



INFORMATION MATCHING BULLETIN

March 2006 Information Matching Newsletter from the Privacy Commissioner

**Don't take action because of a name!
A name is an uncertain thing, you
can't count on it! (Bertold Brecht, *A
Man's a Man*, 1931)**

We're all aware of the many possibilities for mis-spelling our own names. This poses serious difficulties for information matching based on people's names. To see how significant these challenges are, try searching "data matching" in Google. You'll get back many advertisements and entries for companies all saying why they can do the job better than their competition.

Michael Dunkerley wrote an article on Advisor.com¹ highlighting the problem of the non-standard nature of names and their spellings and the challenge of information matching with large databases. We recommend that anyone interested in the subject read the whole article.

When a person's name is entered into a computer system there is a degree of error and variation. Dunkerley says that because there are no rules for names, they cannot be relied on to verify identity. He goes on to say that variations that occur in names are

significant. They include spelling, nicknames, anglicisation, initials, truncation, abbreviation, missing names, extra names and aliases. Dunkerley uses the name Johnstone as an example. Other spellings of the same name include Johnson, Johnsen, Johnsson, Johnston and Jonson.

Dunkerley quotes an interesting statistic, that although there are over 2.5 million first names in use in the United States, as many as 80% of the population may have names from a selection as few as five hundred.

Because of this, it is difficult to maintain a high standard of data quality when it comes to name information. Even with controls in place to limit errors, users can be unaware they are entering incorrect data. Data quality is even more problematic when accepting information from individuals via a web based interface. There is often a reduced level of control over data quality.

Another complication is individuals who purposely change their details in some way to reduce the chances of being matched while still being able to claim a benefit or use a service.

Recognising these problems with names, most of our authorised

¹ Address Critical Searching and Matching Problems in Large Databases <http://www.advisor.com/doc/12857>

information matches use a combination of name, address, date of birth and ID numbers. Some assign a confidence rating or match quality level for the matched records. The Ministry of Education use match quality levels when matching DIA's birth information against their National Student Index (NSI).

Dunkerley talks about the trade off between accuracy and reliability. Some people may know of these as precision and recall. An algorithm that only allows for exact matches may provide excellent accuracy but may miss a number of valid matches. On the other hand, to ensure you get all possible matches, you'll have to sort through lots of false-positives.

The advent of matching traveller names to terrorist watch-lists further highlights the difficulties with matching large volumes of names. Tens of thousands of travellers have been improperly matched. It doesn't matter who you are either. The *New York Times* reported in August 2004 that a watch-list delayed Senator Edward Kennedy at an airport.

Dunkerley's advice is also supported and reinforced in a report produced by the US GAO². They reported last year on difficulties in maintaining the accuracy of eligible voter registrations. State officials were unable to identify which of several voters of the same name had died because of the lack of a birth date on some state death records.

It's too bad we can't do all data matching using people. As Dunkerley points out, the best data matchers are

humans. They are able to use all the available data and intuitively draw a conclusion. Humans are able to compensate for error and variation in a manner that is difficult for a computer programme to replicate.

The trend towards larger databases and online service delivery increases the reliance government and business has on data matching software to correctly identify individuals.

Verifying match results and giving an individual the opportunity to challenge the results of a match before taking adverse action against them are important safeguards built into the Privacy Act.

Inaugural meeting of the Information Matching Interest Group

The first meeting of the IMIG was held on 16 August 2005. About 40 people attended, covering most of the Wellington based agencies involved in information matching.

Geoff Oldham from IRD gave a presentation on the constraints on use of IRD information generally and how information matching has been justified in some cases. He also talked about and diagrammed the typical development of an information matching programme from bright idea to post-implementation review and solicited suggestions from the audience for improvements to the diagram.

Colin Trotter and Lindy Siegert (OPC) gave a presentation on "On-line transfer approvals: what we know now." The audience was encouraged to submit ideas for further meetings such as similar information sessions, field trips, larger meetings including policy, legal and IT staff. Lindy announced

² United States Government Accountability Office, "Elections – Views of selected local election officials on managing voter registration and ensuring eligible citizens can vote", September 2005.

that the OPC was developing a training workshop on information matching and hoped to be able to deliver the first one by the end of the 2005/06 year. She said they expected it to be a half-day workshop and the presentation on on-line transfers were developed as one module for that workshop.

**Privacy Impact Assessment Seminar,
Wednesday 29 March 2006 at BP
House, Wellington**

Blair Stewart, Assistant Commissioner and Bob Stevens, Barrister, will present a seminar on the Privacy Impact Assessment (PIA) process. A CD-Rom of the *Privacy Impact Assessment Handbook* will be provided as part of the seminar fee.

Please note that this is a general introduction to the process and does not cover the specialised area of Information Matching PIAs. For more information and registrations, please visit our website or contact this Office.

**Privacy Issues Forum, Thursday 30
March 2006 at Te Papa, Wellington**

This Office is hosting the above forum. Guest speakers include Australian Privacy Commissioners along with local and international guests with an interest in privacy. Topics include "Identity and Privacy – checking identity" and "Technology developments" For more information and registrations, please visit our website, contact this Office or email to Forum2006@privacy.org.nz

Contacts:

Wellington

109-111 Featherston Street,
gen-i Tower, 4th Floor
PO Box 10-094,
Wellington, New Zealand
Telephone 64-4-474 7590
Facsimile 64-4-474 7595

Internet

<http://www.privacy.org.nz>

Lindy Siegert

Team Leader (Technology)

E-mail: lindy.siegert@privacy.org.nz

Direct Dial: 64-4-494 7082

Colin Trotter

Data Matching Compliance Adviser

E-mail: colin.trotter@privacy.org.nz

Direct Dial: 64-4-494 7087

Jim Whitman

Policy Adviser (Technology)

E-mail: jim.whitman@privacy.org.nz

Direct Dial: 64-4-494 7083

Auckland

Level 8, Gosling Chapman Tower,
51-53 Shortland Street
P O Box 466, Auckland, New Zealand
Telephone 64-9-302 8680
Facsimile: 64-9-302 2305

Blair Stewart

Assistant Commissioner

E-mail: Blair.Stewart@privacy.org.nz