DNA Dataveillance: Protecting the innocent?

Authors

Anna Vartapetiance Salmasi, Department of Computing, University of Surrey, UK.
Lee Gillam, Department of Computing, University of Surrey, UK.

Abstract

Purpose- This paper discusses the National DNA Database (NDNAD) and some of the controversies surrounding it with reference to legal and ethical issues, focusing particularly on privacy and human rights. Governance of this database involves specific exemptions from the Data Protection Act (DPA), and this gives a rise to concerns regarding both the extent of surveillance on the UK population, and the possibility for harm to all citizens. This is of wider importance since every current citizen, and everybody who visits the UK, could become a record in the DNA database. Principally, we explore whether these exemptions would also imply exemptions for software developers from codes of practice and ethics of their professional societies as relate to constructing or maintaining such data and the database.

Design/methodology/approach- We make a comparison between the principles of the DPA, as would need to be followed by all other organizations handling personal data, professional responsibilities based codes of ethics of professional societies, and the current reality as reported in relation to the NDNAD and the exemptions offered through the DPA.

Findings- Primarily, if NDNAD were not exempted from certain provisions in the DPA, the potential for the kinds of data leakages and other mishandlings could largely be avoided without the need for further considerations over so-called “data minimization”. We see how the lack of afforded protection allows for a wide range of issues as relate at least to privacy.

Originality/value- This paper provides the first evaluation of the combination of law, codes of ethics and activities in the real world as related to NDNAD, with concomitant considerations for privacy, liberty and human rights. Originality is demonstrated through consideration of the implications of certain exemptions in the DPA in relation to crime and taxation and national security, and in relating the expected protections for personal data to widely reported evidence that such protections may be variously lacking. In addition, we provide a broad overview of controversies over certain newer kinds of DNA analysis, and other relatively recent findings, that seem generally absent from the vast majority of debates over this kind of analysis.

Keywords- DNA, Data Protection, Computing, Professionalism, Code of Ethics

Paper type- Research paper

1. Introduction

Software systems collect a range of data about individuals in various walks of life. Use of such systems can be wide-ranging, from CCTV-based monitoring of particular scenes and situations, to capturing data about particular purchases and using these data to offer further products and services, to determining whether sufficient numbers of computer keys have been pressed per hour. These software systems exist in the workplace, home, public areas, and transportation (Chen & Park, 2005; Vuokko, 2008). The continual introduction of such software systems often gives rise to privacy issues, with some concerns raised over a "Surveillance Society" (House of Commons, 2008). Of course, it is possible to conceive of somebody who knows everything about an individual, but has purely voyeuristic intentions: "you have complete autonomy, just no privacy" (Moor 1997), and argue a case for such voyeurism as protection from harm: "if every inch of the globe was viewed, recorded and indexed, virtually no deviance from social norms (such as criminal activity) would go undetected" (Lockton & Rosenberg 2005). The extent to which individuals become concerned by such developments will depend to some extent on their perception of the need for such surveillance to ensure collective good versus individual privacy (Palm, 2008). This perception of need of surveillance may be fuelled, further, by the digital-era media-driven characterization of "Person as risk" (Hoven & Manders-Huits, 2008), in which we consider every individual to have potential criminal intent towards us. Such a characterization may lead, subsequently, to having to prove innocence rather than have guilt proven - the comprehensive use of such technologies could lead to increasing numbers of false positives unless accuracy is guaranteed. With increased proliferation of data mining and data sharing, allowing different data to be analyzed to infer new information, and used by others without data subject consent, privacy issues may get accrue (Lockton & Rosenberg, 2005). Over-generation of matches against Terrorist Watchlists suggests significant potential for, at minimum, inconvenience and disruption, particularly when "about half of the tens of thousands of potential matches sent to the [U.S. Terrorist Screening] center between December 2003 and January 2006 for further research turned out to be misidentifications." (USGAO, 2006), and data mining as deployed in recommender systems produces inferences indicative of assumptions that "Tivo thinks I am gay and Amazon thinks I am pregnant" (Contractor, 2007).

This notion of "person as risk", then, not only suggests surveillance (Ward, 2004) but also suggests that such surveillance may be beneficial to us since we can also be ruled out of investigations using the same technologies. One example of where such thinking is evident is the United Kingdom's National DNA Database (NDNAD). DNA samples and associated data can provide a deal of personal information which even its "owner" might not be aware of. Generally, based on European Data Protection Supervisor (EDPS) (OUT-LAW, 2007), the UK's Data Protection Act 1998, should provide safeguards on citizen's data and assure citizens of no infringements on their privacy and human rights (Howley et.al., 2005). Until now, NDNAD has been variously exempt from some of these provisions.

Computer professionals are expected to follow laws of their country of operation. In the online environment, it becomes less clear which laws apply to certain activities unless the relevant jurisdiction is identified in, usually extensive, terms of use, privacy policies (OUT-LAW, 2008b), or other such documents. Moreover, the latest technologies may provide a challenge in interpretation of existing laws, which may only be clarified by test cases. Even then, appeals may alter the interpretation. The computer professional must keep abreast of such interpretations since they may impact on extant business. Conflicts in laws in the international environment, difficulties in interpretation, and potentially even exemptions from law, provide a difficult operating environment. Nevertheless, computer professionals are expected to follow codes of ethics or practice to design systems that will, for example, ensure privacy and human rights are honoured. McGraw (2004) describes a "professional code of Ethics" as the product of a professionally-defined social group's efforts to enforce certain rules of behavior among its members, and there are several that might be referred to here: from the British Computer Society (BCS, 2006), the Association for Computing Machinery (ACM, 1993), and the Institute of Electrical and Electronic Engineers (ACM/IEEE-CS,1999). It is arguable whether they all share the same principles (Gotterbarn, 1997) or how well considered they might be (Gotterbarn, 2007). Occasionally, the codes provide guidance where laws may be deficient in specific instances or countries, though this may be an artifact of reference to national laws in the country of origin of these codes. While there may be an exemption from national law or a part thereof, for a specific application, these codes may not be so clear about whether such exemptions carry over to the codes themselves, and what the priority should be. In addition, such codes apply to the work of the professionally-defined social group, but perhaps consideration of application to the results produced by the systems and software are also vital? It may be difficult to evaluate the potential for negative use of certain software applications, while others may be more obviously apt for negative uses.

This paper discusses surveillance using data – dataveillance - with a focus on the UK's National DNA Database (NDNAD) and some of the controversies surrounding it, and reference to legal and ethical issues. Broader questions exist over privacy and human rights (Johnston, Waterfield, 2007) and viewpoints can be drawn from the application of ethical theories and considerations of negative and positive rights in trading for civil liberties for personal and national security. Consideration of such issues is important to understand the possibility for harm to all citizens, every one of whom, merely by visiting the UK, could become a record in the DNA database. Also, there are calls from certain quarters for a national database of everyone's DNA to avoid discrimination [1] and even a global database [2] - but as yet there are no plans to DNA profile everyone in the UK, though *no-one ever says never*[3]. According to the European Data Protection Supervisor, there is no specific framework proposed for an EU shared database (OUT-LAW, 2007). Those in the UK might have thought they could rely on protection via DPA and related international agreements, though with certain exemptions to the DPA, protection seems to be somewhat limited. The outcome of a recent case in the European Court of Human Rights, S. and Marper v. The United Kingdom, 2008, in which it was deemed that the retention of the DNA of

these subjects represented a breach of their human rights, may force change on how retention of DNA is retained, both for current and future records (ECHR, 2008).

2. Background

In terms of dataveillance, the UK boasts significant, and expanding, infrastructure supporting crime fighting, including 4 million CCTV cameras, a forthcoming Identity Card, already required for non-UK students, and the largest DNA database of any nation. These infrastructures have raised concerns and actions related to privacy, human right and civil liberties including:

1. A politician resigning, in relation to the government's position, in order to the highlight the erosion of civil liberties - only to stand for and win back his seat,. His resignation statement noted: "we will have shortly the most intrusive identity card system in the world. A CCTV camera for every 14 citizens, a DNA database bigger than any dictatorship has, with thousands of innocent children and millions of innocent citizens on it".
2. The UK's Home Affairs Committee releasing results of an inquiry into "A surveillance society?" (House of Commons , 2008)
3. The scientific validity of Low Copy Number (LCN) DNA analysis being challenged in relation to the Omagh bombing (NICC49, 2007), prompting a further investigation into the validity of the approach. (Home Office, 2008)
4. A report into the UK Government losing personal details (names, addresses, dates of birth and bank accounts) of 25 million citizens blames serious institutional deficiencies.

Following events of 2001 and 2005, the call for further laws and technologies supporting a "fight" against terror and crime was made loudly. Some would suggest the response has become disproportionate and overly intrusive, and there are concerns about both legality and scientific rigour. Various UK government departments routinely collect personal data about citizens, and are supposed to be governed under the DPA 1998; indeed, most try not to lose them. Nevertheless, a range of large-scale data collection and population monitoring is underway, with the UK able to boast significant infrastructure supporting the fight:

a) The largest DNA database of any nation state, containing DNA records for over 5% of the population [4], 50% larger than the combination of data in all of the remaining EU member states and growing by more than half a million a year. (Hope, 2008a)
b) Over 4 million CCTV cameras monitoring the movements of citizens, and producing significant volumes of potential video evidence.
c) Discussions being held with potential suppliers over the provision of a National Identity card scheme, and major contracts being signed by IT companies for biometrics databases and registration systems.

Individually, these developments are significant. If we now consider the integration of these initiatives: the Identity Cards Act 2006 provides for a National Identity Register, and registering for the card may require individuals to provide "fingerprints, and other biometric information",

to be photographed, and more broadly to "provide such information as may be required by the Secretary of State". If DNA were considered within this remit as biometric information, and should the photograph be suitable for comparison to captured CCTV data, this Act provides an apparently strong integrating capability with other databases. The future capacity for using such an integrated database for automated, rather than manual, monitoring becomes significant. Projects at this scale, CCTV surveillance, ID cards and DNA databases, may be viewed as based on the inference that: *Criminals are the minority and anybody could be a criminal, therefore we can only prevent and detect crime if we are capable of monitoring everybody: we rule large numbers of people out of having committed the crime and focus resources on the remainder.* Objections to such databases and systems, and to collection of your DNA, imply that you must have something to hide: government spokesmen suggest "people who were innocent had nothing to fear" (Hope, 2008b).

One might expect all of these collections to be subject to the provisions of the DPA which precedes such systems. The DPA reflects the UK implementation of European Directive 95/46/EC for protection of personal data, though there have been debates over the degree to which these are aligned (OUT-LAW, 2004). The DPA includes 8 principles of data protection covering that data should be accurate, up to date, adequate, relevant, not excessive, fairly and lawfully processed in line with rights of data subjects, for specified purposes, secured against loss or accidental destruction, not kept longer than necessary, and not transferred to non-EEA countries without adequate protection. However, Section IV of the DPA outlines a number of exemptions. DPA section 29 (s.29) relates to prevention and detection of crime and collection of taxes, giving an exemption from principle 1 that personal data shall be processed fairly and lawfully; this seems to imply that unfair and illegal processing is acceptable. Additionally, there exists an exemption to s.7 regarding right of subject access. In essence, the person about whom the data are stored is unable to use the DPA to find out how these data are used and by whom (principle 2), whether it is adequate, relevant or excessive (3), accurate (4), is being kept for longer than necessary (5), processed in accordance with their rights (6), and has to assume that it is not being transferred out of the EEA (8). In short, there is little by way of accountability towards the data subjects themselves such that they may have confidence in this system. This becomes particularly relevant given "volunteered" data, as will be discussed later. Furthermore, a second exemption relates to the- apparently ambiguous- notion of national security:

> *"Personal data are exempt from any of the provisions of (a)* ***the data protection principles****, (b) Parts II, III and V, and (c) section 55, if the exemption from that provision is required for the purpose of safeguarding national security".*

Data protection principles which are enforceable, then, would appear to be specific to the context of use of the NDNAD. Extending the arguments above would suggest that anybody could be a terrorist - then, the entire database is for the purpose of safeguarding national security, and protections are largely lifted. In addition, it is evident that "function creep" in one particular law that was intended to ensure legitimacy of surveillance (the Regulation of Investigatory Powers

Act 2000) results in its being used "to conduct surveillance to deal with fly-tipping, littering, dog fouling and the sale of alcohol to those aged under 18"[5]. This, however, is a targeted consideration – surveillance directed towards specific individuals; the processing of data in order to determine groups of people with similar DNA profiles takes a rather less discriminatory approach, and searching through all database records to determine such matches would suggest that the entire database becomes exempt during that processing. It is just such a combination of function creep, differential exemptions, and an inability for individuals to determine, for example, the accuracy of the data held about them, that suggests a difficulty with a statement such as "people who were innocent had nothing to fear".

Once can hypothesize about ownership of personal data and, especially, biological samples and the data that relate them to oneself. Different kinds of value can be ascribed to data of various kinds. For Litman (2000) the privacy-as-property model encourages a market in personal data, and here the proposal is made that breach of confidence/trust would apply more effectively, with tort law being used to "assess the context in which consent [to use the data] was given". This would assume that consent has been given, which we address briefly in section 3. However, without right of access to data, since this is a legal exemption, it may prove difficult to make a case. More widely, the immediacy of invasion of privacy, and comparative glacial pace and disproportionate costs of most legal remedies tends not to discourage a trade in data, and organizations deriving profits in such trade may willingly hedge these against likelihood of action. For the NDNAD, exemption to the right of subject access implies that we may not even know if such data is in the system, suggesting that there can be no challenge over ownership. Bergstrom (2000) considered Lockean, utilitarian and contractarian views relating to ownership of samples, and particularly of medical samples, though little can be discerned by way of conclusion: "we can say that it is very hard indeed to settle political and moral problems of ownership in a principled way". A question explored in this work is: "Should we say that the breast cancer gene BRCA 1 belongs to each person who has that gene?"; an analogy could be drawn to matching sites on the DNA profile. In absence of identifying information, we may make matches amongst samples through common features, but it is in establishing a connection to a named individual, or a group of named individuals, that the issue of privacy arises and for these systems, identifying people is the dominant purpose. For Bergstom, also, the utilitarian view relates to the greater good emergent from use of samples, rather than to the question of sample ownership. The utilitarian view from this consideration would suggest quarantine during an influenza pandemic, for example, takes precedence over liberty. The extension to this consideration would be that individuals may own the samples, but cannot own the knowledge derived from the analysis of these samples; here, by extension, you would have access to DNA samples, but still no rights over your data.

We are concerned, principally, with whether the various exemptions from the DPA would imply exemptions for software developers from the codes of practice and ethics of their professional societies as relate to constructing or maintaining such data and the databases and systems in

which they are housed and used. In particular, we consider the implications of a generic exemption as would exist in the interests of safeguarding national security. This goes further to questions about whether these professionals are fully cognizant of the risks of what they are doing, and whether they are being honest about risks and providing sufficient protection for fundamental human rights and the privacy of others. This is important given doubts about the perceived value of ever-more extensive collections of DNA for fighting crimes, the ways in which the database is being used and controlled, and how the information contained becomes a high-value proposition for private companies and a target for external attacks due to the high-value of information about health, family relationships, appearance, and associated data related to behavior, not just identity. Our intention is to identify the potential conflict for professionals, and ask whether such general exemptions are themselves harmful. Expectations on such professionals may also be compromised by the move from government-supported forensic agency to profit-motivated corporate entity [6] and more widely by the commercial interests of large corporations. In general, perhaps we can only ever limit how much we invade privacy, but we have to balance this against potentially conflicting legal, professional and ethical considerations and the drive towards profitability of such enterprises should not be made to the detriment of civil liberties.

We emphasize that we are broadly in favour of the collection and use of DNA evidence relating to crimes, understand its specific value in solving crimes of a certain nature, and believe in the scientifically well-grounded approaches to the analysis when used to corroborate other evidence, with associated statistics over accuracy of robust testing. However, the role of DNA in crime detection is changing, and previous protections over such data seem to be being eroded. We, as computing professionals, are, concerned about certain approaches being used to populate the database, protections afforded to the data, validity of certain analyses and potential for invasion of privacy of ever larger numbers of individuals.

3. DNA, evidence and analysis

DNA has a number of interesting features, from its structure as discovered by Crick and Watson, to figures regarding its uniqueness. DNA analysis was used initially to relate DNA from a crime scene to DNA of one or more suspects, or to rule specific suspects out. Now, DNA is becoming a principal weapon with profiles obtained from crime scenes used to derive lists of *potential* suspects, or *members of the family* (familial DNA) of a potential offender. The standard test for DNA, SGM+, reportedly offers a 1 in 1 billion chance that an identical sample of DNA could be obtained from a different subject. This is improved over the 1 in 50m chance offered by its predecessor SGM test. These *estimated* probabilities suggest strong accuracy in matching and therefore offender identification. The SGM+ approach may be statistically compelling, but presently concerns are being raised regarding familial DNA, and the accuracy of so-called Low Template analysis techniques such as LCN. There are concerns, also, over increasing reliance on DNA evidence: the so-called “CSI effect” where one should “follow the evidence, it never lies” (Woodsand & Foggo, 2008). Information regarding probabilities for LCN and familial

approaches is not readily available, and the understanding of DNA by both scientists and defense lawyers, is leading to specific questions over *travelling, shedding, contamination* and *analysis.* A number of such risks were identified by Gill (2001) in establishing LCN DNA analysis, an approach we will discuss later.

Shedding: Some are more prone to "shed" DNA than others, and various scientifically controlled experiments have demonstrated that and this can lead to secondary transfer with DNA transferred to an object via another person without their own DNA being placed, depending on whether a particular subject is a good or bad "shedder". If a good shedder shook hands with a bad shedder, and the bad shedder subsequently touched an object, the good shedders DNA could be transferred –

> *"The full DNA profile of one individual was recovered from an item that they had not touched while the profile of the person having contact with that item was not observed" (Lowe et al, 2002).*

This type of secondary transfer was previously thought not to impact standard investigations: "Our data do not support the conclusion that secondary transfer will compromise DNA typing results under typical forensic conditions" (Ladd et al, 1999). Whether this conclusion is consistent with LCN and familial DNA analysis is a different question.

Travelling: DNA goes to places its "owner" has never been, and can remain in the same location for at least a few thousand years (Austin et al, 1999). DNA could be planted at crime scenes to incriminate others, e.g. by transporting objects such as cigarette ends, cutlery, cups or glasses to the scene.

> *"Last year detectives reinvestigating a case of rape got a DNA profile from a strand of hair caught in a ring worn by the victim. The DNA identified Mark Minick, who was charged with the rape. Yet when the case arrived in court, it fell apart. Minick is white, small and slim – while the victim had described her attacker as black, large and tall. She is thought to have picked up Minick's hair by chance from a blanket in the hospital where he had worked".* (Woodsand and Foggo, 2008)

Swabs of saliva may now be taken from buses and used to trace individuals (Lydall, 2007). Aside from the unsavoury nature of the incident, we can identify two causes for concern regarding contamination and analysis.

Contamination: A mixture of DNA from different people may be present in a sample, possibly before or after a crime has occurred. Contamination can occur through incorrect handling, particularly by those with little training in doing so. One police force has estimated a 6% success rate in obtaining a full profile suitable for matching using LCN (Home Office, 2008) due to small sample quantities and impacts of contamination.

Analysis: If a crime were committed on the bus at some distance, or some time, from where the saliva's owner had departed the bus, the DNA trace may or may not be uniquely identifying due to contamination but may implicate one or more owners, and presumption of innocence of the latter crime would become difficult to argue. LCN and familial analysis may result in a number of false positives.

4. Data Collection

The intended purpose of data collection, on such a scale as for the DNA database, is suggested as crime detection and, perhaps, prevention. The populating of this DNA database from samples collected from those who have come into contact with the police suggests that the UK has over 5% of its population with criminal tendencies. NDNAD can be further expanded by continued selection of specific groups of the population, beginning with those undergoing criminal records bureau (CRB) checks for working with children and other groups. In addition, DNA samples can be forcibly taken through supposedly minimally invasive techniques involving removal of a hair or a mouth swab, for example. Voluntary inclusion is also possible, though it has traditionally been difficult to withdraw permission for your data to be used once voluntarily included or to object to its use in genetic research; this may also be a reflection of one of the strategic objectives identified in the National DNA database Annual Report 2005-6 of "maximizing sampling opportunities" (Home Office, 2007). The current shift in thinking is towards an "expectation" that such samples would be destroyed at some undefined future point (OUT-LAW, 2008c) and only the convicted should be retained on NDNAD (OUT-LAW, 2008a). Recent recommendations regarding NDNAD have made little change to what already exists, beyond identifying a need for increased communication (Human Genetics Commission, 2008); much of the debate remains as "divided opinions".

There are discrepancies in retention of data and samples: legislation in Scotland differs slightly from that covering England and Wales, allowing for data to be removed more easily – though there remains a question of how a data subject can know that the data has actually been removed and the sample destroyed. This difference leaves "England and Wales… isolated internationally as the only countries where DNA of thousands of innocent people can be kept permanently" as the recent enrolled law are not applicable for automatic deletion of DNA yet (GeneWatch, 2008). Some suggest "the DNA of people convicted or arrested for violent or sex offences should remain on the database for life, but that need not be the case for minor offences"[7]. Some commentators would like "to expand the database to cover the whole population and all those who visited the UK, even for a weekend"[8]. Others believe that "the larger the databank ... the greater the value of the databank will be in preventing crime and detecting those responsible for crime"[9]. However, GeneWatch have reported that a DNA database containing samples of acquitted suspects may have increased the size of the database but has not increased the likelihood of solving crimes (GeneWatch, 2008). The number of crime scene samples is significant, not necessarily the number of collected DNA profiles. With the exception of Cold

Case Reviews, many publicised convictions appear to be being obtained because DNA data was already available through prior offences rather than serendipitous sampling.

Some use of DNA samples and data has already been made by genetics researchers. Consent to use these samples does not appear to have been sought from, or given by, the data subjects. People are fighting for their DNA and rights in courts (OUT-LAW, 2004; OUT-LAW 2007). Companies in charge of keeping the samples secure (Forensic Science Service) are fighting to retrieve stolen database information and DNA samples (Gallagher, 2007). Some consideration of DNA transmission across Europe has been made, though in relatively small numbers due to a lack of automation [10].

5. Data Analysis: Newer techniques for DNA matching

The standard test for DNA, SGM+, has increased the odds against the sample coming from another offender 1 in 1 billion. This accuracy, and international acceptance and validation, implies strong suitability for offender identification when matched against samples taken from crime scenes. Here, DNA evidence is used to support or refute an existing hypothesis regarding a suspect. Newer approaches have emerged along two particular lines: 1) searching for characteristics in the DNA samples likely to be shared by family members, therefore providing lists of likely suspects; 2) amplification with smaller samples than suited for SGM+. Both techniques are controversial, as will be outlined. Other techniques and variations also exist.

Familial DNA analysis

DNA evidence was previously used for confirmation of a suspect – now it is being used to drive the approach to identifying *potential* suspects. Familial DNA flags what might be termed the "genetically guilty": those contained in the database whose DNA characteristics may be shared by close family members, but the DNA profile of a presumed criminal family member is not present. Here, we may draw analogy with diagnosis and prediction in a medical context: "a difference between predictive genetic testing in which the patient is tested for genetic information that may be indicative of future disease and diagnostic testing in which the patient is tested for genetic information that may confirm a diagnosis of an existing disease" (Moor 1997). Family members may share DNA characteristics that may be indicative of potential guilt, but a full match is needed to "diagnose". The family members will be contacted, with the entire family blanketed by suspicion until all family DNA samples have been "volunteered" and analyzed. Recall that voluntarily given samples may be obtained using force, and cannot be easily removed from the database, so this could, amongst other methods, expand the DNA database further. A familial DNA search may be considered expensive, at "£5,000 for a speculative operation" (Smith, 2006) but with privatization of the forensic service providers, such services can act as revenue generators. Some successes are claimed using familial DNA analysis, but there are various concerns over using familial DNA analysis extensively. These relate to particular ethnic groups, with invasion of privacy noted disproportionately for the Hispanic community of

America (Grimm, 2007). The concern is echoed in the UK also: “Certain groups such as young males and ethnic minorities are over-represented on the database, and the Council will be asking whether this potential for bias in law enforcement is acceptable”[11].

The DNA database annual report suggests that 7 categories of ethnic appearance are associated to the data (and separated by gender); age profile information provides numbers upwards from the 15-24 age group, though records are also included for children under 10. Fortunately: “there were no records on the NDNAD for persons under 10 years of age where the sample has been taken without the consent of a parent or legal guardian” (Home Office, 2007). It would seem a short leap to collecting such data at birth and to use this for research into genetic identification of potential criminals: “You could argue, the younger the better. Criminologists say some people will grow out of crime; others won't. We have to find who are possibly going to be the biggest threat to Society” [12]. The report on “A surveillance society?” (House of Commons, 2008) identifies that “DNA scene-to-scene matches help identify patterns of criminal behaviour that may help solve past, existing and ***future*** crimes”. Howsoever considered, such data may be vulnerable to future misuse, for the racial profiling of suspects for instance, or for future restrictions on civil liberties. So, as posited in the title of this paper, the continued collection of such data could allow for crimes that haven’t yet been committed to be solved – even predicted – and even allow future prosecutions for crimes that haven’t even been invented yet - through retrospective or retroactive (ex post facto) laws enacted by less liberal governments, where parliamentary supremacy exists (e.g. as in the UK), or with whom these data might be shared.

Low Copy Number (LCN) DNA analysis

LCN DNA analysis involves the amplification of smaller samples of DNA than would be used with the standard SGM+ test. This introduces risks of misidentification due to contamination. In proceedings of R v Hoey, (NICC49, 2007) the use of LCN DNA analysis was explored in charges against Sean Hoey relating to a variety of bomb and mortar attacks and the Omagh car bomb on 15 August 1998 that killed twenty-nine and injured hundreds. Evidence collection had not adequately avoided contamination – future use of LCN analysis could not have been expected – and there were other questions over transit of evidence. Concerns raised by the defence experts included that “there is no validation other than the assertion by Drs Gill and Whitaker that two published journal papers they had written amounted in effect to peer review and thereby the necessary validation”. The judge was “not satisfied that the publishing of two journal articles describing a process invented by the authors can be regarded without more as having "validated" that process for the purpose of its being confidently used for evidential purposes”. The LCN approach was criticized for reproducibility: it could be possible to obtain a third contradictory result from smaller parts of the same sample, where the first two results might agree: “Thus the normal approach used in the United Kingdom had unintentionally been demonstrated by its own proponents to be potentially (and in that particular instance actually) misleading”. In the Omagh case, also, LCN analysis produced a Nottingham schoolboy as a

partial match. LCN analysis also suggested that Madeleine McCann had been transported in a hire car rented by her parents after her disappearance (Woodsand and Foggo, 2008).

The Home Office report on the Science of Low Template DNA Analysis (LTDNA) (Home Office, 2008) appears to agree that "This process should then be repeated enough times to obtain a statistically robust measure of the reproducibility of the system". The report identifies that three different providers of LTDNA analysis employ different methods in analysis, and found difficulties obtaining validation data regarding the LCN approach, though they believed that the data, "part of which is from a previous publication and some of which appears as an in-house study" was able to "represent a validation of LCN DNA analyses". Mention is made of a need to provide information regarding the statistical robustness of matching, and the report identifies health warnings that should be applied when results of LTDNA analysis are used in courts:

> *"that the nature of the original starting material is unknown; that the time at which the DNA was transferred cannot be inferred; and that the opportunity for secondary transfer is increased in comparison to standard DNA profiling".*

6. Data Protection Act and Codes of Ethics.

Would exemption from the principles of the DPA for NDNAD, in the interests of safeguarding national security, also automatically provide an exemption for software developers and other computing professionals from the codes of practice and ethics of their professional societies? In constructing or maintaining such a database, and safeguarding the data therein, one would suggest not. This goes further to questions regarding whether these professionals are properly knowledgeable about risks associating to the use of these data, and whether they are providing sufficient safeguards for the privacy of others. We can suggest a potential conflict between how such professionals ought to act, and current NDNAD practices. We relate the 8 principles of the DPA, the ACM Code of Ethics and Professional Conduct (ACM, 1992), and the foregoing discussion in the table below. We could easily have selected other more or less specific codes, including the ACM's Software Engineering Code of Ethics and Professional Practice, but significantly, this Code identifies:

> ***Know and respect existing laws pertaining to professional work:*** *obey existing local, state, province, national, and international laws unless there is a compelling ethical basis not to do so [..] sometimes existing laws and rules may be immoral or inappropriate and, therefore, must be challenged. Violation of a law or regulation may be ethical when that law or rule has inadequate moral basis or when it conflicts with another law judged to be more important. If one decides to violate a law or rule because it is viewed as unethical, or for any other reason, one must fully accept responsibility for one's actions and for the consequences.*

Below we introduce further facts and figures and make suggestions regarding handling such data.

ACM Code of Ethics and Professional Conduct	Analysis, Facts and Figures	Further suggestions
Principle 1: Fairly and lawfully processed		
1.1- protect fundamental human rights and respect the diversity of all cultures; minimize negative consequences of computing systems 1.2 - report any signs of system dangers that might result in personal or social damage. 1.4 – take action not to discriminate on the basis of race, sex, religion, age, disability, national origin, or other such factors 1.7 - Respect the privacy of others 2.5 - Give comprehensive and thorough evaluations of computer systems and their impacts, including analysis of possible risks	Consider LCN DNA analysis and Familial DNA analysis. Estimated 200,300 individuals who have been acquitted; 139,463 neither charged nor cautioned (Home Office, 2007). There are more than a million less than 18 – 44 under 10- many of them never convicted (Home Office, 2007; Hope, 2008b). Concerns over impacts on specific ethnic groups.	Do not retain data regarding sex, gender, age or ethnicity since matching of profiles should be sufficient.
Principle 2: Processed for specified purposes		
1.7 - personal information gathered for a specific purpose not be used for other purposes without consent of the individual(s) 3.5 - Articulate and support policies that protect the dignity of users and others affected by a computing system.	DNA also used being used for scientific (genetic) research: ten agreed requests for access to the database to assisting the research and development of forensic service providers for "R & D papers, for future use in cases, not specific investigations", amongst others (Home Office, 2007). Data subjects not informed and have no control over it	Provide a set of strictly specified and defined purposes that relate only to crime. Offer opt-in for further uses, for ALL samples and profiles. UK has BioBanks for voluntary provision of DNA samples for research proposes.
Principle 3: Adequate, relevant and not excessive		
1.7 Respect the privacy of others; only the necessary amount of personal information is collected in a system. 3.5 - Articulate and support policies that protect the dignity of users and others affected by a computing system.	National DNA database Annual Report 2005-6 refers to "maximizing sampling opportunities". Need to retain samples as well as profiles? Is such a large scale collection required if only 0.36% success rate? What is the need for ethnic profiles if DNA evidence is so robust and accurate?	As Principle 1. Also, specify when it is necessary to keep DNA samples since computerized profiles should be sufficient for suspect matching. No strategies by corporate entities for "maximizing sampling opportunities".

Principle 4: Accurate and up to date		
1.7- ensure the accuracy of data, as well as protecting it from unauthorized access or accidental disclosure	Questions over accuracy of results in relation to LCN DNA analysis and false identification. Also 12% duplication (Home Office, 2007); over 500,000 false names (BBC, 2007b); 100,000 erroneous records (Ballard, 2007)	Ensure approaches used to derive match profiles are technically and scientifically validated. Remove inaccurate data.
Principle 5: Not kept longer than necessary		
1.7 retention and disposal periods clearly defined and enforced	Kept in perpetuity – especially if sample volunteered (see below).	Specify reasonable times given specific provisions.
Principle 6: Processed in line with rights of data subjects		
1.7- allow individuals to review their records and correct inaccuracies 1.8- respect all obligations of confidentiality to employers, clients, and users unless discharged from such obligations by requirements of the law or other principles of this Code	Data removed from database only following substantial appeals, potentially to the Grand Chamber of the European Court of Human Rights. Different law applying to Scotland.	Consent required for continued retention of most current, non-criminal, profiles. Further consent needed for any research purposes. Individuals not charged with crimes should be able to have their data removed from the database and samples destroyed after a specified time, and not have to fight for this if in England or Wales.
Principle 7: Secured against loss or accidental destruction		
1.2- avoid harm to others such as undesirable loss of information (One way to avoid unintentional harm is to carefully consider potential impacts on all those affected by decisions made during design and implementation). 1.7- ensure the accuracy of data, as well as protecting it from unauthorized access or accidental disclosure	Previous FSS employers obtained access to DNA information (Gallagher, 2007); disc containing 2,000 profiles "lost" for a year (Gaedham & Clement, 2008; Watt, 2008)	Security provisions as would be expected for other data under the DPA, and as recommended by the UK's Information Commissioner. More robust supervision of third party forensics providers and audits of activities
Principle 8: Not transferred to other non-EEA countries without adequate protection		
	No secure framework for sharing within 27 European countries.	Specify adequate protections that must exist in relation to other data protection principles outlined here.

7. Conclusion

In this paper, we have discussed some of the issues that arise due to certain kinds of exemptions existing in the UK's Data Protection Act (DPA). We have considered these with respect to the world's largest collection of DNA data, and identified how certain exemptions may cause difficulties for individuals. In addition, we considered how various media reports demonstrate wider lack of adherence to the 8 principles of DPA. Wider understanding of issues relating to DNA, such as travelling, and to the accuracy of the various techniques beyond the widely reported gold standard SGM+ test may enhance the nature of the current debate. As noted, we are broadly in favour of the collection of DNA evidence relating to crimes, understand its specific value in solving crimes of a certain nature, and believe in the scientifically well-grounded approaches to the analysis when used to corroborate other evidence, with associated statistics over accuracy of robust testing. However, with the role of DNA in crime detection changing, protection of such data seems to have eroded, and it would be a concern that such erosion is apparent for other systems similarly DPA-exempted. We, as computing professionals, are concerned generally about protection for data, particularly in large-scale distributed computing systems, and consider this particularly important when the data are gathered and maintained by governments who impose laws over how such data shall be gathered and maintained by others. Further, we would expect that computing professionals would need to abide by data protection and privacy principles as enshrined in codes of conduct, tending here to imply an exemption to exemptions. Such considerations should be foremost when newer technologies are being recommended for uptake, with DPA principle 8 vital when promoting Government use of geographically-distributed systems such as the "G-Cloud" (DCMS, 2009). It would be an expectation that full adherence to the Data Protection Act would exist from the outset for any such data, with specific instances of lack of adherence being documented. This would further remove the need for any future legislation formed to plug a supposed gap, which is likely to introduce its own potential for exploitation. This suggestion is backed, in part, by the report on "A surveillance society": many of the recommendations echo the 8 principles already enshrined in the Data Protection Act - *safeguarding personal information; obtaining consent for collecting and processing data; hold information only as long as is necessary; designed with a focus on security and privacy* - suggesting either that exemptions are already rather more widely applied, or that these principles have been undermined; it would seem wasteful to create new laws because existing ones are not being applied. More generally, development of such technologies with significant potential impacts on privacy place an onus on computing professionals in terms of actions and understanding of increasingly more complex problems and systems. Uses, users, and those impacted by such systems should be well considered, even as far as the nature and veracity of evidential support provided by such systems. In R v Hoey, the judge cited Lord Lowry LCJ in R. v Steenson and others [1986], and this comment bears consideration here:

"Justice 'according to law' demands proper evidence. By that we mean not merely evidence which might be true and to a considerable extent probably is true, but, as the learned trial judge put it, "evidence which is so convincing in truth and manifestly reliable that it reaches the standard of proof beyond reasonable doubt."

8. References

ACM (1992), "ACM Code of Ethics and Professional Conduct", available at: http://www.acm.org/about/code-of-ethics (accessed 01.2008).

ACM/IEEE-CS (1999), "Software Engineering Code of Ethics and Professional Practice" available at: http://www.acm.org/about/se-code#fullhttp://www.acm.org/about/se-code#full (accessed 10.2007).

Austin, J.J. Smith, A.B. and Thomas, R.H. (1999), "Palaeontology in a molecular world: the search for authentic ancient DNA", *Journal of Trends in Ecology & Evolution*, Vol. 12 No.2, pp. 303-306.

Ballard, M. (2007), "100,000 'erroneous' records on DNA database", *The Register*, 17 May 2007.

BBC (2002), "Police can keep suspects' DNA", 12 September 2002.

BBC (2007a), "All UK 'must be on DNA database'", 5 September 2007 .

BBC (2007b), "DNA database call prompts concern", 5 September 2007.

BCS (2006), "Brithish Computer sociwties code of conduct", available at: http://www.bcs.org/upload/pdf/conduct.pdf (accessed 11.2007).

Bergström, L. (2000), "The concept of ownership". *Who Owns Our Genes? Proceedings of an international conference, October 1999, Tallin, Estonia*, The Nordic Committee on Bioethics, 2000.

Chen, J.V. and Park Y. (2005), "The role of control and other factors in the electronic surveillance workplace", *Journal of Information, Communication and Ethics in Society*,Vol.3 No.2, pp. 79 – 91.

Contractor, N. (2007), "From Disasters to WoW: Enabling Communities with Cyberinfrastructure", Australian Partnership for Sustainable Repositories, available at: http://hdl.handle.net/1885/46949 (accessed 07.2009).

DCMS (2009), "*Digital Britain: Final Report*", Department for Culture, Media and Sport and Department for Business, Innovation and Skills, available at: http://www.culture.gov.uk/images/publications/digitalbritain-finalreport-jun09.doc (accessed 07.2009)

ECHR (2008), "S. and Marper v. The United Kingdom" 30563/04, ECHR 1581, available at: http://www.bailii.org/eu/cases/ECHR/2008/1581.html (accessed 1.2009).

Frod. R. (2008), "MPs fear ID cards could be used for spying", *The Times*, 9 June 2008.

Gallagher, I. (2007), "Five civil servants suspended over 'DNA espionage'", *The Daily Mail*, 31 March 2007.

Gardham, D. and Clement, J. (2008), “Inquiry ordered into DNA disc blunder”, *The Telegraph*, 26 February 2008.
GeneWatch (2008), “The database in Scotland”, available at: http://www.genewatch.org/sub-539489 (accessed 06.2008).
GeneWatch (Feb. 2007), “Police Retention of DNA”, available at: http://www.genewatch.org/uploads/f03c6d66a9b354535738483c1c3d49e4/Councillorsbrief07_2.pdf (access data: 10. 2007)
Gill, P. (2001), “Application of Low Copy Number DNA Profiling”. *Croatian Medical Journal*, Vol. 42, pp. 229-232.
Gotterbarn, D. (1997), “Software engineering: A new professionalism”, *The Responsible Software Engineer: Selected Readings in IT Professionalism*. Springer: London, pp. 21–31.
Gotterbarn, D. (2007), “Ethical Decisions: using the back of the envelope“, *Proceedings of the Ninth ETHICOMP International Conference*, 27-29 March, Tokyo, Japan, pp. 226-235.
Grimm, D. J. (2007), “The demographics of genetic surveillance**:** familial DNA testing and the Hispanic community”**,** *Columbia Law Rev*, Vol.107, pp. 1164–1194.
Home Office (2007), “The National DNA Database Annual Report 2005-2006”, available at: http://www.homeoffice.gov.uk/documents/DNA-report2005-06.pdf?view=Binary (accessed 06.2008).
Home Office (2008), “A Review of the Science of Low Template DNA Analysis- Executive Summary”, available at: http://police.homeoffice.gov.uk/publications/operational-policing/Review_of_Low_Template_DNA_1.pdf?view=Binary (accessed 06.2008).
Hope, C. (2008a), “One million children on DNA database”, *The Telegraph*, 21 March 2008.
Hope, C. (2008b), “5,000 children a month added to DNA database”, *The Telegraph*, 7 April 2008.
House of Commons Home Affairs Committee (2008), *“A Surveillance Society?”*, available at: http://www.parliament.the-stationery-office.co.uk/pa/cm200708/cmselect/cmhaff/58/58i.pdf (assessed 06.2008).
Hoven, J. and Manders-Huits, N. (2008), “The Person as Risk, The Person at Risk”, *Proceedings of the Tenth ETHICOMP International Conference* 24-26 September, Mantua, Italy, pp. 408-414.
Howley, R. Rogerson, S. Fairweather, B. Pratchett, L. (2005), “The Data Protection Decade 1995-2005“,Proceedings *of the Eight International Conference*, 12-15 September, Linkoping, Sweden.
Human Genetics Commission (2008), “A Citizen's inquiry into the forensic use of DNA and the National DNA datavase” available at; http://www.hgc.gov.uk/UploadDocs/DocPub/Document/Summary.pdf (accessed 12.2008).
Jho, A. (2005), “Scientist calls for world DNA database”, *The Guardian*, 11 April 2005.
Johnson, S. (2008), “DNA database plans for children who ‘could become criminals’”, *The Telegraph*, 18 March 2008.

Johnston, P. Waterfield, B. (2007), “DNA data deal ‘will create Big Brother Europe’”, *The Telegraph*, 18 February 2007.
Ladd, C, Adamowicz, M.S., Bourke, M.T., Scherczinger, C.A., Lee, H.C. (1999), “A systematic analysis of secondary DNA transfer”, *Journal of Forensic Sciences*, Vol.44 No.6, pp. 1270-1272.
Litman, J. (2000), “Information Privacy/Information Property”, Stanford Law Review Vol. 52 No. 5.
Lockton, V. And Rosenberg, R. (2005), “Technologies of Surveillance: Evolution and Future Impact”, Proceedings *of the Eight International Conference*, 12-15 September, Linkoping, Sweden.
Lowe, A. Murray, C. Whitaker, J. Tully, G. and Gill, P. (2002), “The propensity of individuals to deposit DNA and secondary transfer of low level DNA from individuals to inert surfaces”, *Forensic Science International*, Vol.129 No.1, pp. 25-34.
Lydall, R. (2007), “DNA kits to trace spitting passengers”, *This Is London*, 31 May 2008
McGraw, D.K. (2004), “A social contract theory critique of professional codes of ethics”, *Journal of Information, Communication and Ethics in Society*, Vol.2 No.4, pp. 235-243.
Moor, J (1997), “Towards a theory of privacy in the information age”. *ACM SIGCAS Computers and Society*, Vol. 27 No. 3, pp27-32
NICC49 (2007), *“The queen - v – Sean Hoey”, Crown Court for Northern Ireland Decisions*, available at: http://www.bailii.org/nie/cases/NICC/2007/49.html (accessed 06.2008).
NuffieldBioEthics.org (2006), “Ethical question over police use of DNA” available at: http://www.nuffieldbioethics.org/go/ourwork/bioinformationuse/pressrelease_401.html (accessed 06.2008).
O’Neill. S, (2008),“DNA database under threat from European court, warns police chief”, *The Times*, 7 June 2008.
OUT-LAW (2004), “Police to retain DNA from acquitted suspects”, available at:http://www.out-law.com/page-4740 (assessed 01.2008).
OUT-LAW (2007), “Police will share data across Europe against privacy chief's advice”, available at: http://www.out-law.com/default.aspx?page=8148 (accessed at 10.2007).
OUT-LAW (2008a), “Only those convicted should be on DNa database, says panel”, available at: http://www.out-law.com/page-9316 (accessed 12.2008).
OUT-LAW (2008b), “Average privacy policy takes 10 minutes to read, research finds”, available at: http://www.out-law.com/default.aspx?page=9490 (accessed 12.2008).
OUT-LAW (2008c), “Lords demand amendment to help the innocent get DNA off database”, available at: http://www.out-law.com/page-9564 (accessed 12.2008).
Palm, E. (2008), “Information Security – Security for whom and why? An ethical analysis of conditions for morally defensible IS”, Proceedings *of the Tenth ETHICOMP International Conference*, 24-26 September, Mantua, Italy, pp. 632- 639.
Ryan, J. (2007), “European Union committee- eighteenth report of session 2006-2007 – Prum: effective weapon against terrorism and crime?” available at:

http://www.parliament.uk/documents/upload/LetGovRes2PRUM170507dew.pdf (accessed 01.2008)
Smith, D.J. (2006), "Secret Weapon", *The Times*, 15 October 2006
USGAO (2006), "*Terrorist Watchlist Screening: Efforts to Help Reduce Adverse Effects on the Public*". United States Government Accountability Office.
Vuokko, R. (2008), "Surveillance at workplace and at home: Social issues in transforming care work with mobile technology", *Journal of Information, Communication and Ethics in Society*, Vol.6 No.1, pp. 60 – 75.
Ward, C. (2004), "Privacy and human rights - 1984 revisited or simply the pursuit of a safer society? ", *Proceedings of the Seventh ETHICOMP International Conference*, 14-16 April, Syros, Greece.
Watt. N, (2008), "CPS admits disc of suspects' DNA was 'mislaid' for a year", *The Guardian*, 20 February 2008.
Woodsand, R. and Foggo, D. (2008), "Should Britain have a compulsory DNA database?", available at: http://www.timesonline.co.uk/tol/news/uk/crime/article3423450.ece (accessed 05.2008).

---

[1] Professor Sir Alec Jeffreys, inventor of DNA fingerprinting, according to the BBC: (BBC, 2002)
[2] Professor Sir Alec Jeffreys, again, according to the Guardian: (Jho A, 2005)
[3] Home Office Minister Tony McNulty, according to the BBC: (BBC, 2007a)
[4] According to the Office of National Statistics, mid-2006 the UK population was 60,587,000. "A surveillance society" suggests 4,264,251 individuals represents 5.2% of the population – this would suggest the current UK population stands at around 80 million. The ONS figure, allowing some growth, would put the proportion closer to 7%. The 4 million figure is cited elsewhere also.
[5] Concern expressed by the Commons Home Affairs Select Committee according to the Times: (Ford R, 2008)
[6] "Because of the commercialisation of forensic science provision we are told that some police forces believe that the advice given may be commercially driven". Review of Low Template DNA Analysis (Home Office, 2008)
[7] Tony Lake, Chief Constable of Lincolnshire Police and chairman of the DNA board, quoted in: (BBC, 2007a)
[8] Lord Justice Sedley, quoted in: (BBC, 2007a)
[9] Lord Justice Woold, quoted in: (BBC, 2002)
[10] Government response to House of Lords report: "Prum: an effective weapon against terrorism and crime?" (Ryan J, 2007)
[11] Professor Sir Bob Hepple QC, Chairman of the Nuffield Council on Bioethics quoted in: (NuffieldBioEthics.org, 2006)
[12] Mr Pugh, Scotland Yard's director of forensic services, quoted in: (Johnson S, 2008)