FRAUD AND COMPUTER DATA MATCHING

- *From targeting direct mail to fighting fraud*
- *Privacy aspects*

Technical advances in computers and telecommunications allow targeting in marketing and can also help combat fraud. This raises questions of compatibility with current data protection legislation which have resonated in Parliament.

This note analyses these trends, including data matching, and the policy issues raised.

POST note 93 February 1997

BACKGROUND

Processing most transactions nowadays involves storing some personal records on a computer (details of driving licences, social security payments, etc.). Many of these computer systems are 'stand alone', for example the police keep records for their own purposes, the Inland Revenue (IR) for theirs, but in the private sector particularly, there is also a tradition of sharing computer files. Organisations exchange customer names and addresses, market research companies create profiles (e.g. from surveys) and targeted mailing lists, credit reference agencies compile records of credit behaviour and so on.

While the number and types of public and private data sets continue to grow, more powerful computers and software make it easier to collect and compare information from different sources. Many data sets are available (e.g. those in **Table 1**) and moves underway to encourage more open government will continue to expand them[1]. '**Data matching**' (**Box 1**) makes it possible to put together a composite picture of someone's likes, purchase habits, credit behaviour, etc., or to detect similarities and differences between data collected for different purposes - e.g. someone paying income tax and claiming social security at the same time, or a 'dead' person claiming benefits. Data matching can thus be used both as a commercial tool and as a weapon in the fight against fraud.

The use of personal computer records is regulated under the Data Protection Act (1984) which was drawn up nearly twenty years ago, when computers were relatively slow and isolated, and applications such as data matching were not envisaged. Under the Act, therefore, many data matching initiatives require specific legislation - e.g. to allow data sharing between the Department of Social Security (DSS) and IR to detect fraud. This, coupled with the privacy implications of widespread data matching, raises questions whether the Act needs updating. Since the European Directive on Data Protection must be adopted into law by October 1998, Parliament will have an opportunity to debate these issues over the next 18 months.

Table 1 POTENTIAL SOURCES OF DATA

Open source	*Confidential/commercial*
Electoral Roll	Financial records
Land Register	Application forms
National Census	Health records
Postcodes	Lifestyle surveys
County Court Judgements	Loyalty schemes
Mortality records	Police records
Company reports	National Insurance numbers
	Government records
	Mail redirection database (PO)

Box 1 DATA MATCHING

The term '**data matching**' covers two related but distinct functions of computerised databases. One is the comparison of a given individual's details (e.g. on an application form) with one or more databases (e.g. to check that the address supplied is valid). The other is to perform a 'side-by-side' comparison of two or more (large) databases, to detect trends, anomalies, potential duplicates, etc. In practice, data matching is achieved in three main ways - data sharing, data matching agents and data bureaux.

In **data sharing**, '**data users**' with their own independent databases exchange information directly with other data users for matching. This ranges from furnishing specific information on request, to exchanging subsets or entire databases. Where data users wish to restrict access (e.g. a life insurance company would not be given direct access to medical records), one option is to use a **data matching agent**. Here, the agent receives data files from two or more data users and stores the data only long enough to perform the data matching.

The third way is through **data bureaux**, which are centralised repositories of computer records providing a 'database service' to data users. Some data users rely entirely on data bureaux to store their files, accessing them on-line as needed, while others use data bureaux to supplement local storage and processing facilities. Maintaining a central database allows data bureaux to optimise the way information is stored for data matching, allowing very rapid processing - for instance, the results of an 'instant credit check' typically are available in a few seconds.

A growing range of techniques is available to compare and process information including:

- **neural networks** - 'learn' to detect and predict patterns in data (some banks use them for automatic credit scoring);
- **fuzzy logic** allows concepts like 'a lot' and 'a little' to be used instead of numerical values;
- **phonetic matching**, to check for similar sounding words;
- **'intelligent' systems**, which are programmed with specialist 'knowledge' (e.g. to check that a bank account or national insurance number is valid).

1. It is a primary focus of open government (e.g. the Department of Trade and Industry's Information Society Initiative and the Government Direct Green Paper) that information be made available electronically.

POTENTIAL APPLICATIONS

Marketing While some companies still engage in mass 'junk mail' campaigns, the trend is to target sales at likely customers via personalised messages and products. Data matching gives a more complete picture of customers; it reduces errors in data files (e.g. by matching against the Post Office list of valid addresses); customer databases can be updated automatically with data from other sources; changes in circumstances (e.g. moving house or having children) can be flagged; and customer services can be focused where they are most needed (e.g. a bank may be able to predict when a client needs extra help).

The most comprehensive information about individuals can be found in 'confidential' databases, such as those of banks, health records and government departments. The use of personal data, however, is restricted by the laws of confidentiality and by the Data Protection Act (**Box 2**). Therefore, marketing companies use information from other sources, such as 'lifestyle surveys - around 14M have been completed in the UK and data matching is used to build profiles of other people from open sources (e.g. Table 1).

Databases are also being used in political canvassing - e.g. by matching the electoral register against other data. For example, the Prime Minister has written to registered voters who own shares in privatised companies; in other ways, residents in certain areas may be targeted as possible 'floating voters'.

Data matching is also used to **detect fraud and reduce financial risk in the private sector**. **Credit reference agencies** maintain records of financial behaviour, and one step is to verify the applicant's name and address[2], e.g. from the electoral register. Data matching is also used in credit scoring, where banks share information on accounts, bank cards, credit limits, average balances, etc. Other common checks include matching against the list of County Court Judgements, e.g. for bankruptcy. The required consent for such checks is usually a condition of application.

There are also data **'registers',** such as:

- the **Comprehensive Underwriting Exchange** (**CUE**) to detect previous and multiple claims;
- the **'Possessions Register'** - where repossession details are filed centrally by mortgage lenders;
- the **Credit Industry Fraud Avoidance Scheme** (**CIFAS**), a database of known criminals;
- the **Gone Away Information Network** (GAIN) - debtors who have left no forwarding address, and of 'missing debtors' who have been located.

Box 2 DATA MATCHING AND THE DATA PROTECTION ACT

The Act requires data users to register each purpose for which they hold data, and sets out the 8 'data protection principles' in the Table. Failure to register is a criminal offence, while the Data Protection Registrar can serve legally enforcible notices for non-compliance with the 'principles'.

Some principles are clearly relevant to data sharing and matching, but one key interpretion of the Registrar is that data subjects must be clearly informed about, and consent to, the proposed use of personal records - including disclosure of information. In the private sector, consent usually is obtained through a negative response 'check box' (e.g. "*Tick this if you **do not wish** us to supply your information to other organisations*").

The Act allows some principles to be waived - where a disclosure is required by law, or where their application might prejudice the prevention or detection of crime, the apprehension or prosecution of offenders or the assessment or collection of any tax or duty. Exemptions have, however, been narrowly interpreted by the Registrar and the Data Protection Tribunal and can only be exercised on a case-by-case basis where there is a "real risk" of prejudice, or "strong evidence of the commission of offences".

Table DATA PROTECTION PRINCIPLES

Data shall be:

I. obtained and processed fairly and lawfully
II. held only for lawful purposes which are described in the register entry
III. used or disclosed only for those or compatible purposes
IV. adequate, relevant and not excessive in relation to the purpose for which they are held
V. accurate and , where necessary, kept up to date
VI. held no longer than is necessary for the purpose
VII able to allow individuals to access information held about them and where appropriate correct or erase it
VIII. surrounded by proper security

Other systems focus on **detecting** suspicious irregularities as well as more obvious fraud:

- '**Detect**', which operates alongside a commercial credit referencing system, and uses date of birth, employment type, income, accommodation, etc. to identify anomalies between different applications;
- '**National Hunter**', a specialised anti-fraud system which takes two or more computer records (in principle from any source) and compares them. It is used extensively by mortgage lenders (e.g. identifying multiple mortgage applications) and is being considered for the public sector.

There is a growing **public sector** interest in data matching. Most centres on fraud detection and prevention, and some applications under trial or active consideration are in **Table 2**. **Benefit fraud** takes many forms, but falls broadly into three categories:

- declaring incorrect details (e.g. income);
- deliberate individual fraud e.g. claiming housing benefit at multiple addresses, or education awards and housing benefits simultaneously;
- organised fraud where the benefit system is deliberately and systematically defrauded.

2. Following public complaints in the late 1980s, the Data Protection Registrar served a compliance notice on a credit reference agent forbidding the use of address alone as a measure of creditworthiness.

Many different benefits are vulnerable, including social security payments, housing benefit, educational awards, etc[3], and the Government estimates that two major areas, Income Support and Housing Benefit, lose ca. £1.5B and £1B annually (some 10% of total payments). Data matching can help focus attention on 'suspect' cases. Since April 1995, the Benefits Agency has cross-checked certain benefit payments[4] looking for inconsistencies, multiple claims against the same name or address, etc., leading to 50,000 referrals for further investigation and specific savings of £26M over the first 12 months.

Table 2 ACTUAL AND PROPOSED* PUBLIC SECTOR DATA MATCHING

Initiative	*Purpose*
London Fraud Initiative (33 LAs)	Crossmatch between LAs to detect benefit fraud
Local authorities outside London	
*National Fraud Initiative	Nation-wide implementation of LFI
*DSS National Housing Benefit Register	Pilot completed
DSS Generalised Matching Service	Between DSS agencies
*Social Security Administration (Fraud) Bill	Extend to e.g. Inland Revenue
UCAS	Identify fraudulent multiple applications
Student Loan Company	
*Council for Local Education Authorities	
*Health Service	Organised fraud, e.g.locum, prescription
*DoE, MOD and LAs	Organised contractor fraud

With Housing Benefit, the Social Security Committee noted 48 different types of fraud, in the categories:

- **Property based**, e.g. rent overstated, multiple claims or large scale, organised fraud;
- **Income based**, e.g. earning salary while on benefit;
- **Circumstance**, e.g. fictitious children;
- **Instrument of payment**, e.g. false encashment of cheques, forgery, false claims created by staff.

To counter housing benefit fraud, the **London Fraud Initiative** was established by the **Audit Commission** and the 33 local authorities (LAs). Under this, information is collected from the LAs annually by the Audit Commission for data matching (performed by a contractor). The data matching flags 'suspect' cases which are passed back to the LAs for further investigation. In 1995-96, the London Boroughs saved over £200M through fraud control measures, of which some £3M was detected by data matching of housing benefit. The **National Fraud Initiative** extends this process to around 300 LAs (currently being processed in the 1996 exercise), with the aim of increasing this to 400 by early 1998; £40M is estimated to have been saved by data matching, and a **National Housing Benefit Register** is being set up by the DSS to help LAs in this task.

The success of such initiatives in combating fraud has led to calls for wider implementation, especially between government departments. As mentioned above, while such checks can be made on a case-by-case basis where a crime is suspected, the Data Protection Act may prevent 'wholesale' matching betweeen departments or agencies which have collected data for different purposes, unless there is specific enabling legislation. In this context, the Government has introduced the **Social Security Administration (Fraud) Bill 1996**, which would allow the DSS to cross reference its records with the IR, Customs and Excise and Home Office and make it easier for LAs to share data for fraud prevention.

3. The Social Security Committee focused on Social Security fraud in its Fifth Report (1994-95), and reported on Housing Benefit Fraud in May 1996 (Third Report, 1995-96).

4. Specifically, Income Support, Family Credit, Child Benefit, Disability Allowance and Attendance Allowance.

Such moves are paralleled elsewhere in the world. For instance Australia established a Data Matching Agency within its Department of Social Security in 1990. DSS files are matched with the Australian Tax Office and the Departments of Veterans Affairs, Education Employment and Training and Housing and Regional Development and the results returned to each department for investigation. Costs are around $20M p.a., with net savings of $100M p.a. directly from data matching (plus an estimated $100M p.a. from improved 'voluntary compliance').

ISSUES

There are several **technical factors** affecting data matching. One is the trend towards distributing computers (and data) throughout an organisation, so that data needs to be copied to a central location, or connected via fast and secure computer networks. This makes the format of the data important - centralised data is usually held in well defined, 'structured' files, whereas distributed data may be spread across several different programs and computer systems, which may not be directly compatible. This generates a need to translate between different formats, and to search databases taking account of differences in the way they file records.

A key to solving the latter problem is to ensure that all data sets have a unique personal identifier. Indeed the Social Security Administration (Fraud) Bill includes provision to require benefit applicants to supply sufficient information to allow their National Insurance numbers to be identified, so that cross checks can be carried out against contributions records, etc. If this were not possible, then other techniques would have to be used (Box 1) which allowed individuals to be identified by searching (or 'mining') through data sets and finding the 'threads' which tie information to a particular person. These are inherently less reliable.

Indeed, the **reliability of any data matching system** is a key consideration, whether in commercial or public sector applications. The'errors' may come from information incorrectly assigned to the 'right' person, or two data sets wrongly attributed to the same person (e.g. the two J. Bloggs claiming different benefits are wrongly flagged as the same person).

As commercial users move to extend data matching, the Registrar and others are particularly concerned about:

- **avoiding mistakes** (e.g. people being refused credit because some of their details coincide with a bad debtor; or because errors have been spread by data sharing;)
- **function creep** - where the boundaries between fraud prevention, credit management and marketing become blurred.

In the **public sector**, data matching appears very cost-effective, e.g. the Australian data matching programme shows benefit-cost ratios of 5:1 on direct savings, and 10:1 if voluntary compliance is taken into account. However, whereas the commercial risk on error may be losing a customer, it may be unacceptable if an automated system in the public sector were incorrectly to 'cast suspicion' on a large number of 'innocent' people. Experience of systematic public sector data matching in Australia is that from over 66M DSS records matched[5] in 1994-95, 137,921 cases were passed on for review. Of these, 23% resulted in benefits being reduced or withdrawn. Benefits were increased in 19% of cases. **Any system which flags cases for investigation needs to bear in mind that the majority may be legitimate claimants, or even receiving under payment.**

Another concern about data matching in the public sector is the **risk to 'privacy'** represented by bringing together large quantities of personal data - ranging from the potential for misuse of information by individuals with access to government systems, to visions of a 'big brother' state. While the Government Direct Green Paper rejects the idea of forming a single, centralised government database, interconnecting government computer networks risks **allowing access to the totality of information, unless suitable administrative and technical security is in place.** Thus the use of secure data matching agencies (Box 1) and the development of 'privacy enhancing technologies', such as encryption and smart cards, are seen by many to be required, hand-in-glove with the deployment of data matching in the public sector.

Some countries have legislated variously to allow, control or prevent data matching. In Sweden every data matching exercise must be approved by the Data Protection Commissioner. In the US, a Computer Matching and Privacy Protection Act was introduced in 1988, while in 1989, the Canadian Privacy Commissioner issued a directive under the authority of the 1985 Privacy Act. Similarly in Australia, data matching is regulated under the 1988 Privacy Act and the 1990 Act discussed above.

In the UK, the Data Protection Act does not deal specifically with data matching, but the Registrar interprets the principles to regulate this area. As described in Box 2, wholesale comparison of files may not be allowed unless disclosure is required in law, and a further limitation on government departments sharing data for fraud prevention is the legal concept of "excess of delegated powers", which prevents officials from acting out with their statutory powers.

Public sector data matching exercises so far have thus had to rely either on serendipitous provisions (e.g. the Audit Commission's power to require information from Local Authorities for the purpose of efficiency studies), or occurred entirely within a single department (e.g. DSS). The Social Security Administration (Fraud) Bill, proposes to give the DSS the powers it requires to match data as well as to clarify the rights of its staff to investigate and adjust benefit payments. In terms of the possible wider use of data matching in the public sector, the Government Direct Green Paper states that the Government will consider the need to introduce legislation.

Some observers, however, are concerned that such legislation would inevitably lead to widespread data sharing in government, without the consent of data subjects, and thus be in tension with the philosophy of current data protection legislation. In this context, the need to implement the EC Directive presents an opportunity to look again at data protection legislation in the UK. The Directive offers a degree of continuity with the 1984 UK Act and preserves the 'principles' in broad terms, but introduces more flexibility (e.g. only requiring data users handling sensitive information to register) and strengthens the emphasis on privacy. The key issue is that if the Directive is implemented as a Statutory Instrument then it will only apply to areas of EC competence - excluding many areas of government. The Data Protection Registrar prefers *"a thorough review of data protection legislation, not the minimalist implementation of the Directive proposed by the Government"*, and has also suggested that legislation enabling data matching should include provision for a statutory code of practice to safeguard the rights of the individual.

Overall, the impact of data matching may be shaped as much by public attitudes as by technological developments. Privacy is an abstract concept which is open to different interpretations by individuals and organisations, and Parliament may well need to address question of where to strike the balance between protecting reasonable privacy for the individual and society's need to protect itself from crime such as fraud and to operate efficiently.

5. Each person generally has several records.