Profiling: A Hidden Challenge to the Regulation of Data Surveillance

Roger Clarke
Principal, Xamax Consultancy Pty Ltd, Canberra
Visiting Fellow, Department of Computer Science, Australian National University

Version of 28 November 1993
Published in the Journal of Law and Information Science 4,2 (December 1993)
© Xamax Consultancy Pty Ltd, 1998


Abstract

Profiling is a data surveillance technique which is little-understood and ill-documented, but increasingly used. It is a means of generating suspects or prospects from within a large population, and involves inferring a set of characteristics of a particular class of person from past experience, then searching data-holdings for individuals with a close fit to that set of characteristics.

It is rather different from better-known data surveillance techniques such as front-end verification and data matching. It raises rather different issues, and requires rather different regulatory measures. This paper surveys the limited information available, defines and describes the technique and its social implication, and argues the case for action by regulators.

Contents

- Introduction
- Definition
- Uses
- The Technique
- Benefits of Profiling
- The Negative Impacts of Profiling
- Intrinsic Control Mechanisms
- Extrinsic Constraints
- Conclusions
- Reference List

Introduction

Data surveillance, usefully abbreviated to dataveillance, is the systematic use of personal data systems in the investigation or monitoring of the actions or communications of one or more persons. It is supplanting more traditional forms of surveillance because it is cheap and effective.

Personal dataveillance involves subjecting an identified individual to monitoring, whereas in mass dataveillance groups of people are monitored in order to generate suspicion about particular members of the population. Personal dataveillance techniques include transaction-triggered screening, front-end verification, front-end audit and cross-system enforcement. Mass dataveillance techniques include general use of the above techniques, without any transaction to trigger them, plus additional tools such as profiling.


Profiling is a particular dataveillance technique which is little documented. No active measures have been identified anywhere in the world which have been designed to explicitly subject it to controls. The purposes of this paper are:

- to define and describe profiling;
- to assess its social implications; and
- to establish the need for its use to be regulated.
It has proven extremely difficult to undertake original empirical research into profiling practices. Considerable difficulties were previously encountered by the author during the period 1987-92, in undertaking a study of data matching (Clarke 1992). Particularly in Australia, general enquiries and freedom of information requests were countered by invocation of exemption clauses. There was also very little published, and very little of that was by researchers independent of the organisations concerned. During the last few years, the openness in relation to data matching has improved markedly. In the United States, this resulted from the firm congressional and presidential support for computer matching programs and the extent to which they have become embedded in agency practices, followed by the 1988 Privacy Protection and Computer Matching Act which expressly required a degree of publicity for such programs. In Australia, the decreased secretiveness followed the passage of the Privacy Act 1988, the inclusion in that statute of explicit mention of data matching, and the interest shown in the topic by the Privacy Commissioner since his appointment in 1989.

No such liberalisation has yet occurred in either the United States or Australia concerning profiling. Accordingly, this paper has been developed predominantly by reflection on technological capabilities, anecdotes and unofficial information, and through use of the limited secondary sources. This is clearly unsatisfactory, but so too would be the continued absence of a critical literature on the topic.

**Definition**

The sense in which the term 'profile' is used in this paper is "2. ... [the] schematic representation of [a] person's interests for use in information retrieval" (Concise Oxford, 1976, p.885). The term 'profiling' refers to the process of creating and using such a profile.

There appear to be few authoritative definitions in the literature. One which is oriented specifically toward law enforcement uses is "correlating a number of distinct data items in order to assess how close a person comes to a predetermined characterisation or model of infraction" (Marx & Reichman, 1984, p.429). Another, oriented toward uses by the direct marketing industry, is the application of statistical techniques such as regression analysis, non-responder segmentation and models for recency, frequency and monetary value of purchases to find out which consumers are good prospects for an offer and which are not (Novek et al 1990, p.529, referencing Stevenson 1987).

In order to encompass both public and private sector applications, the following is proposed as a working definition:

> Profiling is a technique whereby a set of characteristics of a particular class of person is inferred from past experience, and data-holdings are then searched for individuals with a close fit to that set of characteristics.

**Uses**

This author's research has been unable to locate any comprehensive review of actual applications of profiling. In particular, most Government publications (e.g. PCIE 1981) are generally unhelpful. One early reference to profiling referred to use by the United States Internal Revenue Service to predict the 'audit potential' of individuals' tax returns (Rule 1974, p.282). Marx & Reichman (1984) provided some evidence of uses in law enforcement. The Office of Technology Assessment of the U.S. Congress noted that most U.S. federal agencies had applied the technique to develop a wide variety of profiles including drug dealers, taxpayers who underreport their income, likely violent offenders, arsonists, rapists, child molesters, and sexually exploited children (OTA, 1986, pp.87-95).

The manifold potential applications in governmental contexts are variously supportive of the interests of the individuals they identify, or of society at large, or benign, or inimical to those interests. To provide a feel for the diversity of its potential, consider the following possible target groups:

- students with high propensities for particular artistic or sporting skills;
- patients with high likelihood of suffering particular diseases or disorders;
- individuals likely to commit a crime of violence against persons;
- adolescents with a significant likelihood of attempting suicide;
- taxpayers who are likely to be materially mis-stating income or expenses;
- travellers likely to be carrying or trafficking in drugs;
- persons likely to be underground activists against an authoritarian government.

In the private sector, profiling can be applied to such matters as the location of employees with particular education, experience and language-skills. The primary use has been, however, in the identification of customers likely to be interested in buying a new product or service. There has been a shift in marketing budgets away from advertising in the mass media towards direct marketing, or 'individualised mass marketing': "sufficiently detailed information on the buying habits and personal preferences of individuals ... enable firms to create individual messages for each consumer."
This need for accurately identifying buyers, combined with the technological capability of ‘massaging’ and manipulating massive quantities of data about thousands of people in a coherent fashion, has spurred a massive reworking of the methods used by marketers in reaching and controlling potential customers” (Mukherjee & Samarajiva 1993, p.52. See also Novek et al 1990, p.526). Much of the direct marketing literature compromises careful analysis with hyperbole, and confuses the present with futures desired by the author, or his employer or clients. See, however, Stevenson (1986, 1987), and reviews of direct marketing practices in Novek et al (1990), Burns et al (1992), Larsen (1992), Mukherjee & Samarajiva (1993) and Gandy (1993).

The Technique

The steps in the profiling process can be abstracted as follows:

- **describe the class of person**, instances of which the organisation wishes to locate;
- **use existing experience to define a profile of that class of person**. This is likely to be based at least in part on informal knowledge, references to the literature of underlying disciplines and professions, and discussions within the organisation and with staff of other organisations with similar interests;
- It is also likely that profile construction will be supported by analyses of existing data-holdings within and beyond the organisation, whereby individuals who are known to belong to that class are identified, their recorded characteristics examined, and common features isolated;
- **express the profile formally**, perhaps including the use of weightings to reflect the degree of correlation between the characteristic and the target, thresholds below which the correlation is low and above which it is high, and complex conditional relationships among the factors (e.g. factor x is indicative, but only if factors y and z are both above particular threshold levels);
- **acquire data** concerning a relevant population; for example:
  - in the case of customers: the organisation's customer database, and the databases of mailing list suppliers;
  - in the case of employees: the personnel database, and the records of staff placement companies;
  - in the case of students' propensities in the arts and sport: school, club and perhaps medical records;
  - in the case of patients: medical records from doctors, clinics, hospitals and registries, and family history information, e.g. from Utah;
  - in the cases of potential criminals and of adolescents: school records, welfare agency databases, and the medical and psychiatric records of doctors, clinics and hospitals;
  - in the case of taxpayers: the historical record of tax returns, cash flow information from employers, financial institutions and other organisations, and statistical analyses of those and other databases;
  - in the case of travellers: inbound and outbound flight and voyage movement records; and
  - in the case of underground activists: associations with known dissidents, and subscription lists to subsersive literature;
- **search the data for individuals whose characteristics comply with the profile**. This is highly likely to involve computer support, especially where the data-set is large (e.g. taxpayers), the processing complex (e.g. psycho-social analyses), or the time available short (e.g. lists of aircraft passengers);
- **take action** in relation to those individuals; for example:
  - mail selected advertising to selected customers or prospects;
  - call students, employees or prospective appointees for interview;
  - counsel patients or students and/or their parents and teachers;
  - counsel adolescents and potential violent criminals, and their families, associates and workmates;
  - subject tax-payers to audit;
  - interview passengers and/or conduct luggage- and body-searches; and
  - impose repressive measures on the activist.

Profiling can be conceived as a sophisticated variant of single-factor screening techniques (which are conducted at the time a transaction is processed), or single-factor file-analysis (conducted at some subsequent time). Screening involves the comparison of just a single characteristic against:

- an *a priori* norm (e.g. a claimed tax deduction against a pre-set limit);
- a population-based norm (e.g. a claimed tax deduction against the 95th percentile of all such claims);
- historical data (e.g. the name of the person's spouse); or
- previous transactions (e.g. the number of dependants against the number declared in the previous return).

Profiling constitutes multi-factor screening (if conducted on transactions) or multi-factor file-analysis (if conducted at some subsequent time).

Profiling may be based entirely on data which the organisation already holds. More commonly, it draws on the data-holdings of multiple organisations using the facilitative techniques of data concentration and/or data matching (Clarke 1988, nn.504-5.

of multiple organisations using the facilitative techniques of data concentration and/or data matching (Clarke 1988, pp.504-5, Clarke 1992). There is an apparent incentive for organisations conducting profiling not only to search out and acquire existing data from various sources, but also to undertake or stimulate new data collection. In some cases, static demographic data is sufficient; more commonly, however, the need is for an ongoing stream of 'transaction generated information' (McManus 1990, Mukherjee & Samarajiva 1993, Gandy 1993). This is a facet of what was referred to by Rule et al (1980) as the increasing 'information intensity' of modern society.

Benefits of Profiling

Like any other technique, profiling is neither good nor evil. It is capable of application for very worthwhile purposes. Unfortunately, as has been the case with data matching (Clarke 1992, pp.41-6), there is little on the public record which evidences serious attempts to assess the real value of profiling by government agencies. The benefits of marketing applications, however, can be summarised as:

- providing consumers with invitations to purchase goods and services that they are likely to be interested in;
- avoiding the provision of invitations to consumers who are unlikely to want the goods or services, or to be able to afford them; and
- improving the economics of marketing activities, by 'selecting out' poor prospects, 'selecting in' good prospects, and choosing the form of promotional material most likely to motivate the prospect to buy.

This paper's primary concern is with the 'downside' of profiling, i.e.:

- uses which society would consider inappropriate and oppressive if they were publicly known and debated; and
- uses which are in principle acceptable to society, but which are undertaken in ways that are unfair, insensitive or discriminatory.

The Negative Impacts of Profiling

Profiling is a mass dataveillance as distinct from a personal dataveillance technique: it does not involve the monitoring of an identified individual for a specific reason, but is instead concerned with the finding individuals about whom to be suspicious, who can then be subjected to personal surveillance. It therefore has all of the potential negative impacts of mass dataveillance techniques generally, as outlined in Exhibit 1. The more general, social impacts are best discussed in the context of dataveillance in general, rather than of one particular technique (see Clarke 1988). This paper accordingly focuses on the dangers to the individual.

In the case of private sector use, concerns exist about selectiveness in advertising. Profiling has considerable potential to improve the efficiency with which companies undertake marketing communications with their customers and prospects. At some point, however, selectivity in advertising crosses a boundary to become consumer manipulation (Packard 1957, Larsen 1992). Novek et al (1990) perceive dangers that go beyond mere moral arguments: "profiles ... allow companies to pre-judge the future behavior of consumers, leading some of these firms to ignore certain types of people, and thereby limiting such persons' access to information about goods and services" (p.533). They suggest that the combination of consumer profiling with 'geodemographic clustering' techniques is inevitably leading to "electronic redlining', where calls from low-income neighborhoods identified by their telephone exchange, can be routed to a busy signal, a long queue, or a recorded message suggesting that the desired information service is not presently available" (p.535).

More generally, "the segmentation and marginalisation of consumer information markets further limits the availability of information necessary for informed consumer choice, while simultaneously increasing consumer dependence upon the direct marketer's tightly managed information stream. The result is a market dominated by sellers ... The wider this information gap, the more difficult it becomes to ensure the equitable and efficient working of the marketplace" (p.536).

Exhibit 1: Real and Potential Dangers of Mass Dataveillance

To The Individual

- selective advertising
- arbitrariness
- acontextual data merger
- complexity and incomprehensibility of data
- witch hunts
- ex ante discrimination and guilt prediction
- inversion of the onus of proof
- covert operations
- unknown accusations and accusers
- denial of due process
- covert operations
- unknown accusations and accusers


4
denial of due process

To Society

- prevailing climate of suspicion
- adversarial relationships
- focus of law enforcement on easily detectable and provable offences
- inequitable application of the law
- decreased respect for the law
- reduction in the meaningfulness of individual actions
- reduction in self-reliance and self-determination
- stultification of originality
- increased tendency to opt out of the official level of society
- weakening of society's moral fibre and cohesion
- destabilisation of the strategic balance of power
- repressive potential for a totalitarian government

From: Clarke (1988), p.505

In the case of public sector applications, several issues are of importance:

- **ex ante discrimination and guilt prediction**. Actions by administrative and law enforcement agencies alike have hitherto been essentially reactive. There are very real concerns about the relationship between the citizen and the State when predictive tools are used proactively to identify potential offenders before they commit an offence;

- **inversion of the onus of proof**. This is closely associated with the proactive attitude inherent in *ex ante* discrimination. Although, in Anglo-American societies, the onus of proof has traditionally been on the accuser, there are several key areas in which that tradition has already been eroded, including taxation administration and the handling of people at border-crossings;

- **covert operations**. It is generally accepted in most societies that State organisations do conduct some kinds of activities out of the public eye (although there are varying views as to whether they should, what operations they should conduct, under what circumstances, and subject to what control mechanisms). Profiling appears to be being conducted not only by such security organisations, but also by many other agencies and corporations, outside the purview, and largely without the knowledge, of the public, its Parliamentary representatives or any statutory watchdog. Organisations remain free to develop and apply the technique as they see fit, without mechanisms to ensure that the interests of all stakeholders are appropriately reflected. This breeds a climate of suspicion among those of the public who are aware of, or suspect, the activities;

- **unknown accusations and accusers**. There is a tendency towards the classical Kafkaesque situation whereby the catalyst for the interview or search is not apparent to the individual. The 'accuser' is in any case disembodied: an abstract algorithm established by persons remote in time and space, and applied by a computer. The authors of the European Commission's Draft Directive were sufficiently concerned about this aspect of dataveillance to propose a right for people "not to be subjected to an administrative or private decision adversely affecting him which is based solely on automatic processing defining a personality profile" (EC 1992, Article 16).

- In the present climate of surreptitious use of the technique, the mystery tends to be compounded by the preference of the organisations involved to not disclose the methods they are using, ostensibly 'for security reasons', but importantly also because of the fear of being subjected to controls, or even having their practices banned;

- **denial of due process**. In order to maintain the smoke-haze surrounding their practices, there is a tendency for organisations using profiling to not bring the real evidence forward into court, but rather to seek out and use other information which they are able to gather during the course of the investigation. Defendants are therefore placed in the position of having to defend against evidence, and even charges, which seem to be beside the point, rather than the real issue.

Associated with these concerns is the extent to which judgmental valuations enter into applications of profiling. Cultural, racial and gender biases, for example, are inevitable, because of the facts of the matter (e.g. the arrest rates of aboriginal people in Australia, and persons of negro and latino origin in the United States, are higher than those for white people), the way in which data is collected, organised and presented (e.g. more data is collected about people of lower socio-economic origins, because they are more commonly applicants for benefits), and the way in which characteristics are inferred (i.e. the people who prepare the profiles bring with them their own theories about which kinds of people are prone to behave in the manner being targeted).

**Intrinsic Control Mechanisms**

A variety of factors might act to prevent unreasonable uses of profiling, and constrain unreasonable practices in relation to such profiling as is done. These factors are summarised in Exhibit 2.

such profiling as is done. These factors are summarised in Exhibit 2.

### Exhibit 2: Intrinsic Control Mechanisms

- self-restraint
  - by the organisation
  - by professionals
- policy or codes of conduct
  - at the political level
  - at government or industry association level
- exercise of countervailing power
  - by other organisations
  - by the public
- infrastructural inadequacy
  - technology
  - historical data
  - contemporary data
- operational incompetence
  - individual skills
  - inter-organisational jealousies
  - inter-jurisdictional difficulties
- economic inadequacy

An organisation might decide on the basis of 'good corporate citizenship', or good faith or fair dealings with their clients, or good relations with their customers, not to use the technique, or to apply particular controls to ensure that the procedures are not unfair. In addition, it is conceivable that employees and contractors who are important to the process may regard themselves as constrained by the code of ethics of their professional body. This might preclude use of the technique at all, use of it for particular purposes, or use of it without particular features protective of the data subjects. There is little evidence, however, of such mechanisms having significant impact on the use of dataveillance practices generally, or of profiling in particular.

Stated government policies regarding such matters as fairness, equity and anti-discrimination might act as constraints, as may codes of conduct, undertakings or policy stances by oversight agencies in the public sector or industry associations in the private sector. There have been instances of constraints on dataveillance arising in this manner, but none is apparent in relation to profiling.

There may be circumstances in which another organisation acts in such a way that an organisation feels itself constrained to not use profiling, or to apply protections. In particular, a competitor may use negative advertising to project its much greater respect for the privacy of its customers; or a government agency whose participation is crucial to the project may decide not to make its data available because of the scheme's privacy-invasive or discriminatory nature. Once again, some instances of this mechanism in operation can be seen in both Australia and the United States in relation to dataveillance practices generally, but not to profiling.

Public opinion can have an effect on profiling practices. Companies which are known to use the technique can be avoided, and publicly vilified, by consumers and their representatives. The passage of the U.S. Video Privacy Protection Act of 1988 (following the publication of Justice Bork's video rental history), the withdrawal of the Lotus/Equifax Marketplace product in 1990, and the cancellation of the Blockbuster video rental profile database in 1991 (see Mukherjee & Samarajiva 1993, p.51) provide evidence that public opinion can be effective in constraining use of profiling and its facilitative mechanisms.

These constraints are, however, entirely dependent upon the practices and their implications becoming common knowledge: they are entirely non-operative if the organisation is successful in suppressing the fact of its use of the technique. Moreover, government agencies are generally much less responsive to pressure from public interest groups through the media. They do, however, tend to appreciate the point better if their Ministers or Secretaries of State, encouraged by the party's constituents and financiers, require them to adapt their procedures.

Another constraining factor may be the insufficiency of the available infrastructure. The hardware, networking and software capabilities to support the project must be in place, or able to be acquired or developed; for example, there may be doubts about the ability of the participants to develop suitable algorithms. Historical data must be available to support the development of a profile; and current data must be available which can be run against the profile in order to generate suspects or prospects. Given the demand powers of government agencies, the serious weaknesses of controls over flows of personal data in the public sector, and the almost complete absence of controls in the private sector, it cannot be expected that these factors have represented a significant brake on profiling activities.
Extrinsic Constraints

Given that profiling is potentially highly intrusive, and that intrinsic control mechanisms appear to act as at best patchy and partial constraints on unreasonable uses of profiling, it is important to give consideration to regulatory mechanisms.

This author's research has to date unearthed no regulatory measures dealing explicitly with profiling. Moreover, long-standing common law protections such as the laws of confidence and defamation, and privacy and data protection measures, were conceived without any understanding of the technique. Such protections as do exist are therefore generic, or accidental and incidental.

One provision which exists in some form in most statutes is that referred to by the OECD Guidelines (1980) as the Openness Principle, i.e. "there should be a general policy of openness about developments, practices and policies with respect to personal data". Unfortunately the implementation of the Principle in most countries falls far short of that aspiration, particularly because of the wording chosen by Parliamentary Draftsmen to implement it, and the manifold exemptions and exceptions provided. The Australian Privacy Act 1988, for example, requires only that "a record-keeper ... shall ... takes such steps as are, in the circumstances, reasonable to enable any person to ascertain ... the nature of [personal information held] ... [and] the main purposes for which that information is used" (Principle 5). Unsurprisingly, disclosure of profiling activities is rare, even in response to direct requests for information.

In most countries, there are constraints on the collection, storage, disclosure and use of personal data, and these would appear to act as controls on profiling, as for other dataveillance techniques. In fact, much of the apparent control is illusory. Many agencies are wholly or partly exempt, and many exceptions exist (for example, in the Information Privacy Principles in the Australian Privacy Act, the qualifier 'reasonable' occurs fifteen times, and 'protection of the public revenue' is a sufficient ground for use and disclosure of personal data). Agencies generally claim practices to be 'authorised by law' simply because they are not prohibited, and are generally consistent with the agency's mission. The private sector in both the United States and Australia is subject to only very limited regulation of dataveillance practices.

Conclusions

Profiling is an important application of information technology, but also one which embodies considerable dangers to individuals and society. On the basis of the limited evidence publicly available, the conclusion is inescapable that the law in most countries provides only very limited means of constraining, or even ensuring public knowledge about, the use of profiling. Profiling is largely being conducted without public knowledge, without justification, and without appropriate safeguards.

Time works in the favour of even the most objectionable uses of dataveillance techniques, because Parliaments, Ministers and regulatory agencies are hesitant to reverse long-established practices: they tend to be convinced by the argument that the activity must have an economic justification because it would not otherwise have been in existence for so long. Meanwhile technological advances are resulting in further improvements to the effectiveness and the economics of the technique, and in increasingly large pools of accessible personal data.

Government agencies and corporations are taking advantage of the lack of regulation to apply profiling as they see fit. Intrinsic controls over dataveillance techniques have been repeatedly shown to be utterly inadequate. Profiling therefore demands far more attention than has to date been given to it by researchers, by executives in both the public and private sector, by regulatory agencies, and by legislators.

Due to the copious weaknesses in existing privacy protections, the inventiveness of agencies in circumventing them, and the
Due to the copious weaknesses in existing privacy protections, the inventiveness of agencies in circumventing them, and the ravages of technological change, many authors have argued the urgent need for 'second-generation' privacy protective legislation (e.g. Laudon 1986, Clarke 1988, Flaherty 1989, Clarke 1992, Bennett 1993). Further piecemeal measures could be considered, including the extension of existing statutes to cope with profiling. The challenge to legislators is, however, much broader than profiling alone. The nature and scope of privacy protections need to be re-examined, and a much more comprehensive framework provided. Such a framework would remove the exemptions and exceptions, address the social and economic justification for privacy-invasive programmes, and empower a suitably resourced 'watchdog' agency to study all uses of dataveillance techniques, and submit each of them to an appropriate and detailed regulatory regime.

Reference List


Stevenson J. (1987) 'The History and Family Tree of 'Databased' Direct Marketing' Direct Marketing (December 1987)


---

**Navigation**

Go to [Roger's Home Page](#).

Go to the contents-page for this segment.

Send an email to Roger

Created: 13 October 1995

Last Amended: 24 Jun 1998

---

These community service pages are a joint offering of the Australian National University (which provides the infrastructure), and Roger Clarke (who provides the content).

**The Australian National University**

Visiting Fellow, Faculty of

Engineering and Information Technology,

Information Sciences Building Room 211

---

**Xamax Consultancy Pty Ltd**

ACN: 002 360 456

78 Sidaway St

Chapman ACT 2611 AUSTRALIA

Tel: +61 2 6288 1472, 6288 6916

---