Joint Committee of Parliamentary Advisory Authority (JCPAA) Report June 2007

Centrelink customer data quality activities undertaken by the Data Quality Team
The Data Quality Team was established in November 2005 in response to an ANAO Audit report into the integrity of Centrelink's electronic customer data (ANAO report 29/2005-2006).

This Team is based in Coffs Harbour, NSW, and is comprised of 7 full time equivalent staff. The Team was originally set up to correct customer records, as indentified in the ANAO report. However, they have expanded their scope to report on the status and condition of Centrelink's customer data. These reports are represented through graphs (attached to the end of this report).

The Team have also commenced educating front of office workers on the importance of collecting customer data correctly the first time round.

Centrelink is dealing with a dynamic customer base and has recently received funding to pursue the ongoing work of improving Centrelink's performance in the area of data quality. Some of the projects the Centrelink has received funding for are identified in this report.

Since the establishment of the Data Quality Team in November 2005, in response to the ANAO audit report no. 29/2005-2006, the Team has accomplished the following activities:

Duplicates
- 33,663 identified - Data Quality Team examined all these records, with the following results:
    - 33,343 logically deleted
    - 15 Centrelink International Services (CIS) cases remaining
    - 294 outsourced to Income Security Integrated System (ISIS) to fix
    - 7 referred to systems to fix
    - 4 must remain - unable to be resolved at this point due to requirement to retain customer data held on in both environments

Multiples
- Audit estimated 500,000 customers with multiple Centrelink Reference Numbers (CRNs), but did not specifically identify all customer records
    - Data Quality Team has identified in excess of 300,000 customer records with multiples
    - Audit identified 4179 cases - these have all been checked, and 4169 have been finalised, with 12 requiring further investigation
    - In total, Data Quality Team has linked or logically deleted 21,958 customer records with multiple CRNs
    - With the implementation of the Delete Person function in September 2006, Data Quality is now able to logically delete (i.e. "lock up") secondary multiple records containing no relevant customer data. To date, Data Quality has deleted in excess of 10,730 secondary multiple CRNs.

Data Quality
- The Audit identified a range of data quality issues, and customer records affected by these issues. These include problems related to Tax File Number (TFN), gender, date of birth

- 170,872 such cases have been fixed to date
- All TFN cases have been examined, and corrective action taken, or files outsourced to existing teams that have responsibility for these functions (i.e. Accelerated Claimant Matching (ACM) team is working on customers with no TFN due to temporary exemption codes)
- Data Quality has prepared a Detailed Business Report relating to the TFN issues, and this report has been delivered to relevant Business Teams including ACM, Data Matching and Identity Management. Identity Management is now working on a project plan to address the recommendations made by the Data Quality Team.

Current Focus:

Data Quality Team is preparing our Action Plan for 2007/2008 with a focus on (funding received to continue with the work):
- Continued review of duplicate records to ensure no further cases occur;
- Ongoing identification of multiple records, using procedures developed in 2006/2007 including changes to CSA instructions to ensure all detected cases of multiple records are referred to Data Quality
- Working with Business teams to identify instances of multiple records where bulk corrective action can be taken - e.g. Data Quality has identified 275,000 customers have both a child (CHI) and a person (PER) CRN - we are working with relevant teams to develop Business Rules to enable to bulk action to delete an estimated 200,000 of these multiples
- Ongoing correction of identified records with data quality issues
- Continual review to identify fragmented payments (i.e. Customers with entitlements on different CRNs)

Data Integrity
- At the time of the ANAO audit in 2005, there were between 8 and 9 million outstanding DI errors detected in ISIS. Importantly, 5 – 600,000 of these were Priority 1 (critical) errors and 5 – 6,000,000 of these were Priority 2 (high) errors.
- As at 4th July 2007 there are 2.7 million outstanding DI errors detected in ISIS. Importantly, 148,910 of these are Priority 1 (critical) errors and 1,121,577 of these are Priority 2 (high) errors.
- These reductions have been brought about in a number of ways:
  - DI check business rule reviews by ISIS application areas to reduce the number of false positives being reported as errors;
  - Fixes to system code by ISIS application areas;
  - Bulk low-level database fixes by the Data Integrity Unit; and
  - Introducing comprehensive pre-Major Release DI checking in test environments to catch system faults before they are released into production.
- Work has been conducted to move the focus on DI activity from a purely technical perspective to one encompassing business activities. To this end, the template for defining a DI error has been significantly revised to capture business metadata, a new prioritisation scheme for DI errors has been developed, and a DI Reporting Committee was formed and chaired by Information Management Branch.
- A measure, the Data Integrity Index, has been developed to represent and report Centrelink's DI as a whole. It is essentially a weighted measure of 'error' on average for each customer (with weights proportional to the priority of each error, ie more weight given to errors of higher priority).

Removing inactive or deceased customers from the production environment

In response to this audit recommendation (3b), Centrelink has received funding to create a separate online production environment to host inactive and deceased customers where there is no continuing business need to retain these records.

Author

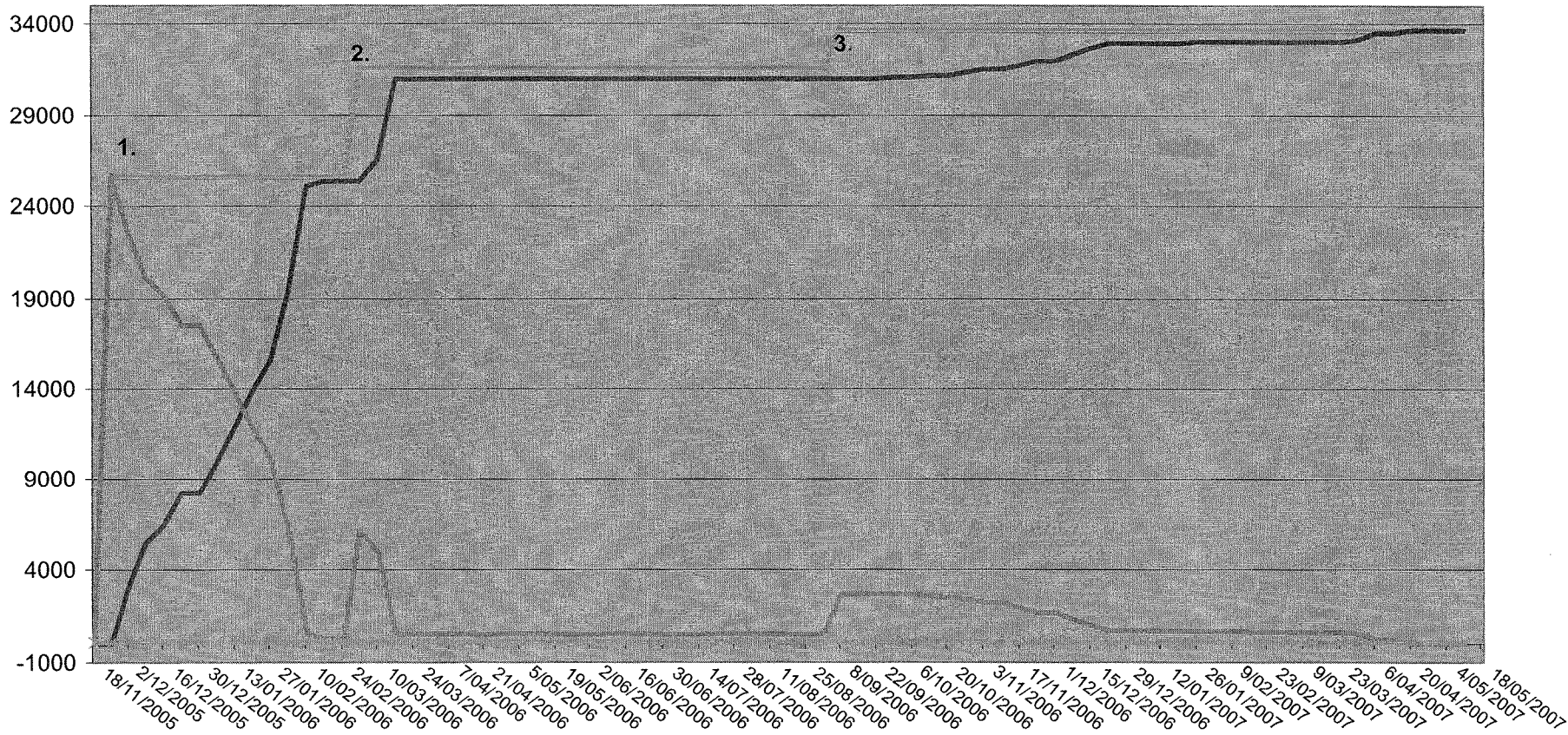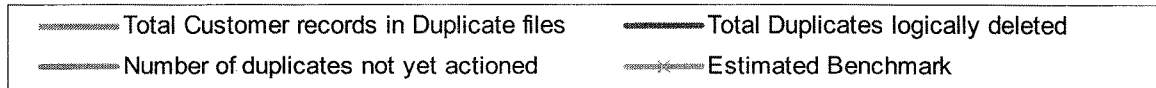Ms Papiya Chakravarti

Director, Records and Data Management Centre

Centrelink.

P: 62447311

M: 0417486669

E: papiya.Chakravarti@Centrelink.gov.au

Data Quality Fixit Team - Duplicate CRN Cleanup History

Total Customer records in Duplicate files —— Total Duplicates logically deleted
—— Number of duplicates not yet actioned —— Estimated Benchmark

To date the **Data Quality Fixit Team** has logically deleted a total of 33,019 duplicate CRNs leaving 651 requiring investigation. We have set the benchmark for duplicates at 50 as there are business rules in place that would allow for an estimated 50 situations where a duplicate CRN is valid, eg CIS protocols.

1. The original Duplicates file was received in December 2005 containing 25,662 CRNs, all but 302 of these, which require further investigation, were logically deleted by February 2006.

2. A second Duplicates file was received late Februrary 2006 containing 5881 CRNs. We were able to logically delete 5,671 of these by the end of March 2006, leaving a total of 566 CRNs requiring further investigation.

3. In September 2006 a third Duplicates file was received, this contained 2,120 CRNs. By February 2007 we had logically deleted 2,035 of these with another 85 requiring further investigation.

Date. Week ending	Total CRNs in files	Total Multiples actioned	Total Multiples not yet fixed
2/12/2005	4179	0	4179
9/12/2005	4179	105	4074
16/12/2005	4179	190	3989
23/12/2005	4179	497	3682
30/12/2005	4179	745	3434
6/01/2006	4179	953	3226
13/01/2006	4179	1979	2200
20/01/2006	4179	3013	1166
27/01/2006	4179	3174	1005
3/02/2006	4179	3235	944
10/02/2006	4179	3235	944
17/02/2006	4179	3360	819
24/02/2006	4179	3717	462
3/03/2006	4179	3717	462
10/03/2006	4179	3717	462
17/03/2006	4179	3717	462
24/03/2006	4179	3904	275
31/03/2006	4179	4099	80
7/04/2006	4179	4167	12

Data Quality Fixit Team - First Priority Multiple CRN Cleanup History

Legend: Total CRNs in files — Total Multiples actioned — Total Multiples not yet fixed
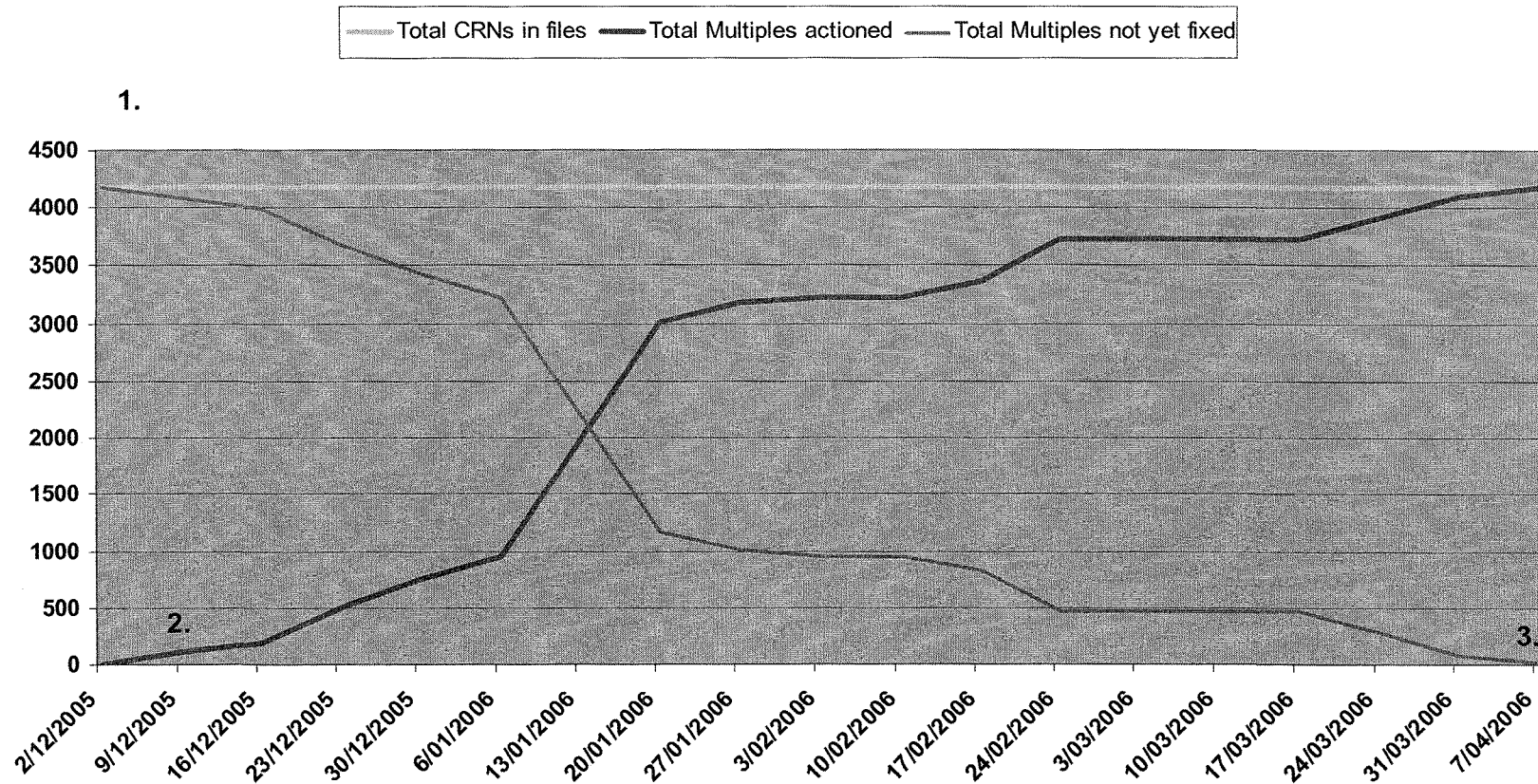
In December 2005 the Data Quality Fixit team received 6 files containing Multiple CRNs which were deemed "priority cases" due to the high risk involved. They were:- Highest Level multi CRNs CUR; Zhigher level Multi CRNs CUR; Zreasonable level Multi CRNs CUR; Post IP one line; Different DOD on 2 CRNs and DODs present Multi CRN. (DOD = Date of Death).

1. Files received in December 2005 containing 4,179 customers idetified as having more than one CRN.

2. Work commenced 9/12/05 to confirm if the CRNs belonged to the same person, confirm correct details coded on the customer record and where a customer had more than one CRN, identify the PRImary record and fix the record by linking the multiples via the MLC link.

3. By April 7th 2006 we had fixed 4,167 customer records leaving 12 requiring addition work. The remaining records that were not fixed have been referred to other teams for their attention. 2 records involved ITO remnants, 1 record has an outstanding debt on one record but is current on the other and 9 records exist where the customer is receiving payment on both CRNs. This does not involve any overpayment as the customer has an entitlement to both payments, eg, Age Pension on one CRN and Mobility Allowance on the other.

Since April 2006 the outstanding cases have been reviewed and fixed at various times. The only two remaining records involve AGE/CUR and MOB/CUR and have been referred to the consolidated Mobility Team.

Data Quality Fixit Team.
Clean up of "YAL/CUR double TFN" file.

—— checked —— completed —— outstanding

All 44 customer records were checked and completed by 24/3/06.
No records remain outstanding.

Total number of CRNs included in file = 44

50
45
40
35
30
25
20
15
10
5
0

17/02/2006 17/03/2006 14/04/2006 12/05/2006 9/06/2006 7/07/2006 4/08/2006 1/09/2006 29/09/2006 27/10/2006 24/11/2006 22/12/2006 19/01/2007 16/02/2007
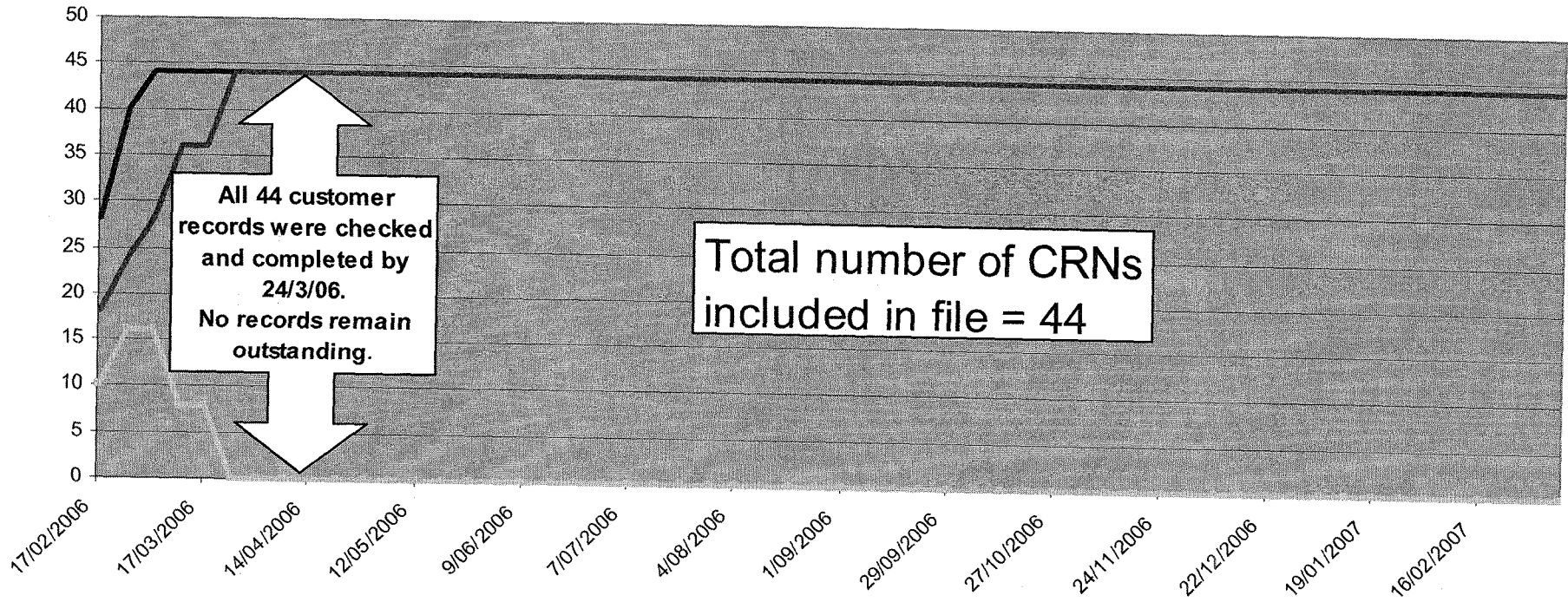
Data Quality Fixit Team.
Clean up of AGE/CUR no TFN file.

━━total checked ━━completed ⋯⋯ outstanding

Total number checked and completed as at 2/3/07 = 116,133

Total number of CRNs included in file = 116,133

1.

Number checked at 10/11/07 = 3,064

Number of records outstanding at 2/3/07 =

(y-axis: 0, 20,000, 40,000, 60,000, 80,000, 100,000, 120,000)

(x-axis: 17/02/2006, 3/03/2006, 17/03/2006, 31/03/2006, 14/04/2006, 28/04/2006, 12/05/2006, 26/05/2006, 9/06/2006, 23/06/2006, 7/07/2006, 21/07/2006, 4/08/2006, 18/08/2006, 1/09/2006, 15/09/2006, 29/09/2006, 13/10/2006, 27/10/2006, 10/11/2006, 24/11/2006, 8/12/2006, 22/12/2006, 5/01/2007, 19/01/2007, 2/02/2007, 16/02/2007, 2/03/2007)

1. On 10/11/06 all 116,133 records were completed by a bulk run which actioned all

Data Quality Fixit Team.   Cleanup
of "Shared TFN same Family name" file.

━━ Total checked ━━ completed ━━ outstanding

All 6,756 customer records checked by 10/3/06 (displayed as a blue line). At that stage we had only completed 6,541cases (displayed as a red line).

As at 9/3/07 6,696 cases have been completed, which leaves 60 outstanding.

Records outstanding as at 9/3/07 = 60. Displayed by yellow line.
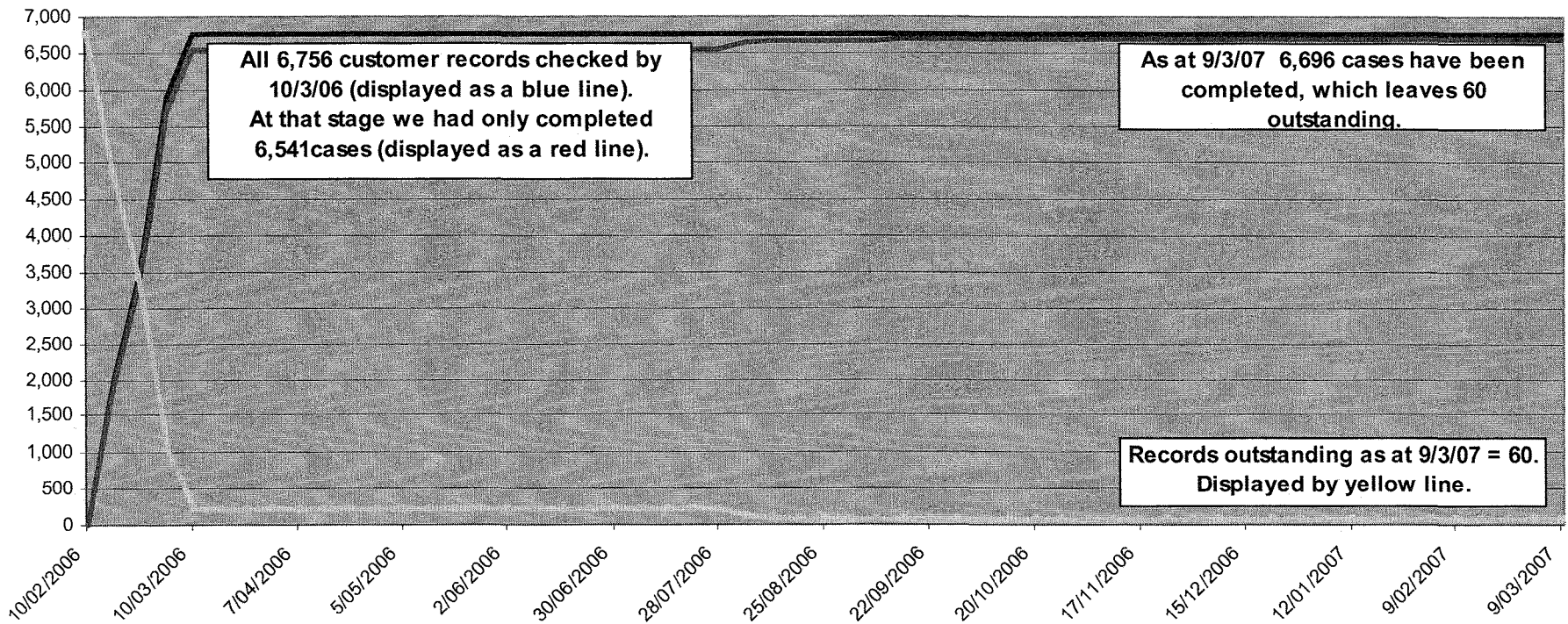
Data Quality Fixit Team.
Cleanup of 'AGE with no TFN' file.

Total checked   ▬ ▬ ▬ completed   ▬▬▬ outstanding
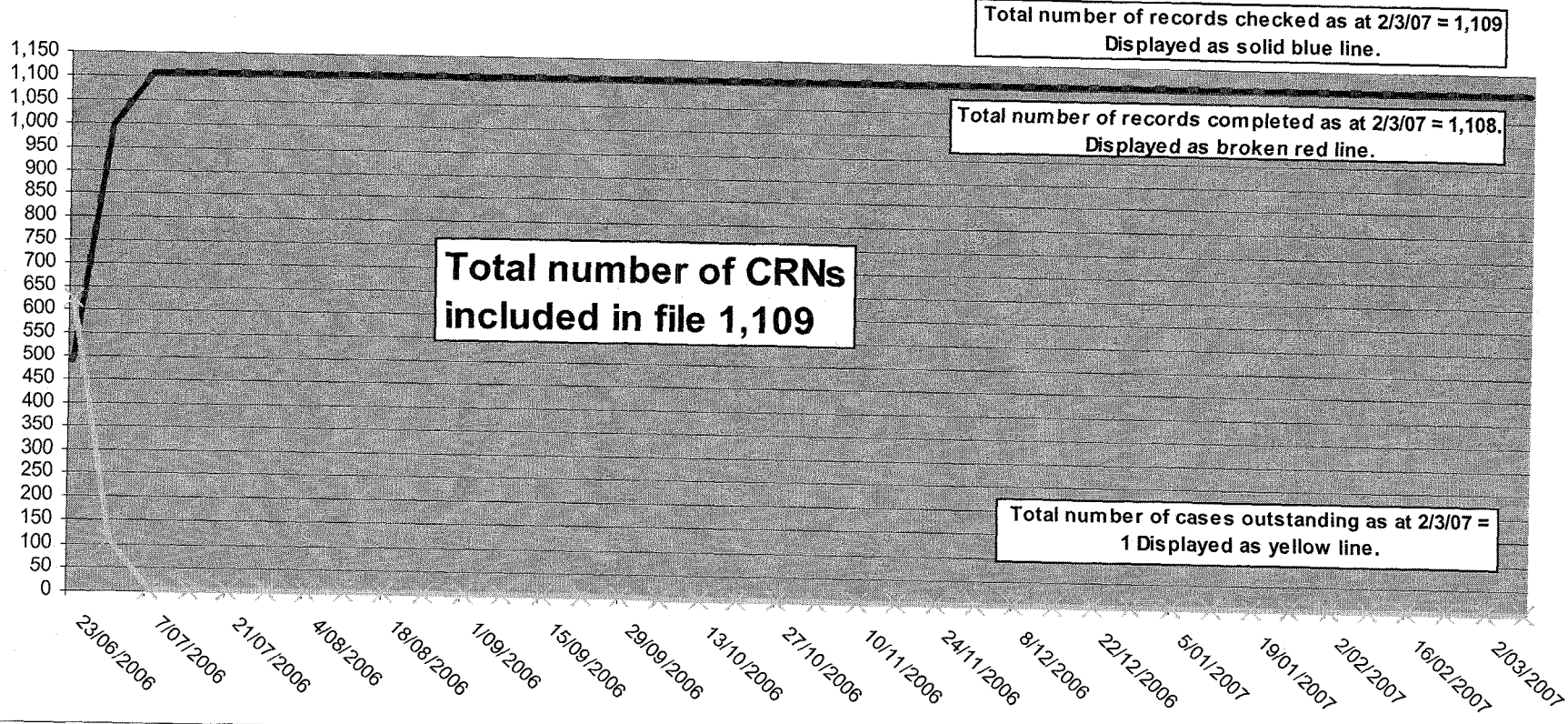
Total number of records checked as at 2/3/07 = 1,109
Displayed as solid blue line.

Total number of records completed as at 2/3/07 = 1,108.
Displayed as broken red line.

Total number of CRNs included in file 1,109

Total number of cases outstanding as at 2/3/07 = 1 Displayed as yellow line.
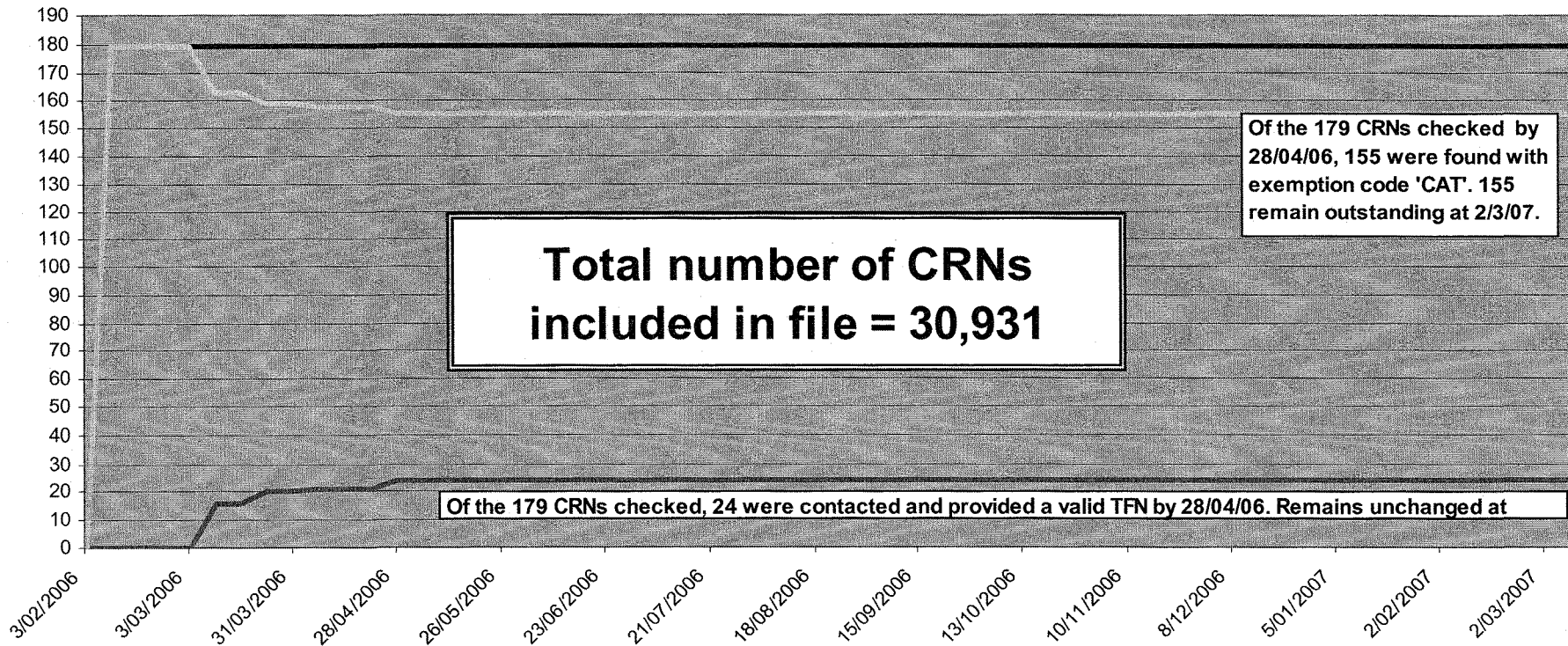
Data Quality Fixit Team.
Cleanup of 'YAL/CUR no TFN' file

— Total checked — completed — outstanding

Of the 30,931 CRNs contained in the file, 179 were checked by 10/2/06. File unchanged since 28/04/06 pending work being conducted by BI Murwillumbah.

Of the 179 CRNs checked by 28/04/06, 155 were found with exemption code 'CAT'. 155 remain outstanding at 2/3/07.

Total number of CRNs included in file = 30,931

Of the 179 CRNs checked, 24 were contacted and provided a valid TFN by 28/04/06. Remains unchanged at

y-axis: 190, 180, 170, 160, 150, 140, 130, 120, 110, 100, 90, 80, 70, 60, 50, 40, 30, 20, 10, 0

x-axis: 3/02/2006, 3/03/2006, 31/03/2006, 28/04/2006, 26/05/2006, 23/06/2006, 21/07/2006, 18/08/2006, 15/09/2006, 13/10/2006, 10/11/2006, 8/12/2006, 5/01/2007, 2/02/2007, 2/03/2007

Data Quality Fixit Team.
Cleanup of 'Shared TFN different 1st name' file.

file size ━Total checked ━completed ▬investigate

Total number of CRNs in file 25,680

Total CRNs checked at 2/3/07 = 22,794
Displayed as Yellow line.

Total Records completed at 2/3/07 = 22,526.
Displayed as Red line

Records remaining under investigation at 2/3/07 = 268.
Isplayed as Blue line.
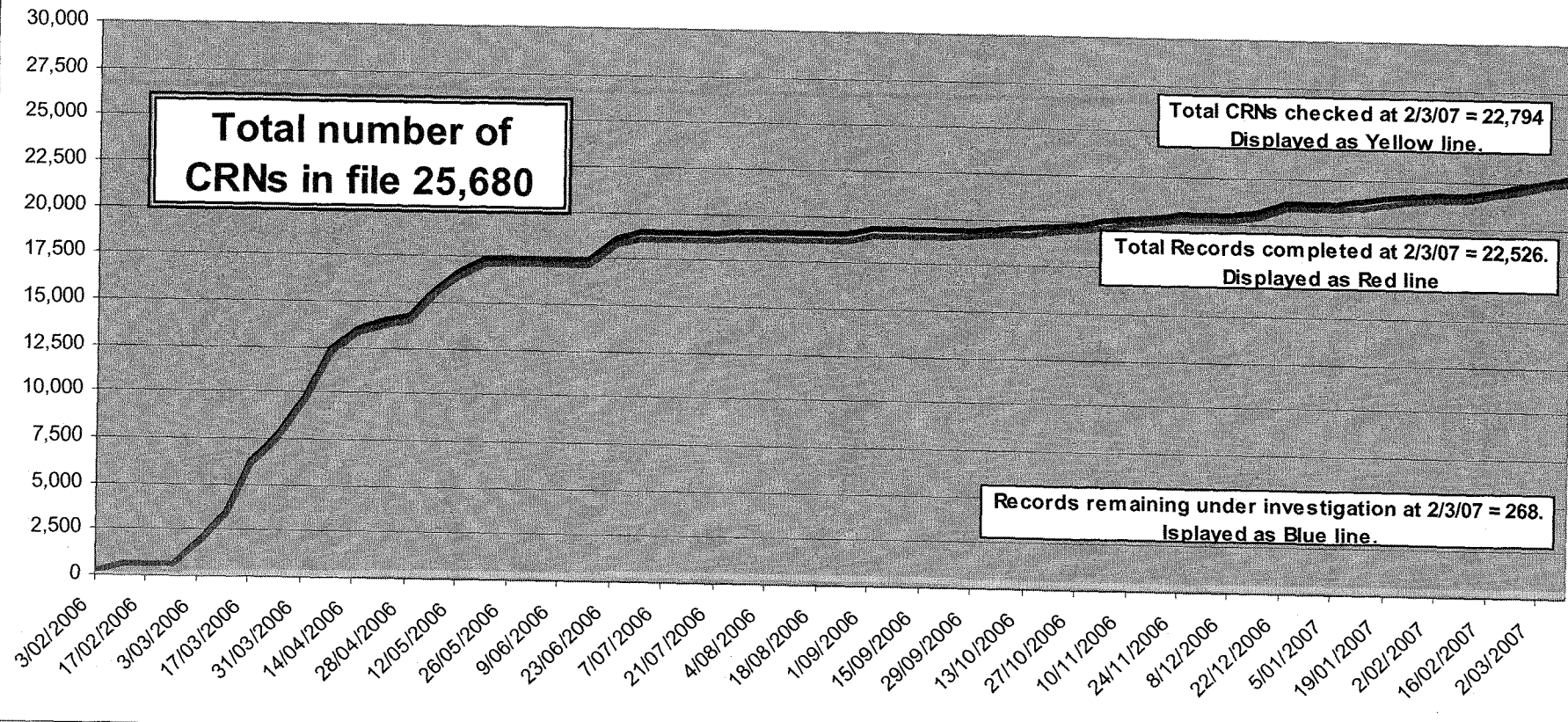
Data Quality Fixit Team
Cleanup of "YAL/CAN TFN exempt" file.

Total checked ——completed ···×···outstanding
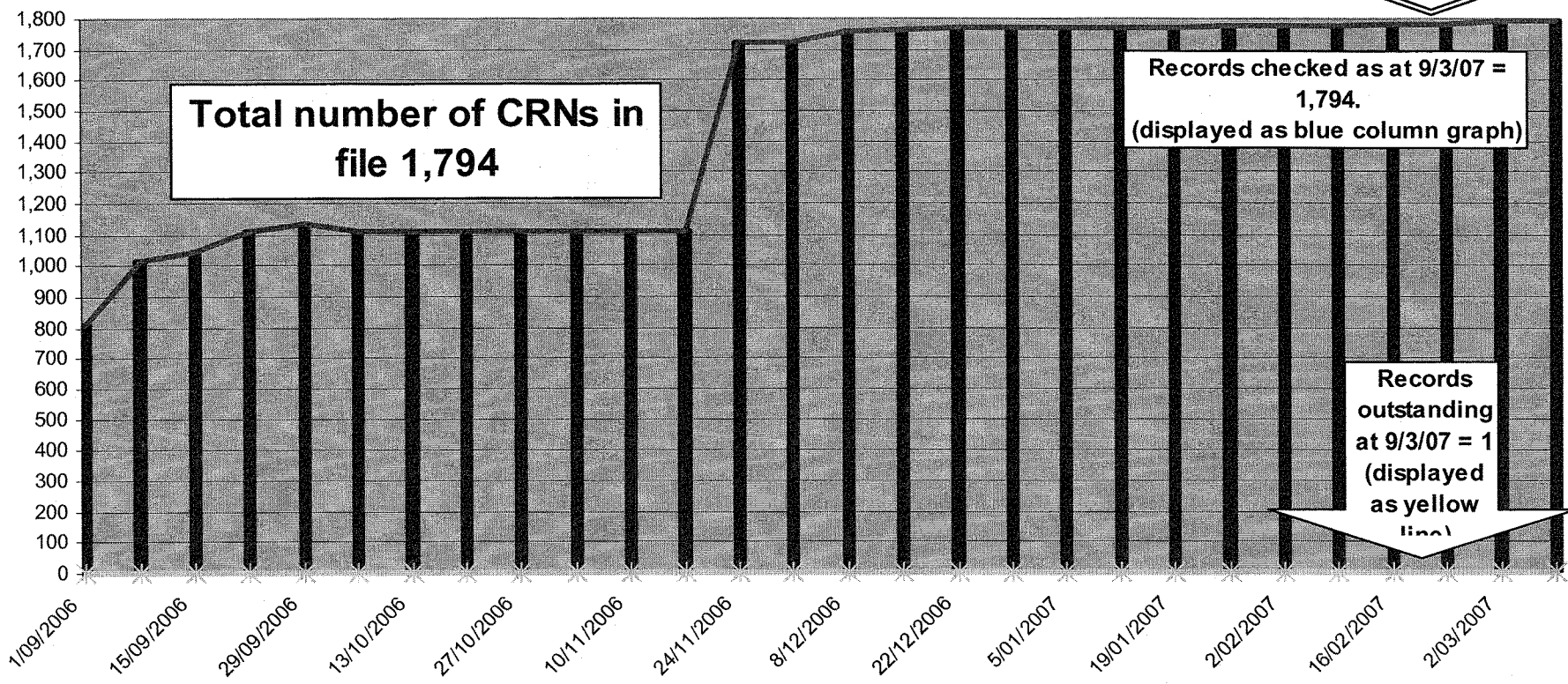
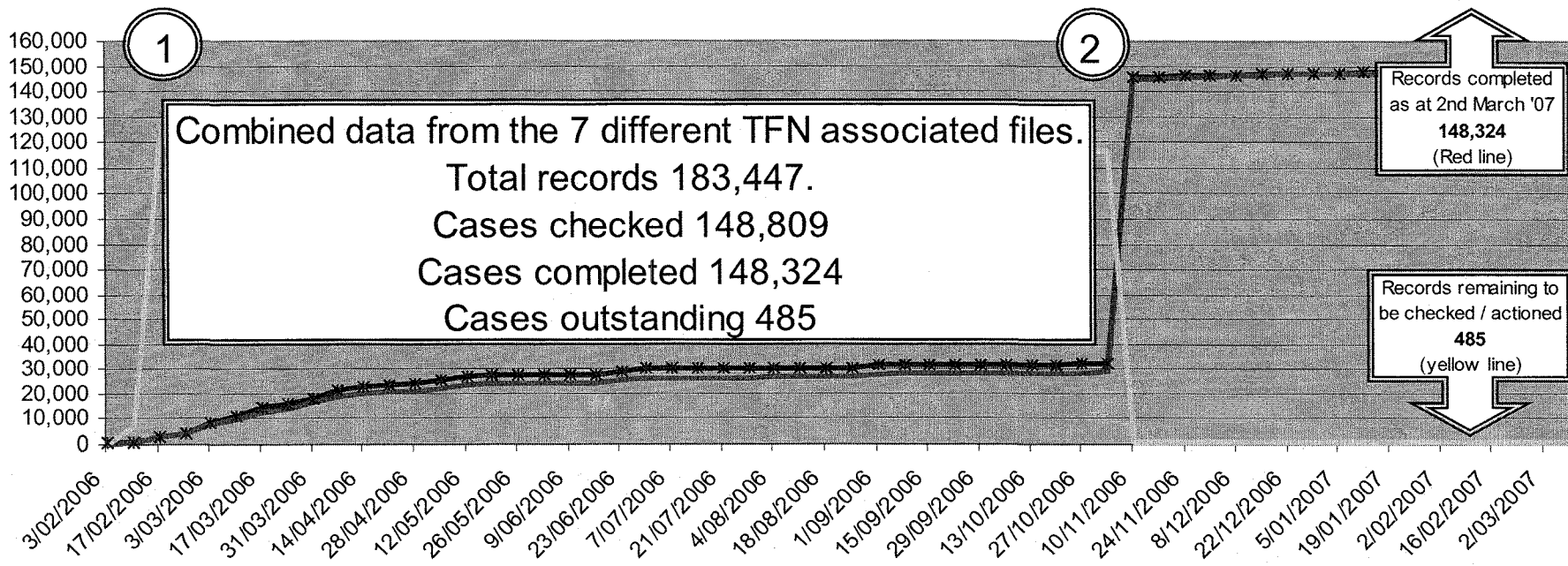Records completed as at 9/3/07 = 1,793. (displayed as red line)

Total number of CRNs in file 1,794

Records checked as at 9/3/07 = 1,794. (displayed as blue column graph)

Records outstanding at 9/3/07 = 1 (displayed as yellow line)
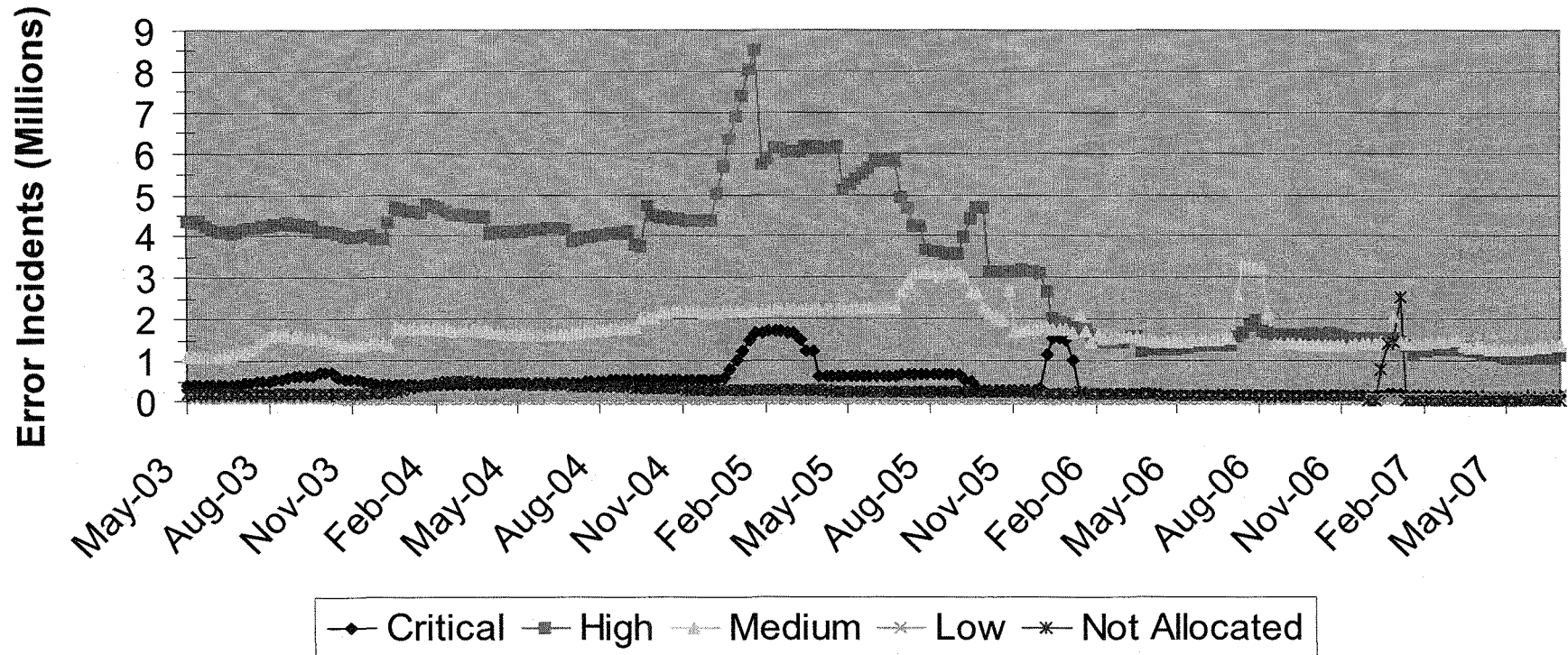
Combined Total of TFN Files

—※— Checked ——— Completed ——— Outstanding

Records checked
as at 2nd March '07
148,809
(blue line)

Records completed
as at 2nd March '07
148,324
(Red line)

Records remaining to
be checked / actioned
485
(yellow line)

Combined data from the 7 different TFN associated files.
Total records 183,447.
Cases checked 148,809
Cases completed 148,324
Cases outstanding 485

160,000
150,000
140,000
130,000
120,000
110,000
100,000
90,000
80,000
70,000
60,000
50,000
40,000
30,000
20,000
10,000
0

3/02/2006 17/02/2006 3/03/2006 17/03/2006 31/03/2006 14/04/2006 28/04/2006 12/05/2006 26/05/2006 9/06/2006 23/06/2006 7/07/2006 21/07/2006 4/08/2006 18/08/2006 1/09/2006 15/09/2006 29/09/2006 13/10/2006 27/10/2006 10/11/2006 24/11/2006 8/12/2006 22/12/2006 5/01/2007 19/01/2007 2/02/2007 16/02/2007 2/03/2007

(1) 17th February 2006 Commenced AGE/CUR no TFN file containing 116,133 CRNs
(2) 10th November 2006 116,133 cases in AGE/CUR no TFN file finalised in bulk as approriate TFN exemption codes.

Data Integrity Statistics Since 2003

Error Incidents (Millions)

Critical —■— High ╌╌ Medium ╌✕╌ Low ╌✱╌ Not Allocated

Data Integrity Index Since 2003

Data Integrity Index

16% 14% 12% 10% 8% 6% 4% 2% 0%

May-03 Aug-03 Nov-03 Feb-04 May-04 Aug-04 Nov-04 Feb-05 May-05 Aug-05 Nov-05 Feb-06 May-06 Aug-06 Nov-06 Feb-07 May-07

DI Index