Fordham Urban Law Journal

Volume 37 | Number 3 Article 5

2010

Databases, Doctrine, and Constitutional Criminal Procedure

Erin Murphy

DATABASES, DOCTRINE & CONSTITUTIONAL CRIMINAL PROCEDURE

Erin Murphy[*]

INTRODUCTION

Around 1964, military information systems employees coined the term "data base" to describe repositories of data accessible by users across time-

[*] Assistant Professor of Law, U.C. Berkeley School of Law. I owe thanks to the Fordham Urban Law Journal for inspiring this Article, and am grateful to the NYU Center on the Administration of Criminal Law for its support of this project. Tracey Maclin also provided incredibly beneficial insight and advice. Lastly, I am indebted to Chris Slobogin, J.J. Prescott, and the members of the Young Criminal Law Professor Workshop at Vanderbilt Law School, as well as to the attendees of the Clason Lecture at the Western New England College of Law, for their helpful comments.

shared computer systems.[1] Today, "database" is a word that requires no definition. In thirty short years, it progressed from two disconnected descriptive nouns to a hazy concept stitched together by a hyphen to a single word instantly understandable to every member of a modern society. Indeed, databases are such an integral component of contemporary life that it is easy to forget just how new they are. Like other currently indispensable technologies—personal computers, cell phones, e-mail—it can be hard to remember just how things worked before they came into existence.

A wide variety of government databases have flourished in the criminal justice field over the past ten to twenty years. Much legal scholarly attention, including some of my own,[2] has been devoted to the impact of these databases on individual privacy, whether as a general normative matter or in relation to specific constitutional doctrines such as the Fourth Amendment. But this Article focuses on something different than general concerns about privacy. Indeed, it takes it as a given that databases affect personal privacy, even while acknowledging that reasonable people disagree about how severe or grave the impact of databases may be.

This Article is a preliminary effort to sketch some of the challenges that large-scale databasing poses to conventional constitutional analysis. In *Herring v. United States,*[3] the Supreme Court engaged in its first head-on confrontation with criminal justice databases in some time. To many scholars, *Herring*'s greatest significance is that it bolsters suspicions that a majority of the Court views the proper application of the exclusionary rule as limited to instances of deliberate malfeasance. Yet *Herring* is a difficult call on the question of malfeasance, given the nature of the constitutional violation claimed. Rather than involve a run of the mill bad call on probable cause or reasonable suspicion, *Herring* dealt with a kind of mistake that increasingly can and does occur in contemporary policing: an error in a computerized database. And, indeed, in their assessments of the nature of the constitutional error, the majority and the dissenters exposed some fundamental problems that arise when claims related to databases are at stake.

This Article takes the fault lines exposed by *Herring* as a point of departure for considering these issues more generally. Specifically, this Article questions whether the practice of databasing comports or conflicts with the assumptions that animate the investigative and adjudicative restraints imposed by the Constitution—generally, the Fourth, Fifth, and Sixth Amendments. Part I begins by sketching, very briefly, the evolution of databases

1. *See* Thomas Haigh, *"A Veritable Bucket of Facts": Origins of the Data Base Management System*, SIGMOD REC., June 2006, at 33, 35.

2. *See, e.g.*, Erin Murphy, *Paradigms of Restraint*, 57 DUKE L.J. 1321 (2008).

3. 129 S. Ct. 695 (2009).

in criminal justice by providing a sampling of databases in existence. Part II then examines the few major instances in which the Supreme Court has commented on databases or databasing in the context of criminal justice, in order to glean different themes that have emerged. Part III then identifies five presumptions that seem to attach to database-related inquiries, while Part IV sets out some thoughts about how constitutional doctrine might evolve to the particular needs of databases, in order to better regulate and safeguard their use in the criminal justice system.

I. THE RISE OF DATABASES IN CRIMINAL JUSTICE

A. Proto-Databases

The criminal justice system's reliance on databases is both old and new. As many know, the formal, organized, public police first emerged as a concept around 1829, when Robert Peel organized the London bobbies.[4] The first detective unit in the United States was formed shortly after in 1846 in Boston,[5] at a time when tracking down criminals largely remained a private sector gig. Dominated by companies such as the (in)famous Pinkertons,[6] the unit's work consisted largely of pounding the pavement (and suspects). Indeed, many of the modern tools of detecting—"[s]ophisticated criminal investigation techniques—well-organized crime records systems, fingerprints, crime labs—did not appear until the twentieth century."[7] Even Alphonse Bertillion's pioneer anthropometrical system of identification in the late 1800s depended largely upon manual recording and comparison of measurements.[8]

The first primitive databases emerged around the same time, at the turn of the century. For instance, as early as 1919, the California State Bureau of Identification introduced a punch-card system for storing and retrieving modus operandi information.[9] But perhaps the watershed moment of government databasing occurred in the early 1930s, around the time that J. Ed-

---

4. For a thorough and encyclopedic account of the evolution of policing, see David A. Sklansky, *The Private Police*, 46 UCLA L. REV. 1165 (1999).

5. LAWRENCE M. FRIEDMAN, CRIME AND PUNISHMENT IN AMERICAN HISTORY 204 (1993). Peel had organized a small detective unit in 1842 in London. Sklansky, *supra* note 4, at 1204 & n.203.

6. Sklansky, *supra* note 4, at 1212-14.

7. SAMUEL WALKER, POPULAR JUSTICE: A HISTORY OF AMERICAN CRIMINAL JUSTICE 160 (2d ed. 1998).

8. SIMSON GARFINKEL, DATABASE NATION: THE DEATH OF PRIVACY IN THE 21ST CENTURY 40 (2000).

9. SIMON A. COLE, SUSPECT IDENTITIES: A HISTORY OF FINGERPRINTING AND CRIMINAL IDENTIFICATION 250-51 (2001).

gar Hoover opened the Federal Bureau of Investigation's first criminal evidence laboratory, which included fingerprint processing capacities, hair, blood, and firearm analysis.[10] As part of the new emphasis on forensic science, FBI implemented its first fingerprint database—a card sorter that capitalized on the technology created to tabulate the census and that led to the formation of IBM.[11]

Just a little over a decade later, the development of the mainframe computer in 1946 and the replacement of punch cards with magnetic tape significantly advanced databasing possibilities,[12] but it remained a largely primitive technology. By as late as 1984, the federal fingerprint database—the most advanced forensic database available—still depended primarily on manual recording and retrieval. At best, it served as an efficient means of organizing cards for retrieval, rather than for generating leads or links.[13] Linking two fingerprints required manual comparison of an unknown scene sample with, for instance, the 23 million criminal cards on file with the FBI.[14]

The 1980s, however, initiated a period of rapid change. Personal computers became commonly available. Law enforcement began to recognize and harness the potential of electronic storage and retrieval. And then, remarkably, the Internet was born. Connectivity became possible in ways previously unimagined, and storage capacity reached new heights. The foundations for the modern criminal justice databases had been set.

The first major advance occurred around 1985, when technology simplified the creation of digital images from physical fingerprint cards, enabling the National Automated Fingerprint Identification System ("AFIS"). AFIS software allowed a technician to conduct both confirmatory matches as well as automated comparisons between known and unknown prints, a landmark advance over the old manual methods of comparison.[15] In a 1987 report on the creation of AFIS, the authors noted that a San Francisco investigation had expended thousands of man-hours searching fingerprint

10. WALKER, *supra* note 7, at 160. Dan Richman's account of the rise of federal law enforcement in the 1930s relates the FBI's efforts to interest generated in the wake of the Lindbergh baby's kidnapping. Daniel Richman, *The Past, Present, and Future of Violent Crime Federalism*, 34 CRIME & JUST. 377, 386-88 (2006).

11. GARFINKEL, *supra* note 8, at 18.

12. DANIEL J. SOLOVE, THE DIGITAL PERSON: TECHNOLOGY AND PRIVACY IN THE INFORMATION AGE 14 (2004).

13. Fed. Bureau of Investigation, Integrated Automated Fingerprint Identification System, http://www.fbi.gov/hq/cjisd/iafis.htm (last visited Apr. 22, 2010) (describing that "[j]ust a few years ago . . . fingerprint cards had to be physically transported and processed." Thus a check "could often take three months to complete.").

14. GARFINKEL, *supra* note 8, at 44.

15. *Id.* at 45.

records for the one clear print an alleged killer had left behind; when that print became the subject of San Francisco's first automated search, a match was found in six minutes.[16] But AFIS still relied upon manual scanning of a card ink stain image, and thus that slowed response time and transmission capacity. The imperfection of the images also diminished automated matching capacity.

Also, in 1983, as a companion to AFIS, the FBI created the Interstate Identification Index ("III"), a pointer index which linked states to criminal records of both arrest and conviction that are associated with uploaded fingerprints.[17] The Index is operated through the National Criminal Information Center ("NCIC"), which has served since the late 1960s as the central federal criminal records database.[18]

But even though the idea of the criminal justice database has existed for over a hundred years—although records and fingerprints have been created and kept for most of that time—it would be wrong to assume that contemporary databases are simply the younger siblings of the databases just described. They are not. Old databases were typically paper files or punch cards that were physically kept and stored in diffused, and at times difficult to access, locations. Even the AFIS and III systems of the 1980s relied heavily on manual inputs and outputs of records. And proactive searching was likewise all but impossible because technology could not automatically sift through huge volumes of standardized material. Thus, an act as simple as switching locations might be effective in obscuring one's identity, since accessing a record created and stored even in the next town over could be prohibitively difficult. In short, until quite recently, the database primarily served an organizational and confirmatory function—if law enforcement had a known suspect, then a database enabled easier access to confirmatory information about that person.

B. Databases Today

In 1999, databasing radically changed when the Integrated Automated Fingerprint Identification System ("IAFIS") became operational.[19] IAFIS is now the largest single biometric database, and it looks nothing like the

16. *Id.*

17. James Jacobs & Tamara Crepet, *The Expanding Scope, Use, and Availability of Criminal Records*, 11 N.Y.U. J. LEGIS. & PUB. POL'Y 177, 181-82 (2008).

18. Congress had authorized the collection of criminal records and fingerprints as early as 1924, 43 Stat. 217, but it did not computerize them centrally until after the mid-century.

19. Fed. Bureau of Investigation, Integrated Automated Fingerprint Identification System, *supra* note 13.

fingerprint databases that preceded it.[20] IAFIS replaced the old ink-and-card system with a new, immediate digital image that could be easily stored, transmitted, and searched. Rather than locate a physical card and conduct a visual comparison, or even—as with the precursor AFIS system—electronically search images of physical cards, IAFIS kept an actual direct digital image of the print itself in high enough resolution to record and classify characteristics. And, because remote locations could immediately upload or retrieve direct records, IAFIS transformed the significance of the FBI's cache of information. Sharing and searching functions were dramatically enhanced. As it currently stands, IAFIS contains the records of over 55 million subjects in its Master File,[21] which can be automatically searched twenty-four hours a day every day of the year, and that number grows daily.

Also in 1999, the FBI implemented a new generation of NCIC technology, which represented a series of advances over the prior system. New features included "the addition of image processing (i.e., mugshots, signatures, and identifying marks); automated single-finger fingerprint matching; and information linking, which provides the ability to associate logically related records across NCIC files for the same criminal or the same crime."[22] Most importantly, the new system "automates functions that employees previously had to perform manually," including collection and evaluation of the system.[23] Whereas in 1967 the National Criminal Information Center served roughly 2 million requests annually, that number has now climbed to 2.5 million per day.[24]

Moreover, the NCIC folds in many additional databases. For instance, because the 1993 Brady Handgun Violence Protection Act (popularly known as the Brady Act) necessitated the creation of a reliable national background check system, Congress allocated funds to help update the over 70 million criminal records held by the fifty states and the District of Columbia.[25] As of 1998, that goal had largely been achieved, and a rapid and quick national records system, called the National Instant Criminal Background Check System ("NICS") had been created.[26] The NCIC 2000 system also added the Convicted Sexual Offender Registry and the Convicted

---

20. *Id.*

21. *Id.*

22. Stephanie Hitt, *NCIC 2000*, FBI L. ENFORCEMENT BULL., July 2000, at 12, *available at* http://www.fbi.gov/publications/leb/2000/jul00leb.pdf.

23. *Id.*

24. *Id.* at 12.

25. Jacobs & Crepet, *supra* note 17, at 180.

26. *Id.* at 181 (estimating that it processes roughly "eight million firearm purchase background checks annually").

Persons on Supervised Release database to augment old standards such as the outstanding warrant files.[27] In addition, around the same time that IAFIS went online, the FBI launched the Combined DNA Index System ("CODIS"). This database is essentially a pointer system to DNA profiles typed by state and federal agents, and it currently contains over 7.9 million offender known profiles and almost 300,000 forensic unknown profiles.[28]

But IAFIS and CODIS are simply the tip of the iceberg. It was estimated in 2001 that federal agencies and departments today maintain roughly 2000 databases.[29] Those databases cover a wide variety of topics, ranging from those directly related to criminal justice purposes to those applicable only in the most specialized circumstances. There are gang databases,[30] terrorist watch lists,[31] violent criminal databases,[32] forensic reference databases,[33] corrections databases, and a wide variety of public and private databases ranging from security industry to identity theft to gaming industry databases.[34]

C. Databases Tomorrow

Databasing is here to stay, and the future promises the creation of more of them, used in increasingly novel ways. To give just one example, in the wake of the September 11, 2001 terrorist attacks and in response to congressional and executive mandates to facilitate information sharing among and within the criminal justice and national security communities, a great deal of attention has surrounded the creation of communication structures across various state and federal entities. Most notably, forty-three "fusion centers" across the nation have received roughly $380 million dollars in

---

27. Hitt, *supra* note 22, at 14.

28. Fed. Bureau of Investigation, CODIS-NDIS Statistics, http://www.fbi.gov/hq/lab/codis/clickmap.htm (last visited Apr. 22, 2010).

29. Daniel J. Solove, *Privacy and Power: Computer Databases and Metaphors for Information Privacy*, 53 STAN. L. REV. 1393, 1403 (2001).

30. *See, e,g.*, James B. Jacobs, *Gang Databases: Context and Questions*, 8 J. CRIMINOLOGY & PUB. POL'Y 705, 705-06 (2009), *available at* http://www3.interscience.wiley.com/cgi-bin/fulltext/122687667/PDFSTART (describing various databases).

31. *See* The Dep't of the Treasury, Office of Foreign Assets Control, Specially Designated Nationals and Blocked Persons, http://www.ustreas.gov/offices/enforcement/ofac/sdn/t11sdn.pdf (last visited Apr. 25, 2010).

32. *See* Fed. Bureau of Investigation, Violent Criminal Apprehension Program, http://foia.fbi.gov/vicappia.htm (last visited Apr. 22, 2010).

33. Robin Bowen & Jessica Schneider, *Forensic Databases: Paint, Shoe Prints, and Beyond*, NIJ JOURNAL NO. 258 (2007).

34. Solove, *supra* note 29, at 1401-05.

federal grants.[35] These centers integrate federal, state, and local officials under one roof, and although their initial mandate was to focus specifically on terrorism-related investigations, "they have increasingly gravitated toward an all-crimes and even broader all-hazards approach."[36] Fusion centers rely intensively on information gathered by other entities. That is, they execute their mission in part by accessing and coordinating an extraordinary number of public and private database systems, including motor vehicle, location, criminal records, corrections, sex offender, public health, and industry databases.[37]

According to some reports, the FBI has also announced plans to replace IAFIS with a new multimodal biometric database that will incorporate DNA, facial imaging, iris scans, palm and voice prints, and identifying information in one integrated system.[38] This announcement follows in the wake of the REAL-ID Act of 2005, which aimed to create a de facto national identity card, but has been met with strong opposition from both states and individual privacy advocates.[39] Again, however, these examples are simply two among many, intended only to provide some insight into the breadth of databasing likely to occur in the future. After all, at present there are even plans for a doggie DNA database.[40]

II. THE DATABASE IN CONSTITUTIONAL CRIMINAL PROCEDURE

The preceding section should, at the least, make two points abundantly clear: first, that there are an enormous number of databases in the criminal justice system, and second, that the database, as it exists today, has really only been around for ten or so years. Any person who has witnessed the past fifteen years of technological advancement knows, without reading a law review article, that online databases have transformed modern life. Yet surprisingly few changes have occurred in actual constitutional doctrine in response to widespread databasing.

---

35. *See* TODD MASSE, SIOBHAN O'NEIL & JOHN ROLLINS, FUSION CENTERS: ISSUES AND OPTIONS FOR CONGRESS app. B (2007).

36. *Id.* at i.

37. *Id.* at 33-34.

38. Ellen Messmer, *FBI Building System That Blows Away Fingerprinting*, NETWORK WORLD, Sept. 23, 2009, *available at* http://www.networkworld.com/news/2009/092309-fbi-biometrics.html?ts.

39. Real ID Act of 2005, Pub L. No. 109-13, 119 Stat. 231, 302; *see, e.g.*, *Four Western States Allow Illegal Immigrant Licenses*, ASSOCIATED PRESS, Dec. 18, 2009 (noting criticisms of REAL ID and efforts to change it).

40. *See* Marc W. Allard, Building a Genetic Reference Database for Dog mtDNA sequences and SNPs (May 2009) (unpublished final report funded by the U.S. Department of Justice), *available at* http://www.ncjrs.gov/pdffiles1/nij/grants/226936.pdf.

Last Term, the Court confronted the applicability of the Fourth Amendment exclusionary rule to a police database error in *Herring v. United States*.[41] Considered a watershed moment in criminal procedure, inasmuch as the Court approved the admission of evidence acquired from an unconstitutional search justified by a police database entry that proved erroneous, the debate in *Herring* presages the kinds of concerns likely to confront a variety of constitutional criminal procedure doctrines in the coming years. Before turning to the Court's reasoning in that case, however, it may be useful to look at previous instances in which the Court addressed—even outside of the criminal context—law enforcement-related databases, in order to see the different ways in which the Court has assessed the constitutional significance of databasing generally.

A. Major Database Cases

The Supreme Court has directly addressed and confronted the significance of databases on only a handful of occasions, few of which were very recent.[42] In fact, perhaps the two most significant cases were issued more than twenty years ago. The few remaining cases only peripherally address the significance of databasing, despite being of more recent vintage.[43] Thus, the current Court's view of databases, and how they might influence or alter the conventional approaches to constitutional regulation, remains largely obscured.

For instance, in the 1977 case *Whalen v. Roe*, the Court confronted a state statute that ordered the creation of a central database containing the names and addresses of persons who had obtained certain prescription drugs in order to combat prescription fraud.[44] The challengers, a group of

41. 129 S. Ct. 695 (2009).

42. In addition to the cases discussed at length below, the Court also addressed or mentioned databases in two additional cases: *Ornelas v. United States*, 517 U.S. 690, 692 (1996) (referring to Narcotics and Dangerous Drugs Information System database) and *Hiibel v. Sixth Judicial District Court of Nevada*, 542 U.S. 177, 196 (2004) (Stevens, J., dissenting) ("A name can provide the key to a broad array of information about the person, particularly in the hands of a police officer with access to a range of law enforcement databases.").

43. The Court has confronted database-related issues peripherally in a handful of non-criminal justice cases as well. *See, e.g.*, Crawford v. Marion County Elec. Bd., 553 U.S. 181, 219 n.23 (2008) (Souter, J., dissenting) (noting problems with Bureau of Motor Vehicles database in case related to voter identification statute); Eldred v. Ashcroft, 537 U.S. 187, 250 (2003) (Breyer, J., dissenting) (acknowledging the utility of "computer-accessible databases . . . to facilitate research and learning" in copyright cases); N.Y. Times v. Tasini, 533 U.S. 483 (2001) (resolving copyright action against publishers operating electronic databases); Printz v. United States, 521 U.S. 898, 930 (1997) (raising likelihood that state officials, rather than federal ones, will be blamed for mistakes in databases related to firearm background checks).

44. 429 U.S. 589, 591 (1977).

doctors and patients, alleged that the statute violated substantive due process because it impinged upon "their interest in the nondisclosure of private information and also their interest in making important decisions independently."[45] The government, in turn, argued that the scheme was necessary to control the illegal market in such drugs.

In upholding the statutory scheme, the Court described the databasing process in detail. It noted that the prescriptions were written in triplicate, and one copy was transported to Albany monthly.[46] They were then sorted, coded, and logged into a computer and stored on magnetic tape. The physical forms were kept for five years, whereupon they were destroyed. The tapes themselves were kept in a locked cabinet and were only run "off-line" in a single room. Only seventeen employees could access the files, and only twenty-four investigators could investigate. In other words, the system in place closely regulated both the physical copy of the information as well as the data storage mechanism, search method, and retrieval. The Court also observed that penalties applied to any person who willfully violated the system.[47] Attentive to the safeguards taken to protect the information from inappropriate disclosure, the Court proclaimed that it "therefore need not, and [will] not, decide any question which might be presented by the unwarranted disclosure of accumulated private data whether intentional or unintentional or by a system that did not contain comparable security provisions."[48]

In a closing passage oft-cited today, Justice Stevens, writing for the Court, declared:

> We are not unaware of the threat to privacy implicit in the accumulation of vast amounts of personal information in computerized data banks or other massive government files. The collection of taxes, the distribution of welfare and social security benefits, the supervision of public health, the direction of our Armed Forces, and the enforcement of the criminal laws all require the orderly preservation of great quantities of information, much of which is personal in character and potentially embarrassing or harmful if disclosed. The right to collect and use such data for public purposes is typically accompanied by a concomitant statutory or regulatory duty to avoid unwarranted disclosures. Recognizing that in some cir-

45. *Id.* at 600. Because the doctrine was still evolving, the litigants framed their interest under several constitutional provisions; *see id.* at 600-02 & nn.23-24. These provisions included the First, Fourth, and Fourteenth Amendments, *id.* at 603-04 & n.32, but the claim sounds in due process today and that was the basis upon which the Court ruled. *Id.* at 604.

46. *Id.* at 593-95.

47. *Id.* at 594-95 & n.12.

48. *Id.* at 605-06.

> cumstances that duty arguably has its roots in the Constitution, nevertheless New York's statutory scheme [adequately protects privacy].[49]

Thus, the Court in 1977 not only contemplated the databasing currently underway in the nation but also anticipated the widespread databasing that would come in the future. It deemed databasing necessary, although also potentially threatening to personal privacy. Most intriguingly, the Court seemed unconvinced by, although open to, the idea that the Constitution would regulate the compilation of databases. Instead, the Court seemed to view the enabling statutes as the proper source of primary regulation—a proposition that assumes that databases are formed largely as a result of statutory mandate.

In contrast, Justice Brennan, concurring, explicitly stated that "[b]road dissemination by state officials of such information . . . would clearly implicate constitutionally protected privacy rights."[50] Justice Brennan also specifically addressed the significance of databasing itself. He identified the troubling aspect of the scheme as "the central computer storage of the data thus collected."[51] Although he deemed it "[o]bvious[]" that "collection and storage of data . . . is not rendered unconstitutional simply because new technology makes the State's operations more efficient," he recognized that the "central storage and easy accessibility of computerized data vastly increase[s] the potential for abuse of that information."[52] Citing, interestingly, to the Fourth Amendment as an example of a means of restriction on information-gathering (as opposed to simple kind-restrictions), Justice Brennan remarked that "future developments" may "demonstrate the necessity of some curb on such technology."[53]

Depending on one's perspective, the Court either made good on its promise or tacked the other direction ten years later in its next major judgment concerning large scale databases, *United States Department of Justice v. Reporters Committee for Freedom of the Press*.[54] In *Reporters Committee*, the Court addressed a claim by a reporter and journalists' association for access under the Freedom of Information Act ("FOIA") to FBI "rap sheets."[55] The FBI had compiled rap sheets that contained the name, birth

---

49. *Id.* at 605.

50. *Id.* at 606 (Brennan, J., concurring). The other concurring Justice, Justice Stewart, sharply disputed this claim, criticizing its citation to *Griswold v. Connecticut*, 381 U.S. 479 (1965), and *Stanley v. Georgia*, 394 U.S. 557 (1969), as inadequate support for the proposition. *Id.* at 608 (Stewart, J., concurring).

51. *Id.* at 606 (Brennan, J., concurring).

52. *Id.* at 606-07.

53. *Id.* at 607.

54. 489 U.S. 749 (1989).

55. *Id.* at 751.

date, physical characteristics, and criminal records (including arrest, charges, convictions, and incarcerations) of over 24 million individuals.[56] The challenge related to the denial of a request by a reporter who sought access to the records of members of a reputed mafia family under the federal FOIA.[57]

As in *Whalen*, the focus of the legal analysis was largely statutory. The enabling statutes that authorized the creation of the rap sheets had been amended to limit disclosure, and the Department of Justice ("DOJ") had a policy of generally treating them as confidential, except for disclosure to the subject of the sheet or for press releases for wanted persons.[58] The FOIA, in contrast, sets forth a presumption of disclosure of agency records.[59] However, three exemptions arguably required the DOJ to withhold the rap sheets: Exemption 3, requiring compliance with other statutory nondisclosure provisions; Exemption 6, protecting certain files that could create "a clearly unwarranted invasion of personal privacy"; and Exemption 7(C), applying to law enforcement records that would likewise invade privacy.[60]

Again writing for the Court, Justice Stevens held that Exemption 7(C) prohibited the disclosure of the rap sheet.[61] Notably, the Court first had to acknowledge the claimed invasion of privacy was to material that, technically speaking, consisted of almost entirely public information. Indeed, the reporter specifically sought only that information that constituted "matters of public record."[62] Of course, the Court observed that a few states prohibited disclosure of out-of-state information or restricted the conditions under which non-conviction information would be disclosed.[63] A larger number had restrictions on issuing criminal history summaries.[64] But individually, each piece of information was almost always technically open and available as a matter of public record.[65] Thus, in order to find that the rap sheet implicated any privacy interest, the Court had to acknowledge that

56. *Id.* at 751-52.

57. *Id.* at 757.

58. *Id.* at 752. The statute had further been amended to permit disclosure of rap sheets to banking institutions, securities industry regulators, and the Nuclear Regulatory Commission. *Id.* at 753.

59. *Id.* at 755.

60. *Id.* at 755-56 (citing 5 U.S.C. § 552(b)(3), (6), (7)(C)).

61. *Id.* at 758-59.

62. *Id.* at 757.

63. *Id.* at 753-54 & n.2.

64. *Id.* at 753.

65. *Id.*

"much rap-sheet information is a matter of public-record,"[66] and specifically draw a distinction between the technical availability of the information and its *actual* availability ("practical obscurity," in the Court's parlance),[67] which it described as "limited."[68]

Rejecting a "cramped notion of personal privacy,"[69] Justice Stevens commented that "[i]n an organized society, there are few facts that are not at one time or another divulged to another."[70] Yet such disclosures did not render that information no longer private, given the passage of time and the practical impediments to broad dissemination. Noting that the "very fact that federal funds have been spent to prepare, index, and maintain these criminal-history files demonstrates that the individual items of information . . . would not otherwise be 'freely available,'"[71] Justice Stevens distinguished between "scattered disclosure" of bits of information and complete and coherent revelation.[72]

While the gravamen of the Court's reasoning rested on a statutory analysis of the provisions involved and the propriety of issuing a categorical rule,[73] the manner in which it assessed the interests implicated by disclosure granted some insight into its view of databasing generally. The Court noted that the Privacy Act had been enacted specifically to counter concerns about "computer data banks on individual privacy."[74] In the Court's estimation, electronic storage altered the nature of the interest affected, in that "the computer can accumulate and store information that would otherwise have surely been forgotten long before."[75] By holding the records unobtainable under the law enforcement exemption, while in no way casting aspersions on the statutory mandate to centralize them in the first place, the

66. *Id.*

67. *Id.* at 762.

68. *Id.* at 753.

69. *Id.* at 763.

70. *Id.* Justice Stevens identified the interest as in "selective" versus total nondisclosure. *Id.* at 763 n.14 (citing Kenneth L. Karst, *"The Files": Legal Controls Over the Accuracy and Accessibility of Stored Personal Data*, 31 LAW & CONTEMP. PROBS. 342, 343-44 (1966)).

71. *Id.* at 764.

72. *Id.*

73. *Id.* at 779-80.

74. *Id.* at 766-67 (citing H.R. Rep. No. 93-1416, at 7 (1974)). The Court also cited to *Department of Air Force v. Rose*, a sort of proto-database case in which law students requested records of disciplined cadets. Despite the publicity of some of the cadets' names, the Court noted that the passage of time and informality of public knowledge continued its character as "private" sufficiently to deny disclosure under FOIA. *Id.* at 768-69 (citing Dep't of Air Force v. Rose, 425 U.S. 352 (1976)).

75. *Id.* at 771.

Court seemed to say that such long memories were properly the province of the state alone.

The holding in *Reporters Committee* contrasts remarkably with that in *Smith v. Doe*, in which the Court upheld the Alaska Sexual Offender Registration Act against an ex post facto constitutional challenge.[76] As in *Reporters Committee*, the case involved a statute that authorized the compilation of publicly available information—here, convictions for sexual offenses.[77] Unlike *Reporters Committee*, however, the purpose of the statute *was* public dissemination.[78] Finding that the statute was civil and regulatory, as opposed to punitive, in character, the Court found no violation of the constitutional bar against retroactive punishments.[79]

Most relevant for this Article, however, was the Court's treatment of the arguments raised by the respondents. The respondents analogized the creation of a centralized public database available online to the public stigma and shaming of the pillories and whippings of old, in order to liken the Act's requirements to "punishment."[80] Rejecting that argument, however, the Court refused to consider as punishment the mere "dissemination of information" or the publication of truthful, formally available information.[81] Writing for the Court, Justice Kennedy distinguished the physical, corporeal "direct confrontation" from the "dissemination of accurate information about a criminal record, most of which is already public."[82] Moreover, to the extent that "adverse consequences" flowed from that dissemination, such results were not the direct purpose of the regulatory scheme.[83]

Thus, unlike the Court in both *Reporters Committee* and *Whalen*, in *Smith* the Court seemed indifferent to the implications of rendering information more easily and widely distributed, choosing instead to focus on its technical availability. That point is particularly salient with regard to the Court's dismissal of the impact of internet publication, which it said served only to "increase[] in proportion" the degree of humiliation.[84] The corres-

---

76. 538 U.S. 84 (2003). *Smith* was also decided with a companion case, *Connecticut Department of Public Safety v. Doe*, 538 U.S. 1 (2003).

77. *Smith*, 538 U.S. at 89.

78. *Id.* at 93.

79. *Id.* at 92-93. For a discussion of the civil/punitive divide as regards this case, see generally Murphy, *supra* note 2, at 1351-58.

80. *Smith*, 538 U.S. at 98-99.

81. *Id.*

82. *Id.*

83. *Id.* at 99.

84. *Id.* Justice Thomas, concurring, took issue with the Court's discussion of Internet distribution as the means of dissemination, as it was not specifically provided for in the statute. *Id.* at 106-07 (Thomas, J., concurring).

ponding simplicity and breadth of disclosure was wholly irrelevant to the analysis; the Court likened the online registry to "a visit to an official archive of criminal records," which was simply "more efficient, cost effective, and convenient for Alaska's citizenry" by virtue of being online.[85] The Court later identified the negative consequences of the statute as flowing from "the fact of conviction, already a matter of public record" rather than from the broad disclosure and broadcast provisions of the statute.[86] In sum, unlike the analysis in *Reporters Committee* and *Whalen*, the Court in *Smith* seemed wholly uninterested in the significance of databasing as an act that changes the nature of information in terms of its meaning and character. Rather, the database was simply another form of the same information, with no particular significance or power merely on account of its technical capacity.

However, two Justices maintained that the compilation and ready availability of the information constituted a significant change from its prior paper-and-archive existence.[87] Justice Souter, concurring in the judgment, disputed that the Act "simply makes public information available in a new way," noting that "[i]ts point, after all, is to send a message that probably would not otherwise be heard."[88] Justice Stevens likewise concluded that "there can be no doubt that the [w]idespread public access . . . to this personal and constantly updated information has a severe stigmatizing effect."[89] Overall, however, *Smith*, marked a fairly sharp departure from *Reporters Committee* and *Whalen* in its seeming indifference to arguments based on the special nature or character of information once collected in centralized databases.

B. *Evans* and *Herring*

The two most significant criminal procedure cases related to databasing occurred more recently, in the context of Fourth Amendment challenges to evidence seized as a result of erroneous information kept in a database. Both *Arizona v. Evans*,[90] decided in 1994, and *Herring v. United States*,[91]

85. *Id.* at 99-100; *see also id.* at 105 (further dismissing the significance of the breadth of disclosure).

86. *Id.* at 101.

87. Justice Ginsburg, writing in dissent for herself and Justice Breyer, found the Act unconstitutionally punitive in an opinion that did not specially address the significance of the creation of a centralized database of information. *See id.* at 115 (Ginsburg, J., dissenting).

88. *Id.* at 109 (Souter, J., concurring).

89. *Id.* at 111-12 (Stevens, J., concurring) (internal citation and quotation marks omitted).

90. 514 U.S. 1 (1994).

91. 129 S. Ct. 695 (2009).

decided in 2009, held the exclusionary rule inapplicable to the fruits of unconstitutional searches conducted in reasonable reliance on a database entry later proven to have been inaccurate. In *Evans*, the error stemmed from the failure of a judicial court clerk to clear a quashed warrant from a computer database;[92] in *Herring*, the same mistake occurred, but this time under the watch of a police department clerk.[93] The Court in both cases ruled the exclusionary rule inapplicable, citing deterrence as the primary purpose for the rule and the absence of a deterrence rationale for reasonable recordkeeping errors.[94]

In both cases, the Court briefly discussed the procedures used to enter and clear warrants in the computer, although in both cases those procedures relied upon factual development, as opposed to statutory or regulatory analysis.[95] In both cases, the Court mentioned the error rate for the database in question, although in both cases that error rate was the subject of dispute.[96] And in both cases, the Justices making up the majority took pains to consider overtly the fact that a database was the source of error, even while ultimately dismissing those implications for purposes of resolving the instant case.

Specifically, in *Evans* three concurring justices (Justices O'Connor, Souter, and Breyer) expressed considerable concern about the nature of the error as a recordkeeping error that occurred as a result of departure from established protocol.[97] They specifically worried about databases—"the advent of powerful, computer-based recordkeeping systems that facilitate arrests in ways that have never before been possible."[98] Cautioning that such technology, while beneficial, ought not to be relied upon "blindly," they warned that reliance on a system with "no mechanism to ensure its accuracy over time" would invoke the exclusionary rule.[99]

---

92. *Evans*, 514 U.S. at 5.

93. *Herring*, 129 S. Ct. at 698-99.

94. *Id.* at 703-04; *Evans*, 514 U.S. at 15-16.

95. *Herring*, 129 S. Ct. at 698; *Evans*, 514 U.S. at 5 (describing testimony).

96. *Compare Herring*, 129 S. Ct. at 704 (citing testimony that no other errors of this kind had ever occurred), *and Evans*, 514 U.S. at 15 (citing testimony that "this type of error occurred once every three or four years"), *with Herring*, 129 S. Ct. at 700 n.1 (debating characterization of error as "negligent" or not) *and Evans*, 514 U.S. at 27-28 (Ginsburg, J., dissenting) (observing both that the witness later admitted to three of the same kind of errors that day, and also recounting prominent cases of similar error). *See also Herring*, 129 S. Ct. at 704 & n.5 (noting dispute in testimony about whether the error occurred with greater frequency than suggested).

97. *See Evans*, 514 U.S. at 16-17 (O'Connor, J., concurring).

98. *Id.* at 17.

99. *Id.*

Similarly, in *Herring*, Chief Justice Roberts stated that a showing that the police were "reckless in maintaining a warrant system" or had "knowingly made false entries to lay the groundwork for future false arrests" would invoke exclusionary rule protection.[100] Demonstrated "systemic errors" in a system might also render reliance on that system objectively unreasonable.[101] But such errors must manifest as substantive mistakes (say, repeated reliance on bad information), rather than on demonstrated procedural infirmities (say, a lack of established protocols). Thus, absent evidence—most obviously offered by the defendant challenging the action—that a system had "routine or widespread" deficiencies, the evidence seized by an officer in reliance on information gleaned from an erroneous database entry would be admissible.

The dissenters in *Evans* each focused particularly on the specific implications of database error. Justice Stevens observed that the exclusionary rule operated to deter "systemically," not just on an individual officer level.[102] He chastised the Court for "overlook[ing] the reality that computer technology has changed the nature of threats to citizens' privacy over the past half century."[103] Justice Stevens also refused to presume that the database was kept in an orderly fashion, calling the testimony "slim evidence on which to base a conclusion that computer error poses no appreciable threat to Fourth Amendment interests."[104] Lastly, he conjured the specter of erroneous arrests attributable to "some bureaucrat" who "has failed to maintain an accurate computer data base."[105]

Justice Ginsburg, dissenting separately and joined by Justice Stevens, opened by warning about the "evolving problem" of "the increasing use of computer technology in law enforcement."[106] She went on to discuss a theme that she picked up again in her dissent in *Herring*,[107] noting that the lower courts found Supreme Court precedent "not helpful" in resolving the issues.[108] Justice Ginsburg agreed that the problem was not just "a court

---

100. *Herring*, 129 S. Ct. at 703.

101. *Id.* at 704.

102. *Evans*, 514 U.S. at 21 (Stevens, J., dissenting). In fairness, Justice Stevens' observation was more motivated by the systemic relationship between the law enforcement and judicial branches, as opposed to the many operators within a law enforcement (or database) system.

103. *Id.* at 22.

104. *Id.* at 21-22.

105. *Id.* at 23.

106. *Id.* (Ginbsurg, J., dissenting).

107. The *Herring* dissenters largely mirrored the justices that either dissented in *Evans* or else concurred with specific reservations related to database entry. Justice O'Connor, who had joined the concurrers in *Evans*, had left the Court by the time that *Herring* was decided.

108. *Evans*, 514 U.S. at 24.

employee's slip" but rather the "'potential for Orwellian mischief' in the government's increasing reliance on computer technology in law enforcement."[109]

In an opinion thoroughly dedicated to assessing the impact of these new technologies, Justice Ginsburg asserted that "[w]idespread reliance on computers to store and convey information generates, along with manifold benefits, new possibilities of error."[110] Moreover, "computerization greatly amplifies an error's effect," because the error can seep out across a wide variety of systems.[111] Citing examples of harassing and dangerous arrests based on erroneous computer data,[112] Justice Ginsburg worried about leaving such databases constitutionally unregulated absent a showing of deliberate error.

Her dissent, this time on behalf of Justices Stevens, Souter, and Breyer, echoed the same themes. Rather than accept the database on its face and assume it optimally operative, she elaborated the many shortcomings of its current configuration.[113] For example, she cited the lack of obvious electronic connections between information terminals, the entirely uncontrolled procedures for inputting and extracting information, and the lack of routine oversight or maintenance for quality control purposes.[114] By comparison, she noted that the national database system, the NCIC, had numerous safeguards in place both to control the quality of information stored and to minimize the adverse consequences of error.[115] In short, Justice Ginsburg demanded some evidence of good database management, rather than trust that such practices were in place absent evidence of bad outputs.[116]

In sum, a variety of lessons emerge from a review of the Court's major efforts to examine the constitutional significance of criminal justice-related databases. Perhaps most evident from the foregoing discussion is its overall failure to articulate a consistent vision of the constitutional significance of the aggregation and easy retrieval and dissemination of otherwise lawfully held—and even publicly available—information. Many of the cases were largely driven by statutory analyses, and yet not all databases are go-

109. *Id.* at 25.
110. *Id.* at 26.
111. *Id.*
112. *Id.* at 27.
113. Herring v. United States, 129 S. Ct. 695, 708-09 (2009) (Ginsburg, J., dissenting).
114. *Id.* at 709 (Ginsburg, J., dissenting).
115. *Id.*
116. *Id.* at 709-10 (Ginsburg, J., dissenting). Indeed, she worried both that it would be difficult to identify the source of error in many cases and that defendants (especially indigent ones) would have trouble obtaining the information necessary to make such a showing. *Id.*

verned by elaborate statutory regimes. Moreover, the degree to which the Court undertook its own assessment of the manner in which centralization and digitalization of records transformed the significance of those records seemed to vary case-by-case. To be sure, much of that variation may be attributable to political or ideological differences among justices and courts over time. But it may also be that databasing is sufficiently new, and sufficiently untested, that a template for a more sensitive inquiry has yet to be developed.

III. THE SLIPPAGE BETWEEN DOCTRINE AND DATABASES

What might databasing mean for rules of criminal justice? Or, more specifically, what does the emergence of the database mean for the procedural rules that have governed how we think about police investigations and criminal adjudication? About evidence and adversary process? How do the Fourth, Fifth, and Sixth Amendments hold up in a newly interconnected world? As shown, the Supreme Court has paid scant (and inconsistent) heed to the peculiar features of databasing or to what special concerns might inform the investigations conducted or evidence collected from them. There is no idea of "database cases" the way there are "informant cases" or "confession cases" or even "wiretap cases."[117]

This Part sketches five presumptions that seem to cloud the assessment of evidence and information gleaned from databasing. The fact that many of these characteristics occur individually with regard to more conventional forms of evidence helps only to obscure the particular havoc they can wreak when aligned together in a single form, such as databasing. But past experience with these presumptions might also help shed light on the best way to analyze database cases going forward. Articulating these presumptions enables a clearer appraisal of the ways in which databases operate, thereby exposing the particular risks that attend the use of databases in criminal justice.

A. The Presumption of Ready Analogy

A strong inclination exists to ignore the ways in which databases transform information and instead characterize them as identical to, or at most as simply more efficient forms of, other kinds of information collection. Typical is the assertion in *Whalen* that the centralized collection of prescription records for databasing was "not significantly different" than that which existed when prescriptions were subject to other forms of limited

117. *But see Evans*, 514 U.S. at 17 (O'Connor, J., concurring) (likening the desire to see standards for databases to a similar approach to informants).

disclosure.[118] The Court in *Smith* similarly analogized the collection and dissemination of conviction records online to a simply more efficient form of historical archiving.[119]

I have elsewhere written at length about the dangers of viewing technological incursions onto privacy as significant only inasmuch as they mirror physical ones,[120] and so will not elaborate on that assertion here. What is significant, however, is that the reliance on non-technological precursors to legitimate technology-based regimes serves to excuse courts from having to seriously contemplate both the actual steps required to construct and operate a database, as well as the new or innovative ways in which information may be used as a result. If an electronic or online repository of convictions is considered identical to a paper one, then courts need not inquire into the complicated questions of algorithms or file maintenance or dissemination or interception methods.

Moreover, databases may simply gain authority by association: if law enforcement has kept fingerprints of criminals for a hundred years, why should any court revisit their continued preservation? Never mind that, for most of those hundred years, the dusty and disintegrating fingerprint cards existed in thousands of drawers in hundreds of decentralized sheriff's offices, whereas now they will be instantly accessible and electronically searchable anywhere in the world. The utility of the analogy is that it bypasses the need for careful consideration of the details of databasing—the nitty gritty operational information about inputs and outputs that would require careful reconsideration and possible intervention in areas of otherwise longstanding familiarity.

B. The Presumption of Demonstrable Harm

Similarly, databases benefit from the general requirement that constitutional violations allege actual and not speculative harms. The moment of judicial confrontation with a database system almost always occurs long after the optimal moment for oversight or control. Thus, it is easy to assume that databasing itself is harmless, or at best generates speculative potential injury to a range of individuals as innumerable as the database's capacity itself. And yet the mere risk of harm has long been deemed an insufficiently worthwhile target of constitutional attention, and it is unwieldy even to imagine defining parameters for the minimization of such risks in the absence of specific instances of injury.

---

118. Whalen v. Roe, 429 U.S. 589, 602 (1977).
119. Smith v. Doe, 538 U.S. 84, 98-100 (2003).
120. *See* Murphy, *supra* note 2, at 1345-64.

This presumption is particularly pernicious when it comes to databasing, however, since the faulty products of a database can go entirely unnoticed under current doctrine even when they are common and recurring. Consider the debate in *Herring* itself: the majority demanded evidence that the database routinely produced bad information, refusing to consider the absence of quality control mechanisms itself a sufficient "harm." Yet a database that generates bad information—say, that falsely reports arrest warrants—may produce many arrests, but little record of those arrests. Unless the arrested person sues civilly, or is found in violation of contraband (as in the case of *Herring*), no formal record of the error may be made. And even if formal suits are filed, it may be difficult to link them to one another as the product of a faulty database. The only proof of the reliability of the database in *Herring* itself were the statements of its keepers—hardly disinterested parties—and yet, even those were contested factually.

Lastly, to wait until harm does result may often foreclose real remedy. The diffused and decentralized nature of databases may make error correction difficult. Back-up records may have to be obtained and destroyed. Inaccurate or obsolete information may have spread virally throughout networks and be virtually impossible to recall. Indeed, a database error, like the database itself, often can be more conceptual than real—it may less be about a server with a bunch of ones and zeros that require erasing than it is about a patchwork of information let loose in a web of systems. Once the information is released, it carries its own momentum, and eradicating it from every system may simply not be possible.

C. The Presumption of Regularity

Databases also appear to benefit from the strong presumption of regularity that attaches to most law enforcement actions. This presumption, as applied to the behavior of individual officers, tends to spring partly from an institutional desire to defer to the expertise of the executive branch in conducting investigations and partly from the recognition that assessing the subjective motivations of individual persons is difficult at best. But even though no equivalent functional or institutional justification impels the same deference when it comes to databases, there nevertheless exists the tendency to exercise it.

Regarding databases, the presumption of regularity means that, absent affirmative evidence that a database is kept in a shoddy or substandard fashion, courts will assume the soundness of the information generated. Notably, this presumption seems to hold even when information about the procedures or practices governing the collection and maintenance of the

database is lacking,[121] either because the record was not developed factually or because the oversight structures for the database are entirely informal. Thus, unlike areas of constitutional criminal procedure in which articulation of formal restraints, and compliance with them, dictates the acceptability of a particular action,[122] database cases can be decided largely ad hoc.

As a corollary to this presumption of regularity, databases also seem to benefit from the fact that they may spring from organic enabling statutes. Courts can therefore rely upon a statutory regime to guide and govern review of the database, rather than grasp at uncertain constitutional doctrine. Conversely, administrative compliance with statutory provisions about database parameters may shield against claims that the database is impeding or contravening constitutional interests.

D. The Presumption of Individual Action

Relatedly, databases can benefit from the difficulty that courts encounter in fitting the actual activity of databasing into conventional molds for police behavior. A database is typically not the product of any single individual actor; it may not even be the product of a single police force. Rather, databases are the ultimate collaborative projects. They may require myriad inputs and enable equally diverse outputs. They may cross jurisdictional borders or require public-private sector cooperation.

By way of example, consider a database as simple even as the one raised in *Whalen*. The database software must be created by computer scientists, and the hardware kept and operated by technologists. Doctors create the source information, which must be transferred to a variety of state officials responsible for inputting it into the machine. The hard copies of the information must then be tended and eventually destroyed. The physical storage mechanism—there, magnetic tapes—must be kept and safeguarded, and then perhaps operated and searched when appropriate. They too must be regularly maintained to ensure compliance with regulatory restrictions and destruction of data when appropriate. Throughout all of this, technology may evolve such that the magnetic tape or machines or the software becomes outdated or obsolete, and the process begins anew. And that is just one state's simple prescription registry, accessible to only a limited number

121. *See, e.g.*, *Herring*, 129 S. Ct. at 698 ("For whatever reason, the information about the recall of the warrant for Herring did not appear in the database.").

122. For example, the Constitution requires suppression of a confession obtained in violation of *Miranda*, even if a particular defendant might be shown to be aware of the rights without warning, *see, e.g.*, Yarborough v. Alvarado, 541 U.S. 652, 668 (2004), or suppression of evidence gathered in violation of the warrant requirement, even if it is clear that one could have readily been obtained, *see, e.g.*, Mincey v. Arizona, 437 U.S. 385 (1978).

of state officials and operated (the Court took great pains to state) out of one room in Albany. A diffused networked system of information like NCIC or the DNA databank or the sex offender registries raises a whole new complex of problems. In sum, a "database" is often a misleadingly singular concept that in fact embodies multifold layers of individuals and objects, that can span both geographical and temporal boundaries.

E. The Presumption of Technological Neutrality

As a companion to the presumption of regularity, databases also remained cloaked in a powerful technological neutrality. The symbolic and actual manifestation of the database as a computerized technology helps to neutralize it. Disembodied from the human beings that define, create, realize, and benefit from its parameters, the database is easily viewed as incapable of bias in the way that human law enforcement agents might be.

That is, although courts of course recognize that a human being might create false entries in a database or intentionally manipulate information, the underlying architecture of the database is largely ignored as a human event. It has long been a familiar cautionary refrain that police officers are engaged in the "competitive enterprise of ferreting out crime" and thus might act with understandable, but impermissible, zeal. But no companion notion exists with regard to databases: we do not conceive of databases as themselves acting "zealously."

Yet databases mirror and replicate human predispositions, and they cannot operate wholly divorced from them. What information is selected for input, how it is input, how it will be searchable—all of these values are defined by humans and therefore replicate human judgments about the relative importance and ordering of information. The database cannot communicate wholly neutrally: it always functions in the shadow of the human hand.

In short, databases are essentially human. Indeed, the way in which information is stored and retrieved may itself communicate predilections and biases. For example, DNA match probabilities are expressly quantified in racial terms that relate to individual self-identification rather than actual biological ancestry. Despite the patina of technological perfection, databases are ultimately compilations of *human* knowledge—created, maintained, and used by humans. It is flawed human beings that collect their information, write their operating codes, input their entries, maintain their systems, and search and retrieve their data. Databases may represent turbo-charged knowledge—but it is still human knowledge, just more powerful.

IV. TOWARD A CONSTITUTIONAL CRIMINAL PROCEDURE OF DATABASES

If these erroneous presumptions represent pits into which constitutional analysis is prone to fall when confronted with database technology, then what considerations might instead better govern or shape constitutional doctrine in this area? This Part offers five characteristics that define databases, and that merit acknowledgment in any honest constitutional inquiry. In outlining these qualities, my goal is to encourage direct thinking about databases, both in how they might conflict with some of the conventional models that guide constitutional doctrine as well as raise new concerns as yet unrecognized. Supplanting faulty stereotypical presumptions with descriptively accurate ones will hopefully facilitate development of doctrine that guides the formulation of appropriate constraints on database use and deployment, while eliminating constraints that unnecessarily impede it. The ideal is to acknowledge the ways in which databases represent a new form of collection, use, and dissemination of information and capitalize on those strengths while minimizing the weaknesses.

By way of clarification, my aim is more to think about the meaning of databases than the meaning of constitutional doctrine. That is, the qualities identified below to some extent transcend the particular constitutional doctrines at issue. Criminal justice database challenges can arise under a variety of constitutional clauses—including most obviously Fourth Amendment search and seizure, Fifth Amendment due process and self-incrimination, and Sixth Amendment confrontation, compulsory process, and right to counsel. But, for purposes of this Article, the specific contours of particular doctrine is less important than the shared features that inhere across databases generally, and which might help illuminate constitutional questions regardless of the particular claim at issue.

A. Structural (vs. Individual) Oversight

Perhaps the most singular trait that differentiates databases is that they require structural, rather than individual, oversight. That is, constitutional doctrines in criminal justice tend to particularize—they assume individual actors and individual actions. The archetypal criminal investigation is of the individual criminal, committing an individual crime (or series of them), investigated by individual officers. Thus, the system is structured accordingly: it restrains individual police officers through the Fourth Amendment, or respects individual rights through the Fifth and Sixth Amendments, and adjudicates individual claims through the Sixth Amendment processes. It seeks to resolve a single case or controversy.

Even to the extent that constitutional criminal procedure wrangles with entities, it does so awkwardly. Courts often reach for analogies to the individual or to personalization of a collective, rather than wholly alter the doctrinal approach. For instance, corporate liability simply substitutes a single, fictional corporate entity for an individual one. Or doctrines of racketeering liability serve to penalize groups of individuals by imputing liability for the actions taken by others with whom they associate. The collective largely evades us. We still depend on the Constitution to regulate investigations by regulating the actions of individual officers (through search and seizure and interrogation), and to regulate adjudication through inquiry of evidence related to an individual case (through case-specific discovery, cross-examination, assistance of counsel, and burdens of proof or evidentiary rules).

Databases, however, require a different tactic. Investigations of one person cannot be neatly disaggregated from that of others in the database; nor can adjudication be relied upon to expose shortcomings or flaws in the collection of evidence. Databases are rarely the product of one individual's action, and rarely contain easily separable individual information. Instead, they tend to be the product of numerous actors and inputs and collate numerous tiers of information. Think about *Herring*: an anonymous person put in the erroneous information, or else failed to remove it; then the information was accessed by one clerk and transmitted to another who failed to undertake any steps to verify it.[123] It is difficult, and maybe impossible, to identify the moment the error occurred or the individual who perpetrated it. It is likewise futile to attempt to regulate databases with reference to only one constituent part or one discrete moment of constitutional significance. The database's very purpose is to derive information from myriad inputs that stretch over time.

Take, for example, the DNA database. The DNA database is in itself largely a fiction; even the name of the federal database, CODIS, reveals as much. CODIS, or the Combined DNA Index System, in fact refers not to a central repository of information, but rather to the software used by the individual law enforcement entities that have met the standards and entered into an agreement to share data.[124] Each local or state entity uploads basic information to a centralized repository, and automated or intentional searches then indicate matches that can be pursued by contacting the uploading agency. Thus, to the extent that CODIS even exists, it incarnates

123. *Herring*, 129 S. Ct. at 698.

124. Erin Murphy, *The New Forensics: Criminal Justice, False Certainty, and the Second Generation of Scientific Evidence*, 95 CAL. L. REV. 721, 739 (2007).

as a pointer system—it tells a user where to look for the source information to which they have generated a match.

Moreover, the stored information itself is a product of a chain of information generation: the chemical and mechanical technologies required to type and analyze a genetic sample, the analyst who must interpret and enter the data, and the engineers that write the software code and maintain and superintend the databases themselves. Within each category are subcategories of data technicians, analysts, and law enforcement agents, each with a particular role to play that is essential to the creation and maintenance of the database. And this description only scratches the surface; the DNA database also depends upon other databases in order to draw meaning. Without sub-databases of profiles collected to gauge population frequencies, for instance, it is impossible to state with certainty what the significance of a DNA match is at all.

Thus, to the extent that oversight of the integrity of such databases is to occur, or that constitutional doctrine intends to regulate their formation and use, any such oversight must inevitably operate structurally, not individually. Yet doctrine is currently ill-suited to such a task. To the extent that there is any notion of database oversight, it ill-accommodates this kind of structural, overarching approach. An individual model focuses on what happens in a particular case, and which individuals interfaced with the database in that instance (who loaded a sample, or ran a search, for example), without considering that person within the larger web of inputs and outputs necessary to the databases' continued operation.

This kind of myopia is understandable, however, given the highly tailored and individualized notion of criminal justice that has typified the Anglo-American system. Indeed, some critics lament that American justice is so intensely myopic that it unnecessarily excludes even highly probative evidence of general propensities, even limited to a particular individual.[125] The current model of adjudication is ill-suited to the task of reviewing and assessing evidence derived from databases. The procedural entitlements of the adversary system depend upon matched opposing forces engaging in a contest to resolve discrete disputed facts. Tools such as cross-examination or compulsory process cannot sweep broadly enough to disclose systemic failings.[126] Discovery, compulsory process, or cross-examination in a sin-

125. Edward J. Imwinkelried, *Reshaping the "Grotesque" Doctrine of Character Evidence: The Reform Implications of the Most Recent Psychological Research*, 36 SW. U. L. REV. 741, 743-45 (2008).

126. *See, e.g.*, Ken Strutin, *Databases, E-Discovery and Criminal Law*, 15 RICH. J.L. & TECH. 6, 14-16 (2009) (discussing United States v. Dioguardi, 428 F.2d 1033 (2d Cir. 1970)).

gle case yields little opportunity to identify and uncover, much less broadly correct, errors apt to occur (and be visible) only from scrutiny on a systemic level.

Any truly effective constitutional doctrine for databases will have to contend with their multifaceted and interdependent, rather than discrete and intimate, operation. It will also have to provide a means of access to information created and held by private entities, yet used for the benefit of public criminal justice actors. A structural approach, rather than a model that presumes individual investigative or adjudicative oversight, is required, and it may be that new monitoring institutions should be developed.

Such a shift requires a major reorientation in constitutional thinking. Rather than ask about individual actors' specific behaviors—a substantive approach, in some respects—a procedural inquiry that asks about the existence of protocols and monitoring systems would be preferred. In other words, it is arguably impossible to regulate databases *substantively*—to truly inquire whether a particular series of tests or entries or searches were accurate, fair, and correct. But it is much easier to impose *procedural* requirements upon databases—to inquire into the existence and thoroughness of protocols for those processes and to presume inadequate or defective any database system maintained without them. Certain structural devices are demonstrably effective in minimizing mistakes, and with greater attention, others would be uncovered. A constitutional doctrine that looks for the signs of good management—think scrutiny of policies for access controls, or regular audits, or blind tests—is far more likely to improve database deployments in society than one that attempts to determine whether a database has failed or not in a case-specific context.

B. Suspicionless (vs. Suspicion-Based) Targeting

Concomitant with recognition of the structural rather than individual nature of database systems is an acknowledgement that, as a result, databases can (and often do) yield information divorced from individual suspicion of a particular person. That is, databases can function both to identify and to confirm. In the confirmatory mode, the search for information is targeted—a user seeks particular information about a particular person, based on suspicion or other information. The information in the database simply confirms—perhaps rapidly or more comprehensively—the suspicions already held. For example, a known suspect's records are searched for a criminal past, or a DNA match, or evidence that she crossed a bridge or made a call.

In the identification mode, in contrast, the user is more or less dispassionate about the ultimately identified individual—the database, in essence,

chooses the suspect. The known information is the input, and its relationship to a criminal act; what the database supplies is a match or list of suspects—the building blocks, or final nail, of the investigation.

By way of example, think of a cell phone call record database. If an investigator seeks all the calls made from a particular phone, or set of phones, then the investigator seeks to confirm or dispel suspicion related to the holders of those numbers. But an investigator might also seek the identity of all callers who contacted the homicide victim, say, or whose calls were routed through a particular tower at a particular time. In this usage, the investigator relies upon the database to identify the suspect, independent from the investigator's own suspicions about the likeliest suspect.

The difference in these two modes of operation matters greatly. Where databases are used to confirm otherwise-held suspicions, they more traditionally resemble other traditional forms of police investigation. But the use of databases to generate suspects represents a new kind of investigation altogether—whether based on particular information (e.g., "who called this number") or upon predefined algorithms (e.g., "who has traveled to these three countries and bought these two items within a one month period").

Suspicion models that constrain investigation, therefore, will both over-regulate and under-regulate the use of databases in this context.[127] The great fear with regard to this kind of databasing is, in Dan Solove's formulation, the fear of the anonymous bureaucrat, not the malevolent inspector.[128] The restraints called for will inevitably falter if they presume the need for the overzealous, overstepping constable. The true risk is a leaping-to-conclusions, or confirmation bias. It is the fear that the individual will be sucked into a morass of suspicion from which escape is arduous or impossible—Kafka's *The Trial*, not Orwell's Big Brother.[129]

Conversely, the need to constrain such searches due to lack of suspicion may in fact constitute unnecessary overreaching. The invasion of privacy occasioned by physical confrontations, or even informational inquiries, conducted by police against suspected individuals raises a different tenor of concern than that represented by search queries into established and regulated databases. Many of these database searches—for the DNA match, the license plate match, the cell phone callers—will be troubling inasmuch as the composition of the database reflects some troubling inequities, or inas-

---

127. Chris Slobogin makes a similar argument with regard to individual versus group searches. *See* Christopher Slobogin, *Is the Fourth Amendment Relevant in a Technological Age?*, *in* TECHNOLOGY AND THE CONSTITUTION 30 (Jeffrey Rosen & Benjamin Wittes eds.) (forthcoming 2010).

128. SOLOVE, *supra* note 12, at 36-42.

129. *Id.* at 31, 41-43.

much as *what follows next* raises issues. The search in itself is not the problem, and questions like reasonable suspicion or probable cause carry little meaning.

A constitutional doctrine obsessed with suspicion standards, however, misses this point. Suspicion standards help winnow the number and degree of interests affected by law enforcement actions, but when a database search is at issue, those are the wrong metrics. And the converse impulse—to simply ignore that there is any special import to a database search—is equally undesirable. Database searches arguably *enhance* privacy in that they are often anonymous, virtual (rather than physical), and suspicionless—those qualities increase the likelihood that they are executed unobtrusively and neutrally. But they do not spread their burdens equitably if they are not fairly composed and adequately monitored. Thus, it is these latter questions that should serve as the gates of entry to relatively unfettered use of databases, rather than to abandon any effort to regulate them at all or to attempt to impose suspicion-based models where they ill fit.

C. Operative Opacity (vs. Transparency)

Relatedly, databases do not just *generate* information anonymously; they tend to *operate* anonymously, as well. Database content is typically shrouded in secrecy. And such secrecy is usually desirable—the mere effort required to compile and organize the information they contain represents a major investment, which the investing parties typically have little incentive or desire to expose. This tendency often serves to help protect information, since databases frequently compile information on sensitive topics, whether it is the criminal history of an individual or the fact that they possess certain characteristics. Generally speaking, the secrecy of databases is desirable—this is precisely why the possibility of "hacking" databases generates such loud objections, and why privacy experts demand myriad safeguards on the integrity of the information they keep.

Databases also are often, by their nature, secret from within. They have multifarious inputs, which means both that the identity of the relevant agent can be difficult to discern, along with their responsibility for particular substance. These ambiguities are exacerbated by the fact that a database can represent the product of public and private cooperation. Thus, the rules and procedures that provide access to one set of database actors (say, the government officials that maintain and run them) may differ from the rules and procedures that govern access to a different set (say, a private company that creates the software).

In some respects, the anonymity and diversity that characterizes databases in turn serves to protect them from abuse. The code that makes them

run, the operators that input data, and even the persons who access that data may all interact with the information in a nameless, faceless way. Unlike the beat officer or public prosecutor, the technician need not have as personal a stake in the construction of the database or the outcome of a search. Each party may have a great interest in ensuring the integrity of their own component part or role, without having a vested interest in any particular outcome. It may defuse some of the potency of the information held in databases if they are primarily composed of separate, neutralized parts.

But the secrecy that protects databases from abuse may also often enable abuse from within. With so many different entities at play, it can be both hard to monitor the quality of each contributor and difficult to use established legal regimes to do so. It is not as though there are procedures in the criminal justice system for a defendant to implead Applied Biosystems in order to gain access to primer sequences used for forensic DNA typing—the only option is an awkward fumble with the jurisdiction's rules of discovery. And given the Sixth Amendment's parsimonious view of criminal discovery, there is no guarantee that those rules will suffice.[130]

Litigation, as a formal and public event, also poorly serves as the vehicle for identifying and correcting problems that may arise in database administration. The secrecy with which most databases are kept is, as previously observed, largely desirable. Compilations of sensitive or proprietary information should not be cavalierly put out for public display. But there is a strong presumption of openness for court records and hearings, and ordering disclosures even in a controlled or sealed environment risks leaks. Understandably, then, courts might hesitate before commanding that Google turn over its billion dollar search algorithms or that the federal government hand over its database to every criminal defendant for research.

It is too much to require courts, or to expect the Constitution, to demand full transparency in the methods of database administrators. But courts, and the Constitution, can demand transparency in the articulation and application of consequences to those administrators' efforts. Courts can require that the database undergo regular, demonstrably effective auditing processes, and ask to see proof of such. They can view the absence of information about the database—such as how often it is used, how often it is audited, what the results are—as a sign that the database is inadequately attended, rather than as confirmation of its reliability.

---

130. The only real constitutional rule is that exculpatory information be turned over. *See* Brady v. Maryland, 373 U.S. 83, 91-92 (1963).

D. Use (vs. Acquisition) Restrictions

Databasing also requires reconsideration of the current focus upon regulating the acquisition of information while ignoring its subsequent maintenance or use. Fourth Amendment doctrine scrupulously attends to how information is acquired, resting largely upon a premise of a physical dimendimension to the moment. The "reasonable expectation of privacy" test once promised to consider "people, not places," but that exhortation has largely reverted to a highly material standard for the constitutional threshold that barely corresponds to individuals' actual subjective expectations.[131] The Sixth Amendment likewise strongly preserves the encounter between accused and evidence—the physical face-to-face confrontation embodied in cross-examination—as represented by the Supreme Court's recent jurisprudence privileging in-court accusation over other, arguably more effective, forms of confronting evidence.[132]

But the import and the impact of a database occurs less with regard to the moment of the information's acquisition than with all the moments that then may follow. Indeed, acquisition may not represent any kind of threat to individual liberty or privacy at all. Recall the criminal records database at issue in *Reporters Committee*—there, the Court acknowledged that the true significance of the database was not its contents, which were all technically a matter of public record, but the act of compiling and rendering that information accessible in a particular way.[133]

Yet the salience of compiling or organizing information all too often remains obscured in constitutional analysis. Consider again, by way of example, the DNA databases. Almost all of the cases that have examined the constitutionality of collecting DNA samples from convicted persons, and some of the recent arrestee cases, have zeroed in on the moment of collection as the relevant point of inquiry.[134] If collecting the sample is permissible, then the constitutional inquiry effectively ends.

A more sensitive inquiry would require a more complicated disaggregation of the steps involved in creating and maintaining a DNA database. It would care less about mere collection—an act which, in all candor, is ra-

---

131. *See, e.g.*, CHRISTOPHER SLOBOGIN, PRIVACY AT RISK: THE NEW GOVERNMENT SURVEILLANCE AND THE FOURTH AMENDMENT 112 (2007).

132. *See* David Alan Sklansky, *Hearsay's Last Hurrah*, 2009 SUP. CT. REV. 1 (2010) (criticizing the romanticization of cross-examination as a tool to uncover errors especially with regard to certain kinds of evidence, such as the scientific evidence that now pervades criminal cases, without inquiring whether other methods might in fact be more effective).

133. U.S. Dep't of Justice v. Reporters Comm. for Freedom of Press, 489 U.S. 749, 764 (1989).

134. Murphy, *supra* note 2, at 1358-62.

ther easily and painlessly accomplished with or without the consent of the person sampled—and focus instead upon the remaining moments in the process.[135] The actual use of the samples—for instance, the kinds of tests used, the information potentially revealed, the validation and accuracy of the methods employed, and so on—might garner greater attention.[136] The structure of the databasing of the information, including the manner of data entry, the steps taken to ensure accuracy, the approved duration of retention, and the means through which information can be retrieved would likewise merit scrutiny. Lastly, the purpose to which the database could be put—not just as a function of the kinds of searches undertaken, but also as a function of the ends such searches intend to serve, would become a critical aspect of any review.

In short, rather than follow an industrial age model reliant upon physical acquisition, constitutional doctrine would transition to an information age approach based on knowledge, creation, and dissemination. Such attentiveness would offer more effective safeguards around the creation and utilization of databases, and be responsive to concerns about insufficient auditing structures and function creep. Viewed as living, evolving organisms rather than as static repositories of discrete bits of information, the lawfulness and constitutionality of a database would more closely correspond to its actual use and deployment.

---

135. As it is, the Supreme Court has suggested that the testing of biological material constitutes a "search" for Fourth Amendment purposes. Ferguson v. City of Charleston, 532 U.S. 67, 76 (2001). However, even in *Ferguson*, three Justices, including currently sitting Justices Scalia and Thomas, deemed the testing of the urine at issue to not constitute a search protected by the Fourth Amendment. *Id.* at 92 (Scalia, J., dissenting) ("I suppose the *testing* of that urine for traces of unlawful drugs could be considered a search of sorts, but the Fourth Amendment protects only against searches of citizens' 'persons, houses, papers, and effects'; and it is entirely unrealistic to regard urine as one of the 'effects' (*i.e.,* part of the property) of the person who has passed and abandoned."); *see also id.* at 92-93 ("Some would argue . . . that testing of the urine is prohibited by some generalized privacy right 'emanating' from the 'penumbras' of the Constitution (a question that is not before us); but it is not even arguable that the testing of urine that has been lawfully obtained is a Fourth Amendment search. (I may add that, even if it were, the factors legitimizing the taking of the sample, which I discuss below, would likewise legitimize the testing of it.)").

136. In the DNA cases, pains are often taken to characterize the typed information as "junk" as a means of assessing the privacy invasion, or to note that penalties attach to unauthorized uses of DNA samples. *See, e.g.*, United States v. Kriesel, 508 F.3d 941, 947-48 (9th Cir. 2007). But no federal appellate court has found the fact that the entire biological sample is retained (rather than just the "junk" numerical profile) relevant to its analysis, and most have relied upon the statutory penalties for "unauthorized" uses without worrying that the entity that defines "authorized" is the very one (law enforcement) that might have incentives to overreach. *See, e.g.*, Banks v. United States, 490 F.3d 1178, 1191-92 (10th Cir. 2007).

E. Benign Neglect (vs. Deliberate Misdeeds)

Finally, regulation of databases require constitutional criminal procedure to focus less upon deliberate or intentional abuses of power than upon unintentional omissions, or mere benign neglect. There is always the risk that a malfeasant actor will corrupt or exploit a database system, to be sure. But constitutional regulation of databases aimed at ferreting out intentional harms will be very thin indeed; it is far easier to do harm, and far greater harm can be done, through mere benign neglect of database systems than through intentional manipulation.

The split among the Justices in *Herring* starkly illustrate this distinction. The majority viewed the sole purpose of the exclusionary rule to be deterrence and concluded that applying the rule yielded little deterrent benefit with regard to a negligent recordkeeping error.[137] In contrast, Justice Ginsburg in dissent argued that more than marginal deterrence was possible, in the specific context of database entry, even when the complained of error constituted mere negligence in care.[138]

Most tellingly, however, the majority of the court essentially presumed the regular operation of the database—even in the face of a factual dispute about the record in this regard[139]—whereas the dissenters questioned whether such evidence could in fact ever be effectively adduced.[140] The dissenters in *Herring* talked about structural reforms and best practices—they could not point to one operator that acted wrongfully or one rule that was flouted. The dissenters, in other words, viewed deterrence as systemic because the database operated systemically, rather than view it as a question of individual deterrence, related to one particular operator's actions.

To be sure, there are political and ideological fault lines that likely separate the majority and the dissenters in *Herring*; their differences rest on more than mere perspective shifts in the meaning of databasing generally. But the debate is nonetheless illuminating in that it demonstrates how the very nature of databases—their opacity, anonymity, and systematic qualities—can in turn enhance their invisibility to conventional constitutional doctrine.

---

137. Herring v. United States, 129 S. Ct. 695, 701, 703-04 (2009).

138. *Id.* at 708 (Ginsburg, J., dissenting) (noting that the tort system is premised on deterrence for negligent actions).

139. *Id.* at 706 & n.2, 709.

140. *Id.* at 708-09.

CONCLUSION

Information databases are an enduring part of the landscape of criminal justice—that much is obvious. They are simply too valuable and too essential to the project of law enforcement to imagine discarding them altogether. But the database model—what goes in them, how they are used, and what comes out—corresponds little to the models of criminal justice that have operated through the ages. Attempts to shoehorn databases into current doctrine have thus largely failed, and simply ignoring the differences risks leaving an important source for investigations and evidence wholly unregulated. By identifying some of the shared, and inaccurate, presumptions that tend to shield databases from closer constitutional scrutiny, and then outlining some of their unique characteristics that merit special attention, this Article underscores the need for deliberate conversation about the significance of this new technology to current criminal justice frameworks.