Privacy Commissioner
Te Mana Matapono Matatapu

# Information Matching Bulletin

News from the Office of the Privacy Commissioner – May 2008

## In this edition



## Team Leader (Technology) appointed

Rosie Byford has recently joined the Office of the Privacy Commissioner as the new Team Leader (Technology). Rosie joined the Office from the Ministry of Research, Science and Technology, and also brings with her UK policy experience (at the Department of Trade and Industry) and a background in manufacturing engineering. Rosie will lead the team responsible for monitoring information matching and policy related to both information matching and the intersect between technology and privacy.

## Encryption required for physical transfers using digital media

**Neil Sanson**

The loss of 25 million personal records being transferred on un-encrypted CDs in the UK prompted us to check with agencies as to how they handle transfer of physical digital media used in authorised information matching programmes.

I began by looking at the technical standards reports as these should have all the relevant details. Then I checked with agencies to clarify or confirm. I hoped that this might prompt agencies to review their security, and one agency did initiate a review.

Delivery methods were listed as by hand or by courier using track-and-trace. These approaches are both useful and make a loss less likely, but in the event of a disc or tape being lost, encryption is a useful extra protection. However, my research suggested that most data was only password-protected.

As departments will be aware, the Privacy Commissioner has announced that she will in future require encryption for data being transferred on digital physical media. GCSB have advised encryption should preferably be at least AES 256 or an equivalent approved algorithm.

I have already heard from a couple of agencies about their plans to achieve this standard, and expect to hear from the other agencies in the near future.

Other methods of transfer such as fax, paper documents and on-line transfers have different characteristics that require different security approaches. These may be reviewed over 2008/09.

Related links:
A collection of stories about the HMRC loss:
www.theregister.co.uk/2007/11/22/hmrc_roundup/
And "recorded delivery" does not guarantee against loss:
www.theregister.co.uk/2008/01/23/court_info_sent_in_post/
Privacy Commissioner requires data encryption
www.privacy.org.nz/privacy-commissioner-requires-data-encryption/

## Information Matching Audit page added to the Privacy website

An information matching audit page has been added to the data matching section of the Privacy website. The page includes a link to the "Information Matching Compliance Auditing Information Pack", www.privacy.org.nz/information-matching-audit/

To view the data matching section, go to: www.privacy.org.nz/data-matching/.

## Privacy breach guidelines

In February, the Privacy Commissioner issued voluntary privacy breach guidelines following several months of consultation with New Zealand organisations.

The guidelines and supporting information paper provide an overview of the privacy breach process and detail key steps for agencies in responding to privacy breaches.

The guidelines can be accessed from: www.privacy.org.nz/privacy-breach-guidelines-2/

## Information matching workshops

The fourth workshop, *The Privacy Act and developing an information matching programme*, held on 7 March, was attended by 18 people from a wide selection of agencies.

The half day workshops are designed to give some practical background knowledge about the Privacy Act along with more detailed information about preparing an Information Matching Privacy Impact Assessment. To register interest in attending the next workshop, contact Sharon Newton on (04) 4747590 or by email to sharon.newton@privacy.org.nz.

## Information Matching Interest Group

The fourth Information Matching Interest Group meeting, held on 20 February, was attended by 14 people representing six agencies. Topics covered in the meeting included the information matching audit pack, encryption of physical data transfers, the information matching shared workspace, followed by an open forum session where agency representatives could give an update on current information matching work projects.

## Technology and Privacy Forums

Details of the next Technology and Privacy Forum are posted at:
www.privacy.org.nz/training-and-education/technology-and-privacy-forums

## Privacy Issues Forum 27 August 2008

Keep your diaries free for Wednesday, 27 August 2008. The Office of the Privacy Commissioner will be hosting a day-long privacy forum, to be held at the Intercontinental Hotel, Wellington. Programme and registration details will be available soon at www.privacy.org.nz. This event is part of Privacy Awareness Week (24-30 August 2008). Departments may like to promote Privacy Awareness Week in their own offices. Please contact us if you have ideas that we could help with.

## Data quality accuracy dimension – part 1 of 2

**Colin Trotter**

My Google alert recently brought up a link to an information matching related blog. Accessible through http://dataqualityaccuracy.blogspot.com/, this is one of 56 technology focused blogs written under the profile of "vijikumar".

The "Data Quality Accuracy Dimension" blog comprises four connected articles under the following headings:

The Data Quality Problem
Definition of Accurate Data
Sources of Inaccurate Data
Implementing a Data Quality Assurance Program.

Much of what vijikumar talked about struck a chord, so much so, that I thought it worthwhile presenting some of those ideas in a short article.

### *The Data Quality Problem*

In *The Data Quality Problem*, Vijikumar writes that data is a precious resource, used as fuel to drive processes of all sorts. Many large organisations, like banks, operate largely as a repository of information and most of what they do is process data. Without their information systems, these organisations could not exist.

Even where organisations are involved in activities that do not appear to be information specific, looking more closely will likely reveal that their decisions and activities are driven by information systems. As Vijikumar says, data is becoming more precious all the time, as it is being used to help organisations make important decisions. New trends in data warehousing, data mining, decision support, and customer management relationship systems all point towards the expanding role data plays in business today.

Most organisations operate a number of databases for different purposes like human resources, creditors, debtors, and suppliers. Whenever someone wants to know something, they use their PC to query a database. As new purposes for data are thought of, there is a tendency to create a duplicate of the primary (or source) data to satisfy the new need.

Not only is most information now in databases, but these databases have been replicated into data warehouses or the like. Vijikumar states that replication often includes aggregating data, combining data from multiple sources, putting data into data structures different from the original, and adding time period information. Often, the original data may not be

recognisable in the aggregations, and as a result errors detected in the aggregations cannot be traced back to the primary instances of data containing the errors.

Data quality problems can be magnified through these replication processes. In the primary or source system, a wrong value may have a small impact. However, if a wrong value is propagated to a higher level decision support system, it may trigger an action that has much greater adverse consequences. Propagating poor quality data from one agency to another is one of the risks associated with data matching. The quality of the data collection process may have been suitable for the original purpose; however relying on that information to make decisions in more critical or sensitive areas may be problematic.

Databases have risen to become one of the most important corporate assets. Despite this, Vijikumar believes organisations tolerate significant inaccuracies in their databases and that data quality is not managed as rigorously as most other areas of an organisation's business. Few organisations have a data quality assurance programme. Vijikumar believes that performing a data quality assessment exercise almost always raises awareness about data quality issues, with a typical response being, "I had no idea the problem was that large".

In Vijikumar's view, the fact that data quality is universally poor indicates that it is the natural result of the evolution of information systems technology. Two major contributing factors are cited by Vijikumar. The first is rapid systems development and change that have made it difficult to control quality. The second is that the standards, techniques, methods, and tools for managing quality have evolved at a slower pace than the systems they serve. Vijikumar believes that the best way to improve the overall data quality of your information systems is to make it a primary requirement for all new projects.

### *Definition of Accurate Data*

In *Definition of Accurate Data*, Vijikumar puts forward that data has quality if it satisfies the requirements of its intended use. Further, that to satisfy the intended use, the data must be accurate, relevant, timely, complete, understood, and trusted.

Imagine a database containing contact details for all GPs in New Zealand (Vijikumar presents a similar example in his article). The database is missing some newly registered GPs, address information for others is incomplete, and some records are obsolete. Overall, it is expected to be 85% accurate.

For the purpose of finding potential customers for a medical supplies manufacturer, this database might be seen to be high quality, offering them an excellent database to market their products. Conversely, this database would be considered poor quality if the database were to be used to notify GPs to immediately stop prescribing a drug found to be dangerous. In fact, it might be negligent to use it for that purpose.

The disclosure of data from one agency to another (possibly in an information matching programme) where the receiving agency is intending to rely on the data for a different purpose is another example where data quality and its suitability for the intended (new) purpose comes into play. When a new use appears, an assessment needs to be completed to determine if the database meets the required quality. If the database is not up to standard, then either the use needs to be discarded or modified, or the database and its data-generating applications need to be upgraded.

Vijikumar writes that data accuracy is one of the components of data quality. It refers to whether data values stored in a record are the correct values. To be correct, a data value must be the right value and must be represented in a consistent and unambiguous form.

Two values can be both correct and unambiguous but still cause a problem. Mt Albert and Mount Albert may both refer to a suburb in Auckland; however the recordings are inconsistent. The problem is that inconsistent values cannot be as accurately aggregated and compared as consistent values, opening up the opportunity for inaccurate use of data.

The recording of dates is a good example of an area where ambiguity can be a problem. For example, the births date of June 5, 1966. A USA representation of this date is 06/05/1966 whereas the European representation is 05/06/1966. A value is not accurate if the user of the value cannot tell what it is.

Vijikumar says that the tendency for data fields that are more important to be more accurate is why quality problems occur less frequently when using data for its primary use. However when data is moved and used for other purposes, data recorded in (originally) less important fields may now become much more important when considering the new use.

Vijikumar points to two methods to determining the accuracy of data: re-verification and data analysis. Re-verification is likely to be too expensive and time consuming and analytical techniques require you to have a good understanding of what is correct in your database. For those interested, Vijikumar goes into further detail about the different analytical techniques in his blog *Definition of Accurate Data*.

**Part 2 will be included in the June 2008 Information Matching Bulletin. The full article is available on the Information Matching Shared Workspace.**

## Publications

There are a number of other publications and reports available from the Privacy Commissioner that may be of interest to those involved in information matching. These are listed on the Privacy Commissioner's website, www.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
www.privacy.org.nz

**Auckland**
Level 13, WHK Gosling Chapman Tower
51-53 Shortland Street
P O Box 466
Auckland, New Zealand
Telephone: 64-9-302 8680

**Neil Sanson**
Data Matching Compliance Adviser
Direct Line: 64-4-474 7592

**Blair Stewart**
Assistant Commissioner

**Colin Trotter**
Senior Adviser, Data Matching Compliance
Direct Line: 64-4-494 7087

**Rosie Byford**
Team Leader, Technology
Direct Line: 64-4-494 7082

You can contact us by email. Our standard email format is first name.surname@privacy.org.nz