was claimed for on one CSID and involved as a relevant dependant on the other CSID, or involved as a relevant dependant twice.

6.26 The six Child—Payee and three Payer—Payee examples represent a logically inconsistent combination of roles—an individual cannot be the payer and the payee in the same case, or a child and the payee in the same case.

6.27 The remaining 320 customers displayed a connection to at least two different case numbers. That is, the same individual was linked to two or more different active CSA cases.[126] Role combinations for these 320 customers were:

- 25 Payer—Payer;
- 49 Payee—Payee;
- 171 Child—Child;
- 63 Payee—Payer;
- 6 Payee—Child;
- 5 Payer—Child; and
- 1 Payee—Spouse.

6.28 Some of these combinations represent extremely unlikely situations—such as the same customer concurrently claimed for as a child on one case (under one CSID) and involved as a payer or payee on another case (under a different CSID). Other combinations—such as payer on two cases—although more representative of a valid situation for an individual (that is, one person may well be involved in two CSA cases as the payer on each), still present a significant business risk to the CSA, as the two customer records are not linked.

Multiple records and case liability determinations

6.29 Where a customer's circumstances involve two or more unlinked records associated with active cases there is a risk of incorrect liability assessments. Within this category, a high risk group is where multiple payer records are linked to active cases. Accordingly, further analysis was

[126] Some of these customers are linked to more than one ACTV case on both CSIDs.

concentrated on the 25 customers who displayed a payer role on two or more unlinked active cases.

6.30 Such records represent a situation where an individual is involved in two CSA cases, on unrelated database records. Where multiple records exist for a payer, and more than one is linked to an active case, there is a risk that each of the case liability assessments will be inaccurate, as the amount of child support payable is modified if a payer is assessed to pay child support to two or more payees.[127]

> If a payer is liable to pay child support to two or more payees the liability to each is calculated according to the basic formula and any modifications that are applicable.
>
> The child support percentage is modified by multiplying the percentage that would otherwise apply by the number of relevant children in that carer's care and then dividing that product by the total number of children for whom the parent is liable to pay child support (section 54).[128]

6.31 For example, the child support formula percentage applicable to one child is 18 per cent of the payer's adjusted income amount. The percentage applicable to two children is 27 per cent.[129] A payer with two child support children in separate cases should pay 13.5 per cent for each child, a total of 27 per cent of his or her adjusted income amount. However, if the payer has two unlinked customer records there is a high risk of the payer being assessed as liable to pay 18 per cent for each child, a total of 36 per cent of the payer's adjusted income amount.

6.32 Records for 24 of the 25 customers, identified above,[130] displayed a different annual case liability amount on each of the customer's two cases.[131] The average difference between the two amounts was $3622.[132] The greatest

---

[127] Subdivision G of Division 2 of Part 5, sections 53 and 54 *Child Support (Assessment) Act 1989.*

[128] CSA, *The Guide—CSA's online law and policy Guide,* [Internet]. Available at: <http://www.csa.gov.au/guide/index.htm>. Section 2.4.8: 'Payer liable to pay two or more payees.'

[129] *The Guide,* Section 2.4.1: 'The basic formula.'

[130] The ANAO cross-referenced the CSIDs for these customers with the customer address file. In 14 instances the same address appeared on both CSIDs for a customer, thereby adding another level of confidence to the identification of these as multiple records for the same customer. In 11 instances the ANAO was unable to directly verify shared addresses across the CSIDs, however, these records displayed the same State across both records. The ANAO had previously removed other customer records that displayed a different State on each of the customer's address records—as probable coincidental matches, rather than genuine multiple customer records.

[131] In one instance the customer's records displayed $333 as a case liability amount on each of the payer's two cases.

[132] This is an annual liability amount sourced from the case liability table.

difference was observed for one customer whose records displayed a case liability of $5600 on one case and $24 500 on the other—a difference of $18 900.

6.33 In the instance noted above—a variation of nearly $19 000 per year in assessed child support liability—the ANAO examined details of the two unlinked cases and found that the payer's first case involved three children and three relevant dependants of the payer. The second case involved three children and one relevant dependant of the payer. The three children identified as relevant dependants of the payer in the first case also appear to have multiple customer records and appeared, under different CSIDs, as the three children for which child support was claimed in the second case.

6.34 Details of this high risk group of payers were provided to the CSA and further information sought in relation to any dollar impact on the payer's overall child support liabilities. In September 2007, the CSA advised that:

> These cases have been reviewed. CSA will further analyse these cases and develop appropriate corrective action for each case.
>
> - 16 pairs were found to have a potential customer impact which will require corrective action;
> - 9 pairs were not multiple customers.
>
> A payer's liability may be lower than what is currently assessed. CSA agrees that these are anomalies and will correct the data accordingly.[133]

6.35 While the other role combinations mentioned in paragraph 6.27 may not represent the same level of risk to the accuracy of child support liability calculations, the CSA should investigate these customer records for inappropriate role combinations and any effect these might have on case liability calculations.

Multiple records and income data

6.36 Case liability calculations use the recorded value for a payer's income. The ANAO examined the 25 customers displaying Payer on each of their CSIDs and mapped these records against corresponding entries in the income table of Cuba. This process revealed that the records for these customers displayed a value, for each of their CSIDs, in the field called 'modified taxable

[133] Advice provided by the CSA—18 September 2007.

income amount'.[134] This is the income amount the CSA uses in calculating child support assessments.

6.37 There are circumstances where a customer, with more than one active case, might validly have a different income used for the liability calculation on each case. For example, where the cases cover different child support periods, income from one tax year may be applicable to one child support period while income from an earlier or later tax year may be applicable to the child support period in the second case.

6.38 However, the application of such business rules in Cuba has been designed to operate on the basis that customers are identified by a single, unique CSID, and that all cases, with which the customer is associated, are linked to that unique CSID. Only then can the appropriate checks be conducted and the business rules effectively enforced. Where customers have undetected multiple CSIDs associated with active cases, there is a risk that the records will contain inconsistent information.

6.39 This risk is compounded by the fact that Cuba does not contain duplicate TFNs. Therefore, if one of a customer's multiple records includes his or her TFN, the other will not.[135] As the CSA's access to customer income information is heavily reliant on TFNs, one of a customer's multiple records could be expected to contain more accurate and/or up to date income information than the other (unlinked) record.

6.40 In the group of 25 payers with active cases on each of their CSIDs, the average difference in the recorded value for modified taxable income was $23 929. The greatest variation in recorded income amounts was $115 000. This particular customer displayed an income of $15 000 on one CSID and $130 000 on the other. This individual is the same individual whose cases were discussed in the example of different case liability calculations (see paragraphs 6.31 to 6.33). This customer also displayed two different TFNs and two different CRNs, across their two CSIDs.[136]

6.41 The ANAO examined the 49 customers displaying Payee on each of their CSIDs and mapped these records against corresponding entries in the

---

[134] CSA, *Entity/Attribute Definition for the entity: INCOME.* Provided to ANAO 23 March 2007.

[135] Or the second CSID will show a different TFN. See the earlier discussion of customers who may have two different TFNs recorded in Cuba.

[136] And yet, there is a high probability that the two records relate to the same person. The records match exactly on first name, middle name, surname, date of birth, they share an address and the names of the three children associated with both cases.

income table of Cuba. This analysis showed that records for 45 customers displayed a value in the field called 'modified taxable income amount', for each of their CSIDs.[137]

6.42 For the 45 customers with two income records, 12 displayed $0 on both records. The average difference in the recorded value for modified taxable income, for the 45 customers, was $16 707 and the maximum difference was $67 520. This customer displayed $67 520 on one CSID and $0 on the other.

6.43 As this amount is well above the disregarded income amount[138] for a formula assessment, it will impact on the child support calculation for both cases this individual is involved in as a payee. Three other records, within this group also displayed a value greater than the 2007 disregarded income amount.

Multiple customer records and TFNs

6.44 Some customers with multiple records displayed two different TFNs—one against each CSID. A customer should only have one TFN.[139] Therefore, the ANAO examined the TFN entries in the dataset of potential multiple customers.

6.45 Within the dataset, 493 customers were identified (986 lines of data) whose records displayed two different TFNs. Mapping these records against the customer address file identified 230 customers, whose records displayed the same address on at least one entry for each CSID.[140]

6.46 If a customer does have two different TFNs, and multiple customer records in Cuba, it would almost certainly lead to inaccuracies or inconsistencies in recording income details for that customer. The ANAO

---

[137] Three customers had no entries in the income table on either CSID; one customer had an entry in the income table under one CSID only.

[138] The disregarded amount for 2007 is $43 654. Source: *The Guide* – section 2.4.2: 'Basic values used in a child support assessment.' Accessed 1 August 2007 at <http://www.csa.gov.au/guide/2_4_2.htm>.

[139] The term Tax File Number has a specific legislative meaning—it is a number that is issued to a person by the Commissioner of Taxation (Section 202A of the *Income Assessment Act 1936*). The original and main purpose of the TFN was to be a numeric, unique identifier of customers of the Australian Taxation Office. The TFN is also used by other government agencies when there is a legislative need to verify customer identity and establish income levels. (See also: ANAO Audit Report No. 37 1998–99, Management of Tax File Numbers. p. 27–28, and ANAO Audit Report No.47 2004–05, Tax File Number Integrity. p. 21.)

[140] Thereby adding to the level of confidence that these are the same customers. While ANAO did not identify an exact match on address details for the remaining 263 customers, through visual inspection, many of these records displayed similar address entries that would have matched using less stringent matching criteria.

invited the CSA to investigate the circumstances surrounding the 493 customers identified above. In September 2007, the CSA advised:

> **Analysis:** A sample of these cases have been reviewed. CSA is already aware of this issue and continues to fix these cases in line with the *Guide to Duplicate Payee/Payer Records.*
>
> 32 customer pairs, where one CSID is not linked to a case, have been corrected.
>
> **Customer Impact:** A customer's liability may potentially be incorrect.
>
> **CSA Response:** CSA agrees that these are anomalies and will correct the data accordingly. CSA would like to note that Cuba currently has a 'dupe function' to alert CSOs of the potential creation of a multiple customer.[141]

6.47 In addition, 41 customers displayed one CSID linked to a case and the other CSID not linked to a case. Of these, 18 displayed a different TFN on each of their CSIDs and 23 displayed a TFN on the CSID not linked to a case. The latter group presents a particular concern associated with TFNs on multiple customer records, in that these TFNs reside on the records not linked to a case.

6.48 As Cuba prohibits the existence of duplicate TFNs, it is not possible to record these customers' TFNs on the CSIDs that are linked to one or more CSA cases. Therefore, the CSA is unable to use a valuable piece of customer information, because it is stored on a redundant record.

Multiple customer records and CRNs

6.49 A customer should only have one CRN—it is assigned by Centrelink as a unique identifier for Centrelink customers.[142] Within the MCR test set, the ANAO identified 99 customers (198 lines of data) whose multiple records displayed two different CRNs. Mapping these records against the customer address file identified 25 customers, whose records displayed the same address on at least one entry for each CSID. If a customer does have two different CRNs, and multiple customer records in Cuba, it would almost certainly lead to difficulties in data matching with Centrelink and possibly other agencies. The ANAO invited the CSA to investigate the circumstances surrounding the 99 customers identified above. In September 2007 the CSA advised that:

> **Analysis:** These cases have been reviewed.
>
> - 48 — have been corrected;

[141] Advice provided by the CSA—18 September 2007.

[142] 2006, ANAO, Audit Report No. 29 2005–06 Integrity of Electronic Customer Records (Centrelink), p.69.

- 2 — no further action, as the case is closed;
- 55 — complex - further remediation required;
- 2 — complex and case closed;
- 5 — relate to international customers;
- 84 — are not duplicates;
- 2 — locked to TSO for further action.

CSA will investigate these cases further and this issue appears to relate to technical aspects of the Centrelink data exchange.

Customer Impact: A customer's liability may potentially be incorrect.

CSA Response: CSA agrees that these are anomalies and will correct the data accordingly. The customers that relate to International cases will not be corrected, as per the processing of International cases.[143]

Multiple records and date of death

6.50 Within the dataset of potential multiple customer records 136 customers, with two CSIDs, displayed a date of death on one record but not the other. In addition, one customer with two CSIDs displayed a different date of death across the two records.

6.51 None of these customers showed a current line of data associated with an active case[144] and, therefore, do not appear to present a significant risk to the accuracy of child support liability determinations. Nevertheless, this type of situation—displaying a date of death on one record and no date of death on the other record—highlights the risks associated with fragmenting information across multiple customer records.

Multiple records for restricted access customers

6.52 In order to expedite the provision of the large data extract to the ANAO, the CSA did not include fields holding information about sensitive or restricted access customers. The actual customer records were included in the extract, but not the fields that would have identified those customers as restricted access customers (RACS). The ANAO agreed to this limitation in the data extract on the understanding that the CSA would undertake to examine

[143] Advice provided by the CSA—18 September 2007.

[144] Although some of the cases they were involved in are still active, these customers exhibit end dated lines of data for those active cases.

the set of potential multiple customer records for any anomalies relating to the RACS fields and advise the ANAO of the results. The particular fields in the individual customer table are:

- RACS rating code;
- RACS reason code; and
- RACS source code.

6.53 In September 2007, the CSA advised the ANAO of the results of its RACS analysis as follows. Within the MCR test set 180 customers (360 lines of data) were associated with RACS records. Of these, the records of 124 customers (248 lines of data) displayed information in the RACS reason code and RACS source code fields, but did not display information in the RACS rating code field. This situation arose from a RACS review process, commenced by the CSA in February 2007. This process saw the RACS rating codes for these customers downgraded, however, the old source code and reason codes data were not removed.

6.54 Of the remaining 56 pairs of customer records, 20 pairs displayed only one CSID linked to a case, while 36 pairs displayed inconsistent RACS information across customer records. The CSA further advised that at least three pairs of records had been corrected, as at September 2007, and that further analysis and remediation was in progress.

Multiple records for children

6.55 Analysis identified 128 lines of data (64 customer record pairs) that appeared to involve multiple records for children, where both CSIDs were associated with ACTV cases.[145] These records represent child associations used in the formula to assess child support liabilities.

6.56 The fact that these children have multiple records, under different CSIDs, represents a weakness in the integrity of their data in Cuba. Multiple records may hold inconsistent information about the child's role—for example, whether the child is claimed for as an eligible child in a case, or claimed as a relevant dependant.

6.57 While some of these combinations may be valid—such as claimed for on one case and relevant dependant of the payee on a different case—the

[145] That is, role type code = CHD, process end date set high, entitlement end date in the future, child liability indicator <> N, Stage 2 eligibility indicator <> E, child role code <> T and, for each CSID at least one case status = ACTV.

business rules in Cuba are predicated on the assumption that the child in question is assigned only one CSID, and that the various associations with cases are linked. However, some of the combinations may not be valid. For example, a child should not be claimed for on two separate child support cases—that is, two cases with two different payers.[146]

6.58 Within the 64 child record pairs, 14 children displayed a child role code of C—claimed for—on at least one active case associated with each of their CSIDs. This represents the highest risk group, as the data suggests that the children are claimed for twice.

6.59 Given the complex interplay of factors determining liability in some child support cases, the ANAO was not able to determine whether errors in child support liability calculations have arisen in these cases. In September 2007, the CSA advised that:

> **Analysis:** These cases have been reviewed (64 customers—128 lines of data).
>
> - 74 have been corrected, where required — 12 were not the same child with no action required and 62 were the same child with a correction required;
> - 36 were international and were same child (refer to note);
> - 18 correction underway.
>
> Note: When a child of the relationship is in the care of the other parent and that parent resides in New Zealand, subdivision E of the Child Support Assessment Act doesn't apply and the case cannot be reflected as a divided care assessment. The payee is entitled to the disregarded exempt income and the payer receives the relevant dependant exempt income amount.
>
> To enable correct processing of this in Cuba, there has been a duplicate record created…
>
> **Customer Impact:** Yes, a customer's liability may potentially be incorrect.
>
> **CSA Response:** CSA agrees that there are anomalies and will correct the data accordingly. The customers that relate to International cases will not be corrected, as per the processing of International cases.[147]

[146] The ANAO is aware that some reverse role cases may see both parents applying for child support and the existence of multiple CHD records could complicate the child role codes and case associations.

[147] Advice provided by the CSA—18 September 2007.

Recommendation No.5

6.60 The ANAO recommends that the CSA resolve all instances of multiple customer records in Cuba.

DHS's response:

6.61 CSA will undertake to resolve the multiple customer records outlined in the audit, in line with current business procedures. CSA would like to note that there are currently controls that exist within Cuba to minimise the creation of multiple records.

6.62 Furthermore, as part of the Quality Assurance system, to be developed, CSA will establish appropriate mechanisms to continuously identify multiple customer records, and remediate accordingly.

7. Implications of Data Integrity Issues

This chapter draws on the overall audit findings arising from the ANAO's analysis and discusses the potential impact of identified data integrity issues on the CSA's administration of the Child Support Scheme.

Data integrity in Cuba

7.1 As a result of the analyses undertaken during this audit, the ANAO concluded that the majority of records in the dataset provided by the CSA are sufficiently accurate, complete and reliable to support the effective administration of the Child Support Scheme. In the majority of examples highlighted in this report, the proportion of anomalous records, across various categories in the dataset, is relatively low.

7.2 However, regardless of the overall impact on the scheme, many of the errors or anomalies associated with individual records may have a significant impact on the particular customers involved. Therefore, these analyses have highlighted a number of areas in which the CSA could significantly improve the quality and reliability of data in Cuba.

7.3 This is particularly important as the CSA positions itself to migrate the current data in Cuba to a new system, in preparation for the introduction of Stage Three reforms to the Child Support Scheme on 1 July 2008. The new system will see Cuba data supplemented by additional information and the introduction of a new formula for the calculation of child support liabilities.

7.4 The following discussion summarises the major themes covered in the audit and identifies the potential impact of poor quality data on the CSA's business.

Reliability of dates in Cuba

7.5 Dates are important in databases. Dates control many business processes within a database and assist database administrators to effectively manage the information. Some date fields in Cuba may be classified as 'system' dates, while other date fields are more directly associated with customers' personal or case information.

7.6 Examples of system date fields are the start date and end date for a line of data within a table. These are used to distinguish between current and historical lines of data within a table—that is, they identify the valid period for

a line of data. In Cuba, a high date of 31/12/4000 in the end date field indicates a current, or open, line of data. Examples of dates relating to customer information include date of birth, case liability end date and entitlement end date.

7.7 The ANAO identified various anomalies in date fields in Cuba. Although the majority of these anomalies involve a relatively small number of records, they have the potential to impact on the management of individual CSA cases and customers. Issues identified during the audit include:

- inconsistent use of low dates and high dates in both system date fields and customer information date fields;
- logical inconsistencies in start dates and end dates within a line of data;
- inappropriate use of high dates in liability end date fields, leading to cases apparently lasting for unrealistically long periods; and
- recorded dates indicating that liabilities commenced or ended before the establishment of the CSA.

7.8 In addition, the following anomalies in recorded dates of birth and dates of death were observed:

- entries in the date of birth field displaying the date 01/01/0001;
- apparently very old customers with no date of death recorded;
- claimed-for children apparently well over the age of eighteen;
- payers and payees apparently well under the age of eighteen;
- customers whose recorded date of death is identical to their recorded date of birth; and
- customers, with a recorded date of death, who are associated with active cases.

7.9 These date anomalies present a risk to the effective management of data in Cuba. As noted previously, many business events within the database are triggered by dates. For example, an approaching liability end date might trigger communication with customers or cessation of employer withholdings beyond that date. It is essential—and mandated by the child support legislation[148]—that children's dates of birth are accurately recorded in Cuba.

[148] *Child Support (Registration and Collection) Act 1988*, Part III, Division 2, Section 26.

Inaccurate dates of birth could result in liabilities ceasing before they should, or continuing when they should have ceased.

7.10 The approach of a child's 18th birthday may also trigger communication with the customers in order to determine whether child support liability should be extended to the end of a school year. Critical communication, such as this, may also be affected if recorded dates of birth are not accurate.

Names and addresses

7.11 Inconsistent recording of customers' names and addresses creates a number of problems and reduces the integrity of customer data generally. Within the dataset provided by the CSA, the ANAO observed instances of:

- customer records with some, or all, name fields blank or 'UNKNOWN';
- names of organisations appearing in the individual customer table;
- invalid, spurious and evidently misspelt names;
- inappropriate entries in the 'Title' field; and
- name element entries appearing in the wrong fields.

7.12 The incorrect recording of customers' names creates some risk to the agency's reputation, through the incorrect addressing of mail. A more serious business impact exists, however, in that these records make the reliable identification of customers, and data matching with other agencies, much more difficult. Furthermore, the child support legislation requires that each entry in the Child Support Register must record the names of the payer, payee and child/ren to whom the entry relates.[149]

7.13 Up to 12 per cent of currently active customer records do not display a valid address. While letters are not the only means of communication available to the CSA,[150] customer address information may also be used in data matching with other agencies. The significant number of active customers without a valid address recorded in Cuba will reduce the effectiveness of such activity.

[149] ibid.

[150] Other communication channels used by CSA include the Internet and telephone.

Multiple customer records

7.14 The ANAO observed a significant number of customer records that had been identified by the CSA as multiple (or duplicated) records—approximately 18 000, or 0.4 per cent of all individual customer records. Many of these had been flagged in accordance with the CSA's guidelines. Others had not been correctly flagged, but are nevertheless easily recognisable as multiple records. The existence of these records in the production environment of Cuba presents a negligible business risk—as long as they are not linked to a case.

7.15 Of greater concern is a larger set of records—approximately 27 500, or 0.6 per cent of all individual customer records—that appear to be multiple customer registrations that have not been identified as such by the CSA. These multiple customer registrations exist for both individual customers and employers.[151]

7.16 Many apparent multiple individual customer records are not linked to a case; however, there is some risk that they might be inadvertently linked to a case in the future. The high-risk category comprises undetected multiple records that are linked to active cases. The existence of these records could result in incorrect liability assessments. There were up to 360 customers in this category.

7.17 Fragmenting customer information across multiple records also presents a number of risks, as customers may:

- appear as deceased on one record and alive on another;
- display different addresses, telephone numbers, and dates of birth;
- display inconsistent associations with CSA cases; and
- display different TFNs and/or CRNs across their records.

7.18 The CSA uses TFNs and CRNs (among other things) to identify its customers. The agency also uses TFNs to gather information on customers' incomes and to intercept the tax refunds of payers who have incurred a debt. This audit has revealed certain weaknesses in the accuracy and reliability of TFNs and CRNs, especially the existence of duplicate CRNs in Cuba and the situation where the same customer displays two different TFNs. Once again, relatively few in number—493 customers, or 1.8 per cent of those customers

[151] In the case of employers, some 38 000 records, or approximately 14 per cent of all employer records, may constitute multiple customer records. Almost 21 000 of these had been identified as multiple records by the CSA.

with multiple records—these inaccurate records may have serious implications for the individuals involved and may also reduce the CSA's capacity to effectively data match these records with other government agencies.

Data migration issues

7.19 The dataset provided by the CSA contained a significant number of records that had been migrated from the CSA's previous IT platform to Cuba in 2002. During fieldwork, a number of CSA staff interviewed by the ANAO referred to a variety of issues related to the migration that still persist—indeed, a group of Technical Support Officers stated that these issues represent their greatest challenge. Specifically, staff cited the loss of (electronic) proof-of-parentage information and details of care arrangements during the conversion as a major, continuing generator of additional work. Many Technical Support Officers interviewed during the audit reported that they were required to rebuild these cases, in order to perform work associated with new CSA cases.

7.20 The planned migration of the existing database to a new version of Cuba, in 2008, involves a risk that such issues could arise once more. This risk is compounded if corrupted records from the previous migration are migrated yet again. During the audit, the ANAO formed the view that the CSA's Technical Support Officers, as a group, represent a valuable source of knowledge and experience in the identification of ongoing problems associated with corrupt records. The ANAO encourages the CSA to draw upon this valuable resource to identify and resolve as many data conversion issues as possible before migrating the existing database to a new system.

Conclusion

7.21 Child support cases involve complicated interrelationships of payers, payees, claimed-for children and relevant dependants, employers, overseas agencies and other parties. A significant number of customers are involved in more than one case, introducing further complexity. Designing, building and maintaining a database capable of tracking these intricate connections, and administering complex legislation, is a difficult challenge.

7.22 Although this audit has highlighted a number of areas in which the CSA could significantly improve the quality and reliability of customer data, the ANAO concluded that the majority of records in Cuba are sufficiently accurate, complete and reliable to support the effective administration of the Child Support Scheme.

7.23 As well as improving the quality of individual customer records, the CSA could also improve the overall management of data in Cuba through a more consistent use of values for system dates. The CSA would also benefit from the introduction of an active programme to regularly test data quality and the application of business rules in Cuba. To this end, the ANAO notes the work commenced by the CSA on its Data Quality Improvement Programme.

7.24 The value to be gained from these improvements takes on added significance as the CSA moves to implement the third stage of reforms of the Child Support Scheme. A comprehensive cleansing of current data and the consistent application of business rules, within the database, should ensure that the CSA commences operation of the new version of Cuba, in 2008, with the highest possible quality dataset.

Ian McPhee
Auditor-General

Canberra ACT
30 November 2007

Appendices

Appendix 1: Cuba data extracts

On 26 April 2007, the CSA provided the ANAO with 16 tables extracted from the Cuba database. These tables were extracted over the period 31 March 2007 to 1 April 2007. The name of each table, the number of records comprising each table and a brief description are presented in the following table.

Table A 1

Cuba tables comprising the data extract

Table name	Number of records	Description
BASIC_VALUE_SET	42	Set of values used in all liability calculations
FORMULA	26 442 258	Actual formula values
INCOME	31 207 814	Customer income details
CASE	1 439 570	Primary key for CASE
CASE_IND	1 623 770	Case indicators
CASE_CLNT_ROLE	7 207 100	Case-Customer-Role relationships
CASE_LIABILITY	27 567 612	Liability calculations for cases
REGISTRATION	2 243 351	Each case may have many registrations
CHLD_SPRT_PERD	11 085 342	Child support period
CUSTOMER	4 829 880	Primary key for CUSTOMER
CUSTOMER_IND	11 247 722	Customer indicators
CUSTOMER_ADDRESS	7 704 095	Customer address details
IDV_CLNT	4 571 020	Individual customer details
IDV-CLNT_NM	5 513 025	History file of individual customer names
NON_IDV_CUSTOMER	258 859	Non-individual customer details
CODE_TABLE	17 464	Lookup table for valid values of many code-type fields used in Cuba

Source: ANAO – data provided by the CSA.

Appendix 2: Customer indicator type codes

Table A 2

Customer indicator type codes identified in the Cuba customer indicator table

No. of records	Type code shown in customer indicator	Legal Values in code table	Description
—		AEWA	ALLOW EMPLOYER WITHHOLDING
59 267	ALWY	ALWY	Always Issue Statement
1 279 115	ARR	ARR	Arrears
1 981	BAN	BAN	Customer has been declared Bankrupt by a Court
5 869	CAT1	CAT1	DMS Category One
6 785	CAT2	CAT2	DMS Category Two
5 318	CAT3	CAT3	DMS Category Three
1 884	CAT4	CAT4	DMS Category Four
—		CMPL	Complaint in Progress
2 681 530	CNLK	CNLK	Customer is also a Centrelink Customer
—		COLL	CSA Collect (Opt In)
26 431	CONT	CONT	Customer has Nominated Contact Person
2 650	CPCT	CPCT	CUSTOMER PREFERRED CONTACT TIME
32 524	DEAD	DEAD	Customer Deceased
1 694 651	DEFT	DEFT	Default Income
4 287	DISE	DISE	DISALLOW ELECTION
38 081	DOFF	DOFF	Customer/Case Debt Off Set
4 992	DVIO	DVIO	Domestic Violence
31 419	ESD	ESD	Customer is a ESD Customer
22 905	EWAR	EWAR	ENDED WITH ARREARS
670	HEAI	HEAI	Customer has hearing impairment
53 119	INOC	INOC	International Other Party Customer
61 171	INPC	INPC	International Primary Customer
9 098	INTR	INTR	Customer requires an interpreter
5 050	LINE	LINE	Low Income Non-Enforcement period
331	LIQ	LIQ	Liquidation Action in Progress for Non-Individual…
911	LITP	LITP	Customer has literacy problems
9 882	LOCI	LOCI	Customers location is an issue
827 890	MCAS	MCAS	Customer has Multiple Cases
1 328	MOBP	MOBP	Customer has mobility problems
—		NCOL	Private Collection (Opt Out)
—		ONC	Overall Non Care

No. of records	Type code shown in customer indicator	Legal Values in code table	Description
140 859	ORGA	ORGA	Active
3 166	ORGC	ORGC	Ceased Trading
21 409	ORGD	ORGD	Duplicate
1 419	ORGI	ORGI	Insolvent
474 515	PARR	PARR	Pay Arrangement
1 150 556	PCC	PCC	Private Collect Candidate
—		PENF	Payee Enforcement in Progress
439 389	RECY	RECY	Action to Recover Outstanding Customer Debts…
69 062	REP	REP	Representative
3 862	REQ	REQ	CSA Private Coll (Registrar initiated opt out)
6 138	RICA	RICA	RICA
9 496	SCSM	SCSM	Multiple Unfinalised SCS Activities
42 421	SCSR	SCSR	Repeat SCS Activities
98	SIGI	SIGI	Customer has sight impairment
252	SIGN	SIGN	Customer requires sign interpreter
99	SPEI	SPEI	Customer has speech impairment
—		SSAT	SSAT REVIEW LODGED
13 726	SULP	SULP	Suspension of Late Payment Penalty Imposition
146 723	SUPP	SUPP	Suppress Statement
269 867	SWHO	SWHO	STOP WITHHOLDING
133 684	TOPD	TOPD	Top Up Debt
759 111	TRAC	TRAC	Tracing Activity being Undertaken in Relation to Cust.
289	TTYM	TTYM	Customer uses TTY machine
—		UTR	Untraceable
27 149	WOFF	WOFF	Customer Debt Written Off
665 293	WTH	WTH	Withholding
—		INCP	INCOMPLETE (**)

Source: ANAO analysis.

Appendix 3: Refining a list of possible multiple customer records

1. As indicated in Chapter 6 of the report, the ANAO examined all 4 571 020 records in the individual customer table. Searching for records with an exact match on first name, middle initial, surname and date of birth identified a set of 102 891 records that could potentially involve multiple customer records.

2. Another 18 records matched on name elements and address history elements but displayed a different date of birth. The ANAO considered these may involve a data entry error on one date of birth record. Another 494 records matched on first name, middle initial, date of birth and some address elements but not on surname. The ANAO mapped these records to the individual customer name history and verified with a reasonable level of confidence that these records related to customers who had changed their surname.

3. Because of the match on middle initial, rather than full middle name, (where recorded) this set of records included potential pairs where the customer's middle name was different. The ANAO identified and removed these pairs from the dataset. For example, a pair of records, with the same date of birth, displaying John Peter Smith and John Paul Smith was removed. However, a pair of records displaying John Peter Smith and John P Smith was retained.

4. The ANAO mapped the dataset against the case_customer_role table and removed any pairs of records where at least one of those records was included in the group of 118 676 unlinked customer records—that is, where a CSID had never been linked to a case. The ANAO also removed all but three pairs[152] of records where at least one record was marked as a known duplicate or displayed a surname of unknown.

5. In this way the ANAO refined the original list of over 100 000 potential multiple customer records to 55 266 records, in which it has a reasonable level of confidence. Each of these records displays an association with a CSA case, albeit with various cases status values.

[152] These three pairs displayed other attributes which supported their inclusion.

Appendix 4: Department of Human Services' Response

Australian Government
Department of Human Services

Helen Williams AO
Secretary

Reference: F2006/469

Mr John Meert
Group Executive Director
Performance Audit Service Group
Australian National Audit Office

Copy: Mr Matt Miller, General Manager, Child Support Agency

Dear Mr Meert

PERFORMANCE AUDIT: DATA INTEGRITY IN THE CHILD SUPPORT AGENCY

I refer to your letter to me on 5 October 2007, enclosing the proposed audit report on Data Integrity in the Child Support Agency for comment pursuant to section 19 of the Auditor-General Act 1997 (the Act).

As required by sub-section 19(4) of the Act, I am providing written comment from the Department of Human Services, including the Child Support Agency, to the Auditor-General.

The following summary response is to be included in the Audit Report summary section and be used in the Brochure to accompany the Audit Report:

The Department of Human Services (DHS) appreciate the assurance provided by the ANAO's Data Integrity Audit report outlining that the majority of records in the Child Support Agency's (CSA) primary database, Cuba, are sufficiently accurate, complete, and reliable to support the effective administration of the Child Support Scheme. However, where anomalies on individual records may potentially affect a customer's child support liability, we will resolve these errors and develop mechanisms and controls to support the continuous improvement of the data holdings in Cuba.

PO Box 3959, Manuka ACT 2603 • Telephone (02) 6223 4410 • Facsimile (02) 6223 4489
Internet www.humanservices.gov.au

Furthermore, DHS values the recognition provided by the ANAO of the work commenced by the Data Quality Improvement Program within CSA and agrees with the benefit that would be gained with incorporating ongoing testing of data quality and the application of business rules, in Cuba, as part of this program.

Recommendation No. 1
Para 3.56

The ANAO recommends that CSA:

- investigate and correct records in the case, case indicator and case liability tables that display anomalies in the various start and end date fields;
- ensure a consistent use by CSOs of high dates and low dates in case liability start date and case liability end date fields; and
- resolve instances of inconsistence between case status and current case liability determination.

DHS's response:

DHS agrees that the anomalies identified will need to be corrected. To this end, CSA will review all anomalies identified and develop appropriate corrective action, in line with material impacts to customers. Furthermore, CSA will review the inconsistent high dates and low dates, and active cases without a current case liability determination, outlined in the audit report, and will undertake corrective action, where required, in addition to on going monitoring of these fields, via CSA's Data Quality Improvement Program.

Recommendation No. 2
Para 4.79

The ANAO recommends that the CSA identify and remove redundant records from the individual customer table in Cuba.

CSA's response:

DHS agrees. CSA will undertake to establish mechanisms to continuously identify redundant records, and remove the records from the production database accordingly.

PO Box 3959, Manuka ACT 2603 • Telephone 02 6233 0411 • Facsimile 02 6233 0489
Internet www.humanservices.gov.au

Recommendation No. 3
Para 4.81

The ANAO recommends that, either separately or as part of its Data Quality Improvement Programme, the CSA will:

- cleanse data in the fields describing customers' names and addresses;
- investigate and resolve anomalies in customer TFN and CRN records;
- develop and implement a quality assurance system to ensure a consistent standard of recording names and addresses in the revised version of Cuba to be introduced in 2008; and
- review and improve the effectiveness of data entry controls to ensure that, for individual customer records:
  - only valid, nine digit TFNs may be entered;
  - only valid values, stored in a Cuba Code Table, are able to be entered;
  - dates of birth are recorded accurately; and
  - all mandatory fields are populated with valid entries.

DHS's response:

It is agreed that CSA should undertake data cleansing of customers' names and addresses, in addition to resolving anomalies in customer TFN and CRN records. CSA would also like to note that controls currently exist within Cuba where invalid TFNs cannot be entered.

As noted by the ANAO, CSA's Data Quality Improvement Program is in it early stages, and the development of a quality assurance system and review of data entry controls will be a focus for this program, to help continuously improve the quality of CSA's data.

PO Box 3959, Manuka ACT 2603 • Telephone 02 6233 0411 • Facsimile 02 6233 0489
Internet www.humanservices.gov.au

Recommendation No. 4
Para 5.19

The ANAO recommends that the CSA:

- cleanse the employer table of redundant records, including occurrences of multiple records for employers;
- resolve the issues surrounding the use of eight digit and nine digit TFNs for employers; and
- ensure that valid data populates all mandatory fields in the employer table.

DHS's response:

DHS agrees. The anomalous records and the identified mandatory fields not populated will need to be investigated and corrected. CSA will also undertake to establish mechanisms to continuously identify redundant employer records and multiple records for employers, and remediate accordingly.

Recommendation No. 5
Para 6.60

The ANAO recommends that the CSA resolve all instances of multiple customer records in Cuba.

DHS's response:

CSA will undertake to resolve the multiple customer records outlined in the audit, in line with current business procedures. CSA would like to note that there are currently controls that exist within Cuba to minimise the creation of multiple records.

Furthermore, as part of the Quality Assurance system, to be developed, CSA will establish appropriate mechanisms to continuously identify multiple customer records, and remediate accordingly.

I note that the conclusion of the proposed report is that, overall, the majority of records in Cuba are sufficiently accurate, complete, and reliable to support the effective administration of the Child Support Scheme, and that there are a number of areas the Child Support Agency could also improve the quality and reliability of data in Cuba.

We accept the recommendations and, and as you can see in the comments above, the Child Support Agency has either taken action, has action in train, or plans to put in place action to address all of the findings.

PO Box 3959, Manuka ACT 2603 • Telephone 02 6233 0411 • Facsimile 02 6233 0489
Internet www.humanservices.gov.au

I would like to thank your staff for their professionalism and co-operative manner in which this audit was undertaken. If there are any questions concerning our comments, please contact Mr Patrick Hadley , DGM CIO ICT CSA, on (02) 627 28785.

Yours sincerely

Helen Williams

9 November 2007

PO Box 3959, Manuka ACT 2603 • Telephone 02 6233 0411 • Facsimile 02 6233 0489
Internet www.humanservices.gov.au

Index

Series Titles

Audit Report No.1 2007–08
Acquisition of the ABRAMS Main Battle Tank
Department of Defence
Defence Materiel Organisation

Audit Report No.2 2007–08
Electronic Travel Authority Follow-up Audit
Department of Immigration and Citizenship

Audit Report No.3 2007–08
Australian Technical Colleges Programme
Department of Education, Science and Training

Audit Report No.4 2007–08
Container Examination Facilities Follow-up
Australian Customs Service

Audit Report No.5 2007–08
National Cervical Screening Program Follow-up
Department of Health and Ageing

Audit Report No.6 2007–08
Australia's Preparedness for a Human Influenza Pandemic
Department of Health and Ageing
Department of Agriculture, Fisheries and Forestry

Audit Report No.7 2007–08
The Senate Order for Departmental and Agency Contracts (Calendar Year 2006 Compliance)

Audit Report No.8 2007–08
Proof of Identity for Accessing Centrelink Payments
Centrelink
Department of Human Services

Audit Report No.9 2007–08
Australian Apprenticeships
Department of Education, Science Training

Audit Report No.10 2007–08
Whole of Government Indigenous Service Delivery Arrangements

Audit Report No.11 2007–08
Management of the FFG Capability Upgrade
Department of Defence
Defence Materiel Organisation

Audit Report No.12 2007–08
Administration of High Risk Income Tax Refunds in the Individuals and Micro Enterprises Market Segments
Australian Taxation Office

Audit Report No.13 2007–08
The Australian Taxation Office's Approach to Managing Self Managed Superannuation Fund Compliance Risks
Australian Taxation Office

Audit Report No.14 2007–08
Performance Audit of the Regional Partnerships Programme:
Volume 1–Summary and Recommendations
Volume 2–Main Report
Volume 3–Project Case Studies
Department of Transport and Regional Services

Audit Report No.15 2007–08
Administration of Australian Business Number Registrations: Follow-up Audit
Australian Taxation Office

Current Better Practice Guides

The following Better Practice Guides are available on the Australian National Audit Office Website.

Public Sector Internal Audit	
An Investment in Assurance and Business Improvement	Sep 2007
Fairness and Transparency in Purchasing Decisions	
Probity in Australian Government Procurement	Aug 2007
Administering Regulation	Mar 2007
Developing and Managing Contracts	
Getting the Right Outcome, Paying the Right Price	Feb 2007
Implementation of Programme and Policy Initiatives:	
Making implementation matter	Oct 2006
Legal Services Arrangements in Australian Government Agencies	Aug 2006
Preparation of Financial Statements by Public Sector Entities	Apr 2006
Administration of Fringe Benefits Tax	Feb 2006
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
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
Some Better Practice Principles for Developing Policy Advice	Nov 2001
Rehabilitation: Managing Return to Work	June 2001
Business Continuity Management	Jan 2000
Building a Better Financial Management Framework	Nov 1999
Building Better Financial Management Support	Nov 1999
Commonwealth Agency Energy Management	June 1999
Security and Control for SAP R/3	Oct 1998
New Directions in Internal Audit	July 1998
Controlling Performance and Outcomes	Dec 1997
Protective Security Principles (in Audit Report No.21 1997–98)	Dec 1997