

Data matching

Prof. Simon Rogerson

**Originally published as ETHicol in the IMIS Journal Volume 7
No 1 (February 1997)**

Whilst fraud is accepted by society as being wrong and should be prevented, the methods and in particular data matching deployed to prevent or detect fraud are open to criticisms regarding the invasion of privacy and the restricting freedom of the individual. There is a difficult balance to be struck between the rights of individuals and the devastating financial losses particularly of public funds when considering anti-fraud measures. It is unclear who should judge whether the cause legitimises the data matching and data sharing activity and how such judgement should be derived.

Data matching is the computerised comparison of two or more sets of records which relate to the same individual. It is primarily used as a method for combating fraud. There is increasing use of data matching by both public and private organisations in an attempt to reduce fraudulent activity which has been estimated to run annually into billions of pounds in the UK. The relative cheapness and availability of sophisticated processing means that data matching is likely to increase even more rapidly.

There are a number of examples of data matching being undertaken by government agencies. The DSS has established a Housing Benefit Matching Service aimed at detecting benefit fraud. The Audit Commission uses data matching across local authorities regarding benefit claims, education awards and activities of local authority employees. The Social Security Administration (Fraud) Bill provides for wider sharing by central and local government and the Post Office for fraud prevention or detection purposes.

There is a difference between the methods of fraud detection used in the past and data matching. Traditional investigation is triggered by some evidence of a wrong doing by an individual, such as tax evasion or bogus benefit claims. Data matching is not targeted at individuals but at entire categories of people. It is initiated not by the suspicion concerning an individual but because the profile of a particular group is of interest. This leads to three issues of concern.

- *Privacy* - Data matching is likely to involve matching personal records compiled for unrelated purposes. Surely an individual has a right to control personal information and prevent its use without consent for purposes unrelated to those for which it was collected.

- *Due process of the law* - Once a match has been undertaken it will result in a number of hits. All those identified are in jeopardy of being found guilty of a wrong doing. It is unlikely that these individuals are given any notice of their situation, since doing so might affect the investigation, or an opportunity to contest the results of the match at an early stage. For these reasons their right to due process of law is curtailed.
- *Presumption of innocence* - The presumption of innocence is intended to protect people against having to prove that they are free from guilt whenever they are investigated. Data matching can reverse this to a presumption of guilt. This is because the technology of data matching is so plausible and the detection of fraud is much applauded. These powerful influences will weigh heavy in favour of the notion that those identified must be guilty.

One of the advances in data matching techniques is the automatic sharing of information. This is dependent on increasingly complex and expanding communication networks that link more and more organisations together. There is a tendency to view people as objects to be moved about these networks. In this instance it is unclear how an innocent citizen might be assured that incorrect information is totally corrected or removed across the complex web of organisational relationships. The security of such a network is also of concern as the network will be as secure as its least secure node. Individual organisations might not be aware of all other organisations linked to the network for if they were they might be wary of sharing information due to insecurity through either technological deficiencies or moral attitudes of some organisations.

In order to detect sophisticated fraud there is need to use complex data analysis techniques which may well involve methods based on partial match interpretation which in turn increases the risks of incorrect hits. Simple fraud detection lends itself to data matching systems that have little or no human intervention and the pressure to use such systems will grow. There needs to be a clear understanding of who has access to this information and how it will be used. Legislative frameworks, codes of practice and operational procedures must be in place to ensure that data matching is applied in a balanced way sensitive to the needs of organisations and society as a whole and to the rights of individuals.

Each of us is probably the subject of some data matching exercise quite frequently. It might be related to a mortgage application or a latest tax return or many other reasons. How do you feel about that? How would you feel if you were on the receiving end of an erroneous data match which led to a mortgage foreclosure or the loss of your credit rating? Next time you are involved in implementing a data matching system think on it could be you incorrectly caught by it!

Please send your views on ethical and social responsibility issues and cases of ethical dilemmas to:

Professor Simon Rogerson
Director
Centre for Computing and Social Responsibility
Faculty of Computing Sciences and Engineering
De Montfort University
The Gateway
Leicester
LE1 9BH
Tel:(+44) 116 257 7475
Fax:(+44) 116 207 8159
Email:<srog@dmu.ac.uk>
Home Page: (<http://www.ccsr.cse.dmu.ac.uk>)