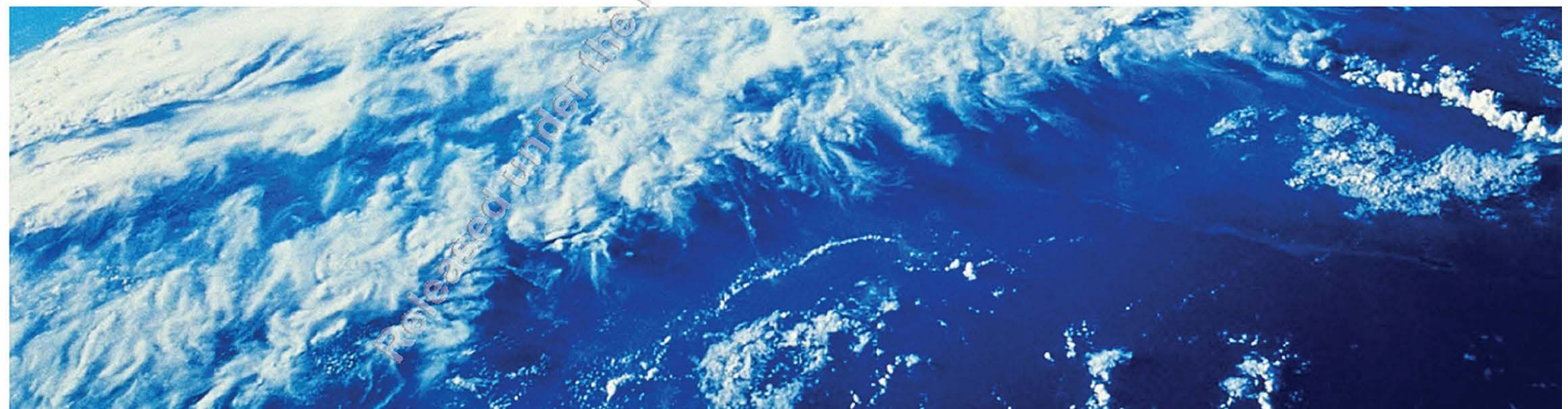




# Scanning Quality Assurance Service – Detailed Report

19 June 2016



# Contents

---

- Background
- Overview of Scanning QA Solution
- QA results summary
- Opinion in relation to the quality of the Senate Scanning Solution
- Recommendations to improve the Senate Scanning Solution
- Appendix
  - Detailed Analysis of exception reports from QA testing
  - QA Test Observations

Released under the Freedom of Information Act 1982

## Background

- Recent changes to the Electoral Act will request voters to record at least six preferences above the line instead of the single preference required under current legislation. AEC will no longer be able to conduct a significant portion of the count manually; therefore preferences from 100% of Ballot Papers must be captured in the AEC computerised scrutiny system. To handle the dramatic workload increase AEC invested in a document scanning ICR/OCR/OMR solution that will significantly automate the process of capturing voter preferences in the computerised scrutiny system. User Acceptance Testing (UAT) of the scanning solution was scheduled for May & June 2016
  
- AEC required an independent 'Quality Assurance' (QA) service to cross-check the output of the AEC vote scanning solution during UAT. AEC engaged IBM to
  - a) Design and build a 'minimal viable product' scanning QA Solution/ Service, to support AEC UAT activities for 1 form type and up to 3,000 image & related XML 'result' files
  - b) Provide the scanning QA service for QA activities for a period of 3 weeks starting on or around 23 May 2016
  - c) Conduct QA tests on supplied images and report all discrepancies between the results of the AEC solution and the those from the QA Service, via batch exception report(s)
  - d) Deliver a detailed report in respect of paragraph (c) (above)

This document is the detailed report from the engagement.

## Overview of scanning QA Solution and how it was used

1. AEC provided IBM with batches of images, together with the XML 'results' file for each batch, on the GovDex portal. IBM manually downloads these.
2. The QA Solution
  - a) Implements a set of ICR/OCR/OMR rules specified by AEC using IBM's IFP Product Set and runs on a secure laptop
  - b) Reads batches of scanned images provided by AEC at (1)
  - c) Outputs a set results (eg. ATL, BTL, formal/ informal) for each ballot in the batch
  - d) Compares the output of the IFP-based solution with the XML 'results' from the AEC solution and produces a 'batch exception report' for each batch
  - e) The batch exception report lists mismatches between AEC solution results and QA solution results
  - f) Each exception is checked by a human by viewing the relevant ballot image and making a determination whether the exception is due to an issue with the AEC system or with the IBM QA solution
3. IBM sends batch exception reports and the related analysis to AEC via email. Any obvious issues are noted to AEC as soon as possible

Released under the Freedom of Information Act 1982

| Batches (most recent first) | # of ballots | # and % exceptions | Summary Analysis (Detail later in report) |
|-----------------------------|--------------|--------------------|---|
| 21                          | 2,449        | 75 ballots<br>~ 3% | 21  |
| 21                          | 2,008        | 62 ballots<br>~ 3% | 21  |
| 21                          | 1677         | 118<br>~7%         | 21  |

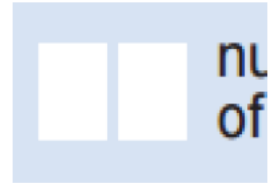
Released under the Freedom of Information Act 1982

- Based on analysis of the supplied ballot images and the AEC XML 'results' our opinion is that the quality of the AEC solution is high
  - Error rate appears to be < 0.5 % (i.e. cases where the AEC result appeared to be incorrect when image viewed manually by IBM)
  - Vast majority of exceptions were due to
    - Slight differences in interpretation between AEC operators and IBM 'operator' for forms requiring viewing
    - Issues in IBM OCR (challenges with numbers that bleed outside boxes) rather than the AEC solution

Released under the Freedom of Information Act 1982

## Recommendations

1. Ensure that *'missing initials'* and *'missing authenticity marks'* are detected by AEC solution
2. Ensure that operators are sufficiently trained and that any signs of operator fatigue are picked up quickly.
3. Review ballot paper format to optimize for OCR
  - Future form design could incorporate preference box outlines that are invisible to the scanner (dropout colour, or through wash/screening of box outline) to ameliorate this issue



Released under the Freedom of Information Act 1982

Released under the Freedom of Information Act 1982



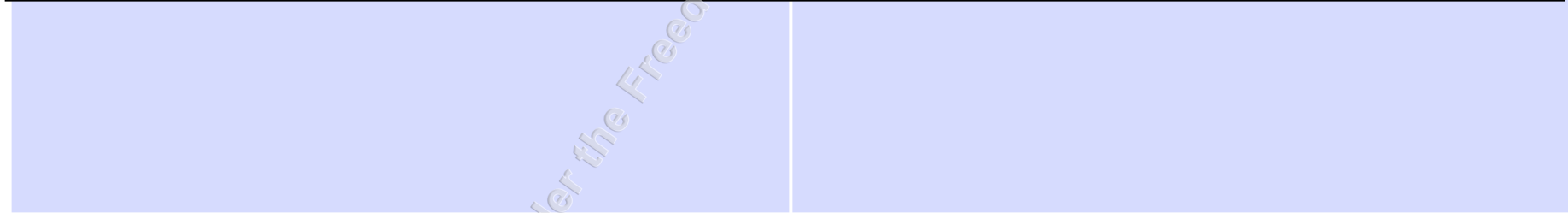
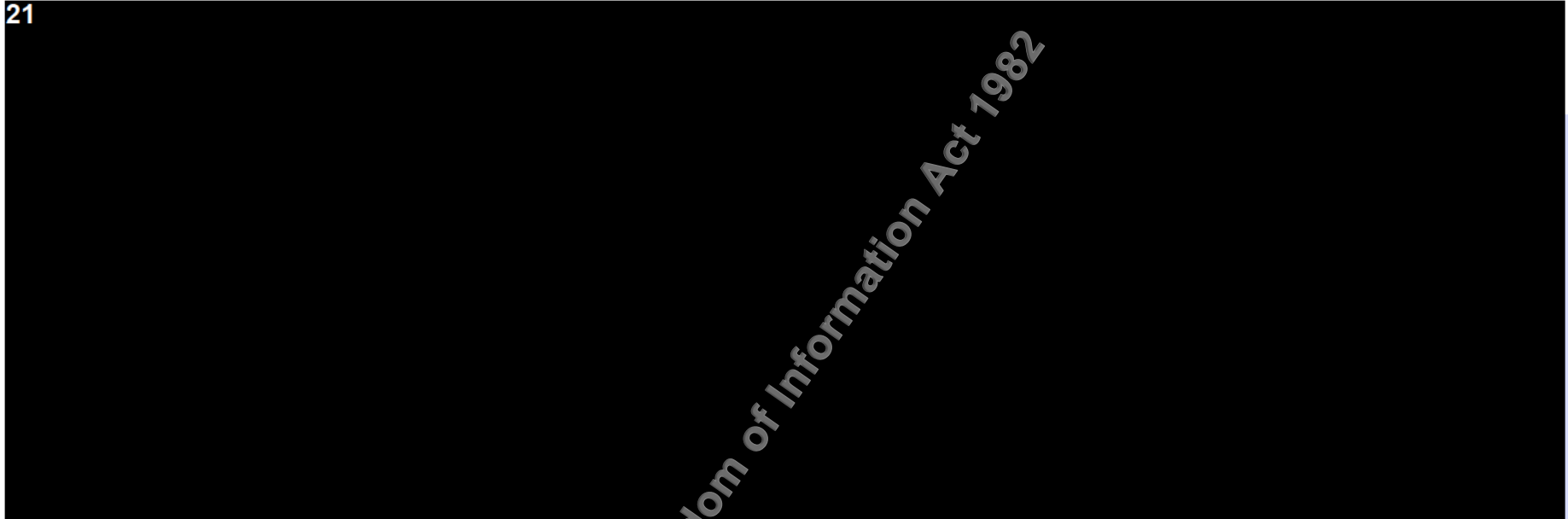
21

Released under the Freedom of Information Act 1982

21

Released under the Freedom of Information Act 1982

21



*Released under the Freedom of Information Act 1982*

21

Released under the Freedom of Information Act 1982

21

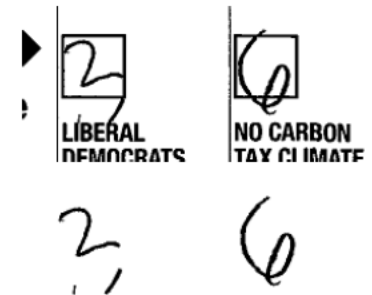
Released under the Freedom of Information Act 1982

21

Released under the Freedom of Information Act 1982

## QA Test Observations

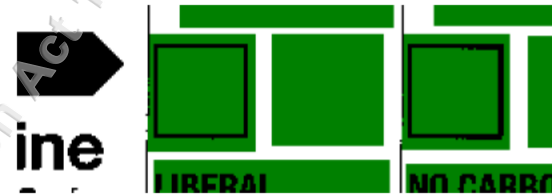
- Current ballot paper not specifically designed for OCR
- Large size unavoidable, but other than issues related to physical size, images processed well in most cases despite some small misalignment with the scanner transport in some batches
- The main issue impacting OCR effectiveness appeared to be the small size of the area for the voter to enter a preference, resulting in characters partly outside the box:
  - OCR is 2 stage process:
    - Segment read area into characters
    - Recognise the character(s)
  - While software will attempt to fill line gaps resulting from removing boxes prior to OCR, the process can still result in segmentation errors as well as recognition errors due to obliteration of parts of characters – e.g. 3 may be recognised as 2:



Released under the Freedom of Information Act 1982

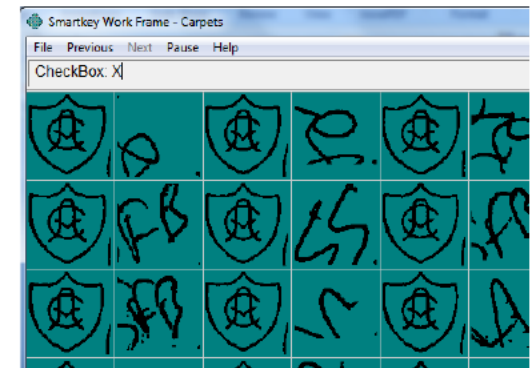
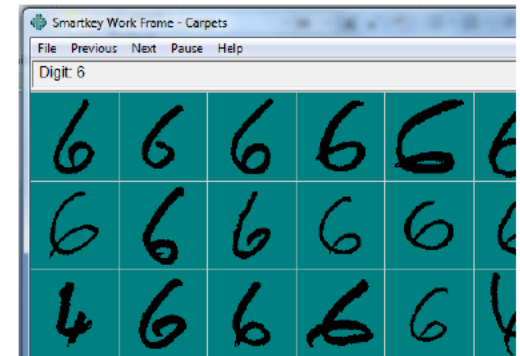
# QA Test Observations . . .

- Initial test defined capture area totally within preference box
  - Poor OCR results for common occurrence of character partially outside box
  - Capture area increased for subsequent tests
    - Much improved results but still further expansion warranted



- “Carpets” & “Triplets” used to confirm/reject preferences
  - Good solution where no other validation mechanism available
  - Validation of preference sequences within OCR/Repair system much better approach where possible:
    - No need to check unsure characters in valid (formal) preference sequence
    - More accurate. and less operator time

(.. But not possible for IBM to do in the time available for the QA Test ..)



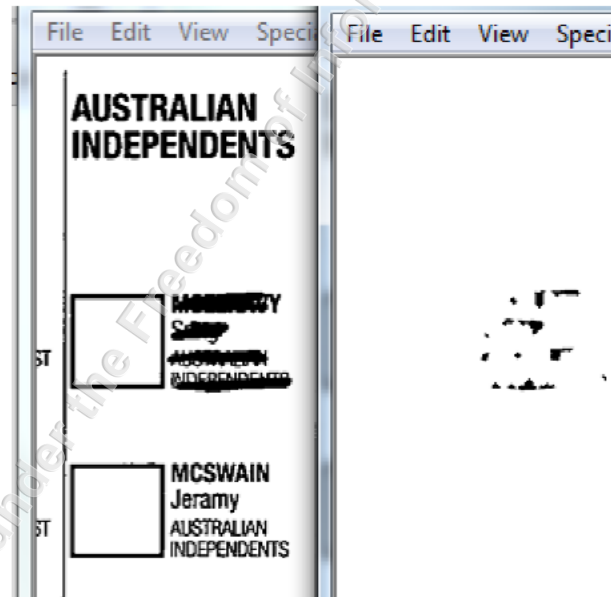
- . . . but useful for checking Official Mark and Authenticity Initials

Released under the Freedom of Information Act 1982



## QA Test Observations . . .

- IFP has option to remove form template prior to OCR
- Facilitates detection of Unusual Marks by doing OCR or mark detection on areas of dropped out image



Released under the Freedom of Information Act 1982

# End of Report

Released under the Freedom of Information Act 1982

Server 34167 Page 22 of 25  
New South Wales - Election of 12 Senators

| Party Name   | Mark                     | Party Name           | Mark                     | Party Name                                | Mark                     | Party Name                         | Mark                     | Party Name   | Mark                     | Party Name                          | Mark                     | Party Name                              | Mark                     | Party Name    | Mark                     | Party Name    | Mark                     | Party Name    | Mark                     | Party Name    | Mark                     | Party Name    | Mark                     | Party Name    | Mark                     | Party Name    | Mark                     | Party Name    | Mark                     | Party Name    | Mark                     | Party Name    | Mark                     | Party Name    | Mark                     | Party Name    | Mark                     | Party Name    | Mark                     | Party Name    | Mark                     | Party Name    | Mark                     | Party Name    | Mark                     | Party Name    | Mark                     | Party Name    | Mark                     | Party Name    | Mark                     | Party Name    | Mark                     |               |                          |
|--------------|--------------------------|----------------------|--------------------------|---|--------------------------|------------------------------------|--------------------------|--------------|--------------------------|-------------------------------------|--------------------------|---|--------------------------|---------------|--------------------------|---------------|--------------------------|---------------|--------------------------|---------------|--------------------------|---------------|--------------------------|---------------|--------------------------|---------------|--------------------------|---------------|--------------------------|---------------|--------------------------|---------------|--------------------------|---------------|--------------------------|---------------|--------------------------|---------------|--------------------------|---------------|--------------------------|---------------|--------------------------|---------------|--------------------------|---------------|--------------------------|---------------|--------------------------|---------------|--------------------------|---------------|--------------------------|---------------|--------------------------|
| ALBION PARTY | <input type="checkbox"/> | ANTHONY LEITCH PARTY | <input type="checkbox"/> | ASSOCIATED AMATEUR FOOTBALL LEAGUES PARTY | <input type="checkbox"/> | ASSOCIATED MUSICAL SOCIETIES PARTY | <input type="checkbox"/> | AUTISM PARTY | <input type="checkbox"/> | AWAY AMATEUR FOOTBALL LEAGUES PARTY | <input type="checkbox"/> | AWAY ASSOCIATED MUSICAL SOCIETIES PARTY | <input type="checkbox"/> | BALANCE PARTY | <input type="checkbox"/> | BALANCE PARTY | <input type="checkbox"/> | BALANCE PARTY | <input type="checkbox"/> | BALANCE PARTY | <input type="checkbox"/> | BALANCE PARTY | <input type="checkbox"/> | BALANCE PARTY | <input type="checkbox"/> | BALANCE PARTY | <input type="checkbox"/> | BALANCE PARTY | <input type="checkbox"/> | BALANCE PARTY | <input type="checkbox"/> | BALANCE PARTY | <input type="checkbox"/> | BALANCE PARTY | <input type="checkbox"/> | BALANCE PARTY | <input type="checkbox"/> | BALANCE PARTY | <input type="checkbox"/> | BALANCE PARTY | <input type="checkbox"/> | BALANCE PARTY | <input type="checkbox"/> | BALANCE PARTY | <input type="checkbox"/> | BALANCE PARTY | <input type="checkbox"/> | BALANCE PARTY | <input type="checkbox"/> | BALANCE PARTY | <input type="checkbox"/> | BALANCE PARTY | <input type="checkbox"/> | BALANCE PARTY | <input type="checkbox"/> |

## **REDACTION CODES**

- 1 Personal Information (name) redacted.
- 2 Personal Information (date of birth) redacted.
- 3 Personal Information (photograph) redacted
- 4 Personal Information (facsimile of signature) redacted.
- 5 Personal Information (facsimile of manuscript initialling) redacted.
- 6 Personal Information (Individual's address) redacted.
- 7 Personal Information (individual's telephone number) redacted.
- 8 Personal Information (individual's opinion) redacted.
- 9 Personal Information (opinion about individual) redacted.
- 10 Personal Information (employment history) redacted.
- 11 Personal Information (qualifications) redacted.
- 12 Personal Information (health) redacted.
- 13 Business information (Bank Account details) redacted.
- 14 Business information (Billing Account details) redacted.
- 15 Legal Professional Communication redacted.
- 16 Deliberative material redacted.
- 17 Irrelevant material redacted.
- 18 Electoral Roll material redacted.
- 19 Tests, examinations or audits material redacted.
- 20 Management or assessment of personnel material redacted.
- 21 Proper and efficient conduct of the operations of AEC material redacted.