Statutory Frameworks for Regulating Information Flows: Drawing Lessons for the DNA Data Banks from other Government Data Systems

David Lazer and
Viktor Mayer-Schönberger

10011101

The above bit string encodes personal information about one of the authors of this essay. Of course, without rules to decode the bit string, it is impossible to say whether it is genetic information, weight, age, fingerprint, religion, etc. Layered on top of that technical decoding process is a social decoding process – how sensitive is this information? How useful is it to the government for various purposes? The objective of this paper is to offer some key lessons for the regulation of genetic information collected by the state for law enforcement purposes. In the first part of the paper, we discuss two fundamental principles of informational privacy theory. Utilizing these, we then examine the statutory regimes that have emerged for the regulation of the information that the government collects in three different domains – fingerprints, department of motor vehicle (DMV) records, and tax records. In the third and final part we use the results of our analysis as analogies for similar regulatory challenges in the regulation of genetic information collected for law enforcement purposes, and make some tentative recommendations.

Data collection about individuals is necessary for the operation of the modern state.[1] By the same token, the governmental collection of personal information has raised justified concerns among citizens: who within government should be able to access this information? How can and should it be used? With whom within government, and in the public (if anyone) may such personal information be shared, and under what circumstances? The origins of these concerns date back to the beginnings of democracy, and have grown in our information collecting bureaucratic welfare state. The rise of digital computing in the 1960s and plans for large governmentally controlled data banks in the 1970s sparked a wildfire of strong and widespread public concern.[2] Reports on the potential dangers associated with the individual's loss of control of her personal information, like Arthur Miller's *The Assault on Privacy,*[3] became bestsellers.

Legislatures in the US as well as in Europe reacted to public sentiment with the enactment of privacy and data protection statutes, from Stockholm to Washington, from Paris to Bonn. Both legislative activity

David Lazer, Ph.D., *is Director of the Program on Networked Governance and Associate Professor of Public Policy at Harvard's Kennedy School of Government, where he researches and teaches about technology, information, and governance.*
Viktor Mayer-Schönberger, M.S., *is Associate Professor of Public Policy at Harvard's Kennedy School of Government, where he is researching telecommunications and information infrastructure law and policy.*

and the new challenges of the computer revolution provided a strong impetus for the development of a robust, yet sophisticated theory of informational privacy. While its original roots are found in claims of personal liberty, the prevailing view is that information privacy is a form of informational self-determination that is grounded in human dignity and the right to participation in society.[4] Much of this early theoretical thinking had been injected into the Organization for Economic Cooperation and Development (OECD, a US-initiated organization for transatlantic cooperation) Guidelines on the Protection of Personal Data, a first set of fundamental principles of informational privacy, formally adopted in 1980.[5] Experiences with informational privacy legislation as well as subsequent technological changes, including the stellar ascent of the Internet as a worldwide information network, have prompted some revision of the original thinking on informational privacy, but almost all of the theoretical foundations remain unchanged and have proven valid through more than three decades of rapid technical and statutory change.[6] It is hence only prudent to turn to these principles when examining the privacy dimension of statutory regimes of information collection and sharing. Two of these principles (context and purpose) are particularly pertinent because of their foundational nature and obvious applicability to the statutory regimes at hand. We discuss each of them in turn.

Context

The sensitivity of information is a malleable thing. What is a completely innocuous piece of information in one context may be terribly harmful in another. One may be willing to disclose personal information when visiting one's physician, but not want one's employer, bank, or the media to know about it. Moreover, individuals may differ in what type of information they consider sensitive. Some personal information may be very important and sensitive to one person, while another person may not much care whether this particular piece of information about herself becomes public or not. Sensitivity may also change over time – what few consider sensitive information when they are young they may not want to have made public when they are older. Social views, too, may change. For example, beliefs about the information contained in fingerprints have varied enormously over the years, where, as Simon Cole convincingly argues, for the first half of the century many viewed fingerprints as incorporating great information about an individual (e.g., race, "criminal predispositions"), whereas now few imbue them with such informational power.[7] Equally important, bits do not wear out with use, and may become more potent over time. Genetic information is an extreme example, where data that are not interpretable today might yield great insights tomorrow. More generally, data about individuals become more powerful in the context of other data about those individuals. The sensitivity of a particular piece of personal information therefore depends on the context as perceived by the person the information relates to.

The context of personal information makes it difficult for statutes to define what types of information warrant special protection, or which situations of information sharing need particular regulation. In general, privacy legislation tends to address this challenge through a combination of two measures: first, by extending protection in principle to *all* personal information rather than affording protection only to certain categories of information not shared. Second, such protection is then limited by a requirement that the individual has to have a *reasonable* interest in having a particular piece of information. This leads to an important balancing between the interests of the citizen and the interests of society, and ensures that privacy concerns of an individual do not necessarily override more important interests of society at large. The principle of context of personal information does not preclude statutes from granting a higher level of protection to certain types of personal information. It mandates, however, that no piece of personal information is *ex ante* outside the scope of privacy protection.

Purpose

While the principle of "context" relates to the particular setting of personal information, "purpose" links to specific intents of the use of personal information. Purpose and context are closely related, yet highlight different aspects of informational privacy. Like context, purpose plays an important role in how individuals decide to share their personal information or not. A person may not be willing to share medical information even with her doctor if it is just for the doctor's personal curiosity. If, however, the doctor explains that the personal information will be used in her treatment, the person may want to share her information with her doctor, and perhaps even other medical specialists. Individuals can only make an informed decision whether to share their personal information with somebody else if they know the exact purpose the information will be used for. On a societal level, citizens can only agree to let government have access to personal information if the purpose of that access is both clear and acceptable.

Digitized information can easily be separated from its original context, and injected in a different context. Moreover, technological capabilities of information processing make it enticing to re-purpose information – that is to use it for other purposes than originally

intended. Why not run a victim's fingerprints against a database of fingerprints from yet unresolved other cases? Why not use motor vehicle registration data to track people's movements around the country over time? More and more frequently, those hard pressed to solve a particular problem – a crime for example – or to protect us from a menace – like terrorism or a flu pandemic – may desire to do what is technically possible in combining and sharing personal information of the citizenry.

While the digitization of information has made it much easier to re-purpose and re-contextualize information, the underlying problem has been with us for a long time. Take as an extreme example the effort in the 1930s by the Netherlands to redesign their population information systems. The clear purpose of this endeavor was to improve administrative efficiency. However, part of the data that they collected, for innocent reasons, was each citizen's religious affiliation. Catastrophically, these data systems fell into the hands of the Nazis, and, arguably, as a result, Dutch Jews were killed at a much higher rate than any other Jews in Western Europe during the Holocaust.[8] This very small amount of data collected on Dutch citizens (representable by a single bit), benign in one context, was re-purposed in deadly fashion in another context.

The essential governance concern that we focus on is that information collected for legitimate public purposes might be re-used or re-constituted in an "unacceptable" fashion. But what is "unacceptable?"

The principle of purpose has led to a number of enormously useful rules of thumb in the field of informational privacy. It requires that the purpose of use of personal information be made explicit and clear, for example through a precise statutory mandate. Similarly, the purpose cannot be changed retroactively without disobeying the principle of purpose, except if those affected agree, and personal information that is no longer necessary for the intended purpose must be deleted. It also leads to the conclusion that statutes ought only to require collecting, storing, and sharing that personal information which is necessary to fulfill the purpose. Collecting information just in case it may become useful at some future date or for some future purpose would be contrary to the purpose principle.

Below we discuss the statutory regimes that have been created in three areas where government collects data from citizens (fingerprinting, at the DMV, and in tax data) before we turn to a comparison with the regulation of government DNA data banks. We focus on statutory frameworks rather than regulatory ones, because we would view the general parameters of access to be of sufficient public interest that it should be specified in statutes rather than delegated to administrative agencies. In each case, we examine (1) what information is collected from citizens; (2) who has access to that information; (3) what is the potential of using that information for another purpose.

Fingerprints

The first fingerprint file in the United States was created in 1902 in New York City, to monitor individuals taking Civil Service exams.[9] Currently, fingerprint identification services are coordinated at a national level by the Criminal Justice Information Service (CJIS) division of the FBI.[10] CJIS was founded in 1992, consolidating criminal recordkeeping, crime-statistics, and fingerprint identification initiatives. CJIS is also responsible for the technological initiatives to network identification and fingerprint resources, the National Crime Information Center (NCIC) and Integrated Automated Fingerprint Identification Service (IAFIS), which began working among states in 1999.[11] Fingerprints are used by law enforcement in two ways. First, they are used to identify those with criminal records. The impression taken of the prints of all ten fingers (the ten-print) allows a search against the Criminal History Database, a database of criminal histories that is organized using fingerprints as a unique identifier (see below). This may be done both with individuals who are entering the law enforcement system (e.g., upon arrest), as well as a variety of non-criminal contexts where fingerprints are searched against the Criminal History Database to determine whether the applicant or subject has a criminal history. This latter category of prints is stored in a Non-Criminal Database. Second, fingerprints from crime scenes are stored in an Unmatched Latent Fingerprint (ULF) database, where prints found at a crime scene are searched against the criminal database for matches (hits).

Law enforcement thus maintains three different kinds of fingerprint databases: criminal history databases, non-criminal databases, and ULF databases. We discuss each below.

Criminal History Database

When a person is arrested for a qualifying offense (a felony or serious misdemeanor) or, in some cases, is incarcerated for any reason, that person's fingerprints are taken and a copy of his record is forwarded to the federal government for comparison with prints already on record in both the Criminal History Database and the Unmatched Latent Fingerprint database. If there is a match in the Criminal History Database the submitting agency is informed and the new information is forwarded to the agency with original responsibility for maintaining that record. In Fiscal Year '05, sixty-eight

percent of criminal submissions yielded a match with a previous record.[12]

The federal government requires that fingerprints be taken of felony arrestees. In addition, every state has its own requirements. Thus, for example, in New Jersey anyone arrested for an indictable offense,[13] shoplifting,[14] or prostitution,[15] as well as all prisoners,[16] are fingerprinted. In Oregon, fingerprints are collected from anyone convicted of or arrested for a felony, sex crime, or serious misdemeanor, as well as any violation of the Uniform Controlled Substances Act.[17]

There are now "more than 51 million" criminal records available electronically.[18] The state databases reported a total of 64,282,700 criminal history records on file at the end of 2001, of which 57,437,800 were automated.[19]

The Non-Criminal Database

In addition to its law enforcement uses, the Criminal History Database is also used to screen candidates for various kinds of employment and licensing. This kind of use seems to be on the rise, in part driven by laws passed after September 11, 2001, which mandated background checks of a much wider array of employees with access to sensitive infrastructure.[20]

At the Federal level, fingerprints are taken for criminal history checks on almost all civil servants, as well as the security-related instances above. The states have varying requirements for fingerprinting and criminal history checks. For example, in New Jersey, a wide array of persons are subject to fingerprinting to check for criminal histories, everyone from police and fire people[21] to pawn brokers[22] to investment advisors.[23] Oregon fingerprints a similarly diverse (if different) set of people, from Members of the Bar[24] to podiatrists.[25] In FY 05, there were 9.8 million non-criminal submissions, eleven percent of which yielded a match to the criminal history database.[26]

Unidentified Latent Prints (ULF)

The FBI offers a latent print matching service over IAFIS. Prints from crime scenes are matched against the Criminal History Database. Unmatched prints are retained in a separate database. As new prints come into the Criminal History Database, these are also checked against the ULF. To date the FBI made 1,301 identifications using the IAFIS. The ULF included 94,000 unmatched latent prints in early 2006.

Context and Purpose Principles

How do existing fingerprint statutes comply with the context and purpose principles discussed above? With respect to the Criminal History Database, federal regulations allow the states to make their own laws regarding the dissemination of criminal history information (of which fingerprints are part) for purposes other than law enforcement. This has resulted in wide variation in what information is available to whom and for what purposes, ranging from some states where any citizen can obtain access to criminal history records including fingerprints, to those states where it is a criminal offense to release such records for unauthorized purposes.[27] Obviously, the former case violates these principles, and the latter aligns fairly well.[28]

With respect to the non-criminal database, there is an obvious potential to re-use these data – to search them against crime scene data. This is theoretically possible because it is federal policy to retain in a non-criminal database all fingerprints submitted for criminal history checks directly, as well as those submitted by states. Again, state policy is critical, because the FBI complies with state requests to destroy or return fingerprint records. For example, in Oregon, most of the statutes that require fingerprints be collected for non-criminal purposes mandate that fingerprints submitted to the FBI be returned or destroyed after the criminal history check has been run.[29] New Jersey statutes are more ambiguous. Some allow the retention of non-criminal fingerprints and from time to time to check them again against the Criminal History Database, while other statutes do not mention retention. None of the statutes specifically authorize or prohibit checking non-criminal prints against latent prints of unsolved crimes.

The retention of these fingerprints (for one time checks) and silence of statutes on potential use for criminal investigations is inconsistent with the privacy principles outlined above – why should a public insurance adjuster in New Jersey be under enhanced lifetime surveillance just because of her initial application to that profession? We could find no statute authorizing or prohibiting the checking of latent prints against the Non-Criminal Database – except to the extent that non-criminal prints that are destroyed after checking against the criminal database obviously cannot be reused in the future for other purposes.[30]

Department of Motor Vehicle Data

Personal information is collected by the government from citizens who are licensed to drive a motor vehicle – such data collected by the DMV may include address, photograph, various physical descriptors, information on vehicular accidents, driving violations, driver's status, medical and disability information, etc.

Context and Purpose Principles

There is significant potential for re-use of DMV data. The murder of the actress Rebecca Schaeffer in 1989

is a tragic case in point. The killer, an obsessed fan, had hired a private detective, who used DMV data to locate the actress's residence.[31] This personal information (address), benign in most contexts, was deadly in this one. This case, and a number of others like it, inspired the passage of a number of state statutes, and, ultimately, in 1994, the passage of the Federal Drivers Privacy Protection Act (DPPA) to regulate access to a driver's license record.[32] Information in a driver's license record is divided into three categories: personal information, highly restricted personal information, and other information, where different levels of access are permitted for each category.[33] Personal information is "information that identifies an individual, including an individual's photograph, social security number, driver identification number, name, address (but not the five-digit zip code), telephone number, and medical or disability information." Highly personal information is "an individual's photograph or image, social security number [and] medical or disability information."[34]

DPPA explicitly limits re-use of DMV data for certain government and private sector purposes. There are exceptions for police, judicial, and government agency uses of the information. Government, including law enforcement, may use all DMV information in carrying out its functions.[35] The information is to be disclosed if needed for any judicial or arbitral process carried out by a federal, state, or local court or agency, or pursuant to an order of any such court.[36]

There are more limited exceptions that allow businesses access to DMV data, in particular, to prevent fraud. There is an exception to allow businesses to check information provided to them. Businesses can submit information provided by customers to verify its accuracy. If the information is inaccurate, and the correct information would be used to prevent fraud by the customer, or to pursue a legal remedy or collect on a debt owed by the customer, then the correct information may be released to the business.[37] The information released in this manner may not include the "highly restricted personal information."[38]

Personal information (but not highly restricted personal information) may also be disclosed for use in connection with

> matters of motor vehicle or driver safety and theft; motor vehicle emissions; motor vehicle product alterations, recalls, or advisories; performance monitoring of motor vehicles, motor vehicle parts and dealers; motor vehicle market research activities, including survey research; and removal of non-owner records from the original owner records of motor vehicle manufacturers.[39]

Private toll collectors can request information if they need it for their operations. Towing companies can request information to notify owners of towed and impounded vehicles.

Some commercial uses of the information include the highly restricted personal information. Employers can seek information on those employees whose jobs require a commercial drivers' license. Likewise, insurance companies may obtain records, as can private investigators seeking to use the records for any other permissible purpose. Both of these groups can access the highly restricted personal information as well as the personal information.

The statute also contains some opt-in provisions. Individual drivers may authorize release of their information. States may pass new laws allowing disclosure of the information, as long as "such use is related to the operation of a motor vehicle or public safety."[40]

Originally, the DPPA allowed states to imply a consent to other releases of information. This was changed by an amendment passed on October 9, 1999.[41] The amendment – in line with the purpose principle – required states to obtain an affirmative consent from individuals before compromising the privacy provisions of the DPPA.[42]

DPPA thus substantially narrows the potential for re-use of DMV data, although some of the language of the statute leaves quite broad possibilities for re-use. In particular, the substantial potential re-use of the data for the prevention of fraud is problematic from the perspective of the purpose principle. For example, the Tenth Circuit ruled in 2004 that digital images of drivers were legally released under the DPPA and the relevant Colorado statute to a company that was developing a point-of-sale image-display technology to help prevent fraud.[43] In finding that the sale did not violate the DPPA, the circuit court wrote that the prevention of fraud generally is encompassed within the DPPA's exceptions. Other contemplated uses, such as preventing "insurance and Medicaid fraud; terrorism; underage drinking; drug crimes; government payments fraud; border-jumping…"[44] would all fall under the exceptions and be permissible uses of image data under the DPPA.[45]

Tax Records

Tax return records offer potentially extensive intimate details about an individual – not just about income, but about health information, debts, donations, religious affiliations, etc.[46] Arguably, tax data are more intrusive than genetic data.

There is significant potential for re-use of DMV data. The murder of the actress Rebecca Schaeffer in 1989 is a tragic case in point. The killer, an obsessed fan, had hired a private detective, who used DMV data to locate the actress's residence. This personal information (address), benign in most contexts, was deadly in this one.

Context and Purpose Principles

The history of control of tax data tells a particularly interesting tale of administrative discretion and individual privacy.[47] The first income tax in the US was introduced in 1861. Tax records at this time were accessible to anyone upon request. The income tax was abolished in 1872, and reintroduced briefly in 1894, this time with prohibitions on disclosure. Shortly thereafter, the income tax was declared unconstitutional, and was then reintroduced with the Revenue Act of 1913, which declared tax records to be "public records" and gave the President wide latitude with whom to share information. In practice, this meant sharing information among government agencies, but not with the public. The tax data were re-used by the Nixon Administration to harass political opponents, which resulted in the Tax Reform Act of 1976, which sharply limited re-use of tax data.[48]

Currently, disclosure may be made to officials for tax administration and enforcement purposes, and to the Department of Justice for the purposes of any ongoing criminal investigation or criminal proceeding *related to enforcement of the tax code.*[49] Other law enforcement uses face a higher hurdle: investigators must obtain an ex parte order of a judge or magistrate for access to tax records.[50] In addition, the Secretary may disclose information to the appropriate officials on his own initiative if he believes the tax return information presents evidence of a crime, or evidence of terrorist activities, or in an emergency situation like danger of death or physical injury to any individual.[51] In addition, agents of any federal law enforcement agency who are directly involved in the investigation or response to a terrorist threat or incident may also request information, and share this information with relevant State agencies.[52]

There are also provisions to allow access to tax data for oversight purposes. Congressional committees on Ways and Means, Finance, and the Joint Committee on Taxation are allowed to review tax return information, although information that may identify a taxpayer, directly or indirectly, must be heard in closed session.[53] The President or his designee may request some tax return information, although in response to the abuses of the Nixon era, the President is required to report to Congress quarterly the names of individuals whose return information he has requested and his reasons for doing so.[54] Also, it is a crime for the President or other executive branch member (aside from the Attorney General) to initiate or terminate any tax investigation.[55]

The current tax data regime thus stands in sharp contrast with DMV and (especially) fingerprint data regimes, and aligns most closely with the context and purpose principles discussed above. Re-use of tax data is limited to making the tax system work effectively, and requires (except for the terrorism exception) some compelling showing to a judge for other government uses. This is notable because the potential for reuse of tax data is considerable – for example, freer access to tax data could be a powerful aid to criminal investigations, as well as useful in reducing fraud (perhaps more useful in these respects than DMV records).

Lessons for DNA Data Banks

The first DNA data bank for law enforcement purposes in the US was created by Virginia in 1989; currently all fifty states have DNA databases, which are linked together in the National DNA Index System (NDIS).[56] These data banks include offender data banks and samples from crime scenes. They were created to develop useful leads for investigators by linking offenders to samples from crime scenes, and linking crime scene samples to each other. The criterion for inclusion has become progressively broader, where currently thirty-nine states include all felons in their database, and six states and the federal government include some arrestees and/or indictees.[57] There are currently approximately three million samples in all of the national and state databases that are eligible for national searches.[58]

Similar to what we highlighted in the cases of fingerprint, DMV, and tax records, the purpose principle would require DNA statutory frameworks to restrict the re-purposing of information. To an extent, DNA statutes have already dealt explicitly with certain types of re-use of genetic data. Most states (thirty-nine) authorize the use of samples for refinement of population statistics (e.g., frequency of particular alleles). Such statistics are important in the interpretation of data in particular cases – e.g., in producing an estimate of random match probabilities. Thirty-one states specifically authorize the use of samples for humanitarian purposes and/or identification of remains.

Far fewer states, however, have statutes dealing with other foreseeable uses. Most states do not have statutes that deal with the use of data banks for research. Only eight states have specific prohibitions on use of the data banks for research on predispositions to disease, physical traits, and/or behavioral predispositions, and one (Alabama) expressly authorizes the use of the data bank for medical research. Even fewer states deal with the use of voluntary samples, where two state statutes authorize the inclusion of voluntary samples in the database, six prohibit such inclusion.[59] This is particularly troubling given the thousands of volunteer samples that have been collected in DNA dragnets in the US. The reuse of these samples beyond the particular case for which they were collected would be a clear violation of the purpose principle. In addition, no jurisdiction, to our knowledge, has statutory authorizations of or prohibitions against familial searching – the searching of offender databases for close relatives of the source of a crime scene sample.[60]

What emerges is a highly heterogeneous statutory landscape, not unlike the landscape of fingerprinting statutes, or the history of statutes regulating access to DMV and tax records. Heterogeneity, and the lack of an underlying coherent framework, has the potential of fostering inequalities and even facilitating abuse, as the example of inappropriate access to tax records in the 1970s highlights. How could this heterogeneity be overcome in the area of DNA data banks – at least as it relates to the purpose principle? We suggest two "rules of thumb" that could guide lawmakers when considering legislation regarding DNA data banks on all levels – federal as well as state.

Hardwired Constraints

The most powerful constraint on re-purposing data is to discard informational elements that are not necessary for the core purpose. What is not needed should be deleted. Some of the state statutes for fingerprinting of non-criminals already require destruction of the records of those fingerprints. Such destruction limits the use of those fingerprints to their core purpose: a one-time verification that a particular individual does not have a criminal record. On the other hand, the retention of the physical DNA samples, which are typically not necessary for matching of the digitized profiles of offenders to crime scene evidence, invites re-purposing at a later stage.

A hard-wired constraint, like the destruction of the physical samples, largely eliminates the possibility of re-examining samples, on either an *ad hoc* or systematic basis. Only one state has a provision to destroy samples once they have been typed (Wisconsin).[61]

All information looks the same in the computer – strings of 1s and 0s. How that information looks to human eyes once it is decoded, however, depends critically on context and contemporary sensibilities.

Such hardwired constraints have the advantage of eliminating the possibility of a rogue agent within government using the information, or of the government's improper sharing of the information with a third party. These constraints also slow down any efforts to radically expand the informational regime. Thus, a short term change in political regime cannot result in an immediate change in the regulation of information.

Hardwired constraints come with important downsides. First, it is often impossible to completely divorce the informational elements for the use and reuse of information. For example, in the case of DNA samples, retention facilitates quality control procedures that insure the integrity of the overall data bank system.[62] Such procedures certainly support the core purpose of the collected data. Further, it is conceivable that additional testing on offender samples would be useful under certain scenarios to confirm a match. Second, to the extent that one wishes to allow future knowledge and sensibilities to redraw that line, hardwiring the system would be undesirable. For example, if it would be useful to incorporate other genetic information into the computerized database, having samples available would make this switch much easier. Where the tradeoffs between potential for intrusion and costs of destroying information are too great, more flexible rules might be desirable, coupled with oversight mechanisms.

Administrative Speed Bumps

Short of hardwiring, statutes may embed procedural hurdles within the processes of information sharing that must be cleared before information is reused. Such speed bumps necessarily create inefficiencies, but offer a degree of protection from a large scale re-purposing of individuals' personal information that would occur if their data were just a mouse click away.[63] Tax data offer a good example of this, requiring a court order for access even for law enforcement purposes.

In the case of the use of DNA, one might imagine needing to clear certain thresholds before conducting a familial search of a database in a particular case. For example, the UK has adopted an informal policy of only using familial searching for serious crimes.[64] One could imagine that such a policy could be implemented

more formally, with statutory guidance as to when such searches are permissible.

To be sure, such speed bumps potentially suffer from two shortcomings: they may not slow information down enough in certain cases, and they may slow information down too much in other cases. Administrative processes do not incorporate the technological conservatism of hardwired constraints, and, on the other hand, they do not offer the advantages of automatic data integration and sharing.

Conclusion

All information looks the same in the computer – strings of 1s and 0s. How that information looks to human eyes once it is decoded, however, depends critically on context and contemporary sensibilities. For that reason, certain key principles have been developed regarding the regulation of government information. In this paper we have described two key principles: the context and purpose principles, which assert that personal data that have been collected for one purpose and in one particular context should only be re-used for other purposes in cases where there is a strong overriding public interest.

With these two principles in mind, we have examined three statutory frameworks for regulating government information (for fingerprint, DMV, and tax data), drawing lessons for how genetic data collected by the government for law enforcement purposes should be managed. A review of these domains highlights the potential for re-contextualization and repurposing information – sometimes to the clear subversion of the public interest (e.g., in the case of Nixon's abuse of tax data). An examination of these frameworks highlights a number of mechanisms to guard against inappropriate repurposing of personal data, including the hardwiring of constraints into the data system (e.g., by discarding of data), and the incorporation of administrative processes that allow a balancing of individual interests in privacy and collective interests in the use of personal data.

Acknowledgements

Thanks go to Caleb Donaldson, for outstanding research support of this project while he was a research fellow at the American Society of Law, Medicine & Ethics (ASLME). This paper also benefited from feedback from Frederick Bieber and David Winickoff, as well as from discussions with participants in the four workshops that ASLME conducted on DNA Fingerprinting and Civil Liberties. This paper was supported by NIH grant # 1RO1 HG2836-01 and NSF grant II0131923. Errors herein are, of course, the authors' alone.

References

1. V. Mayer-Schönberger and D. Lazer, "The Governing of Government Information," in V. Mayer-Schönberger and D. Lazer, eds., *From Egov to Igov: Governance in the 21st Century,* book manuscript, forthcoming, 2006.
2. V. Mayer-Schönberger, "Generational Development of Data Protection in Europe," in P. Agre and M. Rotenberg, eds., *Technology and Privacy: The New Landscape* (Cambridge, MA: MIT Press, 1997): 219-241.
3. A. Miller, *The Assault on Privacy* (Ann Arbor, MI: University of Michigan Press, 1971).
4. P. M. Schwartz, "Privacy and Participation: Personal Information and Public Sector Regulation in the United States," *Iowa Law Review* 80 (1995): 553-618; for a discussion of the (theoretically limited) debate in the United States, see F. H. Cate, *Privacy in the Information Age* (Washington, DC: Brookings, 1997): 19-30.
5. Organization for Economic Cooperation and Development, "Guidelines on the Protection of Privacy and Transborder Flow of Personal Data," 1980, *available at* <http://www.oecd.org/document/18/0,2340,en_2649_34255_1815186_1_1_1_1,00.html> (last visited March 8, 2006); see also V. Mayer-Schönberger, "Strands of Privacy: DNA Databases, Informational Privacy, and the OECD Guidelines," in D. Lazer, ed., *DNA and the Criminal Justice System: The Technology of Justice* (Cambridge, MA: MIT Press, 2004): 225-243.
6. For an assessment in light of European Union legislative activities see V. Mayer-Schönberger, "Operator, Please Give Me Information: The European Union Directive on Data Protection in Telecommunications," in S. E. Gillett and I. Vogelsang, eds., *Competition, Regulation, and Convergence* (Mahwah, NJ: Lawrence Erlbaum, 1999): 121-136.
7. S. Cole, "Fingerprint Identification and the Criminal Justice System: Historical Lessons for the DNA Debate," in D. Lazer, ed., *DNA and the Criminal Justice System: The Technology of Justice* (Cambridge, MA: MIT Press, 2004): 63-90.
8. W. Seltzer and M. Anderson, "The Dark Side of Numbers: The Role of Population Data Systems in Human Rights Abuses," *Social Research* 68 (2001): 481-513.
9. S. Cole, *Suspect Identities* (Cambridge, MA: Harvard University Press, 2002).
10. This is in keeping with a statutory requirement that the FBI shall "Conduct the acquisition, collection, exchange, classification and preservation of fingerprints and identification records from criminal justice and other governmental agencies, including fingerprints voluntarily submitted by individuals for personal identification purposes…" 28 CFR 0.85 (b).
11. CJIS website, *About CJIS, at* <http://www.fbi.gov/hq/cjisd/about.htm> (last visited March 2, 2006). Bureau of Justice Statistics, *Use and Management of Criminal History Record Information: A Comprehensive Report,* 2001 Update, at 44, note 65.
12. J. Laryle of the FBI CJIS division, *available at* <http://fingerprint.nist.gov/standard/presentations/IAFISoverview_Feb_2005.pdf> (last visited March 9, 2006).
13. N.J. Stat. § 53:1-15 (2005).
14. N.J.S.2C:20-11.
15. N.J.S.2C:34-1.
16. N.J. Stat. § 53:1-14 (2005).
17. ORS § 137.074 (2003); ORS § 181.511 (2003).
18. *Supra* note 12.
19. S. J. Barton, *Survey of State Criminal History Information Systems, 2001,* Criminal Justice Information Policy series, NCJ 200343, (Washington, DC: U.S. Department of Justice, Bureau of Justice Statistics, October 2002): at table 1.
20. For example, the FBI reports that between June 1, 2001, and May 31, 2002, most fingerprints submitted for processing were for non-criminal justice background checks, whereas a similar study in 1993 indicated that only 9 percent of fingerprints submitted for non-criminal purposes. S. J. Barton, *Survey of State Criminal History Information Systems, 2001,* Criminal Justice Information Policy series, NCJ 200343 (Washington, DC: U.S. Department of Justice, Bureau of Justice Statistics, October 2002): at 9 (internal citations omitted).
21. N.J. Stat. § 40A:14-9 (2005).
22. N.J. Stat. § 45:22-3 (2005).
23. N.J. Stat. § 49:3-56 (2005).
24. Ore. Admission Rule 4.15 (2004).
25. ORS § 677.265 (2003).

26. *Supra* note 12.
27. *Supra* note 19, at 10.
28. It is difficult to conceive of how criminal fingerprints might widely be used for other purposes at this time; although the point of the principles is to limit unanticipated future uses of personal data.
29. Some of the statutes do require retention of records in Oregon; others do not touch on the issue of retention.
30. Notably, however, we could find no case where the prosecution used a match against the Non-Criminal Database.
31. B. Loving, "DMV Secrecy: Stalking and Suppression of Speech Rights," *CommLaw Conspectus* 4, 203 (1996). Although, oddly, DPPA still allows access to the DMV data by private detectives.
32. P.L. 103-322, § 300002(a), 108 Stat. 2094 (codified as amended at 18 USCS §§ 2721-25).
33. Other information includes zip code and "information on vehicular accidents, driving violations, and driver's status," and this information is not covered by the DPPA.
34. 18 USCS § 2725 (3) & (4). It is unclear why medical and disability information are listed twice.
35. 18 USCS § 2721 (b) (1).
36. 18 USCS § 2721 (b) (4).
37. 18 USCS § 2721 (b) (3).
38. *Id.*
39. 18 USCS § 2721 (b) (2).
40. 18 USCS § 2721 (b) (14).
41. P.L. 104-294, Title VI, § 604(b)(46), 110 Stat. 3509; Oct. 9, 1999.
42. 106 P.L. 69 § 350 (c) & (d).
43. *Miller v. Image Data LLC,* 91 Fed. Appx. 122 (10th Cir. 2004).
44. *Id.,* at 126, note 6.
45. *Id.,* at 126 (citing 18 USCS § 2721 (b) (1) ("Driver's license information may be sold for "use by any government agency…or any private person or entity acting on behalf of a Federal, State, or local agency in carrying out its functions…" and holding that "the permitted uses of DMV information are not limited to the prevention of financial fraud, but include the present and potential applications of Image Data's True ID (R) technology.")
46. Tax return information is defined to include not only all the information on a tax return itself, but also the content of any written determination or investigation file of the IRS, as well as the existence or nature of any agreement, settlement, or disposition of such investigation. 26 USCS § 6103 (b) (2).
47. This history is drawn from J. J. Darby, "Section IV: Confidentiality and the Law of Taxation," *American Journal of Comparative Law* 46 (1998), at 577.
48. 26 USCS § 6103.
49. 26 USCS § 6103 (h).
50. 26 USCS § 6103 (i) (1) (A).
51. 26 USCS § 6103 (i) (3).
52. 26 USCS § 6103 (i) (7).
53. 26 USCS § 6103 (f) (1).
54. 26 USCS § 6103 (g).
55. 26 USCS § 7217.
56. We will adopt the convention that "database" refers to digitized set of profiles, and "data bank" refers to the system that encompasses both physical samples and profiles. We would view the stored samples (essentially the DNA molecule) as searchable data in the sense that for a given individual in the database, one can specify a search that pulls the physical sample and extracts from it genetic information. One may view the physical samples as analogous to paper records that have been placed in storage. Such records are searchable, but at greater expense than digitized information.
57. All statistics regarding state database statutes are drawn from the grid developed, in part based on consultation with one of the authors, by the American Society of Law, Medicine & Ethics, based on research by S. Axelrad. See Survey of DNA Database Statutes *at* <http://www.aslme.org/dna_04/grid/statute_grid.html> (last visited March 9, 2006).
58. For official FBI statistics, see *NDIS Statistics, at* <http://www.fbi.gov/hq/lab/codis/clickmap.htm> (last visited March 20, 2006). Note that there are more profiles that are ineligible for a national search that are available for state or local searches.
59. The federal statute authorizing the creation of the national system does impose certain policy restrictions on states that participate in the national database e.g., a prohibition on research. Also, the federal statute prohibits searching voluntary samples against the national database, and federal rules currently prohibit familial searching of the national database.
60. To date, familial searching has been used more aggressively in the UK than the US. There has only been one publicized case of familial searching in the US – in the case of the murder of Deborah Sykes, where a near miss in the database led investigators to Willard Brown. R. Willing, "Suspects Get Snared by a Relative's DNA," *USA Today,* June 7, 2005, at 1.
61. Our understanding is that even Wisconsin has not destroyed any samples.
62. Note that it is debatable that acceptable quality control procedures could not be implemented without indefinite sample retention.
63. H. Burkert, "Freedom of Information and Electronic Government," in V. Mayer-Schönberger and D. Lazer, eds., *From Egov to Igov: Governance in the 21st Century,* book manuscript, forthcoming in 2006.
64. R. Williams and P. Johnson, "Inclusiveness, Effectiveness and Intrusiveness: Issues in the Developing Uses of DNA Profiling in Support of Criminal Investigations," *Journal of Law Medicine & Ethics* 33, no. 3 (2005): 545-558.